
В. Лави

ОПЛАТА ПО РЕЗУЛЬТАТАМ: ВЛИЯНИЕ НА ОБЪЕМ И ПРОДУКТИВНОСТЬ РАБОТЫ УЧИТЕЛЕЙ И ПРИНЦИПЫ ОЦЕНИВАНИЯ УЧЕНИКОВ¹

Статья поступила
в редакцию
в апреле 2010 г.

Аннотация

Представлены выводы о влиянии материальных стимулов на учителей английского языка и математики в Израиле. Учителя получали денежные бонусы за улучшение результатов учеников на выпускных экзаменах.

В распределении школ на две группы — участвующие и не участвующие в программе — использовалась переменная, значения которой были измерены со случайной ошибкой, что сделало распределение на группы случайным. На этом был основан главный метод исследования.

*Стимулирование учителей способствовало существенному повышению доли сдававших экзамены (*test taking rate*), количества сдавших (среди тех, кто сдавал бы экзамены вне зависимости от введения программы; *conditional pass rate*) и среднего балла сдававших.*

Успеваемость улучшилась благодаря изменению методов обучения, увеличению продолжительности дополнительных занятий во внеурочное время и большей восприимчивости учителей к потребностям учащихся. Никаких манипуляций учителей с результатами экзаменов выявлено не было.

Ключевые слова: учителя, мотивация, оплата по результатам, рейтинги, соревнование, принципы оценивания учеников, социальная политика.

Модели оплаты по результатам внедряются во многих странах в атмосфере острой полемики и противодействия учителей и про-

¹ Lavy V. Performance pay and teachers' effort, productivity and grading ethics. Dep. of Economics, Hebrew Univ., Dep. of Economics, Royal Holloway Univ. of London. NBER and CEPR. August 2008. This is a substantially revised version of NBER Working paper № 10622 (пер. с англ. С. Кучковской, научный редактор перевода Е. Сивак).

фессиональных объединений². Обоснованием для реализации этих программ служит представление о том, что, поощряя учителей материально, можно мотивировать их к улучшению своей работы. Однако данные, свидетельствующие о таком влиянии, малочисленны. В настоящем исследовании представлены результаты экспериментальной программы, в рамках которой учителя имели возможность получить денежные бонусы в зависимости от результатов сдачи их учениками выпускных экзаменов в средней школе по английскому языку и математике. В рамках программы составлялись рейтинги учителей по каждому из предметов на основании успеваемости их учеников, и учителя получали вознаграждение в зависимости от места в рейтинге. В качестве индикаторов результативности работы учителей использовали два показателя успеваемости учащихся: количество сдавших (pass rate) и средний балл (average score) на каждом выпускном экзамене. Общая сумма вознаграждения в каждом из «соревнований» учителей определялась заранее. Размер индивидуальных бонусов зависел от результатов ранжирования и установленных поощрений за верхние места в относительном рейтинге. В проведенном эксперименте наибольший интерес представляло то, как программа повлияла на применяемые учителями педагогические методики, прикладываемые усилия и продуктивность их труда, которая измерялась результатами учеников на экзаменах, а также на принципы оценивания учеников (teachers' grading ethics).

Программа была разработана как экспериментальная, тем не менее для участия в ней школы отбирались не случайно. Несмотря на это, дизайн программы позволяет реализовать квазиэксперимент (quasi randomized trial), такая возможность обеспечивается присутствием в дизайне двух элементов: 1) школы включались или не включались в программу в зависимости от значения определенного показателя: если пороговое значение (threshold function of an observable) было превышено, школа не участвовала в программе; 2) погрешность измерения этого показателя. Школу включали в программу, если в ней доля сдавших выпускные экзамены (matriculation rate) в 1999 г. не превышала критического значения (45%). Это правило может быть представлено формулой

$$T = 1\{S \leq 45\},$$

где T — индикатор включения в экспериментальную группу (assignment to treatment);

S — переменная, в зависимости от значения которой школа попадает под воздействие программы (assignment variable, далее переменная-критерий).

² В качестве примера можно привести системы оплаты по результатам в Денвере (2006 г.) и Хьюстоне (2006 г.). Еще раньше, в середине 1990-х годов, такие программы были реализованы в Дейд Каунти (Флорида) и в Далласе; на уровне штата — в Айове, Аризоне и Калифорнии в 2002 г.; в Цинциннати, Филадельфии и Ковентри (Род Айленд). Некоторые из этих программ рассмотрены в публикациях [27; 21; 25; 17].



Однако S была измерена с определенной погрешностью, т. е.

$S = S^* + \varepsilon$, где:

S^* — реальное значение переменной (true rate);

ε — погрешность измерения.

Администраторы программы, не зная, что используемая переменная измерялась с погрешностью, включили некоторые школы в выборку ошибочно. Как показано ниже, ε , по всей видимости, является в значительной степени случайной величиной и не влияет на потенциальный результат. Поэтому группа школ вокруг порогового значения включалась в обследование случайным образом. И поэтому, контролируя S^* потенциально полностью непараметрическим способом, фактически получали условное случайное включение в программу (conditional random assignment).

При реализации рассмотренной стратегии были также использованы имеющиеся панельные данные (данные измерений до и после внедрения программы), что позволило выполнить оценивание методом «разность разностей» (differences-in-differences, DID) в условиях естественного эксперимента³.

Основной метод исследования (identification strategy) описан выше. Были также использованы два других подхода. Первый из них основан на дизайне «разрыв регрессии» (regression discontinuity), в рамках которого предполагается, что школы, значение переменной-критерия (S) в которых чуть выше или чуть ниже порогового, имеют равную вероятность попадания в контрольную или экспериментальную группу. Другими словами, распределение школ, в которых значение S близко к пороговому, на контрольную и экспериментальную группы случайно. Если эта предпосылка выполняется, то подвергнутые и не подвергнутые воздействию программы школы, значение S в которых близко к пороговому, неразличимы по потенциальному результату, т. е. баллы учеников на экзаменах были бы примерно равными, если бы программы не было или если она не оказывает никакого влияния на успеваемость. Однако более слабое допущение основано на контроле параметрических функций от S . Другими словами, при данных значениях S не ожидается изменений T . Эта разрывность (discontinuity) в механизме включения в программу была использована в настоящем исследовании для оценивания эффекта от стимулирования учителей. Здесь, как и в методе с погрешностью измерения, использовались панельные данные и разрывный дизайн (regression discontinuity design, RD) в оценивании методом «разность разностей». Второй подход основан на сопоставлении всех участвовавших в программе школ со всеми школами, которые не были отобраны для участия в программе, с использованием панельных данных для контроля фиксированных эффектов

³ Метод «разность разностей» (difference-in-differences — DID) — это неэкспериментальный метод, применяемый в эконометрике для измерения эффекта, оказанного определенным воздействием, в данный период времени. — Примеч. пер.

школ (school fixed effects). Он также снижает риск влияния неучтенных характеристик учеников, так как в качестве контроля используются многомерные данные об успеваемости за предыдущие годы. Потенциально расчеты, основанные на двух указанных альтернативах, все равно могут быть смещенными, но фактически метод разрывности регрессии дает результаты, схожие с теми, что получаются при использовании естественного эксперимента с рандомизацией, основанного на погрешности измерения переменной, по которой школы включались или не включались в программу.

В разделе 1 настоящего исследования приведена базовая информация об израильской школьной системе, дано описание программы стимулирования учителей и обсуждается теоретический контекст моделей оплаты по результатам преподавания. В разделе 2 рассмотрены идентификация (разные способы деления школ на две группы для сравнения), расчеты и результаты; в разделах 3 и 4 содержатся материалы, подтверждающие влияние поощрения учителей на прикладываемые ими усилия, на методики преподавания и принципы оценивания учеников (начинают ли учителя завышать оценки ученикам). В разделе 5 описан контекст эксперимента со стимулированием учителей в Израиле и возможности его применения в других странах. В разделе 6 сформулированы выводы.

Полученные результаты дают основания полагать, что материальное стимулирование учителей способствует улучшению успеваемости учащихся: повышается уровень участия в тестировании, увеличивается условная доля сдавших экзамен (т. е. среди тех, кто сдавал бы его независимо от введения программы стимулирования учителей) и растут баллы, полученные учащимися в ходе тестирования по математике и английскому языку. Рост этих условных результатов, рассчитанных на основе оценок по внутришкольным экзаменам и внешним тестам, обуславливает более чем наполовину улучшение безусловных результатов по математике и несколько меньше — по английскому языку. Мы предполагаем, что это связано не с искусственным завышением баллов тестирования или манипулированием ими, а с изменениями в методах преподавания, увеличением времени на внеурочные занятия и с возросшим вниманием учителей к потребностям учащихся. Данные о том, что стимулирование увеличило прилагаемые учителями усилия и обусловило совершенствование педагогических методик, имеют большое значение в связи с высказываемыми в последнее время многими специалистами опасениями, что стимулирование может привести к таким непреднамеренным результатам, как «натаскивание» на тесты и манипуляции с баллами при тестировании, т. е. что стимулирование учителей не способствует полноценному обучению.



С точки зрения экономической теории есть основания ожидать, что стимулирование будет способствовать возникновению мотивации к эффективной работе. Так, заключение контракта, предусматривающего вознаграждение за качество работы, обычно приводит к изменению в профессиональном поведении работника. Однако в частном секторе только для некоторых видов профессиональных обязанностей в контракт включается четко сформулированный пункт о вознаграждении в зависимости от характеристик работы. Главное ограничение в применении индивидуального вознаграждения заключается в том, что соблюдение этого условия влечет за собой дополнительные риски для тех работников, которым работодатели будут вынуждены повышать зарплату и из-за которых они, следовательно, будут нести дополнительные издержки. Еще одну проблему порождает несовершенство контрактов, которое может приводить к дисфункциональным изменениям в поведении работников — перераспределению внимания и усилий с тем, чтобы улучшать только те аспекты своей деятельности, за которые они получают вознаграждение. Учитывая эти обстоятельства, легко понять, почему частные компании стимулируют работников в основном либо путем продвижения по службе, либо в рамках систем группового поощрения, а не вознаграждением отдельных лиц за их достижения [24].

В сфере образования групповое стимулирование также превагирует над схемами материального поощрения отдельных работников. Обычно это объясняют особенностями самого образовательного процесса: обучение предполагает групповую работу, объединение усилий коллектива учителей и многих других заинтересованных сторон, а также комплексный характер деятельности, которая отличается большим числом решаемых задач. В такой ситуации трудно измерить вклад конкретного работника. Группа — в данном случае коллектив учителей — часто бывает лучше, чем работодатель, информирована об отдельных работниках, участвующих в учебном процессе, и об их индивидуальном вкладе, что позволяет членам группы осуществлять мониторинг работы своих коллег, поддерживать в них стремление к увеличению прикладываемых усилий или проявлению каких-либо других необходимых форм поведения. Также отмечается, что работники, перед которыми стоит общая цель, с большей вероятностью будут помогать друг другу, а также будут усерднее работать при необходимости подменить отсутствующего члена группы. С другой стороны, безусловное существование «проблемы безбилетника» дает повод сомневаться в том, действительно ли схемы вознаграждения групп учителей предоставляют значительные стимулы, особенно если группа многочисленна [8].

Первоначально считалось, что соревнование как инструмент материального стимулирования целесообразно в ситуациях, когда отдельные работники проявляют усердие в надежде получить более высокооплачиваемую должность, когда за этой должностью

1. Соревнование как стимул к совершенствованию

закреплено определенное материальное поощрение и когда имеет место конкуренция между работниками за эту должность [19; 5]. Единственное, что действительно имеет значение в соревновании, так это то, как человек работает в сравнении с другими, а не абсолютные показатели его деятельности. Хотя в преподавательской среде повышение по службе не считается важной составляющей карьеры, в схемах материального стимулирования учителей также уместен акцент на относительных достижениях, а не на абсолютных. Этому есть две причины. Во-первых, стимулирование, основанное на относительных показателях работы, и фиксированный размер выделенных для вознаграждения сумм позволяют не выйти за рамки бюджета. Во-вторых, при отсутствии четких стандартов, которые можно было бы использовать в качестве основы для абсолютных показателей, предпочтительной альтернативой является оценка того, насколько хорошо работает учитель по сравнению с его коллегами. Поэтому в описанном ниже эксперименте по материальному стимулированию учителей было применено соревнование с ранжированием.

1.1. Среднее образование в Израиле

Учащиеся старших классов средних школ в Израиле сдают несколько выпускных экзаменов *Bagrut*. Это набор государственных испытаний по основным дисциплинам и дисциплинам по выбору. Выпускные экзамены начинаются в 10-м классе, продолжаются в 11-м и заканчиваются в 12-м, когда проводится большая часть тестов. Учащиеся сами выбирают, тесты какого уровня сложности они будут сдавать по каждому из предметов. За каждое тестирование по каждому предмету дается от одного до пяти кредитов (зачетных единиц)⁴. Итоговый балл выпускного экзамена по предмету равен среднему значению двух промежуточных результатов. Первый промежуточный балл присваивается на государственных экзаменах, являющихся внешними для школ. Задания для них готовит независимое агентство, которое также организует и проводит тестирование, осуществляет контроль и выставляет баллы. Оценивание результатов является анонимным: внешнему экзаменатору не сообщают ни фамилии ученика и учителя, ни номер школы. Экзамены проводятся в июне и в январе, тестирование всех учащихся по данному предмету проходит в один и тот же день. Проверку результатов на государственных экзаменах осуществляют централизованно два независимых внешних экзаменатора, и конечный результат равняется средней величине выставленных ими баллов. Второй промежуточный балл

⁴ Во многих странах (Германия, Франция, Италия) применяется схожая система выпускных экзаменов в средней школе. Это касается также регентских экзаменов (*regents examinations*) в штате Нью-Йорк и системы комплексной оценки (*comprehensive assessment system*) в штате Массачусетс. Учебный план и система внешних экзаменов *Bagrut* в Израиле очень напоминают существующую в США систему *advanced placement courses*. (Университетские курсы, которые ученики могут проходить еще в школе; кредиты по этим курсам засчитываются ученикам, если они после школы учатся в этом университете. — *Примеч. пер.*



присваивается на внутреннем экзамене. Этот экзамен проводится в школе, и он аналогичен государственному экзамену по содержанию и формату, но учитель сам проверяет тесты и выставляет оценки своим ученикам.

Некоторые дисциплины являются обязательными, по ним требуется набрать по крайней мере три кредита. Тесты, за выполнение которых присваивают больше кредитов, являются более сложными. Английский язык и математика относятся к основным обязательным дисциплинам; их необходимо изучить на одном из трех уровней: базовом (3 кредита), промежуточном (4 кредита) или продвинутом (5 кредитов). Для получения выпускного сертификата требуется минимум 20 кредитов. В 1999 и 2000 г. выпускных сертификатов добились около 45% старшеклассников средних школ, т. е. они сдали достаточное число экзаменов, чтобы набрать 20 кредитов, и сделали это к моменту выпуска или сразу после него [10]. Наличие выпускного сертификата средней школы является обязательным условием для поступления в университет и одним из самых важных с экономической точки зрения образовательных достижений.

В начале декабря 2000 г. Министерство образования развернуло в 49 израильских средних школах эксперимент по апробации новой бонусной системы оплаты труда учителей⁵. Его главной особенностью стало то, что учитель получал персональный бонус, размер которого определялся в зависимости от успеваемости его учащихся. В эксперименте участвовали все учителя английского языка, иврита, арабского языка и математики, преподававшие в 10–12-х классах в период до выпускных экзаменов по этим предметам в июне 2001 г. В декабре 2000 г. при участии Министерства образования проводился инструктаж для директоров и административных работников школ. Программа позиционировалась как добровольный трехлетний эксперимент⁶. Его основные положения были с энтузиазмом встречены всеми директорами школ, за исключением одного человека, который решил отказаться от участия в эксперименте.

Школам было разрешено вместо учителей иврита и арабского языка допустить к участию в программе учителей по другим основным дисциплинам, по которым сдаются выпускные экзамены (Библия, литература или гражданское право). Участие в эксперименте учителей иврита и арабского языка было необязательным — право принимать решение было предоставлено школам.

1.2. Израильский эксперимент по стимулированию учителей

⁵ Одновременно в группе из других школ осуществлялась программа, предусматривавшая бонусы учащимся. Школы, участвовавшие в этих разных программах по материальному стимулированию, не пересекались — это касается как групп, участвующих в программе, так и контрольных групп.

⁶ Вследствие изменений в правительстве марте 2001 г. и последовавших за этим сокращений бюджета Министерство образования объявило летом 2001 г., что эксперимент не будет проводиться на второй и третий год, как планировалось ранее.

Данный выбор мог оказать влияние на данные эксперимента, т. е. на вероятность успешного участия учителей в рейтинговании, что могло привести к смещениям в результатах по тому или иному предмету. Поэтому, оценивая программу, мы рассматривали только учителей английского языка и математики.

Каждый из четырех рейтингов (английский, иврит и арабский, математика, другие предметы) составляли учителя 10–12-х классов, ученики которых должны были сдавать экзамен по одному из этих предметов в июне 2001 г. Каждый учитель участвовал в рейтинговании столько раз, в скольких классах он преподавал, и его работу каждый раз оценивали путем ранжирования по средней успеваемости учеников из его класса. Ранжирование основывалось на разнице между фактическим результатом и значением, прогнозируемым с помощью регрессии, в которой контролировались социально-экономические характеристики учащихся, уровень программы обучения по соответствующей дисциплине (базовый, промежуточный или продвинутый), класс (10-й, 11-й и 12-й), численность класса и фиксированные эффекты школы и уровня обучения (*fixed school-level effect*)⁷. Использование фиксированных эффектов школы предполагает, что прогнозируемые значения строились на основе дисперсии показателей среди учителей в школе, и учителей четко информировали о том, что их будут сравнивать с другими учителями, преподававшими тот же предмет в той же школе. Отдельные регрессии были применены для расчета прогнозируемой доли сдавших экзамен (*pass rate*) и среднего балла (*mean score*), и каждого учителя оценивали дважды, рассматривая разницу между предсказанным с помощью регрессии значением с фактическим результатом. Школы предоставляли списки учащихся, которые должны сдавать экзамены, с указанием классов, предметов и фамилий учителей. В изучаемую группу вошли те, кто попал в списки на дату начала реализации программы — 1 января 2001 г. Всем учащимся, включенным в эти списки, но не сдавшим экзамен, независимо от причины присваивали экзаменационный балл, равный нулю.

Все учителя с положительной разницей между фактическими результатами учеников и предсказанными в регрессии значениями для обоих показателей успеваемости — и для доли сдавших экзамен, и для оценок за экзамен — подразделялись на четыре группы. Учителя получали баллы в зависимости от группы, в которую они попадали: 16 баллов в первой группе, 12 — во второй, 8 — в третьей и 4 — в четвертой. Администраторы программы присваивали больший вес результатам по показателю «доля сдавших экзамен», увеличивая на 25 % количество баллов за каждое

⁷ Стоит обратить внимание на то, что использованная для прогнозирования регрессия не включала баллы за прошлые годы, чтобы у преподавателей не было стимула мошенничать, например предлагая учащимся не проявлять старание на более ранних экзаменах, не учитываемых в соревновании. Если бы программа имела продолжение, эта характеристика была бы важна.



из четырех мест в рейтинге (вместе с «надбавкой» получалось 20, 15, 10 и 5 баллов соответственно). Итоговое количество единиц по двум ранжированиям использовалось для определения позиции учителя в рейтинге, выявления победителей и их награждения по следующей шкале: 30–36 баллов — \$7500; 21–29 баллов — \$5750; 10–20 баллов — \$3500 и 9 баллов — \$1750. Эти суммы следует признать значительными, если сравнить их со среднегодовым доходом учителей средней школы — \$30 000 без налогов — и учесть то обстоятельство, что в одном цикле соревнования учитель мог выиграть несколько наград, подготовив не один, а несколько классов к выпускным экзаменам⁸. Поскольку программа была представлена учителям только в середине года, маловероятно, что имел место отбор учителей на основе ожидания повышенного дохода или что учителя манипулировали составом учащихся в своих классах.

При включении школ в программу руководствовались тремя формальными правилами: допускались только полные средние школы, реализующие обучение с 7-го по 12-й класс; предыдущие результаты учеников школы на выпускных экзаменах по математике и английскому языку должны были быть сравнительно низкими⁹; результат самых последних выпускных экзаменов (доля получивших сертификаты) должен был равняться или быть ниже, чем средний по стране (45%). Первым двум критериям соответствовали 106 школ, но семь из них были исключены из эксперимента, поскольку они уже участвовали в других коррекционных образовательных программах. Из оставшихся 99 школ — потенциальных участниц программы 49 удовлетворяли третьему критерию. Однако, как упоминалось выше, одна школа отказалась участвовать в программе, и таким образом фактически в эксперименте участвовали 48 школ (изучаемая выборка)¹⁰.

В программе приняли участие 629 учителей, из которых 207 преподавали английский язык, 237 — математику, 148 — иврит или арабский и 37 — другие предметы, выбранные школами вместо иврита. 302 учителя выиграли награды: 94 учителя английского языка, 124 учителя математики, 67 учителей иврита и арабского языка и 17 учителей по другим предметам. Три учителя английского языка и двенадцать учителей математики получили по две награды каждый, а один учитель иврита занял дважды первое место и получил суммарно \$15 000.

⁸ Подробнее см. [9].

⁹ В качестве показателя успеваемости использовалась средняя доля сдавших экзамены по математике и английскому языку за предшествующие четыре года (1996–1999). Если уровень был ниже 70% в двух или более случаях, считалось, что показатели работы школы были слабыми. Английский язык и математику выбрали потому, что по этим предметам были самые высокие показатели не сдавших экзамены.

¹⁰ Поскольку в выборку школ, которые могли бы участвовать в программе, было включено относительно большое число религиозных и арабских школ (выше их доли в выборке), пороговое значение для них было установлено на уровне 43%.

Мы провели опрос учителей, участвовавших в программе, во время летних каникул по окончании учебного года. Удалось опросить 74% учителей. Причиной несостоявшихся интервью в большинстве случаев оказались неправильно указанные номера телефона или невозможность дозвониться после нескольких попыток. Опрос показал следующее. 92% учителей знали о программе. 80% были осведомлены о деталях, почти всех инструктировали директора школ и координаторы программы. 75% учителей считали, что информация была полной и удовлетворительной. Почти 70% были знакомы с критериями награждения победителей и около 60% из них полагали, что попадут в число победителей. Только 30% опрошенных не верили, что смогли бы победить; остальные же считали это вероятным. Две трети выборки полагали, что программа стимулирования учителей повысит успеваемость учащихся.

1.3. Данные

В исследовании использовались данные по учебному году, предшествовавшему проведению программы — с сентября 1999 г. по июнь 2000 г., и году, когда проходил эксперимент (с сентября 2000 г. по июнь 2001 г.). По каждой школе была получена информация о том, арабская это школа или еврейская; если еврейская — то светская или религиозная; какова доля сдавших выпускные экзамены в 1999–2001 гг. По каждому учащемуся был получен полный отчет о результатах на выпускных экзаменах *Bagrut* за период обучения в старших классах средней школы (10–12-й классы) и персональные данные: пол, образование родителей, размер семьи, иммиграционный статус — для тех, кто приехал в Израиль недавно. По каждому выпускному экзамену была указана дата, предмет, уровень сложности (количество кредитов) и полученный балл. Основу выборки составили все обучавшиеся в 2000 и 2001 г. ученики 12-х классов. Очень незначительная часть учащихся выполнила все требования по математике и (или) английскому языку к окончанию 11-го класса — эти учащиеся в исследовании не рассматривались; соответственно выборки по математике и английскому языку не идентичны.

Были установлены три показателя результатов летних экзаменов (июнь 2001 г.) по каждому предмету: участие в тестировании, прохождение экзамена (балл больше или равен 55) и фактический балл за тестирование (от 1 до 100). По двум последним критериям ранжировали учителей для отбора победителей. В табл. 1 представлена описательная статистика для когорт старшеклассников средних школ 2000 и 2001 г. по двум выборкам: 48 школ, включенных в программу, и все оставшиеся 50 средних школ, не включенных в программу, но удовлетворявших первым двум критериям¹¹.

¹¹ Школа, отклонившая предложение участвовать в программе, была исключена из всех выборок, поскольку она не предоставляла необходимые данные. Изъятие этой школы из анализа не влияет на основные результаты исследования, поскольку с учетом доли сдавших выпускные экзамены за 1999 г. в этой школе ее в любом случае не включили бы в выборки.



Таблица 1

**Описательная статистика: участвовавшие в программе
школы в сравнении со школами, не участвовавшими в ней (удовлетворявшие
только двум первым критериям включения в программу)**

	2000 г.			2001 г.		
	Школы, участвовавшие в программе	Школы, не участвовавшие в программе	Разница	Школы, участвовавшие в программе	Школы, не участвовавшие в программе	Разница
А. Характеристики школ						
Религиозные школы*	0,199	0,269	-0,070 (0,084)	0,182	0,258	-0,076 (0,080)
Арабские школы*	0,260	0,099	0,161 (0,081)	0,284	0,107	0,176 (0,087)
Уровень сдачи экзаменов Ваgrut в предыдущем году	0,369	0,620	-0,251 (0,027)	0,377	0,607	-0,230 (0,035)
В. Характеристики учащихся						
Образование отца	9,062	11,386	-2,324 (0,639)	9,029	11,357	-2,329 (0,582)
Образование матери	8,817	11,486	-2,669 (0,709)	8,551	10,846	-2,295 (0,751)
Число родных братьев и сестер	3,463	2,580	0,883 (0,422)	3,472	2,481	0,991 (0,425)
Пол (мужчины = 1)	0,495	0,466	0,028 (0,032)	0,508	0,492	0,016 (0,029)
Иммигранты	0,031	0,027	0,004 (0,015)	0,022	0,010	0,012 (0,009)
Азиатское/африканское этническое происхождение	0,190	0,208	-0,018 (0,031)	0,170	0,190	-0,020 (0,030)
С. Успеваемость учащихся в прошлые годы						
Кредиты, полученные по математике	0,290	0,499	-0,209 (0,132)	0,320	0,571	-0,251 (0,130)
Кредиты, полученные по английскому языку	0,127	0,194	-0,067 (0,047)	0,116	0,183	-0,067 (0,050)
Полное количество кредитов, которое могло быть получено	4,292	5,283	-0,991 (0,320)	4,502	5,464	-0,962 (0,341)
Полное количество полученных кредитов	3,388	4,591	-1,203 (0,301)	3,633	4,773	-1,140 (0,303)
Средний балл	56,580	69,555	-12,974 (2,296)	58,381	69,699	-11,318 (2,010)
Количество наблюдений	6,250	5,931	12,181	6,084	5,820	11,904
Количество школ	48	50	98	48	50	98

Примечание: Стандартные ошибки, приведенные в круглых скобках, даны с учетом группировки на уровне школ.

* Статус школ по национальной и религиозной принадлежности не меняется. Любое изменение средних значений по годам отражает относительные изменения в численности учащихся в когорте.

* Данная таблица основана на выборке по математике.

Приведенные в табл. 1 стандартные погрешности даны с поправкой на кластеризацию внутри школ. Из панели А следует, что в обследуемой выборке было больше арабских школ, чем в группе не участвовавших в программе школ, и что эта разница значительно отличалась от нуля¹². Панель А демонстрирует также большой разрыв между результатами выпускных экзаменов в 2001 г. в участвовавших и не участвовавших в программе школах. В 1999 г. этот разрыв составил 25,1 %, в 2000 г. — 23,0 %. Панели В и С показывают, что в год, когда вводилась программа, и в предшествующие годы средние показатели по разным характеристикам учащихся в попавших в программу школах отличались от соответствующих средних показателей в не попавших в программу школах. Например, среднее количество лет обучения родителей более чем на два года выше в необследуемых школах (в школах, которые удовлетворяли только первым двум критериям). Значительная разница между двумя группами школ наблюдалась также в среднем уровне успеваемости учащихся за прошлые годы. Таким образом, выборка 48 обследуемых школ не является репрезентативной для всех школ, которые могли бы участвовать в программе. Поэтому простое сопоставление попавших и не попавших в программу школ не может служить основанием для выводов.

2. Идентификация, оценивание и результаты

2.1. Естественный эксперимент вследствие случайной ошибки измерения

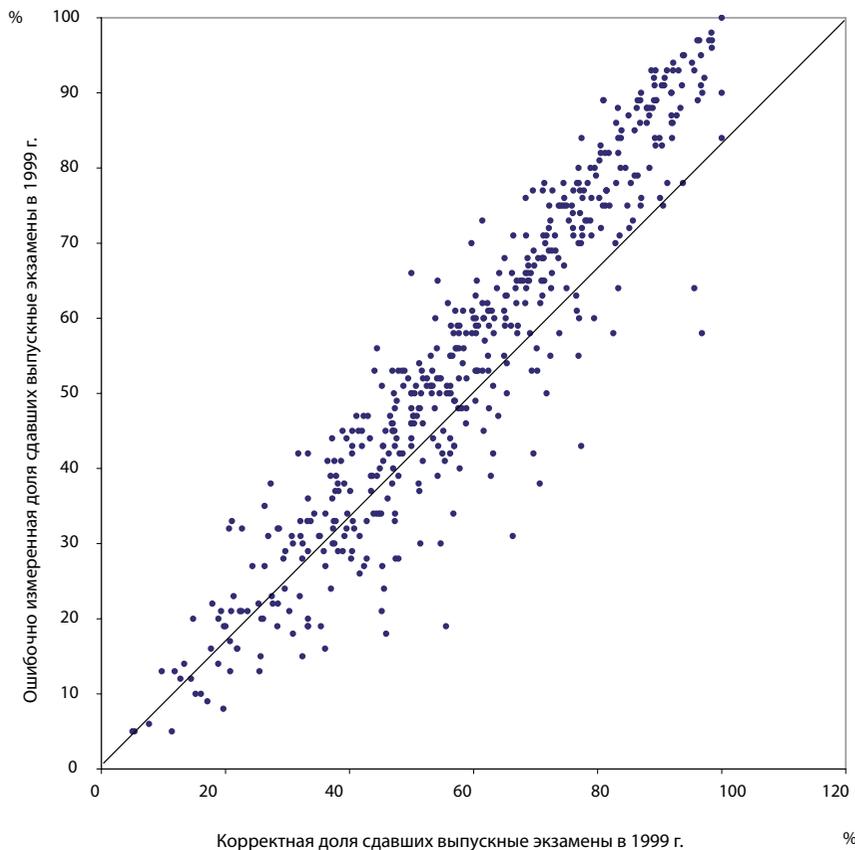
В программу по правилам включались только те школы, в которых доля сдавших выпускные экзамены в 1999 г. равнялась или была ниже 45 % (43 % для религиозных и арабских школ). Однако результат выпускных экзаменов (*matriculation rate*), на который ориентировались при отборе школ в программу, был измерен неточно. Данные, предоставленные администраторам программы, были взяты из предварительного и неполного файла, содержащего результаты сдачи выпускных экзаменов. По некоторым учащимся информация была некорректной. Позже министерство подкорректировало данный предварительный файл, как это делается ежегодно¹³. Оказалось, что результаты выпускных экзаменов, использованные для включения школ в программу, по большинству школ были некорректными. Погрешность измерения могла быть полезной для определения эффекта реализации программы. В частности, потенциально зависящий от реального результата выпускных экзаменов статус участия в программе мог быть по факту присвоен случайным образом вследствие указанных неточностей в предварительном файле.

¹² Эти характеристики школ постоянны во времени, но имеются незначительные различия их средних значений для 2000 и 2001 г., потому что их рассчитывают на разных выборках учащихся.

¹³ Требования к процессу сдачи выпускных экзаменов различаются для школ разного типа и для тестов разного уровня сложности по каждому предмету. Проверка информации администраторов национального тестирования и школ является длительным процессом. Первая версия данных о результатах выпускных экзаменов бывает доступна в октябре и окончательно формируется в декабре.



Рис. 1. **Связь между корректным и ошибочно измеренным результатами выпускных экзаменов в 1999 г. Выборка = 507 школ**

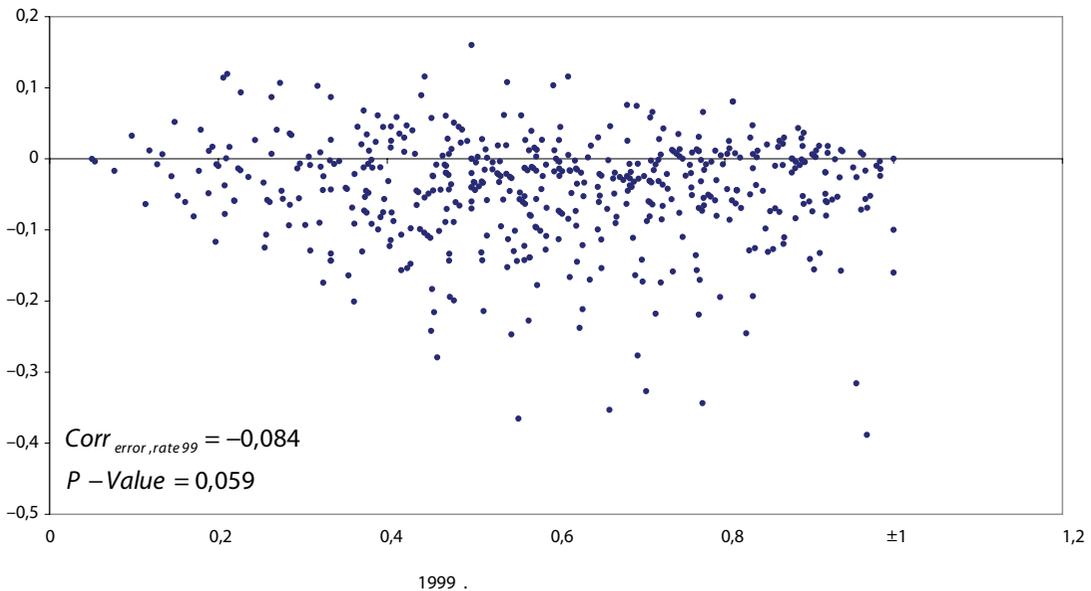


На рис. 1 показана связь между корректными и ошибочно измеренными результатами выпускных экзаменов для выборки 507 средних школ Израиля в 1999 г.¹⁴

Большая часть (80%) погрешностей измерения была отрицательной, в 17% случаев погрешность была положительной, и в оставшейся части школ погрешность отсутствовала. Отклонение от уровня 45%, по-видимому, никак не коррелирует с настоящим результатом сдачи выпускных экзаменов — долей сдавших экзамен учеников в школе. Это лучше видно на рис. 2. Здесь показано, что погрешность измерения и результат сдачи выпускных экзаменов меняются не единообразно; их коэффициент корреляции очень мал ($-0,084$) и p -value отлично от нуля ($0,059$). Однако если исключить несколько экстремальных значений (пять школ), коэффициент корреляции фактически равняется нулю. Хотя рисунок, вероятно, показывает, что дисперсия погрешности измерения меньше при низких долях сдавших выпуск-

¹⁴ В выборку вошли только школы с положительными ($> 5\%$) истинными значениями долей сдавших экзамены.

Рис. 2.

Корректные результаты выпускных экзаменов в 1999 г. и погрешность измерения. Выборка = 507 школ.


ные экзамены, это, скорее всего, объясняется так называемым floor effect, ограничивающим величину отрицательной погрешности: чем ниже доля сдавших выпускные экзамены, тем меньше абсолютная максимальная величина отрицательных погрешностей¹⁵.

Дальнейший контроль случайности характера погрешности измерений может быть основан на статистической связи погрешности измерения с другими характеристиками учащихся или школы. В табл. 2 приведены коэффициенты регрессий, в которых зависимой переменной была погрешность измерений, а независимыми — характеристики учащихся, результаты экзаменов в 2001 г. и характеристики школ для выпускников 2001 г. Каждая запись в таблице основана на отдельной регрессии, которая проводилась на средних на уровне школы значениях по всем переменным, отдельно для всей выборки и для выборки школ, подходящих только по двум критериям. Эти расчеты представлены в колонках 1 и 2 соответственно.

¹⁵ Схожие результаты наблюдались в случаях, когда в выборку были включены только школы с долей сдавших экзамены, превышающей 40%. В этой выборке устранена проблема границ для величины погрешности измерения в школах с низким показателем успешности на выпускных экзаменах. Была исследована также выборка, в которую вошли школы, которые могли бы участвовать в программе (97 школ, а не 98, так как у одной из школ отсутствовали данные об истинном результате выпускных экзаменов в 1999 г.). Результаты, представленные на рис. А1–А2 в онлайн-приложении (онлайн-приложение можно найти по ссылке <http://www.aeaweb.org/aer/data/dec09/20060532_app.pdf>), идентичны результатам на рис. 1–2.



Таблица 2 **Оценка регрессии для изучения связи погрешности измерения доли сдавших выпускные экзамены в 1999 г. с характеристиками учащихся и школ в 2001 г.**

	Все школы	Школы, не участвовавшие в программе	Школы, участвующие в рандомизированном эксперименте (RT)
А. Характеристики школ			
Религиозные школы	-0,008 (0,007)	-0,026 (0,017)	-0,065 (0,031)
Арабские школы	-0,022 (0,009)	-0,013 (0,021)	0,038 (0,058)
В. Характеристики учащихся			
Образование отца	0,001 (0,001)	0,001 (0,003)	-0,011 (0,007)
Образование матери	0,000 (0,001)	-0,002 (0,002)	-0,008 (0,006)
Число родных братьев и сестер	-0,008 (0,002)	-0,003 (0,005)	-0,007 (0,014)
Пол (мужчины = 1)	0,023 (0,013)	-0,001 (0,035)	-0,087 (0,060)
Иммигранты	0,068 (0,046)	-0,364 (0,173)	-0,674 (0,247)
Успеваемость учащихся в прошлые годы			
Кредиты, полученные по математике	0,010 (0,004)	0,009 (0,013)	-0,023 (0,034)
Кредиты, полученные по английскому языку	0,025 (0,008)	0,117 (0,042)	0,100 (0,091)
Кредиты, полученные по истории	0,007 (0,007)	0,026 (0,019)	0,055 (0,032)
Полное количество кредитов	0,003 (0,002)	0,002 (0,005)	-0,011 (0,011)
Средний балл	0,001 (0,000)	0,001 (0,001)	0,000 (0,002)
Количество	508	98	36

Примечание: Представленные в таблице коэффициенты основаны на отдельных регрессиях для изучения связи погрешности измерения с характеристиками учащихся, результатами экзаменов Bagrut в прошлые годы и характеристиками школ. Использовались средние данные по школам. Стандартные ошибки даны в круглых скобках.

По каждой выборке имеется 12 коэффициентов, и лишь несколько из них значимо отличны от нуля: три — среди коэффициентов в расчетах по полной выборке и два — по выборке школ, не участвовавших в программе. Кроме того, значимые переменные различны в рамках разных выборок. Это позволяет предположить, что это временные и случайные различия. Например, в выборке школ, не участвовавших в программе, имеется диспропорциональное количество учащихся-иммигрантов (коэффициент равен $-0,364$, стандартная ошибка $0,173$), но в выборке, включающей все школы, это различие между контрольными и обследуемыми школами положительное и практически нулевое (коэффициент равен $0,068$, стандартная ошибка $0,046$). На основании данных, приведенных на рис. 1–2 и в табл. 2, можно сделать вывод об отсутствии доказательств наличия существенной связи между погрешностью измерения в 1999 г. и исследуемыми характеристиками. Поэтому вероятность того, что погрешность измерений коррелирует с невыявленными переменными, также очень низка. Надо отметить, однако, что допущение о том, что погрешность измерения не коррелирует через некоторые невыявленные переменные с результатами 2000–2001 г., не может быть проверено.

Идентификацию, основанную на случайной погрешности измерения, можно представить следующим образом.

Пусть $S = S^* + \varepsilon$ — это содержащая погрешность доля сдавших выпускные экзамены 1999 г., использованная для включения школ в программу, где:

S^* — реальная доля сдавших выпускные экзамены 1999 г.;
 ε — погрешность измерения.

T обозначает статус участия, при этом $T = 1$ для участников программы и $T = 0$ для школ, не участвующих в ней. Поскольку $T(S) = T(S^* + \varepsilon)$, если контролируется S^* , то включение в программу является случайным («случайное включение» в программу при данных реальных значениях доли сдавших выпускные экзамены).

Наличие погрешности измерения приводит к естественному эксперименту, когда школа попадает в обработку случайно при данных S^* в подвыборке из 98 школ, удовлетворяющих двум критериям из трех. Для 18 школ в этой выборке правильная доля сдавших выпускные экзамены 1999 г. была выше порогового значения. Эти школы были ошибочно выбраны для участия в программе. Для каждой из этих школ могла существовать школа, имевшая идентичный реальный уровень сдачи выпускных экзаменов, но с небольшой отрицательной погрешностью измерения, достаточной для того, чтобы значение показателя стало ниже порогового значения и школа не попала под действие программы. Такое объединение школ в пары эквивалентно непараметрическому подбору школ

Рис. 3.

Определение выборки школ, ошибочно включенных в обработку или контроль. Выборка = 97 школ

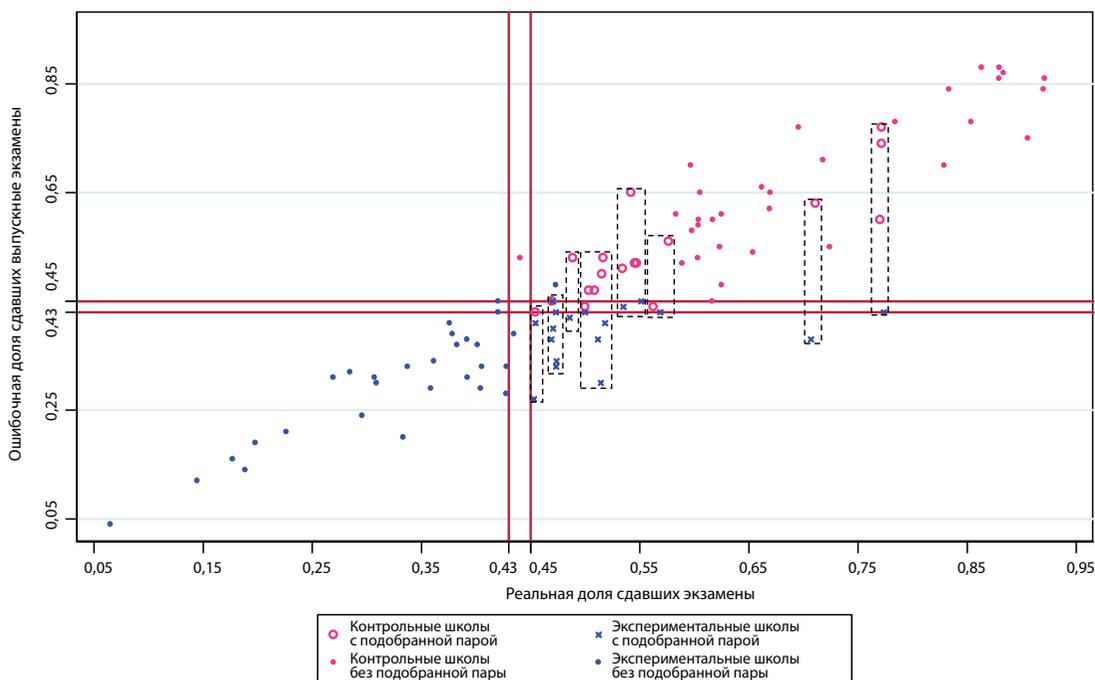




Таблица 3

Сравнение экспериментальных и контрольных школ: выборка рандомизированного эксперимента (RT)

	2000 г.			2001 г.		
	Экспериментальные	Контрольные	Разница	Экспериментальные	Контрольные	Разница
А. Характеристики школ						
Религиозные школы	0,330	0,219	0,110 (0,163)	0,324	0,214	0,110 (0,164)
Арабские школы	0,158	0,000	0,158 (0,088)	0,155	0,000	0,155 (0,087)
Уровень сдачи экзаменов Bagrut в предыдущем году	0,467	0,509	-0,042 (0,032)	0,474	0,475	-0,001 (0,053)
Уровень сдачи экзаменов Bagrut два года назад	0,490	0,519	-0,029 (0,049)	0,0527	0,0528	-0,002 (0,034)
В. Характеристики учащихся						
Образование отца	10,685	10,586	0,100 (0,821)	10,539	10,332	0,207 (0,838)
Образование матери	10,624	10,764	-0,140 (0,849)	10,519	10,539	-0,020 (0,947)
Число родных братьев и сестер	3,009	2,026	0,983 (0,410)	2,912	1,662	1,250 (0,384)
Пол (мужчины = 1)	0,513	0,414	0,098 (0,066)	0,556	0,431	0,125 (0,061)
Иммигранты	0,016	0,029	-0,013 (0,017)	0,025	0,012	0,013 (0,018)
Азиатское/африканское этническое происхождение	0,218	0,325	-0,107 (0,062)	0,325	0,276	-0,041 (0,054)
С. Успеваемость учащихся в прошлые годы						
Кредиты, полученные по математике	0,337	0,277	0,061 (0,172)	0,256	0,453	-0,197 (0,118)
Кредиты, полученные по английскому языку	0,155	0,077	0,078 (0,051)	0,107	0,079	0,028 (0,061)
Полное количество кредитов, которое могло быть получено	5,251	4,594	0,657 (0,674)	5,322	5,342	-0,020 (0,498)
Полное количество полученных кредитов	4,308	3,761	0,547 (0,601)	4,218	4,482	-0,264 (0,393)
Средний балл	63,131	64,774	-1,643 (2,591)	62,121	67,710	-5,589 (2,217)
Количество наблюдений	2,654	2,369	5,023	2,598	2,236	4,834
Количество взвешенных наблюдений	4,095	3,818	7,913	3,812	3,679	7,491
Количество школ	18	18	36	18	18	36

Примечание: Стандартные ошибки, приведенные в скобках, даны с учетом группировки на уровне школ. Данные были перевзвешены, чтобы получить одинаковую численность учащихся в контрольных и обследуемых школах внутри каждой группы школ, характеризующихся близкими значениями реальной доли сдавших экзамены. Статус школ по национальной и религиозной принадлежности не меняется. Любое изменение средних значений за несколько лет отражает относительное изменение численности учащихся в когорте. Данная таблица основана на выборке по математике.

по величине S^* . Поэтому был принят следующий алгоритм подбора: включать не включенную в программу школу в контрольную группу, если в ней доля сдавших выпускные экзамены примерно равна ($\pm 1\%$) реальной доле сдавших экзамены хотя бы в одной школе, включенной в программу ошибочно. На рис. 3 показан результат этой процедуры подбора пар: 18 школ из контрольной группы объединены прямоугольником со школами, попавшими в программу. В онлайн-приложении на рис. А3 показан более точный подбор пар. Внутри таких пар включение в программу может

рассматриваться как случайное. Поэтому 18 не включенных в программу школ можно использовать в качестве контрольной группы для идентификации эффекта от программы. Поскольку некоторые из контрольных школ соответствуют по параметрам более чем одной школе в экспериментальной группе и наоборот, взвешенные регрессии используются для объяснения различий между экспериментальными и контрольными школами в размерах выборки внутри выравненных групп (веса приведены в таблице А4 онлайн-приложения). С точки зрения внешней валидности результатов важно отметить, что допрограммная доля сдавших экзамены в 18 экспериментальных школах в 2000 г. составляет широкий диапазон — от 32 до 79%, что почти полностью перекрывает соответствующий диапазон для всех 48 попавших в программу школ.

Важно воспроизвести регрессионный анализ связи между погрешностью измерения и характеристиками школ и учащихся, взяв за основу выборку из 36 школ. Эти расчеты приведены в столбце 3 табл. 2; они очень схожи с расчетами по выборке всех школ и по выборке школ, не попавших в программу по третьему критерию. Два из 12 коэффициентов значимы, но ни один из результатов по успеваемости учащихся за 2001 г. в данной выборке не демонстрирует какой-либо дисбаланс между экспериментальными и контрольными школами. Фактически некоторые из различий в успеваемости учащихся за прошлые годы положительны, а некоторые отрицательны; это свидетельствует о том, что любые различия такого рода случайны.

В табл. 3 представлены усредненные характеристики школ и учащихся — допрограммные (2000 г.) и постпрограммные (2001 г.) — для 18 включенных в программу и 18 контрольных школ выборки рандомизированного эксперимента (RT). Первая панель отражает связи с характеристиками школ. Различия между экспериментальными и контрольными школами и стандартные ошибки приведены в колонках 3 и 6. Экспериментальная и контрольная группы выравнены по религиозному статусу, но не по национальности, поскольку в контрольной группе нет арабских школ. Как и ожидалось, результаты сдачи выпускных экзаменов год и два года назад хорошо сбалансированы для когорты старшеклассников 2001 г., поскольку были использованы значения S^* 1999 г., чтобы выравнять контрольную и экспериментальную группы. Вторая панель представляет связи с характеристиками учащихся. Имеются различия в числе родных братьев и сестер у учащихся и различия по гендерному составу, но какие-либо значимые различия в уровне образования родителей, иммиграционном статусе и этническом составе отсутствуют. Например, среднее количество лет обучения матери в 2001 г. практически идентично в двух группах, что заметно контрастирует с соответствующим различием в 2,3 года обучения, рассчитанным для выборки школ, которые могли бы по двум критериям участвовать в программе (табл. 1). Третья панель представляет связи с результатами учеников за прошлые годы, накопленные кредиты на выпускных экзаменах за 10–11-й классы, которые



для когорты 2001 г. следует рассматривать как допрограммные результаты. Большинство этих различий малы и незначимы, за исключением среднего балла в 2001 г. Особый интерес представляют кредиты за прошлые годы по математике и английскому языку. Ни в том, ни в другом году не наблюдалось каких-либо существенных различий между экспериментальной и контрольной группами. Усредненные значения итоговых кредитов за прошлые годы, включающие кредиты по всем предметам, по которым учащихся тестировали в течение 10–11-го класса, также идентичны для двух групп за оба года.

Приведенные в табл. 3 данные демонстрируют баланс между экспериментальными и контрольными школами по большей части характеристик школ и учащихся, что резко контрастирует с существенными различиями, показанными в табл. 1. Почти идеальная тождественность двух групп по результатам экзаменов прошлых лет (показано в нижней панели в табл. 3) — дополнительное основание для сравнения двух групп, образованных в ходе естественного эксперимента. Как будет показано в следующем разделе, также имеет место почти полная идентичность результатов экзаменов в 12-м классе когорты учащихся, окончивших школу до начала программы: доли участвовавших в тестировании и доли сдавших экзамены по математике и английскому языку в когорте учеников, окончивших школу до реализации программы, практически идентичны в контрольных и экспериментальных школах. Полученные данные дают возможность установить причинно-следственные связи на основании различий в результатах учеников после введения программы, что рассматривается в следующей части работы.

Для расчета регрессии с помощью выборки RT была применена следующая модель:

$$(1) Y_{jt} = \alpha + X_{jt}\beta + Z_{jt}'\gamma + \delta T_{jt} + \phi_j + \eta D_t + \varepsilon_{jt},$$

где i — индекс учащегося;

j — индекс школы;

t — индекс года, 2000-го и 2001-го;

T — статус участия/неучастия в программе; X и Z — объясняющие переменные, связанные с характеристиками учеников и школ;

D_t обозначает эффект года с факторной нагрузкой η .

Индикатор участия в программе T_{jt} равен взаимодействию между дамми-переменной для обследуемых школ и дамми-переменной для 2001 г. Регрессионный анализ проводился с использованием объединенных данных по обоим годам (для двух расположенных рядом когорт 2000 и 2001 г.), сформированных как панельные данные для школ с фиксированными эффектами школ (ϕ_j), включенными в регрессию. Результирующие расчеты могут быть описаны как взвешенная по количеству учеников процедура определения «разности разностей», обеспечивающая сопоставление эффекта от программы за несколько лет. Расчеты

2.2. Оценка и результаты



Таблица 4 **Оценивание методом «разность разностей» влияния программы на результаты по математике и английскому языку по выборке рандомизированного эксперимента (RT)**

	Математика						Английский язык					
	Все квартили		Оценивание по квартилям				Все квартили		Оценивание по квартилям			
	Ограни- ченный контроль	Контроль всех пере- менных	1-й	2-й	3-й	4-й	Ограни- ченный контроль	Контроль всех пе- ременных	1-й	2-й	3-й	4-й
Доля участвовавших												
Среднее значение для контрольной группы	0,802	0,419	0,815	0,903	0,971	0,865	0,529	0,903	0,972	0,977		
Эффект программы	0,046 (0,027) [0,038]	0,041 (0,021) [0,029]	0,133 (0,051) [0,068]	0,055 (0,035) [0,047]	0,037 (0,021) [0,030]	-0,021 (0,029) [0,039]	0,040 (0,017) [0,025]	0,033 (0,013) [0,019]	0,129 (0,045) [0,060]	0,013 (0,024) [0,033]	0,004 (0,014) [0,019]	-0,003 (0,019) [0,026]
Доля сдавших экзамен												
Среднее значение для контрольной группы	0,637	0,258	0,503	0,726	0,928	0,795	0,455	0,770	0,906	0,959		
Эффект программы	0,110 (0,036) [0,051]	0,087 (0,028) [0,040]	0,146 (0,048) [0,065]	0,209 (0,063) [0,087]	0,106 (0,035) [0,047]	-0,026 (0,029) [0,041]	0,047 (0,022) [0,031]	0,039 (0,020) [0,028]	0,107 (0,040) [0,052]	0,071 (0,035) [0,048]	-0,011 (0,021) [0,028]	-0,009 (0,034) [0,045]
Средний балл												
Среднее значение для контрольной группы	55,046	21,232	46,917	63,946	77,710	59,496	35,464	59,608	68,901	73,751		
Эффект программы	5,469 (2,292) [3,249]	5,307 (1,950) [2,739]	9,798 (3,497) [4,768]	10,920 (4,104) [5,686]	6,352 (2,122) [2,927]	-0,861 (2,493) [3,443]	3,240 (1,666) [2,359]	2,527 (1,452) [2,040]	5,889 (2,295) [3,150]	1,790 (2,840) [3,932]	-0,648 (1,708) [2,341]	0,344 (2,077) [2,828]
Условный эффект обследования												
Доля сдавших экзамен		0,052	0,051	0,161	0,073	-0,007		0,009	-0,001	0,059	-0,015	-0,006
Безусловный эффект, %		59	35	77	69	-		23	-	84	-	-
Средний балл		2,323	2,465	7,006	3,541	0,839		0,238	-2,255	0,901	-0,915	0,550
Безусловный эффект, %		44	25	64	56	-		9	-	50	-	-
Количество	9,857	2,421	2,365	2,424	2,647	10,111	2,506	2,390	2,500	2,715		

Примечание: 1. Стандартные ошибки, приведенные в круглых скобках, сгруппированы на уровне «школа — год». Стандартные ошибки, приведенные в квадратных скобках, даны для уровня школ. 2. Данные были перевзвешены, чтобы получить одинаковую численность учащихся в контрольных и экспериментальных школах внутри каждой группы школ, характеризующихся близкими значениями реальной доли сдавших выпускные экзамены. 3. В колонках 3–6 и 9–12 эффекты программы меняются по квартилям распределения балла за предыдущие тесты. Оценивание выполнено на основании четырех отдельных регрессий, по одной для каждого квартиля. 4. Контрольные переменные — характеристики студентов во всех колонках, кроме 1 и 7, включают набор дамми-переменных: для числа родных братьев и сестер; образования отца и матери; средней по школе доли сдавших экзамены год и два года назад; азиатского/африканского происхождения; иммиграционного статуса; пола; количества кредитов, которое могло быть получено; среднего балла; полного количества присужденных кредитных единиц и кредитных единиц, присужденных только по рассматриваемой дисциплине. Во все уравнения включены фиксированные эффекты школы. 5. В колонках 1 и 7 контролируемыми переменными являются дамми-переменные для школ — доля сдавших выпускные экзамены в школе год и два года назад, а также количество кредитных единиц, которое могло быть получено учащимся. 6. Строка, в которой указаны средние значения показателей для контрольной группы, показывает среднее значение для учащихся в контрольных школах в 2001 г.



взвешивают по числу учащихся в каждой школе. Включение фиксированных эффектов школ контролирует влияние неучтенных переменных, постоянных во времени, а также обеспечивает альтернативный контроль кластеризации на уровне школы. Однако стандартные ошибки также кластеризуются на уровне школы для каждого среза. Фиксированные эффекты школ также поглощают часть изменений в усредненных результатах выпускных экзаменов по школе, что, возможно, приводит к повышению точности.

В табл. 4 приведены результаты оценки уравнения (1) для выборки рандомизированного эксперимента (RT). Представлены расчеты для двух видов этого уравнения. Первый вид характеризуется ограниченным набором контролируемых переменных, в который включены только фиксированные эффекты школ, дамми-переменная для года, доля сдавших выпускные экзамены за один или два предыдущих года и количество кредитных единиц, которые пытаются получить ученики на экзаменах. Последняя переменная объясняет характерную дисперсию результатов экзаменов в школах от года к году. Второй вид уравнения дополнительно контролирует следующие характеристики учеников: число родных братьев и сестер; дамми-переменную для пола; образование отца и матери; дамми-переменную для иммиграционного статуса; дамми-переменную для азиатской или африканской этнической принадлежности; количество кредитов, полученных по соответствующему предмету (т.е. по математике или английскому языку) до начала программы; полное количество кредитов, накопленных на экзаменах к началу программы, и средний балл по этим экзаменам.

Оцененный эффект от участия в программе на основании ограниченного набора контрольных характеристик по математике (колонка 1) и английскому языку (колонка 7) является положительным и значимым для всех трех показателей успеваемости. Расчеты, основанные на полном наборе контрольных характеристик, приведены в колонках 2 и 8. Они очень близки к тем, которые основаны на ограниченном наборе контрольных характеристик. Это подтверждает ожидаемый результат при полученных данных сравнения контрольной и экспериментальной групп школ, представленных в табл. 3. Поэтому рассматриваются только расчеты по второму виду уравнения¹⁶. Эффект от программы применительно к доле сдававших тест по математике составляет 0,041,

¹⁶ В табл. А1 онлайн-приложения приведены расчетные коэффициенты по всем переменным, включенным в полные уравнения по математике и английскому языку, для всех трех результатов по каждому предмету. Однако при этом не приводятся расчеты коэффициентов для образования отца или матери, так как они содержатся в уравнении в виде набора дамми-переменных (отдельный индикатор для каждого интервала лет обучения). Предпочтение было отдано этому виду уравнения, чтобы исключить «вменение» образования родителей для недостающих значений. Учащиеся, для которых нет данных об образовании отца или матери, сгруппированы под двумя отдельными дамми-переменными. Однако оценки влияния не меняются, когда интервальная переменная образования родителей заменяет дамми-переменные образования родителей.

Таблица 5

Оценивание влияния программы на результаты по математике и английскому языку по выборке рандомизированного обследования (RT) по данным срезов 2001 и 2000 г.

	Математика					Английский язык				
	Все квартили	Оценивание по квартилям				Все квартили	Оценивание по квартилям			
		1-й	2-й	3-й	4-й		1-й	2-й	3-й	4-й
2001										
Доля участвовавших										
Среднее значение для контрольной группы	0,802	0,419	0,815	0,903	0,971	0,865	0,529	0,903	0,972	0,977
Эффект программы	0,030 (0,032)	0,092 (0,051)	0,087 (0,028)	0,011 (0,025)	-0,096 (0,037)	0,026 (0,020)	0,076 (0,037)	0,054 (0,019)	-0,015 (0,016)	-0,043 (0,025)
Доля сдавших экзамен										
Среднее значение для контрольной группы	0,637	0,258	0,503	0,726	0,928	0,795	0,455	0,770	0,906	0,959
Эффект обследования	0,037 (0,052)	0,089 (0,048)	0,139 (0,070)	-0,005 (0,062)	-0,127 (0,045)	0,025 (0,028)	0,063 (0,047)	0,068 (0,033)	-0,016 (0,024)	-0,057 (0,029)
Средний балл										
Среднее значение для контрольной группы	55,046	21,232	46,917	63,946	77,710	59,496	35,464	59,608	68,901	73,751
Эффект программы	2,561 (3,706)	4,818 (3,489)	10,216 (4,745)	-0,338 (4,155)	-9,327 (4,154)	1,769 (1,947)	3,196 (2,894)	4,328 (2,053)	-1,019 (2,054)	-2,953 (2,666)
Количество	4,834	1,234	1,161	1,196	1,243	4,964	1,281	1,163	1,227	1,293
2000										
Уровень участия в тестировании										
Среднее значение для контрольной группы	0,786	0,492	0,771	0,909	0,945	0,846	0,587	0,834	0,974	0,964
Эффект программы	-0,022 (0,022)	-0,027 (0,055)	0,055 (0,040)	-0,018 (0,023)	-0,057 (0,021)	-0,005 (0,019)	0,015 (0,045)	0,066 (0,042)	-0,019 (0,015)	-0,037 (0,016)
Доля сдавших экзамен										
Среднее значение для контрольной группы	0,654	0,305	0,619	0,785	0,880	0,744	0,438	0,717	0,870	0,926
Эффект программы	-0,023 (0,035)	0,008 (0,047)	-0,005 (0,046)	-0,034 (0,037)	-0,072 (0,036)	0,002 (0,038)	0,000 (0,073)	0,067 (0,052)	0,032 (0,031)	-0,060 (0,020)
Средний балл										
Среднее значение для контрольной группы	51,895	24,490	46,606	62,379	72,783	56,821	34,430	53,404	66,227	72,019
Эффект программы	0,185 (2,537)	-0,274 (3,293)	4,900 (3,159)	0,866 (2,891)	-4,372 (3,267)	1,929 (2,519)	1,828 (4,615)	8,422 (4,018)	2,505 (1,898)	-2,375 (1,715)
Количество	5,023	1,187	1,204	1,228	1,404	5,147	1,225	1,227	1,273	1,422

Примечание: 1. Стандартные ошибки, приведенные в скобках, сгруппированы на уровне школ. 2. Данные были перевзвешены, чтобы получить одинаковую численность учащихся в контрольных и экспериментальных школах внутри каждой группы школ, характеризующихся близкими значениями реального уровня сдачи выпускных экзаменов. 3. В колонках 2–5 и 6–10 эффекты обследования меняются по квартилям распределения балла за предыдущие тесты. Оценивание выполнено на основании четырех отдельных регрессий, по одной для каждой квартиры. 4. Контролируемые характеристики учеников включают набор дамми-переменных: для числа родных братьев и сестер; образования отца и матери; средних значений уровня сдачи выпускных экзаменов в прошлые годы; азиатского/африканского происхождения; иммиграционного статуса; пола; количества кредитных единиц, которое могло быть получено (credit units attempted); среднего балла по заявленному количеству кредитов, полного количества присужденных кредитных единиц и кредитных единиц, присужденных только по рассматриваемой дисциплине. Во все уравнения включены фиксированные эффекты школ.



т. е. имеет место 5 %-е улучшение по сравнению со средним значением для контрольных школ (0,802). Влияние программы на долю сдавших выпускные экзамены составляет 0,087, т. е. относительное увеличение равняется 14 %. Эффект программы применительно к среднему баллу по математике состоит в увеличении этого показателя на 10 %. По английскому языку влияние более скромное: на 4 % повысился уровень участия в тестировании, на 5 % — доля сдавших выпускные экзамены и на 4 % — средний балл.

Интерпретация этих улучшений успеваемости как следствия введения программы основана на случайном включении школ в программу из-за погрешности измерений при заданной фактической доле сдавших экзамен в 1999 г. Это условие обеспечивалось подбором к каждой экспериментальной школе контрольной школы, имеющей точно такую же долю сдавших выпускные экзамены в 1999 г. (S^*) в выборке рандомизированного эксперимента. В результате такого выравнивания высокая отрицательная корреляция ($-0,689$) между S^* и T , рассчитанная по выборке школ, которые удовлетворяли первым двум критериям, была уменьшена практически до нуля на выборке рандомизированного эксперимента. Таким образом, непараметрический контроль S^* делает распределение на контрольную и экспериментальную группы случайным в выборке RT. Тем не менее в панельные данные была включена величина S^* за 1999 г. в качестве контрольной для результатов в когорте выпускников 2001 г., и при данной симметрии в структуре сформированных панельных данных была также использована S^* за 1998 г. в качестве контрольной для когорты 2000 г. Отсутствие этого контроля снижает оценку эффекта от программы применительно к доле сдававших тестирование по математике до 0,037 (по сравнению с 0,041) и по английскому языку до 0,029 (по сравнению с 0,033) и повышает эффект программы применительно к доле успешно сдавших экзамен по математике до 0,092 (по сравнению с 0,087), но оставляет оценки других эффектов без изменения. Таким образом, введение в уравнение реальной доли сдавших выпускные экзамены двухлетней давности не изменяет существенно оценки эффектов от программы, поскольку эта доля не коррелирует с участием/неучастием в выборке рандомизированного обследования и поскольку ее коэффициент в итоговом уравнении незначим (см. табл. А1 онлайн-приложения). Отсутствие в уравнении показателя доли сдавших выпускные экзамены год назад также не изменяло результаты.

До сих пор анализ основывался на выборке RT с панельными данными, позволяющими проводить оценивание методом «разность разностей». В табл. 5 представлены перекрестные результаты для выборок RT 2000 и 2001 г., для того чтобы понять, насколько эффект от программы чувствителен к контролю фиксированных эффектов школ.

В столбцах 1 и 6 приведены оценки, основанные на полной выборке и полном контроле переменных, для математики

и английского языка соответственно. В панели А приведены результаты по 2001 г., в панели В — аналогичные результаты по 2000 г. Эти данные показывают, что бóльшая степень повышения успеваемости, показанная в табл. 4 (столбцы 2 и 8), обусловлена положительным и значимым влиянием программы на результаты 2001 г., поскольку эффекты для 2000 г. в большинстве своем крайне малы, а иногда даже отрицательны. Например, влияние на долю сдававших тест по математике в 2001 г. составило 0,030, а в 2000 г. — (-0,022); разница между двумя значениями равнялась 0,052 по аналогии с соответствующим значением 0,041, приведенным в столбце 2 табл. 4. Однако в большинстве случаев представленные в табл. 4 результаты, полученные методом «разность разностей», выше и точнее по сравнению с результатами расчетов по срезам, приведенных в табл. 5. Это позволяет предположить, что введение фиксированных эффектов школ необходимо для того, чтобы получить более последовательные и точные оценки. Однако, как будет показано ниже, эта ситуация меняется, если допустить, что оценка эффектов от программы может варьировать в зависимости от способностей учащихся.

2.3. Допущение гетерогенности в эффекте от программы в зависимости от способностей учащихся

Для дополнительной проверки причинно-следственной интерпретации результатов, представленных в столбцах 1 и 7 табл. 4, были оценены модели, допускающие вариативное влияние программы в зависимости от успеваемости учеников в предшествующие программе годы. В частности, допускалось наличие связи между масштабом влияния программы и средним количеством набранных на предыдущих экзаменах баллов (взвешенным на количество кредитов; при этом тем, кто не участвовал в сдаче экзаменов, присваивалось значение 0). На основании этого среднего балла, являющегося мощным инструментом прогнозирования успеха учащихся на тестировании по математике и английскому языку, были созданы дамми-переменные для каждого квартиля распределения среднего балла. Применяя эти дамми, можно оценить следующую модель для каждого из трех представляющих интерес результатов по английскому языку и математике:

$$(2) Y_{ijt} = \alpha + X_{ijt}'\beta + Z_{jt}'\gamma + \sum_q d_{qi} \mu_q + \sum_q \delta_q d_{qi} T_{jt} + \Phi_j + \eta D_t + \varepsilon_{ijt},$$

где d_{qi} ($q = 2, 3, 4$) — квартили распределения среднего балла за тестирования, которые проходили до января 2001 г., когда была введена программа;

δ_q — эффект от программы в зависимости от квартиля; μ_q — основной эффект квартиля.

В альтернативной стратегии допускается, что влияние всех переменных может различаться по квартилям, и этому варианту было отдано предпочтение, хотя результаты не слишком отличаются, если оценивать уравнение (2).

Учащиеся, получившие очень высокие баллы в прошлые годы, с большей вероятностью могли участвовать в итоговом тестировании и успешно сдать выпускные экзамены по каждому из предметов



и без содействия программы. Это утверждение подкрепляется тем фактом, что средняя доля сдавших выпускные экзамены в данном квартиле в 2000 г. равнялась 90 %. Преподаватели, вероятно, сами поняли, что предел для учеников самого высокого квартиля практически достигнут, и поэтому направили свои усилия на более низкие квартили, где имелись большие возможности для улучшения показателей. Поэтому и не ожидалось, что в самом высоком квартиле будет наблюдаться влияние программы стимулирования учителей на успеваемость учащихся. Напротив, учащиеся, баллы которых приближались к среднему уровню или были ниже, составляли категорию, где дополнительные усилия — и со стороны учителей, и их собственные — давали эффект. Именно поэтому в данном исследовании следовало ожидать значимых показателей влияния программы особенно для учащихся в двух последних квартилях.

В столбцах 3–6 и 9–12 табл. 4 приведены результаты оценивания эффекта от программы стимулирования учителей математики и английского языка, проведенного на разделенной на квартили выборке. Представленные данные свидетельствуют о том, что общий эффект программы применительно к результатам по математике (столбец 2 табл. 4) достигнут преимущественно за счет влияния на учеников в трех первых квартилях (чей балл немного выше или ниже среднего балла), тогда как для самых успевающих учащихся (квартиль 4) существенного эффекта от программы не наблюдалось. Влияние программы на результаты экзаменов по английскому языку можно отметить только в двух первых квартилях. Неудивителен нулевой эффект в четвертом квартиле: большинство учащихся этого квартиля должны были и вне зависимости от введения программы принять участие в экзаменах (о чем свидетельствует средняя доля сдававших экзамены в четвертом квартиле в контрольной группе — 97 % по математике и 98 % по английскому языку) и сдать их (93 % сдавших экзамен по математике и 96 % по английскому языку). Заслуживает внимания также тот факт, что установленный эффект программы применительно к доле принявших участие в экзаменах велик и значим только для первого квартиля, а во втором и третьем квартилях программа главным образом влияет на долю сдавших экзамен и на средний балл тестирования и не влияет на долю принявших участие в тестировании. В третьем квартиле по результатам экзамена по английскому языку отсутствует существенное улучшение какого-либо из трех показателей успеваемости, вероятно, вследствие очень высоких средних показателей в этой группе, которые (и это интересно) гораздо ближе к средним показателям в четвертом квартиле по данному предмету, нежели соответствующие показатели по математике в третьем и четвертом квартилях.

Улучшение всех трех показателей успеваемости в первом квартиле весьма велико по сравнению со средними значениями в контрольной группе: доля принявших участие в экзамене увеличилась на 32 % по математике и на 24 % по английскому языку, доля

успешно сдавших экзамен выросла на 57 % по математике и на 23 % по английскому языку, а средний балл увеличился на 46 % по математике и на 17 % по английскому языку. Прирост по математике значительно выше, чем по английскому языку, но и в последнем случае он все же немалый. Аналогичная картина имеет место во втором квартале. Например, доля сдавших экзамены повышается на 42 %, и на 23 % увеличивается средний балл по математике; по английскому языку эти показатели выросли соответственно на 9 и 3 % во втором квартале.

Важно отметить: тот факт, что в основном программа влияет на учеников, показатели успеваемости которых ниже медианы распределения учеников по способностям, может быть связан со спецификой предусмотренной программой схемы стимулирования. Для ранжирования учителей использовались два количественных показателя успеваемости, а именно: доля сдавших экзамены и средний балл на экзаменах, при этом бóльший вес (выше на 25 %) присваивался первому показателю¹⁷. Интересно не только то, что эффект от программы в основном проявляется на учениках, успеваемость которых ниже медианного значения, но и то, что относительный эффект применительно к доле сдавших экзамены значительно больше, чем применительно к полученному при тестировании баллу. Например, в первом и втором кварталах программа повысила долю сдавших экзамены по математике на 57 и 42 % соответственно, а прирост среднего балла за экзамен для этих двух кварталов составил всего 46 и 23 %. Схожая картина относительного прироста этих показателей имеет место и по английскому языку, хотя разница между двумя результатами меньше. Тот факт, что улучшение результатов наблюдалось в основном у учеников с успеваемостью ниже медианного значения, а также то, что значительно ярче это улучшение проявилось на таком показателе успеваемости, как доля сдавших экзамены успешно, позволяет предположить, что учителя осознали бóльшую денежную выгоду от повышения доли сдавших экзамен по сравнению с повышением среднего балла за тестирование. Кроме этого, данный факт говорит о том, что учителя отдают себе отчет в том, что потенциал улучшения и тех, и других показателей успеваемости значительно выше у учеников, способности которых находятся в области ниже медианы в распределении учеников по способностям¹⁸.

В отношении рассмотренного выше общего эффекта программы также были выполнены оценки для срезов 2000 и 2001 г. (выбор-

¹⁷ Эта схема «дифференциального взвешивания» была рассчитана с учетом увеличения на 25 % количества пунктов за каждое место в рейтинге (с 1-го по 4-е), построенном на основании доли сдавших экзамены. Более подробно вопрос рассмотрен выше в разделе 1.2.

¹⁸ В опубликованном недавно материале [22] приведены аналогичные результаты, свидетельствующие о том, что в Чикаго реакцией учителей на усиление подотчетности стала концентрация времени и ресурсов на учащих, которые с большой вероятностью могут провалить экзамен.



ки разбиты на квартили). Полученные данные приведены в столбцах 2–5 (для математики) и столбцах 7–10 (для английского языка) табл. 5. Эти расчеты с большей очевидностью, чем результаты для полной выборки (столбцы 1 и 6 табл. 5), свидетельствуют, что большая часть прироста результатов по панельным данным, представленным в табл. 3, объясняется положительным и значимым влиянием в срезе за 2001 г., так как оценки по срезу 2000 г. большей частью малы, иногда даже отрицательны и незначимы. Что касается второго квартиля, то здесь влияние на долю сдавших экзамены в 2000 г. составило $-0,005$ (стандартная ошибка $0,046$), тогда как в 2001 г. — $0,139$ (стандартная ошибка $0,070$), и соответствующая оценка по панельным данным составила $0,209$, что не слишком отличается от простой разности результатов по срезам для 2001 и 2000 г. Такая картина характерна и для других результатов, хотя не все оценки по срезам так четко выстраиваются относительно оценок по панельным данным.

Приведенные выше результаты показывают, что изменения в стимулировании учителей повысили долю учащихся, участвовавших в тестировании, долю успешно сдавших экзамены и средний балл за тестирование. Улучшение последних показателей могло происходить двумя путями. Во-первых, вследствие изменения доли сдававших экзамены, которая повышалась, если те ученики, которые в других условиях (если бы программы не было) не сдавали бы экзамен, все-таки принимали в нем участие благодаря программе. Это повысило бы безусловную долю сдавших экзамен успешно (т. е. долю сдавших среди всех учеников, включая тех, которые сдавали экзамен только благодаря программе) «механически», так как часть из этих прибавившихся учеников сдала бы экзамен успешно. Второй механизм повышает условную долю сдавших (т. е. среди тех, кто и без программы сдавал бы экзамены). Оценку влияния программы на условную долю сдавших экзамены и на условный средний балл за тестирование можно выразить следующим способом. Пусть Y_i обозначает результат для школы i и P_i будет доля учащихся школы i , участвующих в экзаменах. Ожидаемый результат для школы i равняется:

$$E(Y_i) = P_i \cdot Y_{i1} + (1 - P_i) \cdot Y_{i0} = P_i \cdot Y_i^1,$$

где Y_i^1 — доля сдавших экзамены среди учащихся школы i , принявших участие в экзаменах; Y_i^0 — соответствующий результат для учащихся той же школы, кто не пошел сдавать экзамены.

Согласно правилам организации соревнования (см. раздел 1), значение 0 присваивается всем учащимся, решившим не сдавать экзамены, поэтому ожидаемый средний результат экзаменов в школе i равняется $E(Y_i) = P_i \cdot Y_i^1$. Средний эффект (average treatment effect, ATE) при условии, что школы были включены в экспериментальную и контрольную группы случайным образом, равняется:

$$ATE = P_1 \cdot Y_1^1 - P_0 \cdot Y_0^0 = (P_1 - P_0) Y_1^1 + (Y_1^1 - Y_0^1) P_0,$$

2.4. Оценивание влияния программы на условную долю сдавших экзамены и на средние баллы

где P_1 и P_0 — доля участвовавших в тестировании; Y_1^1 и Y_0^1 — условные доли сдавших экзамены успешно для экспериментальных и контрольных школ соответственно.

Средний эффект можно было бы также выразить в терминах безусловных результатов: $ATE = Y_1 - Y_0$, где Y_1 — безусловная доля сдавших экзамен, $Y_1 = P_1 \cdot Y_1^1$ и $Y_0 = P_0 \cdot Y_0^1$.

Поэтому общий эффект от программы может быть разложен на средний эффект от повышения доли участвовавших в тестировании $(P_1 - P_0) \cdot Y_0^1$ (безусловный эффект) и условный эффект — $(Y_1^1 - Y_0^1)P_0$. В нижней части табл. 4 приведены соответствующие данные для доли сдавших экзамены успешно и для среднего балла тестирования по английскому языку и математике на основании расчетов, представленных в верхней части таблицы. В составе 8,7% повышения безусловной доли сдавших экзамены по математике 5,2% обусловлено повышением условной доли сдавших экзамен и, следовательно, объясняет 59% безусловного увеличения доли сдавших экзамены. Оставшиеся 3,5% улучшения показателя объясняются увеличением доли участвовавших в сдаче экзамена. Аналогичный анализ, основанный на расчетах по квартилям, дает такой же результат, хотя доля безусловного влияния выше во втором (77%) и третьем (69%) квартилях, поскольку здесь почти отсутствует влияние программы на уровень участия в тестировании, и ниже всего (35%) в первом квартиле. Аналогичное разделение влияния программы на средний балл означает, что в среднем почти половина увеличения роста безусловных результатов (по баллам тестирования) происходит благодаря увеличению условных результатов (по баллам за тестирование). Результаты по квартилям здесь такие же, что и в случае с долей успешно сдавших экзамены: самые высокие значения во втором и третьем квартилях (64 и 56% соответственно) и самое низкое — в первом квартиле (25%).

Разделение безусловного влияния, оказываемого программой на результаты по английскому языку, позволяет предположить, что повышение уровня участия в экзаменах по этому предмету было важнее и оно объясняет 23% общего прироста доли успешно сдавших экзамен. Однако во втором квартиле условная доля сдавших экзамен «вложила» в совокупный прирост 84%. Прирост среднего балла в тестировании по английскому языку связан главным образом с повышением уровня участия в тестировании по данному предмету, но во втором квартиле условное повышение составило половину от совокупного прироста балла за тестирование. Следует еще раз подчеркнуть, что эти связи нельзя рассматривать как причинно-следственные, поскольку участие в тестировании представляет собой эндогенный результат, и поэтому оценки влияния программы на условную долю сдавших экзамен и на средний балл могут отражать смещение. Однако данный метод «разложения» результатов, по всей вероятности, показывает нижнюю границу вклада условных результатов в общее изменение показателей, поскольку успеваемость учащихся, которые благодаря



программе приняли участие в экзаменах, видимо, была ниже, чем успеваемость тех, кто принимал бы участие в тестировании независимо от программы¹⁹.

С целью проверки надежности результатов, основанных на выборке RT, были применены два дополнительных альтернативных метода идентификации эффекта от программы поощрения учителей.

Первый метод — разрыв регрессии (regression discontinuity, RD). Так как отбор школ для включения в программу основывался на пороговом значении одной переменной, вероятность включения в программу изменяется дискретно как функция этой переменной. Дискретность в этом случае характеризуется резким уменьшением (до нуля) вероятности попадания в программу в случае, если превышено пороговое значение 45 % для нерелигиозных еврейских школ и 43 % для религиозных школ и арабских школ²⁰. Такая резкая прерывность использовалась для выделения экспериментальной выборки, в которую вошли школы, расположенные прямо под пороговым значением (до -5 %), а также контрольной группы, которая включала не попавшие под воздействие школы, расположенные прямо над пороговым значением (до +5 %). Временной ряд результатов школьных выпускных экзаменов показывает, что год от года наблюдается флуктуация доли сдавших экзамены по причинам, не зависящим от трендов или в связи с изменениями численности контингента учащихся. Некоторые из этих флуктуаций являются случайными. Поэтому близкие к пороговому значению школы — участники программы могут иметь результаты, схожие с результатами школ, также близких к пороговому значению, но не вовлеченных в программу. Степень схожести зависит от ширины околороговой области. Чтобы размер выборки не сокращать слишком сильно, ширина этой области была ограничена 10 %, так как более широкая область вокруг порогового значения не позволила бы исходить из предположения, что разброс вокруг порогового значения по большому счету случаен (т. е. что школы, расположенные рядом с пороговым значением, случайным образом либо немного превышают это значение, либо не дотягивают до него). Поэтому в нашем случае выбор области вокруг порогового значения, равной 10 %, представляется обоснованным. Главный недостаток данного подхода заключается в том, что он обеспечивает расчет данных только применительно к близким к пороговому значению школам. Однако такая выборка может быть особенно интересной, поскольку «пороговые» школы

2.5. Альтернативные методы идентификации

¹⁹ Декомпозиция безусловного влияния может быть также выполнена и другим способом: с помощью уравнения $ATE = P_1 \cdot Y_1 - P_0 \cdot Y_0 = (P_1 - P_0) Y_0 + (Y_1 - Y_0) P_1$. Однако это уравнение дает верхнюю границу для условных оценок, так как $(Y_1 - Y_0) P_0 < (Y_1 - Y_0) P_1$.

²⁰ Министерство образования решило снизить порог для еврейских религиозных школ и арабских школ, чтобы уменьшить их долю среди школ — участников программы.

Таблица 6

Сравнение экспериментальных и контрольных школ: выборка рандомизированного эксперимента (RT)

	2000 г.			2001 г.		
	Обсле- дуемые	Контроль- ные	Разница	Обсле- дуемые	Контроль- ные	Разница
А. Характеристики школ						
Религиозные школы	0,100	0,301	-0,201 (0,142)	0,095	0,290	-0,195 (0,140)
Арабские школы	0,131	0,000	0,131 (0,094)	0,132	0,000	0,132 (0,096)
Доля сдавших экзамены Bagrut в предыдущем году	0,448	0,495	-0,047 (0,017)	0,458	0,470	-0,012 (0,041)
В. Характеристики учащихся						
Образование отца	11,027	10,219	0,808 (0,591)	10,835	10,081	0,753 (0,643)
Образование матери	11,095	10,526	0,570 (0,659)	11,027	10,527	0,501 (0,711)
Число родных братьев и сестер	2,622	2,288	0,335 (0,352)	2,605	1,902	0,703 (0,383)
Пол (мужчины = 1)	0,493	0,425	0,068 (0,058)	0,499	0,451	0,048 (0,052)
Иммигранты	0,014	0,054	-0,031 (0,021)	0,013	0,009	0,004 (0,007)
Азиатское/африканское эт- ническое происхождение	0,215	0,313	-0,097 (0,052)	0,214	0,273	-0,060 (0,054)
С. Успеваемость учащихся в прошлые годы						
Кредиты, получен- ные по математике	0,185	0,364	-0,180 (0,131)	0,185	0,452	-0,267 (0,128)
Кредиты, полученные по английскому языку	0,207	0,053	0,155 (0,061)	0,183	0,101	0,083 (0,088)
Полное количество кредитов, которое могло быть получено	4,788	4,944	-0,156 (0,476)	5,064	5,346	-0,283 (0,489)
Полное количество по- лученных кредитов	4,008	4,066	-0,058 (0,376)	4,188	4,394	-0,206 (0,384)
Средний балл	61,671	64,548	-2,877 (2,932)	61,797	65,770	-3,973 (1,973)
Количество наблюдений	2,471	1,638	4,109	2,401	1,519	3,920
Количество школ	14	13	27	14	13	27

Примечание: Стандартные ошибки, приведенные в скобках, сгруппированы на уровне школ. Статус школ по национальной и религиозной принадлежности не меняется. Любое изменение средних значений за несколько лет отражает относительное изменение численности учащихся в когорте. Данная таблица основана на выборке по математике.

репрезентативны именно для выборки учебных заведений, на которые с наибольшей вероятностью подействуют программы такого рода.

Всего было 13 не участвовавших в программе школ с долей сдавших выпускные экзамены, лежащей в диапазоне 0,46–0,52, и 14 участвовавших в программе школ с соответствующей долей, находящейся в диапазоне 0,40–0,45 (см. рис. А4 онлайн-приложения)²¹. Диапазон 0,40–0,52, возможно, слишком велик, но значение переменной (средний уровень сдачи выпускных экза-

²¹ На рисунке показано, что имеется одна школа-участник с ошибочной долей сдавших выпускные экзамены в 1999 г., равной 48%, и поэтому ее не следует включать в программу. Эта школа не была включена в выборку RD, так как она не попадала в диапазон выборки (40–45%).



менов) можно контролировать в ходе анализа. Стоит отметить, что наблюдается некоторое пересечение данной выборки с выборкой естественного эксперимента. 11 из 14 школ — участниц программы и 8 из 11 контрольных школ в выборке для метода разрывности регрессии (RD) являются также частью выборки естественного эксперимента. В результате остается только шесть школ (3 контрольные и 3 среди участников), включенных в первую, но не включенных во вторую выборку. Однако в выборке рандомизированного обследования (RT) имеется 17 школ (7 участников и 10 контрольных), не включенных в выборку разрывности регрессии (RD). Это позволяет предположить, что расчеты по каждой выборке дают достаточно высокую прибавку информации.

Табл. 6 повторяет табл. 3 для выборки разрывности регрессии (RD). Различия между участвовавшими в программе и контрольными школами и стандартные ошибки по переменным, отражающим характеристики ученика (колонки 3 и 6), показывают, что две группы были очень схожи в оба года по всем переменным, кроме этнической принадлежности (в 2000 г.) и числа родных братьев и сестер (в 2001 г.). Однако и эти различия почти незначимы. Третья панель в таблице показывает некоторые различия между участвовавшими в программе и контрольными школами в кредитах по математике в прошлые годы и в среднем балле для когорты 2001 г., а также в кредитах по английскому языку в прошлые годы для когорты 2000 г. Однако различия в кредитах за прошлые годы имеют противоположные знаки по математике (отрицательный) и по английскому языку (положительный), и по каждой из этих дисциплин они значимы только для одной из двух когорт.

В табл. 7 представлены результаты оценки уравнений (1) и (2) для выборки RD²². Все оценки эффекта программы положительны и значимо отличаются от нуля для всех средних результатов по английскому языку и математике. Оценки с ограниченным набором контролируемых показателей очень схожи с теми, которые были получены при контроле всех показателей. Обратив внимание на результаты, основанные на расчетах с контролем всех переменных (столбцы 2 и 8), можно увидеть, что они качественно схожи с представленными в соответствующих столбцах табл. 4, но при этом имеется одно важное отличие. Оценки влияния программы на долю участвовавших в тестировании (и для английского языка, и для математики) методом RD выше, чем те, которые были получены на выборке RT. По английскому языку, например, влияние программы на долю принявших участие в тестировании на выборке RD почти на 50% выше по сравнению с оценками, полученными

²² В принципе идентификация на основе метода RD зависит от контроля ошибочно измеренного уровня сдачи выпускных экзаменов, который использовался для включения школ в программу. Однако фиксированные эффекты школ, включенные в каждую регрессию, контролируют ошибочно измеренный уровень сдачи выпускных экзаменов в 1999 г., и поэтому последний показатель из уравнения исключен.



Таблица 7

Оценивание методом «разность разностей» влияния программы на результаты по математике и английскому языку по выборке RD

	Математика						Английский язык					
	Все квартили		Оценивание по квартилям				Все квартили		Оценивание по квартилям			
	Ограни- ченный контроль	Полный контроль всех пере- менных	1-й	2-й	3-й	4-й	Ограни- ченный контроль	Полный контроль всех пере- менных	1-й	2-й	3-й	4-й
Доля уча- ствовавших												
Среднее зна- чение для контроль- ной группы	0,767	0,407	0,775	0,880	0,947	0,826	0,489	0,859	0,939	0,955		
Эффект про- граммы	0,072 (0,034) [0,049]	0,055 (0,029) [0,040]	0,112 (0,068) [0,090]	0,027 (0,051) [0,070]	0,034 (0,037) [0,052]	0,031 (0,024) [0,032]	0,053 (0,022) [0,032]	0,048 (0,018) [0,026]	0,092 (0,066) [0,089]	0,028 (0,025) [0,035]	0,028 (0,013) [0,018]	0,040 (0,017) [0,023]
Доля сдав- ших экзамен												
Среднее зна- чение для контроль- ной группы	0,602	0,248	0,503	0,699	0,889	0,745	0,377	0,725	0,882	0,923		
Эффект про- граммы	0,111 (0,037) [0,052]	0,088 (0,028) [0,040]	0,091 (0,049) [0,064]	0,141 (0,056) [0,077]	0,086 (0,035) [0,049]	0,056 (0,026) [0,036]	0,039 (0,029) [0,041]	0,033 (0,022) [0,031]	0,109 (0,071) [0,096]	0,009 (0,039) [0,055]	-0,041 (0,018) [0,025]	0,036 (0,027) [0,036]
Средний балл												
Среднее зна- чение для контроль- ной группы	51,219	20,388	44,620	60,009	75,259	55,243	30,143	56,915	65,432	69,825		
Эффект про- граммы	6,733 (2,415) [3,437]	5,790 (1,812) [2,555]	4,408 (2,710) [3,711]	4,687 (3,299) [4,636]	6,608 (2,054) [2,845]	6,293 (2,566) [3,536]	2,975 (1,858) [2,642]	2,671 (1,421) [2,000]	4,144 (3,600) [5,014]	-1,530 (2,762) [3,856]	0,709 (1,446) [2,011]	4,419 (1,826) [2,462]
Условный эффект про- граммы												
Доля сдав- ших экзамен	0,040	0,010	0,118	0,055	0,026	-0,011	0,027	-0,017	-0,067	-0,002		
Безусловный эффект, %	46	11	84	64	47	-	25	-	-	-		
Средний балл	1,752	-2,037	2,785	3,968	3,743	-0,712	-1,886	-3,482	-1,295	1,341		
Безусловный эффект, %	30	-	59	60	59	-	-	-	-	30		
Количество	8,029	2,002	1,983	2,032	2,012	8,264	2,065	2,066	2,066	2,067		

Примечание: 1. Стандартные ошибки, приведенные в круглых скобках, сгруппированы на уровне «школа — год». Стандартные ошибки, приведенные в квадратных скобках, даны на уровне школ. 2. В колонках 3–6 и 9–12 эффекты программы меняются по квартилям распределения балла за предыдущие экзамены. Оценивание выполнено на основании четырех отдельных регрессий, по одной для каждого квартиля. 3. Контролируемые характеристики учащихся во всех колонках, кроме 1 и 7, включают набор дамми-переменных: для числа родных братьев и сестер; образования отца и матери; средних значений доли сдавших (по школе) выпускные экзамены год назад; азиатского/африканского происхождения; иммиграционного статуса; пола; количества кредитных единиц, которое могло быть получено; среднего балла, который мог бы быть получен по этим заявленным кредитам; полного количества присужденных кредитных единиц и кредитных единиц, присужденных только по рассматриваемой дисциплине. Во все уравнения включены фиксированные эффекты школ. 4. В колонках 1 и 7 контрольными элементами являются эффекты, фиксированные для школ, уровень сдачи выпускных экзаменов в школе год назад, а также количество кредитных единиц, которое могло быть получено учащимся. 5. Ряд средних значений для контрольной группы показывает среднее значение для учащихся контрольных школ в 2001 г.



с использованием естественного эксперимента. Эта разница обусловлена главным образом положительным эффектом в третьем и четвертом квартилях. Эти эффекты полностью отсутствовали в результатах по выборке RT. С другой стороны, оценки влияния программы на долю успешно сдавших экзамены и балл за тестирование в первом и втором квартилях, рассчитанные с помощью обоих методов, очень схожи.

Для интерпретации разницы результатов, представленных в табл. 4 и табл. 7, важно иметь в виду, что главным концептуальным различием методов рандомизированного обследования (RT) и разрывности регрессии (RD) является то, что последний не контролирует S^* и если бы погрешность измерений отсутствовала, то в дизайне разрывности регрессии (RD) необходимо было бы сравнить еще учеников или школы с разными S^* . Поэтому оба метода могут дать одинаковые результаты, либо если показатель S^* слабо связан с результатами, либо если дисперсия погрешности измерений велика по сравнению с дисперсией S^* вокруг порогового значения (это означает, что точки выше или ниже порогового значения S имеют приблизительно одинаковое значение S^*). Первое условие выполняется, поскольку S^* характеризуется очень малой положительной связью с тремя показателями результатов по каждому предмету. Фактически самая сильная корреляция наблюдается между S^* и долей участвовавших в тестировании по английскому языку (0,043) и по математике (0,069), и для этих показателей действительно наблюдается самая большая разница между результатами, полученными с использованием методов рандомизированного обследования (RT) и разрывности регрессии (RD). Второе условие не выполняется, так как в диапазоне 0,40–0,52 вокруг порогового значения переменной — критерия для включения в программу (S) размеры обеих релевантных дисперсий очень схожи: 0,075 для погрешности измерений и 0,078 для S^* .

Второй метод идентификации эффекта от программы поощрения учителей, использованный с целью проверки надежности результатов, основанных на выборке RT, базируется на оценивании методом «разности разностей». В нем использованы панельные данные (замеры до и после введения программы) по всем школам, которые по двум критериям могли бы участвовать в программе, и он основан на контроле фиксированных эффектов школ для преодоления различий между включенными и не включенными в программу школами, представленными в табл. 1. Оценки эффекта программы, полученные с помощью метода «разности разностей» (DID), приведены в табл. 2А онлайн-приложения. Они положительны и существенно отличаются от нуля, но значительно ниже оценок, полученных с использованием рандомизированного эксперимента (RT). Это позволяет предположить, что они существенно занижены. Указанные различия рассмотрены в онлайн-приложении.

3. Меняются ли педагогические методики и усилия учителей в ответ на финансовое стимулирование?

Данные предыдущего раздела со всей определенностью показывают, что программа стимулирования учителей значительно повысила успеваемость учащихся по английскому языку и математике. Насколько тесно эти улучшения связаны с увеличением усилий, прикладываемых учителями? Свидетельствуют ли полученные результаты о применении новых педагогических методов? Ответив на эти вопросы, можно отчасти снять опасение, состоящее в том, что введение финансового стимулирования учителей в первую очередь будет способствовать натаскиванию на тесты. В таком случае повышение успеваемости свидетельствует лишь о более тщательной подготовке к тестированию, а не о долгосрочном обучении или реальном повышении качества человеческого капитала²³. Чтобы получить ответ на поставленные выше вопросы, были использованы результаты проведенного Министерством образования телефонного опроса учителей английского языка и математики, участвовавших в программе²⁴. Для сравнения аналогичный опрос проводился среди сопоставимого числа учителей английского языка и математики из не участвовавших в программе школ, в которых доля сдавших выпускные экзамены в среднем примерно совпадала со значением соответствующего показателя для попавших в программу школ. Поэтому выборка школ, включенных в данный анализ, не совпадает с выборками рандомизированного эксперимента (RT), метода разрывности регрессии (RD), а также с выборками всех участвовавших и не участвовавших в программе школ. Поскольку число опрошенных учителей из выборки рандомизированного эксперимента (RT) достаточно мало, в настоящем разделе были проанализированы три выборки. Первая включала учителей из выборки школ рандомизированного эксперимента (RT), вторая — учителей школ, не участвовавших в программе по третьему критерию, третья — всех опрошенных учителей. Как будет показано ниже, результаты, полученные по трем выборкам, в целом не противоречат друг другу, однако их следует рассматривать только в качестве предположения.

В таблице А3 электронного приложения представлены данные для сравнения характеристик учителей, участвующих и не участвующих в программе, отдельно для каждой из указанных трех выборок и по каждому предмету. Они свидетельствуют о том, что учителя по обоим предметам во всех трех выборках имели в среднем очень схожие характеристики. В каждой колонке приведены 13 переменных, и во всех колонках имеется максимум по два значимых различия между обследуемыми и контрольными школами. Исключения составляют учителя математики в полной выборке (колонка 2), где

²³ Например, см. [4].

²⁴ Возможно, учителя знали о том, что исследование являлось частью эксперимента по стимулированию, и это повлияло на их ответы. Для минимизации такого хоуторнского типа смещения данный опрос был представлен респондентам как обычное обследование Министерства образования, посвященное выпускным экзаменам и достигнутым результатам, а вопросы, относящиеся к программе, были размещены в конце анкеты.



Таблица 8

Влияние оплаты по результатам работы на методики преподавания и усилия учителей

	Все опрошенные учителя				Учителя школ, не участвовавших в программе				Учителя школ, включенных в рандомизированный эксперимент			
	Учителя математики		Учителя английского языка		Учителя математики		Учителя английского языка		Учителя математики		Учителя английского языка	
	Среднее значение для выборки	Разница между экспериментальными и контрольными школами	Среднее значение для выборки	Разница между экспериментальными и контрольными школами	Среднее значение для выборки	Разница между экспериментальными и контрольными школами	Среднее значение для выборки	Разница между экспериментальными и контрольными школами	Среднее значение для выборки	Разница между экспериментальными и контрольными школами	Среднее значение для выборки	Разница между экспериментальными и контрольными школами
Методики преподавания												
Обучение в маленьких группах	0,661	0,007 (0,051)	0,631	0,085 (0,054)	0,557	0,111 (0,068)	0,574	0,143 (0,073)	0,525	0,193 (0,078)	0,467	0,175 (0,127)
Индивидуальное обучение	0,614	-0,028 (0,060)	0,583	0,112 (0,060)	0,600	-0,014 (0,087)	0,574	0,122 (0,078)	0,600	-0,008 (0,125)	0,633	0,103 (0,109)
Группировка по способностям	0,397	0,130 (0,059)	0,417	0,221 (0,058)	0,471	0,055 (0,073)	0,471	0,168 (0,087)	0,500	-0,035 (0,102)	0,367	0,256 (0,113)
Адаптация методик преподавания к способностям учащихся	0,942	0,011 (0,023)	0,925	0,068 (0,021)	0,914	0,038 (0,037)	0,956	0,037 (0,030)	0,900	0,030 (0,055)	0,933	0,067 (0,057)
Действия учителей												
Дополнительное обучение в течение всего года или перед экзаменом Bagrut	0,831	0,015 (0,036)	0,564	0,223 (0,054)	0,871	-0,025 (0,045)	0,544	0,243 (0,068)	0,825	-0,022 (0,072)	0,567	0,207 (0,098)
Дополнительное обучение только перед экзаменом Bagrut	0,296	0,071 (0,048)	0,207	0,211 (0,057)	0,300	0,067 (0,069)	0,147	0,271 (0,069)	0,150	0,160 (0,082)	0,167	0,192 (0,102)
Количество дополнительных учебных часов в неделю	2,038	1,987 (0,600)	1,144	1,458 (0,440)	2,959	1,066 (0,809)	1,655	0,946 (0,767)	2,382	1,246 (0,965)	1,040	1,148 (0,451)
Увеличение учебных часов по инициативе учителя	0,709	-0,017 (0,051)	0,463	0,098 (0,064)	0,714	-0,022 (0,074)	0,456	0,104 (0,082)	0,625	0,093 (0,105)	0,533	0,089 (0,101)
Дополнительные усилия учителей были нацелены на:												
всех учащихся	0,587	-0,025 (0,059)	0,330	-0,004 (0,064)	0,614	-0,052 (0,079)	0,338	-0,012 (0,086)	0,575	-0,012 (0,106)	0,433	-0,131 (0,098)
слабых учащихся	0,212	-0,058 (0,042)	0,197	0,129 (0,054)	0,214	-0,060 (0,059)	0,176	0,150 (0,068)	0,200	-0,045 (0,066)	0,133	0,225 (0,083)
средних учащихся	0,011	0,043 (0,018)	0,016	0,020 (0,025)	0,029	0,025 (0,025)	0,015	0,021 (0,027)	0,025	0,017 (0,033)	0,000	0,057 (0,055)
сильных учащихся	0,000	0,006 (0,006)	0,000	0,028 (0,013)	0,000	0,006 (0,006)	0,000	0,028 (0,013)	0,000	0,000 (0,000)	0,000	0,000 (0,000)
Число учителей		358		329		239		209		111		83
Число учащихся		109		105		68		64		27		25
Число школ		46		42		46		42		17		15

Примечание: Стандартные погрешности, приведенные в скобках, сгруппированы на уровне школ.

значимы три из 13 различий. В полной выборке и в выборке рандомизированного эксперимента (RT) для учителей английского языка наблюдается дисбаланс между экспериментальными и контрольными школами по доле учителей, родившихся в Израиле, но в выборках школ, участвовавших и не участвовавших в программе, значения этой переменной примерно равны. Как будет показано ниже, результаты по всем трем выборкам очень схожи, и поэтому можно заключить, что они не различаются в зависимости от доли учителей, родившихся в Израиле.

В табл. 8 приведены данные, показывающие влияние программы на три аспекта поведения участвующих в ней учителей: методики преподавания, прилагаемые учителем усилия и фокусирование внимания на работе со слабыми или сильными учащимися. Для интерпретации полученных результатов следует заметить, что подготовка к выпускным экзаменам в конце 12-го класса — это основная задача в учебном плане выпускного класса средней школы. Учащиеся выпускного класса и их учителя завершают обычный учебный год в середине марта, а оставшуюся часть времени посвящают подготовке к выпускным экзаменам. Обычным явлением в этот период является, например, специальный учебный марафон во время уик-эндов, который проводится вне стен школы.

Результаты, представленные отдельно по учителям английского языка и математики, позволяют судить о наличии двух эффектов внедрения программы стимулирования учителей. Программа модифицировала методы преподавания, а также способствовала увеличению прилагаемых учителями усилий, что выражается в дополнительных занятиях после уроков в школе. Такие занятия ведут и учителя, не участвующие в программе, но те, кто участвует, прибегают к ним значительно чаще. Данные результаты более заметны, точнее измерены у учителей английского и схожи во всех трех выборках этой категории учителей.

Доля участвующих в программе учителей английского языка, проводивших занятия в маленьких группах, применявших индивидуальный подход в обучении, разделявших учеников на группы по способностям и адаптировавших обучающие методики с учетом способностей учащихся, выше, чем соответствующая доля среди учителей, не участвующих в программе стимулирования. Наиболее показательным и значимым является различие групп участвующих и не участвующих в программе учителей по размеру доли учителей, выделяющих группы учеников по способностям в классе. Например, среди учителей английского это 64% против 47% в выборке школ, которые могли бы участвовать в программе по двум критериям, и 62% против 37% в выборке RT. Среди учителей математики выявлено только одно значимое различие в методиках обучения: проведение занятий в маленьких группах. В выборке RT этот подход применяли 71,8% учителей, участвовавших в программе, по сравнению с 53,5% учителей из контрольной груп-



пы; схожие результаты наблюдались в выборке всех школ, которые могли бы по двум критериям участвовать в программе.

Чуть больше половины учителей английского языка из группы для сравнения и 77% участвующих в программе учителей в выборке RT сообщили, что на протяжении учебного года они увеличили время для дополнительных занятий. Более 80% участвующих и не участвующих в программе учителей математики также увеличили время, затрачиваемое на занятия с учениками вне обычной учебной нагрузки. Однако ответ на вопрос относительно увеличения учебного времени только на период подготовки к экзаменам — с середины марта до конца июня — выявил существенные различия между экспериментальными и контрольными школами, причем это касалось как математики, так и английского. Результаты по выборке рандомизированного обследования (RT) свидетельствуют об очень больших различиях: 19,2 процентных пункта (16,7% против 35,9%) среди учителей английского языка и 16,0 процентных пунктов (15,0% против 31,0%) среди учителей математики. По обоим предметам эти различия значимо отличаются от нуля. Существенные различия были выявлены среди учителей английского языка в объеме дополнительного учебного времени — более одного часа в неделю. Схожий эффект от участия в программе очевиден и среди учителей математики во всех трех выборках, хотя он измеряется точно только для выборки школ, которые могли бы участвовать в программе по двум критериям²⁵. В табл. 8 также видны некоторые особенности в направлении усилий в работе учителей английского языка, участвующих в программе: 22,5% учителей из выборки RT больше внимания уделяли слабым учащимся, среди не участвующих в программе таких оказалось 13,3%. Что касается учителей математики, среди них такие различия не выявлены.

Помимо демонстрации того, что программа способствовала увеличению прилагаемых учителями усилий и изменению методик преподавания, ее результаты важны еще в одном отношении. Программа улучшила методы преподавания и усилила направления в работе, широко практиковавшиеся учителями и до ее начала. Поэтому достигнутое в результате стимулирования учителей улучшение показателей на выпускных экзаменах по математике и английскому языку с меньшей вероятностью можно расценивать как следствие натаскивания на тестирование — скорее всего, оно отражает реальное накопление человеческого капитала. Однако следует обратить внимание на то, что эксперимент проводился в течение одного года, и если бы он продолжался в течение двух или трех лет, учителя могли бы применить другие, более эффективные, с их точки зрения, способы повысить свои шансы на победу в этом соревновании.

²⁵ Следует заметить, что учителя были осведомлены о том, что программа представляет собой трехлетний эксперимент, и поэтому нельзя игнорировать вероятность того, что изменения в действиях учителей отражали простое перераспределение усилий между текущим и следующим годом.

Представленные выше результаты свидетельствуют о том, что отдельные учителя, а не только группы учителей, могут повлиять на успеваемость учащихся и результативность педагогической работы. Можно ли, используя обычные методики измерения качества работы учителей, прогнозировать, кто из них окажется лучшим? Взаимосвязь между ранжированием учителей в соревновании и их персональными характеристиками можно использовать для того, чтобы выявить хороших учителей даже при наличии определенного «шума» ранжирования, вызванного стохастическими составляющими в успеваемости учащихся. Оценка таких взаимосвязей в отношении учителей математики и английского языка подтверждает точку зрения, что качество преподавания не очень сильно зависит от таких персональных характеристик учителей, как возраст, пол, образование, профессиональная категория (teaching certification) и преподавательский стаж [6]. Ни одна из этих переменных не была значима в объяснении результатов ранжирования учителей. Ни одна из персональных характеристик учителей не имела значимой связи с каким-либо из рассмотренных в предыдущем разделе количественных показателей качества их работы. Однако по другим переменным в регрессиях проявились существенные взаимосвязи. Факт рождения и получения образования за пределами Израиля положительно повлиял на эффективность работы учителей английского языка. Среди учителей английского языка, получивших образование в Израиле, те, кто обучался в самых престижных университетах (в Еврейском университете Иерусалима или Тель-Авивском университете), работали эффективнее, чем обучавшиеся в других университетах или педагогических колледжах. Что касается учителей математики, то единственной персональной характеристикой, существенно повлиявшей на эффективность преподавания, явилось образование матери: учителя, чьи матери окончили среднюю школу или высшее учебное заведение, работали значительно эффективнее по сравнению с другими. Схожее влияние образования отца выявлено не было.

4. Влияет ли стимулирование на принципы оценивания учеников?

Модели стимулирования могут привести к нежелательным смещениям в поведении, поскольку люди пытаются обойти правила [8]. В. Jacob и S. Levitt [11] обобщили свидетельства о манипуляциях учителей с баллами за тестирование в рамках программ, которые усиливают подотчетность учителей и школ, а также факты откровенного мошенничества со стороны учителей и административных работников, завышавших разными способами баллы за тестирование. Модель оплаты по результатам работы, которая изучалась в эксперименте, могла побудить учителей завышать внутришкольный балл, поскольку итоговые баллы на выпускных экзаменах являются взвешенным средним значением двух показателей: балла за внутришкольное тестирование и балла за государственный выпускной экзамен, а результаты ранжирования учителей основывались на итоговых (взвешенных) баллах. При этом



учителя самостоятельно оценивают школьные выпускные экзамены по своему предмету. Однако в бонусной программе предусмотрены четкие санкции в отношении учителей, у которых будет наблюдаться большой разрыв между результатами внутришкольных и государственных экзаменов. Для обеспечения сопоставимости баллов за школьные и государственные экзамены Министерство образования с 1996 г. применяет так называемую модель дифференциального взвешивания, в которой предусмотрен контроль за разрывом между внутришкольными и государственными баллами и установлены правила, определяющие санкции в отношении школ, где были выявлены большие расхождения²⁶. По условиям эксперимента, учителя, нарушающие правила дифференциального взвешивания, исключались из программы и соответственно лишались права на получение бонусных доплат. Однако и в рамках этих правил у учителей сохраняется возможность манипулировать внутришкольными баллами. Например, выбросом считается среднее расхождение в 20 единиц или больше между внутришкольным и национальным баллами, и поэтому учителя вполне могут завышать внутришкольные баллы, не нарушая правил дифференциального взвешивания.

Представленные в настоящем разделе данные свидетельствуют о том, что учителя, участвующие в программе, не манипулировали школьными баллами и не завышали их. Эмпирические результаты основаны на сравнении расхождений между результатами внутришкольных и государственных экзаменов (за каждый экзамен) и на сравнении величины таких расхождений в экспериментальной и контрольной группах школ, причем эти расхождения сопоставлялись в год перед началом программы (2000) и в год реализации программы (2001). Сравнение двух результатов можно рассматривать как естественный эксперимент, поскольку балл за государственный экзамен является объективной, несмещенной оценкой знаний учащихся, а внутришкольный балл может быть искажен из-за мошенничества и других форм манипуляций баллами за тестирование.

В табл. 9 представлены результаты для трех выборок: рандомизированного обследования (первая панель), разрывности регрессии (вторая панель) и всех школ, которые могли бы по двум критериям участвовать в программе (третья панель). Как было упомянуто

²⁶ В документе Министерства образования подробно рассмотрена модель дифференциального взвешивания. Если средний школьный балл за данный экзамен выше среднего балла на государственном экзамене на 20 единиц или более или если он ниже на 10 единиц или более, этот случай рассматривается как выброс. Если вероятность такого события составляет 1:10 000, то веса двух баллов устанавливаются равными 30 и 70% соответственно вместо 50% каждому из них. Если вероятность такого события составляет 1:1 000 000, то обоим баллам присваиваются веса 10 и 90% соответственно. Если выбросы имели место в 8% экзаменов или более, по крайней мере по двум предметам и за два года из трех последовательных лет, в отношении школы принимаются административные меры.

Таблица 9

Влияние программы на принципы оценивания учеников

	Математика						Английский язык					
	2000 г.			2001 г.			2000 г.			2001 г.		
	Участие в программе	Контроль	Разница									
А. Выборка рандомизированной обработки												
Разница внутришкольных и государственных баллов	0,096 (0,068)	0,159 (0,046)	-0,063 (0,080)	0,094 (0,079)	0,021 (0,049)	0,072 (0,092)	-0,013 (0,093)	0,247 (0,204)	-0,260 (0,221)	0,026 (0,076)	0,047 (0,081)	-0,021 (0,110)
Количество	6,646	6,110	12,756	6,606	5,288	11,894	5,368	4,714	10,082	4,840	3,952	8,792
В. Выборка разрывной регрессии												
Разница внутришкольных и государственных баллов	-0,004 (0,061)	0,113 (0,082)	-0,117 (0,100)	-0,038 (0,067)	0,168 (0,085)	-0,206 (0,106)	-0,154 (0,060)	0,122 (0,119)	-0,277 (0,130)	-0,099 (0,060)	0,031 (0,070)	-0,130 (0,090)
Количество	6,374	3,912	10,286	6,130	3,248	9,378	4,672	3,080	7,752	4,264	2,720	6,984
С. Школы-дублиры												
Разница внутришкольных и государственных баллов	-0,017 (0,047)	0,056 (0,035)	-0,073 (0,058)	-0,040 (0,045)	0,032 (0,036)	-0,072 (0,057)	-0,066 (0,055)	0,052 (0,052)	-0,118 (0,075)	-0,097 (0,055)	0,040 (0,037)	-0,137 (0,066)
Количество	14,414	16,130	30,544	13,768	14,102	27,870	12,546	12,286	24,832	11,326	10,006	21,332

Примечание: Стандартные ошибки, приведенные в скобках, сгруппированы на уровне школ. Каждая запись в таблице представляет собой оценку разницы между внутришкольными и государственными баллами на выпускных экзаменах. Данные в столбцах «Разница» представляют собой разницу между соответствующими различиями средних по экспериментальной и контрольной группам, приведенными в двух предыдущих столбцах той же строки.

выше, данные относятся к учебному году, в котором реализовывалась бонусная программа, т. е. к 2001 г., а также к предпрограммному 2000 г., что позволяет провести сопоставление «разность разностей». Каждое значение показывает разницу между внутришкольным баллом и баллом за государственный экзамен. Баллы были стандартизированы — приведены к распределению со средней, равной нулю, и стандартным отклонением, равным единице. Стандартизация была выполнена отдельно для внутришкольного экзамена и государственного экзамена.

Из шести установленных различий в средних баллах по математике во всех трех обследуемых выборках в 2001 и 2000 г. четыре имеют отрицательное значение и два — положительное, но ни одно из них статистически значимо не отличается от нуля. Все различия, рассчитанные для трех контрольных групп, положительные, но только одно из них статистически значимо отличается от нуля. Все, кроме одной, «разности разностей» между средними значениями для экспериментальных и контрольных школ в данном году имеют отрицательные значения, из чего можно заключить, что учителя в экспериментальных школах в среднем недооценивают достижения своих учащихся по математике на внутренних экзаменах



в сравнении с их результатами на государственных экзаменах, а также с успеваемостью учащихся контрольных школ. Однако все эти различия, полученные методом «разность разностей», значимо не отличаются от нуля, за исключением одного случая (колонка 6, выборка RD), и никакой тенденции методом «разности разностей» для 2000 и 2001 г. также не выявлено. Результаты тройного сравнения для обоих годов (колонка 6 минус колонка 3) подтверждают этот результат.

Ситуация с учителями английского языка практически такая же. Поэтому из данных, приведенных в табл. 9, можно заключить, что рассмотренная в настоящей работе модель стимулирования по результатам деятельности не провоцировала учителей искусственно завышать внутришкольные баллы своих учащихся на выпускных экзаменах по математике и английскому языку по сравнению с государственными экзаменами. Возможно, сыграли свою роль санкции, предусмотренные в правилах дифференциального взвешивания, которые были приняты министерством, или угроза исключения из программы, а может быть, учителя оценивали знания своих учащихся, ориентируясь на высокие этические стандарты. Схожесть до- и постпрограммных результатов позволяет предположить, что беспристрастность при оценивании учащихся имела место и до введения бонусной программы.

Растущий интерес к программам стимулирования учителей в США, Европе и в других регионах способствовал появлению множества новых инициатив. В США действуют такие системы надбавок по результатам работы для учителей, как программа QComp²⁷ на сумму \$86 млн в Миннесоте, Pro-Comp²⁸ на сумму \$25 млн в Денвере [21], E-Comp и STAR²⁹ во Флориде. Система государственных школ Чикаго получила в 2006 г. федеральный грант на сумму \$27,5 млн на пилотную программу оплаты труда учителей по результатам их работы; Чикаго стал крупнейшим регионом в стране, внедряющим оплату по результатам³⁰. Несколько отличается от других интересная программа Advanced Placement Incentive Program (APIP) в Далласе (штат Техас), предусматривающая финансовые поощрения и для преподавателей, и для учащихся за каждый балл, полученный на экзамене по математике, естествен-

5. Израильский эксперимент в более широком политическом контексте

²⁷ Программа Q-Comp была предложена губернатором Tim Pawlenty и узаконена в июле 2005 г.

См.: http://education.state.mn.us/MDE/Teacher_Support/QComp/index.html.

²⁸ Анализ пилотного проекта Denver Pro-Comp приведен в [2].

²⁹ Подробнее о E-Comp см.: http://www.fldoe.org/news/2006/2006_04_05/ValueTable.pdf, о STAR см.: <http://www.fldoe.org/PerformancePay/>.

³⁰ Среди других последних программ такого рода следует отметить программы, осуществленные в Толедо, Коламбусе, Хьюстоне, Шарлотт-Мекленбурге и Далласе, а также более десятка экспериментов по оплате учителей в зависимости от результативности их работы, осуществляемых по всей стране. Многие из этих программ финансируются за счет федеральных грантов.

ным наукам и английскому языку [12]³¹. Недавно газета N.Y. Times отметила: «В политических кругах зреет консенсус относительно того, что поощрение учителей бонусами или прибавками к зарплате за повышение успеваемости учащихся... может стимулировать учителей-ветеранов, а также привлечь в профессию учителя подающих надежды новичков»³².

Оплата труда учителей по результатам работы внедрялась и в других странах, например Pay performance and management reform в Великобритании в 2000 г. [1], Victorian government schools agreement 2001 г. в Австралии³³, Carrera magisterial program в Мексике и SNED в Чили [28]. Рандомизированные эксперименты меньшего масштаба были осуществлены в Индии [3; 26] и в Кении [4].

Многие из этих программ обладают уникальными характеристиками, что свидетельствует о многообразных попытках найти оптимальные варианты стимулирования в разных системах образования. Эксперимент, представленный в настоящей публикации, также имеет особенности, которые больше нигде не воспроизводились. Несмотря на это, полученные данные могут представлять интерес и за пределами Израиля, поскольку израильская программа имеет много общего с инициативами по оплате за результативность работы, апробируемыми в других странах. Например, так же как в недавних американских экспериментах, в израильской программе в качестве главного индикатора результата используются достижения учащихся, а не те или иные характеристики учителей, такие как профессиональные знания и преподавательская подготовка. Как и программы в Денвере, E-Comp во Флориде и APIP в Далласе, израильская программа фокусируется на основных учебных предметах. Для оценивания успеваемости учащихся не применялись специальные тесты, разработанные для целей программы, вместо этого за основу были взяты государственные тесты, используемые в настоящее время для измерения достижений учащихся и соответствующие критериям надежности и валидности. Аналогично этому применяемая в Chattanooga программа основана на системе добавленной стоимости, с помощью которой в Теннесси оценивают, насколько повысились учебные достижения учащихся³⁴; программа E-Comp заимствовала применяемую во Флориде систему Comprehensive assessment test (FCAT)

³¹ Подробнее об этой программе см. на сайте: <http://www.collegeboard.com/student/testing/ap/about.html>.

³² Dillon S. Long reviled, merit pay gains among teachers // N.Y Times. Education. 2007. June 18 <http://select.nytimes.com/gst/abstract.html?res=F10715FA395B0C7B8DDDAF0894DF404482>.

³³ Это соглашение было достигнуто в декабре 2001 г. 78% принявших участие в голосовании учителей. Оно подразумевает, что все продвижения учителей будут связаны с улучшением результатов учеников, что проверялось государственным тестированием учеников по математике и английскому языку в 2003 и 2005 г. См. <http://www.wsws.org/articles/2001/jan2001/edu-j11.shtml>.

³⁴ Более подробно о том, как оценивался прогресс учащихся и квалификация учителей, см. <http://mb2.ecs.org/reports/Report.aspx?id=1131>, а также [7].



для измерения успеваемости учащихся; в австралийской программе использована действующая система Achievement improvement monitor (AIM), включающая тестирование по математике и английскому языку³⁵; программа, применяемая в Великобритании, опирается на баллы за тестирование, проводимое в рамках действующей государственной программы. Как в и других похожих программах³⁶, при оценке успеваемости учащихся в израильском исследовании были проконтролированы переменные, которые могут влиять на успеваемость, такие как социально-экономические характеристики учащихся.

В израильской программе используется широко известный метод отбора учителей, подлежащих вознаграждению. Так же как и в программе E-Comp, применяемой во Флориде, в израильской программе учителей математики и чтения ранжировали согласно результатам (очкам), которые они получили в зависимости от успехов их учащихся, т. е. размер бонусов определялся повышением уровня знаний. По аналогии с флоридской программой больше очков присваивалось за те результаты, которые ценятся выше, а шанс достичь их ниже³⁷. Как и в большинстве других программ, полученные по результатам работы вознаграждения являлись единовременными, а не постоянными бонусами, т. е. они не прибавлялись к базовой оплате. Средние бонусы были относительно щедрыми аналогично ряду программ в США, где они составляли до \$20 000 [23]. Например, в рамках программы APIP некоторые учителя получили более \$11 000 к годовому заработку [12]. Как упоминалось в разделе 3, в израильской программе предусматри-

³⁵ Австралийская система мониторинга повышения успеваемости (Australian Achievement Improvement Monitor system, AIM) предусматривает тестирование учащихся по математике и английскому языку в третьем и пятом классе. В 2001 г. ее распространили на седьмой класс — это первый класс средней школы.

³⁶ В Северной Каролине аналогичная программа действовала более пяти лет [13]. Альтернативный подход предусматривает сопоставление абсолютного роста успеваемости и установленного стандарта, как это было в Кентукки в 1990-е годы [14].

³⁷ Основное требование программы E-Comp заключалось в том, что все школьные округа во Флориде должны были определить 10% лучших преподавателей по каждому предмету и выдать им 5%-ю надбавку к зарплате. Учителей математики и чтения ранжировали исключительно с учетом того, насколько их учащиеся повысили свои баллы по сравнению с предыдущим годом. Учителя получали очки, когда их учащиеся переходили с одного уровня знаний на другой. В 2006 г. во Флориде программа E-Comp была заменена аналогичной моделью оплаты по результатам работы, получившей название Special teachers are rewarded (STAR) с годовым бюджетом \$147,5 млн. Эта программа дала возможность окружным школам и чартерным школам осуществлять выплаты согласно положениям раздела 1012.22 Собрания законов штата Флорида и получать свою долю из ассигнованных средств. Подробнее см. http://www.fldoe.org/PerformancePay/pdfs/STAR_SuptMemo.pdf. (Чартерные школы — это начальные или средние школы в США с государственным финансированием (допускаются также и частные субсидии), но освобожденные от некоторых форм регулирования, применяемых в отношении государственных школ, в обмен на предоставление определенной отчетности по некоторым показателям деятельности, предусмотренным их уставом. — Примеч. пер.)

валось несколько уровней бонусов аналогично другим американским программам³⁸. Еще одной общей характеристикой израильской и других программ является то, что все учителя имели возможность получить бонусы, хотя на практике их получила только группа учителей: если вознаграждается слишком много учителей, учителя могут потерять стимул к увеличению прилагаемых усилий.

Социально-политические параметры израильского эксперимента представляют особый интерес и могут стать моделью для других исследований. Например, израильская программа основана на сравнении результатов внутришкольного и государственного тестирования и предусматривает жесткие санкции в случае большого расхождения между ними, как это разъяснено выше в разделе 5. Это снижает вероятность появления таких практик преподавания, которые имеют место в государственных школах Чикаго [11]. Кроме того, израильская программа оставляет за учителями большую часть решений относительно того, как обеспечить выполнение поставленных целей. Похоже, что такая гибкость себя оправдала. Весьма эффективным оказался еще один аспект программы, который поэтому можно рекомендовать и для других программ: использование нескольких показателей для измерения результативности работы учителей и для «сигнализирования» учителям о том, какие изменения в их работе и успеваемости учеников наиболее желательны. Полученные результаты свидетельствуют о том, что учителя реагировали на особенности модели стимулирования, и в первую очередь на сигналы о дифференцированном поощрении за разные результаты. Главное, что удалось вынести из апробирования этой программы, заключается в том, что с помощью стратегических денежных сигналов в рамках систем стимулирования учителей в школе можно побудить их наращивать свои усилия в направлении, желательном для тех, кто занимается социальной политикой. Данный вывод может способствовать повышению эффективности программ стимулирования учителей.

6. Заключение

Данные проведенного исследования подтверждают, что оплата по результатам может приводить к появлению у учителей тех же интересов, что и у школьной системы в целом, не вызывая при этом искаженных форм поведения, таких как манипулирование баллами за тестирование или практика натаскивания учеников на тесты. Этот результат очевиден, несмотря на бытующее мнение о коллективном характере обучения в школе, т. е. о том, что успехи учащихся являются результатом работы не одного учителя, а группы учителей. Размер установленного влияния оплаты по результатам на успеваемость учеников и вывод о больших усилиях учителей, которые участвовали в программе, позволяют заключить,

³⁸ Израильская программа предусматривала четыре уровня бонусов, а например, Cincinnati and Charlotte-Mecklenburg program — два уровня вознаграждений [23].



что стимулирование учителей можно рассматривать как многообещающий способ повышения качества школьного образования³⁹. Полученные данные по индивидуальному стимулированию учителей и имевшиеся ранее примеры того, насколько эффективно стимулирование групп учителей [15], важны в контексте социальной политики по вознаграждению учителей и качеству школьного образования. Однако к представленным результатам следует относиться с осторожностью, поскольку эксперимент продолжался всего один год и поэтому не позволил изучить воздействие программы по стимулированию учителей на другие когорты учеников, а также выявить ее долгосрочный эффект. Можно предположить, например, что в результате проведения таких программ школы со временем станут более привлекательными для учителей, способных повышать успеваемость своих учеников [20]. В то же время в долгосрочной перспективе программа такого рода может оказаться и контрпродуктивной. В рамках настоящего исследования оценка потенциальных позитивных и негативных долгосрочных эффектов не предусматривалась. Это может быть предметом будущих исследований.

Неслучайное включение школ и учителей в данный эксперимент повлекло за собой альтернативные способы выявления контрольной и экспериментальной групп. Естественный эксперимент, являющийся следствием погрешности измерения переменной, в зависимости от значения которой школы включались или не включались в программу, оказался привлекательным решением проблемы неслучайного включения школ в программу стимулирования. Однако большая схожесть результатов оценивания с использованием такого «условного» случайного распределения на группы (RT) и метода разрывности регрессии (RD) позволяет предположить, что метод разрывности регрессии работал так же хорошо, тогда как метод «разности разностей», применявшийся на всех включенных в программу школах, давал смещенные результаты.

В настоящем исследовании показано, что индивидуальное стимулирование учителей сработало в двух направлениях: побудило большее число учащихся участвовать в сдаче выпускных экзаменов, чем было бы без программы, и повысило количество сдавших экзамены и средний балл за экзамены среди учащихся, которые стали бы участвовать в экзаменах независимо от программы. Хотя представляется, что учителям легче манипулировать первым показателем (числом сдававших экзамен), тот факт, что по обоим дисциплинам и безусловный, и условный (т. е. среди тех, кто сдавал бы экзамен вне зависимости от введения программы) уровень

³⁹ В публикации [18] представлены убедительные данные об издержках и выгодах от использования программы стимулирования учителей как в сравнении с другими программами, так и с точки зрения расходов в расчете на одного учащегося по сравнению с возможными экономическими выгодами от повышения результативности обучения.

успешного прохождения экзаменов и средний балл за экзамены повысились, позволяет предположить, что улучшилось также качество знаний и навыков тех учащихся, которые участвовали в сдаче экзаменов благодаря программе. Это подтверждается тем, что повышение безусловного уровня успешно сдавших экзамены по математике примерно наполовину (а по английскому языку — примерно на четверть) обусловлено более высоким уровнем участия в сдаче экзаменов.

На основе сравнения участвовавших и не участвовавших в программе учителей были найдены свидетельства связи улучшения успеваемости учеников по математике и английскому языку с изменениями в методах и педагогических методиках и объеме дополнительных усилий учителей, участвовавших в программе.

Например, участвующие в программе учителя значительно чаще применяли такие подходы, как обучение в маленьких группах и выделение групп учеников по способностям. Учителя активизировали также и обычную в их среде практику выделения дополнительного учебного времени в течение 4-месячного периода для подготовки учащихся к выпускным экзаменам.

Структура израильской системы выпускных экзаменов, основанная на обязательном тестировании по окончании средней школы и определенном количестве кредитов, необходимых для сдачи экзаменов, очень близка к системам, действующим во Франции, Германии, Италии, Нью-Йорке, Массачусетсе (США) и в других регионах. Формат израильской программы стимулирования учителей имеет также много общего с программами оплаты по результатам, апробируемыми в США и во многих других странах. Соответственно полученные результаты и уроки, извлеченные в ходе данного эксперимента, релевантны для многих систем образования в Европе и США.

Литература

1. Atkinson A. et al. (2004) Evaluating the impact of performance-related pay for teachers in England/CMPO DP № 04/113, Bristol University, England.
2. Community training and assistance center (2004) Catalyst for Change, Boston.
3. Duflo E., Rema H. (2006) Monitoring works: Getting teachers to come to school. Department of Economics and Poverty Action Lab, Massachusetts Institute of Technology.
4. Glewwe P., Nauman I., Kremer M. (2003) Teacher Incentives/NBER WP No. W9671.
5. Green J., Stokey N.L. (1983) A comparison of tournaments and contracts // Journal of Political Economy. Vol. 91. P. 349–364.
6. Hanushek E. (2002) Publicly provided education / NBER Working Paper No. 8799.
7. Handley C., Kronley R.A. (2006) Challenging myths: The Benwood initiative and education reform in Hamilton County. Kronley & Associates.



8. Holmstrom B., Milgrom P. (1991) Multitask principal-agent analysis: Incentive contracts, asset ownership and job design // *Journal of Law, Economics and Organization*. Vol. 7. P. 24–52.
9. Israel Ministry of Education, High School Division (2000) Individual teacher bonuses based on student performance: Pilot program. December, Jerusalem (Hebrew).
10. Israel Ministry of Education, Bagrut Test Data (2000) Jerusalem: Ministry of Education, Chief Scientist's Office, April 2001.
11. Jacob B., Levitt S. (2003) Rotten apples: An investigation of the prevalence and predictors of teacher cheating // *Quarterly Journal of Economics*. Vol. 118. No. 3. P. 848–878.
12. Jackson C. K. (2007) A little now for a lot later: A look at a Texas advanced placement incentive program. Harvard University, Department of Economics.
13. Kelley C., Heneman H., Milanowski A. (2002) Teacher motivation and school-based performance awards // *Education Administration Quarterly*. Vol. 38. P. 372–401.
14. Kelley C., Conley S., Kimball S. (2000) Payment for results: The effects of the Kentucky and Maryland Group-based performance award programs // *Peabody Journal of Education*. Vol. 75. P. 159–199.
15. Lavy V. (2002) Evaluating the effect of teachers' group performance incentives on students achievements // *Journal of Political Economy*. Vol. 10. No. 6. P. 1286–1318.
16. Lavy V. (2004) Performance pay and teachers' effort, productivity and grading ethics/NBER Working paper No. 10622.
17. Lavy V. (2007) Using performance-based pay to improve the quality of teachers // *The Future of Children*. Spring. P. 87–110.
18. Lavy V., Schlosser A. (2005) Targeted remedial education for under-performing teenagers: Costs and benefits // *Journal of Labor Economics*. Vol. 23. No. 4. P. 839–874.
19. Lazear E., Rosen S. (1981) Rank-order tournaments as optimum labor contracts // *Journal of Political Economy*. Vol. 89. P. 841–864.
20. Lazear E. (2003) Teacher incentives // *Swedish Economic Policy Review*. Vol. 10. P. 179–214.
21. Mitchell N. (2005) Denver teachers opt for merit pay // *Rocky Mountain*. December 29.
22. Neal D., Schanzenbach D. W. (2007) Left behind by design: Proficiency counts and test-based accountability/NBER Working Paper. No. 13293.
23. Odden A., Wallace M. (2007) Rewarding teacher excellence. Wisconsin Center for Education Research, University of Wisconsin, Madison, Wisconsin.
24. Prendergast C. (1999) The provision of incentives in firms // *Journal of Economic Literature*. Vol. 37. P. 7–63.
25. Sadowski C. (2006) Huston district OKs teacher merit pay plan. July 23 www.heartland.org/article.emf?artId=18637.
26. Sundararaman V., Muralidharan K. (2008) Teacher performance pay: Experimental evidence from India. Presented at the Ninth Neemrana Conference on the Indian Economy, January.

27. Wakelyn D. J. (1996). The politics of compensation reform: A Colorado case study/American Educational Finance Association conference paper, March 23.
28. Vegas E., Ed. (2005) Incentives to improve teaching lessons from Latin America. The World Bank, Washington, DC, USA.

Автор выражает особую благодарность Alex Levkov, Yannay Spitzer, Roy Mill и Katherine Eyal за ценную научно-исследовательскую помощь, а также Josh Angrist, Abihijit Banerjee, Eric Battistin, Esther Duflo, Caroline M. Hoxby, Andrea Ichino, Hessel Oosterbeek, Yona Rubinstein и участникам семинара European University Institute (EUI), Массачусетского технологического института, Еврейского университета Иерусалима, Принстонского университета и Тель-Авивского университета, The Tinbergen Institute⁴⁰ и Labor Studies Summer Institute of National Bureau of Economic Research⁴¹ за полезное обсуждение и комментарии и, наконец, издателю и рецензентам — за полезные комментарии. Программу стимулирования учителей 2001 г. финансировало Министерство образования Израиля, административное управление осуществлял отдел среднего образования. Автор признателен также Falk Institute for Economic Research в Израиле за предоставленное финансирование. Выводы, изложенные в настоящей публикации, принадлежат исключительно автору, они не были рекомендованы спонсорами программы или финансирующими организациями. Настоящая публикация является существенно пересмотренной версией препринта NBER (Working paper 10622).

⁴⁰ The Tinbergen Institute — научно-исследовательский центр в Нидерландах, один из ведущих в Европе в области экономики и финансов, названный в честь экономиста Jan Tinbergen <http://www.tinbergen.nl> — *Примеч. пер.*

⁴¹ National Bureau of Economic Research (NBER) — Национальное бюро экономических исследований, частная некоммерческая исследовательская организация США. — *Примеч. пер.*