

Педагогическое тестирование*

Джон РАВЕН

* Печатается с разрешения «Когито-Центра». Исключительные права на публикацию книги на русском языке принадлежат издательству «Когито-Центр».

Доктор Джон Равен — известный британский психолог, в сферу научных интересов которого входят проблемы ценностей, природы и развития компетентности, эффективности образования. Он разработал новую концептуальную схему измерения системы качеств и способностей, входящих в состав компетентности. Соавтор и продолжатель дела своего отца — Дж. К. Равена-старшего, разработчика всемирно известного теста «Прогрессивные матрицы Равена». Среди написанных им книг отметим: «Образование, ценности и общество» (1977); «Компетентность в современном обществе» (1984); «Управление образованием для эффективного школьного обучения» (1994); «Новое богатство наций» (1995).

Предлагаем вниманию читателей фрагмент из книги Дж. Равена «Педагогическое тестирование: проблемы, заблуждения, перспективы», в которой рассматривается ограниченность доминирующей в образовании и психологии измерительной парадигмы, а также то негативное влияние, которое она оказывает на образование. Система оценивания, существующая за рубежом, имеющая свою историю и большой опыт проведения тестирования, уже подвергается критике за необъективность и низкую прогностическую валидность. Пройдя длительный путь развития, она обнаружила столько просчётов и ошибок, что готова к трансформации и доработке. Сегодня, когда мы только приступаем к централизованному и, по существу, массовому тестированию, весьма полезно узнать, как относятся к нему наши зарубежные коллеги.

Рассмотрим причины того, почему людям, связанным с проблемами образования, принципиально важно найти способы выявления таких индивидуальных качеств, как инициатива и желание участвовать в общественной жизни, — качеств, ускользающих от психометристов на протяжении всей истории психологии? Мы затронем здесь семь таких причин, которые можно объединить в четыре группы: 1) то, чему уделяется основное внимание в школах и что служит предметом дискуссий о политике в сфере образования, в действительности определяется самими показателями тестов учебных достижений; 2) учителя и ученики нуждаются в адекватных методических средствах для контроля за своим продвижением к поставленным целям и успешностью их достижения; 3) принятым способам оценивания учебной деятельности недостаёт конструктивной и прогностической валидности, и они не позволяют нам исключить случаи, когда некомпетентные люди с неадекватными амбициями занимают влиятельное положение в обществе; 4) современные процедуры педагогического измерения необъективны.

1. Содержание выпускных экзаменов и квалификационных испытаний при приёме на работу обуславливает происходящее в школе и последовательно вытесняет из неё всё, что достойно называться образованием.

В исследовании, которое проведено в Англии, Шотландии, Северной Ирландии, Ирландии, Бельгии и США, ученикам, их учителям, родителям, выпускникам школ и работодателям предлагали оценить важность различных потенциальных целей образования.

Полученные данные интересны во многих отношениях, однако здесь мы обсудим только три их аспекта.

По мнению мальчиков, наиболее важные цели обучения включают развитие уверенности и инициативы, необходимые для реформаторской деятельности; формирование независимо-

сти и способности применять знания для решения новых проблем; воспитание личности и характера. Эти данные бросают вызов широко распространённому мнению, что школу не должны волновать социальные изменения. Похоже, что непосредственные потребители услуг школы (ученики) хотят, чтобы она в большей степени содействовала выработке именно таких качеств, в развитии которых большинство педагогов не хочет принимать участия.

Ученики считают очень важной помощь школы в выборе профессиональной карьеры. Когда мы спросили их, должна ли школа делать больше для достижения этих целей, чем делает сейчас, оказалось, что помощью в выборе карьеры школа пренебрегает более всего. Теряется масса времени, возникают тяжёлые стрессы, пока наконец в своих скитаниях на рынке труда люди не найдут ту нишу, которая удовлетворит их личные интересы и склонности. Поступление на работу для многих становится, по существу, первой возможностью определить и *развить* свой истинный потенциал. Однако школьники (как и многие исследователи) ошибочно формулируют эту проблему как отсутствие помощи в *выборе карьеры*. На самом деле в любой профессиональной группе есть множество людей с абсолютно разными мотивами и способностями, которые занимаются самыми разными видами деятельности. Так, среди психологов есть те, кто руководит компаниями, и те, кто занят исследовательской работой, а также преподаватели, редакторы журналов, политики и секретари.

Далее исследования показали, что хотя мальчики и отводят успешной сдаче экзаменов второе-третье место, они в то же время и помещают такие ценности, как более широкое знание предмета и получение внепрограммных знаний, в конец списка своих приоритетов. Факт такого расхождения в сочетании с другими данными — такими, как беспокойство, которое они испытывают на информативных и интересных, но не предусмотренных программой уроках, даёт основание считать, что, заявляя о желании получить хорошее образование, они всё же имеют в виду высокие отметки. При этом содержание изучаемого предмета не так важно.

Исследования убедительно показывают: мальчики-подростки хотят, чтобы школа помогла им реализовать себя, помогла выявить и развить их индивидуальные способности и обеспечила признание этих способностей; при этом они не считают таким уж важным *содержание того*, что называется академическим образованием. Говоря, что школы должны сосредоточиться на академических целях, они имеют в виду, что хотят получить сертификаты об образовании, которые обеспечат им доступ к работе, а эти сертификаты могут быть получены только через так называемые академические занятия (впрочем, очень редко заслуживающие такого определения).

Следующий вопрос очевиден: разделяют ли это мнение мальчиков-подростков девочки, а также родители, учителя и работодатели? Данные опроса девочек-подростков показывают, что девочки отвечают на только что сформулированный вопрос дружным «да».

У нас нет данных по опросам родителей тех детей, которые участвовали в этих исследованиях, но есть результаты более ранних исследований и опросов родителей, проведённых в Великобритании и США, которые показали, что, несмотря на некоторые важные различия, ответы родителей и детей похожи.

Были выявлены, однако, существенные расхождения между данными, полученными от детей и родителей, с одной стороны, и от учителей и директоров школ — с другой.

Учителя отмечают, что это *единственная* цель, которой, как полагает большинство педагогов, они добиваются довольно успешно.

Из этих данных, полученных как от учеников, так и от учителей, следует, что те качества, которые оцениваются на выпускных экзаменах и при приёме на работу, как бы предопределяют всё происходящее в школе. Таким образом, если мы хотим влиять на содержание и формы обучения в школе, нам нужно прежде всего попытаться изменить процедуры аттестации и приёма на работу.

Но хотим ли мы, чтобы школы сосредоточились на других целях? Мы видели, что школьники, а также их родители и учителя считают именно так (при условии, что эти изменения не помешают ребятам в будущем найти работу). Правы ли они?

В ряде исследований (анализировавших обширную литературу по данной проблеме) мы

показали, что такое мнение действительно справедливо. Именно те качества, развить которые обычно планируют в школе, отличают наиболее успешно работающих механиков, водителей автобусов, рабочих, продавцов, инженеров, учёных, армейских офицеров, дипломатов и политиков.

Они правы не только в распределении своих приоритетов; они правы и в том, что школа пренебрегает такими целями и поэтому не обеспечивает их достижения. Многочисленные исследования, подтверждающие такой вывод, использовали методы внутришкольного наблюдения, изучали образовательную практику либо анализировали результаты обучения. Они проводились как в Великобритании, так и в США.

Есть множество причин для столь очевидного теперь расхождения между желаемым и действительным. Лишь две из них, однако, напрямую касаются обсуждаемого здесь предмета. (Для того, кто заинтересован в поиске путей переориентации школы на достижение её главных целей, полезно было бы рассмотреть и остальные причины этого расхождения: они изложены в моей книге «Самая важная проблема образования — соответствие ценностям» («The most Important Problem in Education is to Come to Terms with Values»).

Одна из причин, по которым школа редко способствует развитию инициативы, готовности к сотрудничеству, стремления к взаимопониманию и влияния на окружающую жизнь состоит в том, что эти качества приобретаются людьми только на практике в процессе интересной или важной для них деятельности. К тому же нет соответствующих оценочных процедур, способных помочь учителю в определении индивидуальных интересов, приоритетов и компетентностей его учеников, в планировании индивидуального плана развития для каждого ученика и в контроле за возможными последствиями таких педагогических воздействий.

Другая важная причина пренебрежения этими целями состоит в том, что нет механизма поощрения этих качеств у ребят при их школьной аттестации и при поступлении выпускников на работу. Учителя, как правило, не получают признания за развитие у своих учеников этих качеств. Они говорят, что ухудшили бы жизненные перспективы учеников, если бы уделяли больше времени формированию этих качеств в ущерб академическим занятиям, тем самым уменьшив их шансы получить высокие баллы по тестам достижений.

Очень важно понять все последствия установленных нами фактов: во-первых, школы не достигают главных целей *образования*, во-вторых, содержание их деятельности определяется в основном критериями оценки при сертификации и приёме на работу. Эти факты говорят о том, что основной функцией школы является не образование *как таковое*, а, по выражению Jencks, «легитимизация табели привилегий». Другими словами, не сертификация определяет происходящее в школе, но то, чему она служит, а именно: распределение социальных положений и статусов.

Наши наблюдения и весь предыдущий опыт в целом показывают, что учителя всё равно будут «натаскивать» в решении тестов и стремиться к достижению тех целей, по которым даётся оценка, независимо от того, кто это оценивание проводит.

Многие люди становятся учителями, потому что хотят помочь детям взрослеть, потому что им нравится общаться с ними и они хотят делать для общества что-то стоящее. Мысль о том, что их *главная* задача сводится к ранжированию жизненных возможностей учеников, звучит для них кощунственно. Моё предложение состоит в том, чтобы ввести в процедуры школьной аттестации и отбора при приёме на работу измерение тех самых качеств, развитию которых и должна служить система образования. Мы получили бы тогда в сфере образования «паровой двигатель», по-прежнему использующий энергию социальных сил, но уже под контролем педагогов.

2. Принятая система тестирования ограничивает содержание и круг проблем, рассматриваемых при выработке политики в сфере образования.

Использование тестов, сконструированных в соответствии с традиционным критерием, согласно которому «хороший» тест должен давать «нормальное» распределение, имеет ещё и побочный результат. Он состоит в том, что такие тесты крайне *неэффективны* для оценки

качества системы образования в целом или какой-либо её части (подсистемы) с точки зрения достижения стоящих перед ней важных целей. Невозможно ведь получить нормальное распределение по всей популяции, если на практике лишь немногие учителя развивают названные способности.

В результате лишь единицы из имеющихся в настоящее время тестов могут быть использованы для того, чтобы показать, способствует или нет данная система обучения или данный конкретный учитель появлению большого числа учеников с разнообразными типами одарённости. Чтобы сделать это, необходимо отказаться от нынешних представлений о различиях между учениками и использовать тесты, способные продемонстрировать, в чём данный конкретный ученик особенно силён, а вовсе не то, силён ли он в решении задач, отобранных учителем или внешним экспертом без учёта его способностей.

Но это не единственная причина, по которой использование тестов, предполагающих «нормальное» распределение результатов, ведёт к недооценке важнейших проблем образования. Принятие этих тестов в качестве адекватных методов измерения *результатов* обучения побуждает исследователей использовать многофакторные схемы анализа. При таком подходе за аксиому принимается допущение, что прилагаемые *педагогические усилия*, равно как и результаты обучения, могут существенно различаться *количественно*. Следовательно, можно будет ожидать желательных (или нежелательных) результатов в одних классах и их отсутствие в других. Следование статистическому (строго научному) стандартному многофакторному исследовательскому плану, освобождает специалиста, оценивающего процесс обучения, от обязанности ясно формулировать предполагаемую им цепочку причинно-следственных связей между воздействием, например, финансовой поддержки образования и результатами, к которым это финансирование должно привести, — таким, например, как улучшение показателей школьников по тестам. По этой причине оценочные исследования редко сообщают о том, как можно улучшить педагогический процесс. Ещё меньше они способны показать, как перейти от того, что *есть*, к тому, что *должно быть*.

Скажем ещё об одном важном последствии, к которому ведёт неспособность общепринятых процедур оценивания выявлять существенные характеристики педагогического процесса и особенности учрежденческого и социального контекста, в котором этот процесс происходит. Исследователи часто проходят мимо того, что конкретные нововведения «проваливаются» не столько из-за неверно избранных средств и методов, сколько из-за недостаточного внимания к тому влиянию, которое они оказывают на другие аспекты системы. Например, довольно часто приходится иметь дело с проектами по совершенствованию учебных программ, в которых группы учителей обучаются неким «новым методам». Эти учителя, возвращаясь работать в свои школы, не предпринимают никаких шагов, чтобы другие педагоги, родители и ученики поняли смысл и новизну. Необходимые пособия и оборудование чаще всего отсутствуют, не делается ничего, чтобы помочь учителям с максимальной отдачей осуществить эти нововведения.

На самом же деле никому не интересны результаты нововведений, если они были плохо организованы, если они не были обеспечены соответствующей поддержкой. Такие результаты будут иметь слишком малое значение как для отдельного ученика, так и для общества. Гораздо важнее узнать о положительном или отрицательном влиянии педагогического нововведения в долгосрочной перспективе (вроде экономического и социального развития общества). Не менее интересно также понять, какие изменения — собственно педагогические и внешние, ситуационные — необходимы для получения желательных и предотвращения нежелательных результатов.

Излишняя сосредоточенность исследователей, оценивающих содержание и методы обучения только на узком спектре результатов (а именно на тех, которые проверяются «надёжными и валидными» тестами), часто мешает им глубоко изучить те формы педагогической деятельности, которые не оказывают существенного влияния на контролируемые результаты. Это особенно часто бывает в тех случаях, когда сам процесс обучения рассматривается как нечто неосознаваемое и не поддающееся измерению. К примерам такого рода относится боль-

шинство программ и методов обучения, разработанных для обеспечения личностного роста школьников (и часто достигающих поставленной цели). Это справедливо для большинства образовательных программ, направленных на духовное или нравственное развитие, а также на выявление и развитие скрытых способностей и одарённости. Такие изменения личности, как правило, происходят в результате взаимодействия с людьми, которые преуспевают в том, в чём сам испытуемый хотел бы добиться успеха, или с людьми, на которых он хотел бы быть похожим. (Такие ролевые модели могут быть представлены как на примере конкретных людей, так и с помощью различных рассказов или притч.) Эти изменения также могут возникать в результате стрессовых или экстремальных ситуаций. Некоторые из тех, кого мы интервьюировали, отмечали, что такого рода изменения были вызваны у них встречей с профессионалом более высокого класса, чем они сами, в то время, когда они были заняты очень важной для них деятельностью. Или источником таких изменений может быть эмоциональный опыт, связанный с переживанием восторга и изумления, охватывающих, например, при проникновении в суть устройства сложных организмов.

Ряд других процессов, которые игнорируются исследованиями в рамках факторно-аналитического подхода, связан с естественным стремлением человека активно выбирать и конструировать среду, в которой он живёт и работает.

Многие исследователи заявляют, что такие процессы не поддаются научному изучению. Это привело некоторых из них к заключению, что и деятельность в сфере образования, и её оценка — это искусство, в котором науке (по крайней мере, позитивизму) делать особенно нечего. Обе эти позиции свидетельствуют о глубоко неверном понимании существа науки. Вся история науки — это отчёт о процессе превращения невидимого в видимое: Ампер использовал магнитную стрелку, чтобы сделать видимым электричество.

Принятие «надёжных и валидных» тестов как единственной правомерной основы для научной оценки в образовании порождено таким представлением о науке, которое недостаточно ориентировано на развитие новых концепций, новых способов мышления, новых путей формализации интуитивного знания. В свою очередь ошибочное представление о науке порождает нежелание властей финансировать (независимо от величины расходов) работу, направленную на углублённое понимание содержания и методов обучения и воспитания.

Подводя итоги, подчеркнём, что применение традиционных тестов, считающихся отличными инструментами по таким критериям, как различительная способность, внутренняя согласованность, надёжность и валидность, привело к неадекватным схемам проверки и оценивания, к дезориентирующим исследовательским выводам, к социально и нравственно несправедливым решениям в области образовательной политики и практики, к распространению неверных взглядов на исследовательский процесс и науку в целом. Похоже, вместо того чтобы судить о совершенстве этих тестов, нам надо оценить их по тому, помогают ли они людям обнаруживать и развивать свои таланты, получать их признание; способствуют ли новым взглядам на проблемы образования; помогают ли понять, что работает, а что нет — и почему; позволяют ли учителю выявлять способности и таланты каждого из его учеников, способствуют ли их развитию и признанию.

3. Существующая система оценивания не распознаёт талантов учеников, тормозя этим их развитие, сужая их возможности, а также лишая общество отдачи от их потенциальных достижений.

Факторно-аналитические исследования, которые ведут психологи с начала века, усилили тенденцию учителей рассуждать и анализировать в терминах «общей способности», а не многообразия «способностей» (множества талантов). По общему мнению, имеющиеся тесты в целом позволяют различить «способных» и «менее способных» учеников. Недовольство же тем, что существующие тесты не только не позволяют распознать способности учеников, но и активно подавляют их развитие, не слишком распространено. Здесь я хочу показать, что для такого недовольства есть существенные основания.

В ходе нашего расследования мы обнаружили не так много учителей, которые были в

состоянии помочь своим ученикам обнаружить и развить важные для них способности. Работу этих учителей отличало то, что они продолжали настойчиво работать с теми учениками, которые подозрительно долго не справлялись с заданиями. Но чтобы сделать этот процесс эффективным, им приходилось формировать у своих учеников — у каждого в отдельности и у группы в целом — соответствующие представления, необходимые для анализа самых разных способностей, а также создавать развивающую образовательную среду, в которой эти способности могли бы обнаруживаться, развиваться и проявляться на пользу всей группы. Вот несколько примеров способностей и компетентностей, которые удалось обнаружить и поддерживать таким энтузиастам: способность убеждать других людей, острая наблюдательность в отношении биофизических процессов, людей или общественных явлений, способность выносить разумные суждения, способность снижать напряжённость в группе.

В исследованиях, авторы которых изучали, что происходит с людьми в начале их трудовой деятельности, также можно обнаружить свидетельства того, что используемые школами тесты измеряют лишь крохотную часть тех способностей, которые школы могли бы выявлять, поддерживать и оценивать, и, соответственно, внимание педагогов сосредоточено лишь на этой части. Vachman с коллегами показали: 80% опрошенных ими молодых людей считают, что смогли бы выявить и развить свои способности на работе (по сравнению с 13% в школе), притом опыт работы привёл к наиболее значительному личностному росту и росту их компетентности за всё время жизни этих молодых людей. Мы обнаружили, что один из наиболее привлекательных аспектов работы для молодых людей состоит в том, что она позволяет им выполнять то, с чем они успешно справляются, а не только одинаковые для всех рутинные задания, обязательные в школе, в которых они вряд ли могли проявить себя с лучшей стороны.

Вывод очевиден: человеческая одарённость чрезвычайно разнообразна, и это разнообразие слабо связано с характером способностей, необходимых для успешной учёбы в школе. Мы, таким образом, несправедливы в отношении большинства школьников из-за неумения помочь им выявить и развить их способности. Цена этому огромна. Имеющиеся в настоящее время тесты не только не поощряют использование учителями обучающих программ, развивающих различные таланты, но отвращают их от этого, заставляя натаскивать учеников на получение высоких баллов, способных обеспечить дальнейшее обучение в вузах или поступление на хорошую работу. Если мы хотим решить эти проблемы, если мы понимаем, что должны помочь школам и учителям достичь подлинных *целей образования*, сделать так, чтобы школы развивали таланты наших детей, если мы хотим дать признание этим талантам, нам нужна совершенно иная модель измерения способностей.

4. Конструктивная валидность: большинство тестов не в состоянии измерить ни «академические», ни «интеллектуальные» способности.

Инспектора школ в Шотландии показали, что большая часть школьного дня заполнена скучной, неувлекательной, рутинной деятельностью, которая редко включает анализ, оценивание, суждение, критическое мышление, столкновение разных мнений, общение и редко ведёт к новому пониманию исторических, литературных, научных или социальных вопросов, не говоря уже о постановке и решении новых проблем. Практически не наблюдается сколько-нибудь ощутимой поддержки учеников в развитии их индивидуальных способностей. У детей очень мало возможностей поразмышлять над теми или иными вопросами, планировать, изобретать, проверять результаты, сотрудничать с другими, формировать собственные представления о функционировании общества и попытаться как-то повлиять на него. У детей недостаточно возможностей даже для написания объёмных текстов, не говоря уж о текстах, которые долго перерабатываются и переписываются как результат серьёзной работы.

Обучение языку часто сводится к подчёркиванию слов в предложениях и заучиванию сформулированных учителем правил грамматики. Арифметика в основном ограничивается механическим применением правил. Все эти операции, конечно, измеряются и поощряются используемыми в настоящее время тестами. Те же учителя, чья успешная работа описана в книге «Открытая начальная школа», отмечают другие навыки: способность отбирать из

множества предметов те, что действительно необходимы для достижения цели; способность привлекать прочитанное для размышления над проблемой, которую ученик пытается решить, даже если оно прямо с ней не связано; способность использовать намёк, метафору, жестикляцию для передачи сообщения — ничего этого не выявляют традиционные тесты.

Не выявляют они и овладение стратегиями, необходимыми для правильного выполнения арифметических действий, если вдруг забывается результат механического заучивания (вроде $7 \times 9 = 63$), а он забывается, в этом нет сомнения, и, как показывают исследования, забывается очень быстро, если не повторяется ежедневно. И наоборот, многие ученики оказываются не в состоянии применить даже известное им действие умножения $7 \times 9 = 63$, чтобы подсчитать в магазине, сколько будут стоить 7 предметов, если цена каждого 9 центов. Таким образом, большинство имеющихся тестов вовсе не измеряет способности к какой-либо содержательной академической, интеллектуальной и даже практической деятельности, а следовательно, лишено валидности. Они вовлекают учителей и учеников в такую деятельность, которая не позволяет считать сами школы академическими или профессиональными институтами, а происходящую в их стенах деятельность — интеллектуальной или практической.

Вывод о том, что в большинстве случаев тесты не позволяют измерять ни академическую, ни интеллектуальную, ни практическую компетентность, подкрепляется результатами анализа, показавшего, что эти тесты измеряют только сиюминутное владение небольшими и произвольно отобранными фрагментами информации низшего уровня, выхваченными из обширной системы знаний, сложившейся в соответствующих научных дисциплинах. Показано, что 50% запоминаемой школьниками информации забывается уже через год, до 80% — через два года. Элементарные знания, достаточные для успешного прохождения школьного тестирования, иногда тут же устаревают. Они не накапливаются и даже при прочном запоминании вряд ли понадобятся когда-нибудь в будущем. Когда людям требуется какое-то знание, им, как правило, нужна специальная и современная информация, которой по большей части ещё не существовало, когда они учились в школе. В одном из наших исследований половина участвовавших в интервью представителей среднего класса сказали, что образование (т.е. экзаменационные результаты) помогло им получить работу, и лишь 13% из той же группы отметили, что оно помогло им развить необходимые для этой работы способности.

Ответы респондентов из рабочей среды были, ещё менее позитивны. Их «образование» *не позволило* им даже получить хорошую работу. За пределами заданий на чтение, письмо и счёт конструктивная валидность имеющихся на сегодняшний день тестов крайне низка. Нет никакого смысла в том, чтобы выдавать сиюминутное знание небольшого числа устаревших научных фактов за валидный показатель научного образования: такая подмена создаёт впечатление, что достигнутое знание в некотором смысле составляет репрезентативную часть целой области научного знания. Ещё в меньшей мере баллы по подобным тестам могут считаться показателями способностей научно мыслить, быть на современном уровне в конкретной специальной области, собрать информацию по интересующей проблеме или провести наблюдения, помогающие решить эту проблему. Фактически дела обстоят даже ещё хуже, потому что существует по меньшей мере 12 различных типов выдающихся учёных (не говоря уж об учёных вообще), обладающих разными интересами и типами компетентности. Более того, ни один из них не относится к числу людей, получивших высокие баллы по традиционным тестам достижений. Таким образом, совершенно неверно считать используемые тесты средством оценки научных знаний, ещё меньше их можно считать средством оценки способностей к научной деятельности.

Многие традиционные тесты достижений отличаются не только низкой надёжностью, но и слабой различительной способностью: 60% разброса оценок в классах, где все дети учатся по единой программе, зависит от экзаменатора, и только 40% — от различий между учениками. Абсолютная разница в сырых баллах между тестовыми оценками классов ступеней А и Д (классами высокого и низкого уровней) обычно не превышает 8 пунктов. Большая часть этой незначительной разницы может быть отнесена за счёт различий в оформлении работы, а не за счёт различий в знании предмета. Wolf удалось показать следующее:

1) «достижения» при выполнении конкретных операций очень сильно зависят от того, как задаётся вопрос или как подаётся задание (так, например, число учеников, способных выполнить простое деление, существенно снижается, если вместо привычного « $112:7=?$ » написать выражение « $112/7=?$ », не говоря уже о представлении задачи в форме реалистичной жизненной проблемы);

2) среди экзаменаторов и проверяющих часто нет согласия даже в том, следует ли отнести данный ответ к верхней или нижней половине шкалы оценок, не говоря уже о более тонком ранжировании;

3) некоторые экзаменаторы в зависимости от ситуации нередко выставляют разные баллы за одно и то же качество работы;

4) на экзаменах ученики нередко выполняют одно и то же задание по-разному в зависимости от ситуации.

Из этого Wolf делает вывод, что оцениваемые конструкторы не дают оснований для сколько-нибудь обобщённых заключений и что надёжность, устанавливаемая повторным тестированием, чрезвычайно низка, даже если тесты обладают высокой внутренней согласованностью. Следовательно, эти тесты не только не измеряют никакой академической или интеллектуальной способности в собственном смысле этого слова, но не обладают достаточной надёжностью и конструктивной валидностью даже в качестве показателей различий между учениками по уровню их знаний или умений. Дальнейшее повсеместное использование этих тестов для оценивания учеников ничем не оправдано. То же самое можно сказать и относительно оценки образовательных институтов, учебных программ и самих педагогов. Ошибочно полагать, что это обеспечивает потребителям возможность выбора услуг в сфере образования и контроля за их качеством и, в свою очередь, способствует улучшению этих услуг. Такая оценка не может быть полезной и при отборе лиц, подходящих для выполнения важных общественных функций.

5. Нынешняя практика оценивания не обладает прогностической валидностью.

Если конструктивная валидность академических тестов исследовалась нечасто, то этого нельзя сказать о прогностической валидности.

Несмотря на то что работодатели достаточно часто используют результаты тестов достижений для отбора персонала, проведённый супругами Hunter анализ результатов около 90 исследований показывает, что прогностическая валидность этих тестов даже по отношению к успешности профессионального обучения не превышает 0,2. Данный показатель не намного (хотя и значимо) меньше прогностической валидности тестов интеллекта. Но этим, однако, проблема не исчерпывается, поскольку корреляция между школьными и более поздними «достижениями в обучении» — даже по одному и тому же предмету — падает до нуля, если соответствующие курсы изучаются не сразу после окончания школы и особенно если изучаемые курсы оказываются действительно необходимыми ученику для его работы.

В то время как выводы исследователей в целом подтверждают, что по достижениям в обучении можно предсказать профессиональный успех, подлинное объяснение полученных результатов состоит в том, что работодатели, *принимая на работу* выпускников школ и университетов, предоставляют обладателям более высоких баллов возможность более быстрой карьеры (и, соответственно, возможность более качественной профессиональной подготовки), а также более высокие и лучше оплачиваемые должности. Если сделать поправку на эти обстоятельства, мнимая прогностическая валидность тестов достижений исчезает вовсе.

Исследования показали, что, несмотря на явную связь между жизненным успехом и *уровнем* образования, баллы, полученные при окончании любой ступени обучения, не обладают прогностической валидностью. Если школьники не имеют характеристик, свидетельствующих о способности продолжать обучение, его фактическое продолжение не улучшает перспектив их карьеры. Все вместе эти факты приводят к выводу, что тестовые оценки не только не предсказывают жизненного успеха в какой-нибудь социально значимой области, но что система образования как таковая вносит очень незначительный вклад в формирование у

людей определённых компетентностей.

Но каким образом традиционные тесты всё же *могли бы* давать надёжный прогноз профессионального успеха? Любая профессиональная группа состоит из множества людей, выполняющих совершенно разные функции. Профессиональные группы следует рассматривать не как психологический, а как социальный феномен. Одна из задач профессионального сообщества состоит в контроле над межгрупповой конкуренцией. Те сообщества, которые допустили бы дефицит «квалифицированных» новых работников, потеряли бы престиж. Требования при приёме на работу или в учебное заведение повышаются с увеличением числа соискателей, имеющих высокие баллы, и это происходит вне зависимости от изменений условий труда или характера решаемых данной группой задач.

60% изменений социального статуса (как в сторону повышения, так и понижения) в США и Шотландии могут быть объяснены (в статистическом смысле) на основе IQ. Если признать, что тесты на IQ измеряют фактор, существенно влияющий на оценки академических достижений, то как можно соотнести эти данные с приведёнными выше? Частично это связано с уже отмеченным фактом, что *уровень* образования не имеет решающего значения. Действительно, дети, закончившие своё формальное образование в разных возрастах, различаются по многим характеристикам, выходящим за рамки IQ. Они различаются своими ценностями, компетентностями, которые они имеют и которые хотели бы обрести; различны их взгляды и намерения в жизни, их планы относительно работы и учёбы, жизненные приоритеты.

Различия между ценностями и показателями IQ в разных социально-экономических группах вовсе не определяются (как думают многие) преимущественно фактором происхождения. Не являются они и результатом профессионального опыта. Они формируются в результате ещё плохо изученного нами процесса социализации, ориентированного на будущее. В процессе социализации молодые люди формируют ценности и способы поведения, характерные для тех групп, в которые они позднее войдут. Это справедливо даже для представлений, отношений и ценностей, которые подростки никогда не обсуждали, а также поведения, которого они никогда не наблюдали и никогда не смогут наблюдать. (Из этого следует, что надо иметь гораздо больше уважения к людям, компетентно делающим нечто отличное от того, что мы сами высоко ценим.)

Другая часть объяснения состоит в том, что учебные оценки, несмотря на их чрезвычайно низкую прогностическую валидность, — единственное социально приемлемое средство контроля за приобретением специальности. Надо искать пути расширения принципов отбора, чтобы учесть дополнительные качества, которые кажутся нам необходимыми. Мы могли бы не просто платить бешеные деньги за систему образования, которая предоставляет нашим детям лишь примитивную опеку и способы распределения привилегий, но управлять этой системой, в то время как сегодня она заполняет вакансии на основе неадекватных (в лучшем случае) и, что точнее, социально деструктивных критериев.

6. Существующая система оценивания не позволяет подбирать достойных кандидатов на социально значимые должности.

В предыдущем разделе было высказано предположение, что традиционные тесты достижений плохо соответствуют своему назначению — задаче профессионального отбора. При этом мы лишь слегка коснулись того, что на самом деле отбираем на влиятельные позиции в обществе не тех людей, каких следует. Мы сталкивались с менеджерами, которые, стремясь ускорить собственную карьеру, пытались продемонстрировать, что они могут эффективно управлять подразделением в условиях «снижения затрат». На деле они просто развалили работу в своём учреждении, снизив потенциал развития своих подразделений и своих подчинённых, не говоря уже о прямом ущербе образу жизни и настроению людей.

Некоторые школьники, пользуясь существующей системой и заискивая перед учителями, заранее готовили и репетировали правильные ответы для экзамена. Интересно, однако, что хотя многие дети считают, что и они «могли бы» или «должны были бы» так поступать, чтобы «успевать», большинство из них всё же отказывается от такого поведения, считая его без-

нравственным. Наше исследование, проведённое среди взрослых, показывает, что слишком многие из тех, кто так поступают, занимают затем влиятельное положение в обществе. При отборе и приёме на работу нужно руководствоваться качествами, необходимыми для успешного и качественного выполнения работы. Повышение по службе не должно быть наградой ни за хорошо выполненную работу в другой области, ни тем более за персональные услуги кому-то находящемуся у власти. Чтобы изменить существующий порядок вещей, необходимо разработать комплекс методов оценивания, однако их было бы трудно внедрить в практику не только потому, что они сильно отличались бы от применяемых сегодня, но также потому, что они представляли бы особую опасность для людей, ориентированных на личную власть и собственное продвижение.

7. Существующие методы оценивания необъективны.

Наконец, наиболее распространённое возражение против любых изменений в методах оценивания состоит в том, что предлагаемые альтернативы «субъективны». Однако и существующие оценочные процедуры, по существу, необъективны. Как может быть объективной оценка ученика или служащего, если она игнорирует наиболее значимые интересы и приоритетные ценности этого человека, те способности, которые могли бы у него развиваться, следуй он им? Как может быть объективной экспертиза или оценка обучающей программы, если она игнорирует важнейшие вредные последствия её внедрения? Тем не менее именно такие — в высшей степени избирательные — экспертные заключения нам, как правило, и предлагают с оговоркой, что их односторонность и обеспечивает объективность предоставляемой информации.

На самом же деле субъективные моменты содержатся во *всех* экспертных заключениях, но их прячут за фасадом псевдонаучности. Выбор того, что конкретно оценивается и включается в отчёты, *существенно* зависит от предпочтений, ценностей и целей тех, кто создаёт и отбирает используемые тесты, а не от качеств людей или программ, подвергающихся экспертизе. Экспертиза и оценка, которые не зависят от ценностного аспекта, как и свободное от ценностей образование, — это взаимоисключающие понятия, то, чего быть не может. Невозможна не только внеценностная экспертиза, но и диагностика: допущение, что тестирование должно быть внеценностным, само основано на определённых ценностях. Оно опирается на субъективные критерии выбора подходящих случаю методик. При таком подходе тестирование оставляет ценности обследуемого человека скрытыми, пусть даже это делается не намеренно. Чем скорее этот факт будет осознан и *учтён для изменения наших представлений о тестировании, оценках и образовании*, тем скорее мы сможем продвинуться вперёд.

Однако ещё более важным ограничением объективности результатов тестирования является то, что проявляемые человеком качества существенно зависят от обстоятельств, в которых он находится, или от возможностей, которые у него были для развития оцениваемых качеств. Довольно трудно, скажем, найти примеры проявления инициативы школьников на уроках латинского языка. Следовательно, выбор урока латыни в качестве контекста для оценки инициативы даёт основание заключить, *какого рода* инициативу ценит сам проводящий тестирование специалист. В результате выводы об инициативе на уроке латыни не обладают прогностической валидностью не только по отношению к другим ситуациям, но они вообще не отражают способности человека проявлять инициативу. Это никак не лишает содержательности понятие инициативы, это просто показывает, что её проявление и признание зависят от ценностных установок как оцениваемого, так и оценивающего.

Таким образом, настаивая на стандартизации тестовых заданий и ситуаций, мы, может быть, и добиваемся воспроизводимости результатов на уровне поведения, но уж, безусловно, не достигаем объективной оценки компетентности. Даже если аттестуемый заинтересуется самими заданиями или теми преимуществами, которые даёт их успешное решение, ситуация тестирования может воспрепятствовать проявлению настойчивости, убеждённости и организованности, которые совершенно необходимы для реализации любой инициативы. Подобным же образом различия между кандидатами нередко больше говорят нам о возможно-

стях, которые они в прошлом имели для проявления инициативы в определённых ситуациях, чем о том, что они объективно способны делать в подобных ситуациях или что они обычно делают в важных для них ситуациях.

Соображения, приведённые в последнем разделе, составляют настоящий обвинительный приговор любому тестированию ввиду его необъективности. Действительно, процедуры тестирования не позволяют учитывать:

— контекст, в котором ведётся обследование, и характер поведения, которое данные условия обычно стимулируют либо подавляют;

— обстоятельства, которые в прошлом могли влиять на развитие тех или иных способностей аттестуемого;

— реальные возможности аттестуемого при условии предоставления ему достаточного времени для работы над заданием, которое его действительно интересует;

— разновидности ситуаций, которые могут затронуть ценности, мотивы и способности аттестуемого.

Очень немногие из используемых и претендующих на объективность измерительных процедур выявляют такого рода сведения. Типичное оправдание, что результат тестирования показывает только то, как ученик выполнил конкретный тест за отведённое время, по сути только перекладывает ответственность за несправедливость, порождаемую оцениванием, с разработчика теста на его пользователя. Но это отнюдь не придаёт тестированию большей объективности. Тем более что у пользователя нет ни информации, ни опыта, чтобы понять, что было упущено при оценивании и интерпретации оценок. Не имеет он также и информации о социальных последствиях полученных оценок, в то время как профессиональная ассоциация диагностов могла бы всю эту информацию предоставить. Отказ от обязанности придавать гласности факт ограниченности принятых оценочных процедур и того потенциального ущерба, который они способны нанести, мы считаем крайне безнравственным.

Выводы

Итак, существует, по крайней мере, семь веских причин, по которым разработка альтернативной измерительной парадигмы в образовании жизненно необходима. Играя такую важную роль при распределении должностей и социальных позиций, современная система оценивания отвлекает школы от их истинных целей, сталкивает дискуссии по политике в сфере образования на обсуждение тривиальных вопросов и приговаривает людей с ценными для общества способностями к убогому и нищенскому существованию, хотя школы вполне могли бы содействовать их развитию. Такого рода оценки необъективны, лишены конструктивной и прогностической валидности. Они не помогают нам отбирать кандидатов на влиятельные посты в обществе, не исключают допуск к этим должностям людей неподходящих. Не помогают эти оценки и учителям при диагностике трудностей в обучении детей, при определении необходимых мер для их преодоления и при руководстве индивидуальными обучающими программами, ориентированными на развитие компетентностей. Не помогают они также политикам и менеджерам в создании более эффективной системы образования и управлении ею, не говоря уже о том, чтобы способствовать развитию разных образовательных систем. Принятая система держит пользователей тестовых методов в плену ложных представлений об оценке, а педагогов-исследователей в целом — в плену ложных представлений о науке. Они ставят психологов, работающих в системе образования, в двусмысленное положение. И поскольку в долгосрочной перспективе последствия использования этих оценочных процедур деструктивны как для отдельных людей, так и для общества в целом, такие методы можно рассматривать только как безнравственные.