

ВВЕДЕНИЕ: ТЕСТЫ МОГУТ БЫТЬ ПОЛЕЗНЫ И... ВРЕДНЫ

К тестам и тестированию отношение у разных людей разное. Да и у одного и того же человека это отношение зависит от ситуации, а также от того, что именно тестируется – знания и умения или личностные качества. Например, вряд ли кто в наше время будет оспаривать тесты на знание правил дорожного движения. Выбор из предлагаемых вариантов в данном случае вполне оправдан, ведь вариантов у водителя на перекрестке на самом деле немного: стоять либо двигаться – вперед, направо, налево или на разворот... Да и во многих других видах деятельности вариантов возможного поведения вовсе не так много, как кажется. Во всяком случае правильных способов поведения гораздо меньше, чем ошибочных. Так что тесты на знания и умения решать задачи более чем естественны в эпоху мощных автоматизированных систем управления транспортными потоками, потоками людей, потоками информации, потоками денег.... Они позволяют допускать к работе только подготовленных людей, а неподготовленных отправлять на доучивание, то есть не допускать в данный момент к ответственной деятельности (работе), связанной с риском для жизни и здоровья людей.

Рассмотрим для начала отношение к тестам с позиции испытуемого – того, кто выполняет тест. Одно дело – пройти тест “для себя”, чтобы получить только самому какие-то сведения о себе. Таких возможностей с появлением различных Интернет-сайтов все больше у каждого человека. Как правило, на этих сайтах достаточно лишь весьма условной регистрации (можно вводить какие-то псевдонимы, ники и т.п.) или ее вовсе не требуется. В этих случаях само по себе прохождение теста не вызывает острой тревоги, а скорее провоцирует любопытство, поэтому каждый день такие тесты “для самопознания и развлечения” в Интернете проходят буквально десятки или даже сотни тысяч людей только в России, а по планете в целом – это миллионы. Совсем другое дело, когда результаты тестирования начинают на что-то влиять вне зависимости от желаний испытуемого. В этом случае тест превращается в инструмент... экзамена. А кто же любит экзамены? Точнее, экзаменуемые не любят такие экзамены, когда возникает риск “провалиться” – показать низкий результат и ухудшить для себя жизненную ситуацию. Но ведь если тестирование организовано грамотно, «без дураков», то есть, без сознательных послаблений со стороны экзаменующих, то риск “провалиться” всегда возникает, поэтому настоящее тестирование, как и настоящий экзамен (с привлечением независимых и строгих экзаменаторов), всегда вызывает тревогу, а нередко – скрытый и даже открытый протест¹.

¹ Чтобы прочувствовать этот протест, предлагаю читателю простейший социальный эксперимент: предложить своим коллегам с помощью не шуточного, а достаточно серьезного

В этой книге мы будем не раз рассматривать вопрос об отношении к тестам и к самой ситуации тестирования со стороны испытуемых, но главный адресат книги – это тот человек, который проводит тестирование. Назовем его для простоты “тестолог”. Чаще всего в роли “тестолога” в разных организациях выступают разные специалисты, которые сами себя тестологами не считают. Это педагоги, врачи, психологи, управляющий и вспомогательный персонал (в частности, специалисты по человеческим ресурсам – “эйчары”) и т.п. В западной тестологической литературе тех, кто проводит тестирование, называют, как правило, не “тестологами”, а “пользователями” или “администраторами” тестов. Но мы чаще будем использовать термин “тестолог”, или «менеджер-тестолог», хотя бы для того, чтобы читателю было все более и более понятно само название данной книги. Одним из ключевых принципов для нас является требование к квалификации проводящего тестирование – этот пользователь тестов должен быть профессионально подготовленным тестологом, иначе возникает масса ошибок в применении тестов.

Так вот у самих потенциальных “тестологов” (у тех, кому предписывается роль организаторов процесса тестирования) отношение к тестам тоже часто оказывается разным. Далекое не все врачи и учителя относятся к тестам позитивно, а если их самих отчуждают от проверки результатов, то большинство относятся к тестам просто негативно (автор этих строк проводил массовые опросы учителей об их отношении к ЕГЭ и другим тестовым проектам). Да и среди психологов очень распространенным является предубеждение против тестов: мол, тесты не позволяют учесть все тонкости человеческой уникальной индивидуальности, поэтому являются несправедливым и негуманным инструментом.²

Уже здесь, во введении мы сформулируем тезис, в котором очень важно разобраться с самого начала. Это тезис о том, что тесты могут быть И ПОЛЕЗНЫ, И ВРЕДНЫ.

Тезис 1. Тесты могут быть и полезны, и вредны в разных ситуациях, но... различие этих ситуаций осложняется ролевой позицией испытуемого (тестируемого) и тестирующего.

теста проверить уровень знаний иностранного языка (или своим однокурсникам) и пронаблюдать за реакцией, при этом результаты каждого предложить сделать публичным достоянием. – *Прим. автора.*

² Правда, на фоне этого пафосного осуждения тестов с позиции Высокого Гуманизма редко ставится вопрос о том, а в какой мере два наугад взятых независимых друг от друга гуманиста ОДИНАКОВО учитывают “уникальную человеческую индивидуальность” в конкретном случае, или каждый все-таки по-своему? Но сейчас мы отодвинем на время в сторону этот вопрос – о согласованности экспертных оценок в неформализованных и нестандартизированных оценочных процедурах. По крайней мере, это вопрос не для введения. – *Прим. автора.*

Мне представляется принципиальным найти читателей этой книги не только среди тех, кто считает, что тесты полезны, но и среди тех, кто уверен в обратном и относит себя скорее к противникам тестов. Понимаю, как это непросто, но мне очень хотелось бы с самого начала найти взаимопонимание если уж не с яростными противниками тестов и тестирования, то хотя бы с теми критически мыслящими читателями, которые видят в этом методе, как и я сам, не только достоинства, но и недостатки. А для этого я собираюсь детально описать такие ситуации, в которых тесты действительно ВРЕДНЫ.

Итак, всякий метод имеет свои сильные стороны и слабости, так и метод тестов. Написав несколько лет назад статью “Тест как оружие” (Шмелев, 2004), я хотел предельно ясно высветить утверждение: “Тесты – это острый инструмент, и он может быть опасным, когда применяется неправильно”. Впрочем, оружие – это возможно даже избыточная гипербола. Возьмем обычный бытовой колющий или режущий инструмент – такой как ножницы или обыкновенная иголка. И этими инструментами можно больно порезаться и уколоться, если держать неправильно (обратным концом), или применять неумело. Тесты привлекают кажущейся простотой неопытных специалистов и всевозможных дилетантов: начинающих и просто слабых педагогов, неуверенных в себе врачей, застенчивых психологов, малосодержательных журналистов, плохих чиновников, привыкших скорее имитировать управленческую деятельность, чем осуществлять ее по-настоящему. Многие любители тестирования подобного рода имеют крайне поверхностные представления о тестах и процедурах тестирования и не знают о существовании науки тестологии, которую необходимо изучать. Поэтому они совершают грубейшие ошибки, которые, в свою очередь, приводят к массовой дискредитации самой идеи тестирования. В настоящее время не существует никаких особых квалификационных требований, которые ограничивали бы или пресекали непрофессиональное и неграмотное использование тестов. Остается лишь надеяться, что когда-нибудь эта ситуация изменится, и для проведения массового тестирования, влияющего на судьбы людей, в нашей стране потребуется особый «сертификат тестолога» (чтобы понять, почему диплом о высшем психологическом образовании ныне не может заменить такой сертификат, надо прочесть эту книгу).

Полезно сравнить позицию “тестолога”, проводящего тестирование, с позицией... парикмахера-брадобрея. Может ли неумелый парикмахер порезать своего клиента, сидящего в кресле? – Может. Это происходило в истории парикмахерского дела, наверное, много тысяч раз. Но почему же против опасной бритвы и ножниц не так много протестующих, как против тестов? Я имею в виду протестующих среди исполнителей. – Вот в этом надо бы разобраться, это надо бы понять.

А разница в этих инструментах в том, что тест – это стандартизированный интеллектуальный инструмент, в то время как ножницы и опасная бритва – это “ручные

инструменты”.³ “Ручные инструменты”, как правило, повышают требования к квалификации мастера (исполнителя). Поэтому сами мастера любят именно нестандартизированные инструменты: они подчеркивают их незаменимость, их ценность на рынке. Не всем можно доверить опасную бритву, не всем можно доверить стричь и брить клиентов. В то же время стандартизированный инструмент нередко воспринимается профессионалами как своеобразная «безопасная электробритва», которой бреют себя сами миллионы (наверное, миллиарды) людей. Разве могут ценить электробритвы мастера-парикмахеры? Поэтому-то профессиональные психологи видят в тестах инструмент, который снижает планку профессионализма. Поэтому они выше ставят методы беседы и наблюдения, или уж такие методики психологической диагностики, которые тестами можно назвать с определенной натяжкой – это так называемые “проективные тесты”, результаты которых подлежат экспертной (ручной, точнее “глазной”), а не автоматизированной обработке. (Впрочем, про проективные тесты мы еще поговорим подробнее, но не во введении).

Итак, стоит сразу же подчеркнуть, что “тестология” – наука о стандартизированных измерительных тестах, а не вообще обо всех возможных тестах. Кстати, само слово “тест” дословно переводится на русский очень просто – это “испытание”, поэтому в общем случае можно говорить о том, что существуют нестандартизированные и неизмерительные тесты. Но тестология их не изучает, или, скажем точнее, изучает не в первую очередь. Она пытается для всяких неизмерительных тестов придумать емкое и общепринятое название, но... оно все еще не утвердилось среди самих тестологов. Иногда о «других тестах» говорят как о «развлекательных» тестах, иногда их называют «авторские» (имея в виду, что ключи к этим тестам автор базирует на собственной интуиции и не проверяет статистически, так как не умеет или не хочет этого делать), иногда говорят про «развивающие тесты самопознания». Есть и такие словечки, как «салонные тесты», «пляжные тесты», «эвристические тесты» (тесты, в основе которых лежат нечеткие, то есть эвристические правила и алгоритмы). Я предпочитаю в этой книге использовать термин «спекулятивные». Однако в этом введении мы говорим не о недостатках «спекулятивных» тестов, но о пользе или вреде от «измерительных тестов» – как раз от тех, которые являются предметом тестологии.

Разумеется, не вполне корректно называть «спекулятивными» обширный класс образовательных тестов, которые в настоящее время в массовом порядке производят вузы и техникумы в рамках своих программ по созданию ФОС – «фондов оценочных средств». Крайне редко тесты в рамках этих фондов достигают уровня психометрических инструментов (измерительных средств). Вероятно, в контексте

³ Мы дальше будем говорить, конечно, не о “ручных инструментах”, а об “экспертных процедурах”, но пока достаточно понятной будет эта аналогия.

тестологии для этих тестов требуется ввести особый термин. Один из вариантов – «контрольно-оценочные тесты» (ибо они входят в состав КОС – «контрольно-оценочных средств»). При этом мы имеем в виду, что не всякая «контролирующая процедура» является измерительной. Сможет ли тестология в ближайшем будущем содействовать доведению КОС до уровня измерительных инструментов? – Это во многом зависит не только от баланса численности сторонников и противников тестовых технологий как таковых, но и от качества самих создаваемых тестов.

Сколько раз мне доводилось слушать от потенциальных пользователей или испытуемых такие суждения о тестах (сразу после знакомства с конкретными тестами): «И кто такие глупые задания придумывает?», «Если вдуматься, то в любом задании невозможно найти ни одного правильного ответа!», «Ваш скоростной тест ставит в невыгодное положение более глубоких людей, которые думают долго, но в конце дают более точные и адекватные решения», и тому подобное! Но когда начинаешь анализировать, чем отличаются люди, которые больше критикуют тесты, от людей, которые спокойно их выполняют, то выясняется, что критики – это чаще всего либо такие испытуемые, которые очень тревожатся (не без оснований), что их реальный невысокий уровень знаний будет обнаружен, либо такие специалисты по оценке (в частности, преподаватели-экзаменаторы), которые боятся потерять часть своего заработка от внедрения тестов.

В этой книге мы подробнее будем обсуждать, какие еще бывают оценочные процедуры, какие складываются взаимоотношения между ними и тестами в разных ситуациях. А во введении мы приближаемся к другому важнейшему тезису, который также считаем необходимым зафиксировать уже сейчас.

Тезис 2. Стандартизированные тесты по-разному оцениваются теми, кто по своей квалификации может дать высококачественную оценку человека (они ревниво относятся к тестам), и теми, кто не может дать такую оценку (они видят в тестах полезный лично для них инструмент, повышающий качество их услуг).

Тесты, конечно, гораздо более интеллектуальный инструмент, чем бритва и ножницы. Поэтому давайте сравним тесты и... шахматные компьютерные программы. До изобретения компьютеров у новичка не было никаких других шансов сыграть с сильным противником, кроме того, чтобы обратиться к живому мастеру-шахматисту. Платные и благотворительные сеансы одновременной игры с мастерами и гроссмейстерами были в докомпьютерную эпоху весьма и весьма востребованными. Но теперь можно найти в Интернете и установить на любом домашнем компьютере такую

сильнейшую бесплатную шахматную программу, для которой большинству пользователей в настройках нужно сразу же “понижать” уровень (установленный по умолчанию), на котором она играет, чтобы надеяться у этой программы выиграть хотя бы раз. Хотя гроссмейстерам-профессионалам такая бытовая шахматная программа, скорее всего, проиграет, но у всех любителей выиграет надежно и с большим запасом прочности. Как известно, теперь программами, оснащенными базами данных по известным дебютам, пользуются даже самые сильные гроссмейстеры (втайне от соперников и судей и с помощью различных уловок, которые изредка всплывают в форме скандалов).

Пример с шахматными программами, помогающими шахматистам, нам нужен, чтобы ввести важный концептуальный термин – термин “интеллектуальный инструмент”. Что такое шахматная программа? Это автоматизированный эквивалент мастера-исполнителя, это инструмент искусственного интеллекта, который заведомо полезен в руках всех слабых и средних специалистов (ибо усиливает их возможности), но с определенной ревностью воспринимается сильными и тем более великими, чей естественный интеллект до сих пор достоин восхищения. Итак, формулируем важный тезис номер три.

Тезис 3а. Полезный стандартизированный тест – это интеллектуальный инструмент, который ПОДНИМАЕТ качество оценки в конкретных обстоятельствах, то есть качество оценки в результате внедрения теста оказывается выше, чем качество, которое могут обеспечить другие специалисты и другие (нетестовые) процедуры оценки в данной организации.

Тезис 3а можно сформулировать и в виде обратного тезиса 3б:

Тезис 3б. Вредный стандартизированный тест – это интеллектуальный инструмент, который ПОНИЖАЕТ качество оценки в конкретных обстоятельствах, так как его качество ниже, чем качество, которое могут обеспечить другие специалисты и другие (нетестовые) процедуры оценки в данной организации.

Итак, попробуем теперь просто перечислить, в каких ситуациях (при каких условиях) измерительные тесты могут быть полезны, то есть эффективны:

1) Тест полезен, если он разработан профессионально – соответствует нормативным психометрическим параметрам научной тестологии (о которых пойдет речь в книге). В некоторых книгах по тестологии перечень требований к тестам

ограничивается, увы, фактически только этим пунктом. Но знакомство с другими пунктами нашего списка позволит читателю понять, что этого совершенно недостаточно.

2) Тест полезен, если он адекватен ситуации, то есть по самому содержанию заданий и рекомендаций (интерпретирующих сообщений на основе результатов тестирования) соответствует ожиданиям и заказчиков (ЛПР – лиц, принимающих решение), и самих испытуемых. Плохим следует считать тест, который вызывает скрытый (или явный) саботаж, так как содержит задания, кажущиеся большинству испытуемых «искусственными», «надуманными», «глупыми», «бессодержательными», «не имеющими отношения к делу». Хороший тест требует от испытуемого выполнять задания, которые моделируют его реальную производственную, учебно-научную или другую деятельность.⁴

3) Тест полезен, если организаторами тестирования проделана необходимая подготовительная работа по созданию атмосферы сотрудничества, причем если и не со всеми потенциальными испытуемыми (этого добиться в принципе практически невозможно!), то с большинством так называемых «высокоочковых» испытуемых, способных успешно выполнить тест.

4) Тест полезен, если принципы обработки тестовых результатов и выработки решения на основе этих результатов понятны и прозрачны для людей, то есть соответствуют имплицитной (подразумеваемой, обыденной) теории справедливости, принятой в данной социальной группе.

5) Тест полезен, если подавляющее большинство испытуемых соблюдает правила тестирования при его выполнении и не получает односторонних преимуществ в результате определенных приемов фальсификации (люди не списывают, не подсказывают друг другу, не подделывают и не переделывают уже созданные протоколы и подсчитанные баллы и т.п.).

6) Тест полезен, если его предметом (тестируемыми знаниями, умениями, качествами) оказываются такие свойства людей, которые тест выявляет и измеряет лучше,

⁴ Недаром на рынке оценочных услуг в наше время столь ценится так называемый «ассесмент» (или метод Центра оценки): метод, предполагающий наблюдение оценивающего (ассессора) за тем, как выполняет оцениваемый узкоспециальные упражнения, моделирующие (имитирующие) реальные процессы – коллективного обсуждения проблем, анализа документов, структурирования информации и т.п. (Базаров, 2011). Метод Центра оценки нередко противопоставляют методу тестирования, банально сводя тесты к выбору из готовых вариантов ответов на вопросы - причем вопросы, касающиеся весьма отвлеченных теоретических знаний. По убеждению автора этой книги, практическая тестология и методология ассесмента должны быть интегрированы в едином концептуально-терминологическом поле, чтобы эти методы не исключали, а дополняли друг друга. – Прим. автора.

чем это делает человек-специалист традиционными методами (беседа, анализ документации, интуитивная оценка), и при этом у традиционных специалистов остается свой фронт работы – они выполняют такую оценочную деятельность, которую не может выполнить тест.

7) Тест полезен, если исполнители альтернативных оценочных процедур (например, интервьюеры-эйчары, которые принимают кандидатов на работу по результатам собеседования с ними) очевидно уступают ему по критерию «цена-качество», то есть тест дает прирост в точности оценок при незначительных затратах на его создание (приобретение) и применение, в то время как традиционные оценочные процедуры приводят к большим расходам.

Последние пункты номер 6 и 7 нуждаются в иллюстрации. Возьмем обыкновенных преподавателей иностранного языка. В ходе своей работы они постоянно проверяют достигнутый уровень знаний и языковых умений учащихся и часто применяют тесты. Но в разных аспектах языковых способностей тесты в разной степени эффективны. Разумные преподаватели на уровне здравого смысла понимают, что для проверки словарного запаса или знания грамматики любой тест сработает лучше, чем они сами. Почему? По сравнению с устным опросом или нестандартизированной письменной работой тест по грамматике или на знание слов обеспечит больше разнообразных заданий, расширит охват материала, обеспечит более быструю проверку – путем простого подсчета числа правильных ответов. А в каких случаях тест уступает преподавателю? – В случае проверки коммуникативных навыков устного общения. Так вот, на этих навыках и нужно сконцентрировать усилия преподавателя в плане оценочной (контролирующей) деятельности. В отношении этих навыков тесты не очень нужны, ибо их преимущество совершенно не очевидно, а нередко оно просто отсутствует, ибо тест НЕ МОДЕЛИРУЕТ живого общения учащегося с человеком, не моделирует особого эмоционального фона, который характерен именно для живого общения (требующего от учащихся преодолевать собственную застенчивость, нерешительность и т.п.).

Резюмируем перечисленные выше 7 критериев эффективности в виде кратких названий (некоторые могут показаться терминологически сложными и непривычными, но зато эти термины подготавливают читателя не только к восприятию этой книги, но и для чтения другой профессиональной литературы по этой тематике):

**1) Критерий психометрической обоснованности
(надежности, валидности, репрезентативности)**

2) Критерий моделирования (очевидной валидности)

- 3) **Критерий кооперативной готовности**
 - 4) **Критерий субъективной справедливости**
 - 5) **Критерий достоверности (информационной безопасности)**
 - 6) **Критерий системной дополнителности**
- (в отношении других оценочных процедур)**
- 7) **Критерий рентабельности (конкурирующей эффективности)**

И вот теперь мы можем лучше определить и описать ситуации, когда тесты ВРЕДНЫ. Мы просто «перевернем» предложенный список – опишем такие ситуации, когда эти критерии не соблюдаются:

1) Тесты вредны, если допускаются грубые методические ошибки, то есть если используются низкокачественные тесты или высококачественные тесты, но с методическими ошибками. Если вы взяли какие-то тесты, в отношении которых никто никогда не проверял их психометрические свойства, будьте настороже – скорее всего, они подведут вас. Это все равно, что взять какой-то компьютер, собранный «на коленках» неизвестным производителем, про который нет точной технической информации о том, что этот инструмент прошел технические испытания. Но... вы можете взять и хорошие в психометрическом плане тесты, но не соблюдать при их проведении то, что называется, «условия эксплуатации». Например, все тесты требуют, чтобы испытуемый тщательно ознакомился с инструкцией, а не «пролистывал» ее. Некоторые тесты требуют контроля за пониманием инструкции на пробных заданиях. Если при проведении за этим не следить, то ждите беды – результаты теста будут ошибочными. Казалось бы, ничего необычного в этом условии номер 1 нет, ведь точно такие же требования применяются ко всем инструментально-техническим средствам человеческой деятельности: надо использовать качественные инструменты и использовать их правильно. Но... не следует забывать, что именно тесты вызывают гораздо более сильное сопротивление их внедрению, чем другие инструменты. Почему? – Потому что оценке подвергается самое главное для людей – их собственные профессиональные и личные качества. Поэтому-то любые методические ошибки – это колоссальный козырь для тех, кто пытается, прикрывая истинные причины сопротивления, оправдать необходимость отказа от внедрения тестов. Поэтому при наличии методических ошибок тесты оказываются не просто вредны, а вредны «вдвойне», ибо сама тема возможного применения теста после ярких ошибок в организации нередко закрывается надолго – вплоть до прихода нового состава руководителей.⁵

⁵ Кстати, методические ошибки бывают самые разные, но во введении мы их не будем подробно классифицировать. Коротко назовем, например, такой распространенный тип ошибок, как выбор неправильного теста. Например, берут тест, который измеряет временный, динамически изменяющийся

2) Тесты вредны, если они плохо моделируют деятельность, ради прогноза эффективности которой они созданы. Даже если имеются определенные экспериментально-статистические доказательства, что какой-то косвенный психологический тест высоко коррелирует с эффективностью деятельности (например, эффективные менеджеры статистически чаще рисуют «проективное дерево» с ветками, которые бодро торчат вверх, а низкоэффективные – с ветками, которые понуро опускаются вниз), нельзя всерьез привязывать результаты такого тестирования к каким-то организационным решениям. Это вызовет гарантированный протест у большей части испытуемых (а чаще всего и непонимание у ЛППР, то есть у заказчиков). Такие тесты можно осторожно включать в процедуру консультативно-психологической помощи, игрового тренинга, но никак нельзя включать в программы производственной аттестации. Даже если сотрудники смирились под административным давлением и внешне послушно выполняют подобные тесты, их скрытое сопротивление может приводить к тому, что результаты будут серьезно искажены и будут тем самым не просто бесполезны, а именно вредны: в лучших условиях окажутся те самые испытуемые, которые пошли на скрытый саботаж и обман, а в худших – те, кто добросовестно и честно старался выполнить тест. Любой скрытый саботаж против программы тестирования рискует обернуться этими чудовищными результатами – давать обратный эффект за счет конкурентных преимуществ, которые получают менее лояльные сотрудники по сравнению с более лояльными. Следующий критерий в каком-то смысле можно рассматривать как более обобщенное продолжение этого правила.

3) Тесты вредны, когда большинство испытуемых не настроены честно сотрудничать с организаторами тестирования, а главное – не настроена в пользу тестирования ключевая группа наиболее эффективных и влиятельных (авторитетных) работников в организации (успешных студентов в студенческой группе и т.п.). Тут определенный нюанс заключается в следующем. От тестов не могут выигрывать все, кто-то обязательно проигрывает. Тест – это не «миллион в наследство», чтобы всем нравиться. Но главное, чтобы тесты не вызывали отторжения у «элиты», чтобы от внедрения тестирования не считали себя в проигрыше те самые эффективные работники, на которых держится производство. Если по каким-то причинам (иногда в результате неадекватных иллюзий или заблуждений) среди работников распространилось убеждение, что тест – это

фон настроения (от эмоционально-позитивного до эмоционально-негативного), а использовать данные этого теста пытаются для долгосрочного прогноза. Другой тип методических ошибок – это неправильное принятие решения на основе результатов в виде «сдвига в критерии отсечения». Например, деятельность оператора, управляющего опасным производством (ядерным реактором), требует безошибочного надежного исполнения, а при отборе применяют не абсолютный, а статистически-относительный критерий отбора – берут высокую группу как 15 процентов лучших (по правилу «среднее плюс сигма»). – Прим. автора.

инструмент незаслуженной дискриминации, если не удалось настроить сотрудников еще до массового тестирования по-другому, то результаты тестирования будут скорее вредны, чем полезны, ибо люди будут выполнять тесты с искаженной мотивацией, применяя сознательно или чаще даже полубессознательно определенные тактики фальсификации (уклонения от достоверного выполнения теста). На лекциях в МГУ для иллюстрации этого социально-психологического механизма я нередко привожу для студентов такой грубоватый, зато понятный пример. Это анекдот еще советских времен. Проходя врачебную комиссию в военкомате, призывник послушался совета друзей и очень вяло дунул в трубочку, чтобы изобразить объем легких меньше, чем на самом деле. В результате врачебная комиссия приняла решение: «Годен в подводники в силу малого потребления кислорода». На этом примере мы видим, каким тонким психологическим инструментом являются на самом деле даже казалось бы такие простые медико-физиологические тесты, если учитывать, что объект тестирования – это человек, который применяет рефлексивную активность и по-своему пытается управлять ситуацией, то есть отражает цели и возможные последствия тестирования и ведет себя не так, как вел бы себя в обычной ситуации. Тысячи специалистов – изобретателей тестов – потратили, наверное, миллионы человеко-часов, чтобы изобрести такие тесты, которые бы давали объективные результаты против воли людей, сопротивляющихся тестам. Но... как известно, даже самые изощренные детекторы лжи («полиграфы», как их сейчас чаще называют) создают парадоксальные преимущества для тех, кто их не боится и твердо знает, что любой прибор можно обмануть. В результате тест угрожает обратным эффектом – тем, что будет производиться «отрицательная селекция» (анти-отбор) – к работе, как уже говорилось выше, будут привлечены люди, которые не больше, а меньше солидарны с организацией. Применение тестов в режиме волюнтаристического стиля управления «через колено», как правило, создает огромный риск вредного, а не полезного эффекта тестирования. Самым красноречивым свидетельством того, как плохо работают определенные формальные критерии и инструменты отбора, являются антиэлитные механизмы, приводящие к выдвиганию на руководящие должности (в масштабах крупных корпораций и даже государства в целом) людей с паразитарными установками коррупционеров⁶. Тесты

⁶ Кстати, вполне возможно, что большинство читателей до этого самого момента вовсе не задумывалось над тем, что механизмы появления коррупции могут быть последовательно рассмотрены с позиции именно тестологии, которая призвана изучать как полезные, так и вредные формализованные процедуры оценивания и отбора кадров. По дороге «наверх» людям приходится преодолевать в своей карьере такие испытания («тесты в широком смысле слова), которые дают эффект «отрицательной селекции»: приходится угождать начальству, преувеличивать или имитировать полезный эффект своей деятельности (ибо реальный скромный эффект нереалистичных начальников не устраивает), вместе с непосредственным начальником (или по его указке) обманывать подчиненных и «широкие массы», а также... начальство еще более высокого уровня. И если весь этот обман вполне устраивает обманутых, не желающих «портить себе настроение» и докапываться до истины, отрицательная селекция торжествует... наверх восходят ТОЛЬКО изощренные лжецы, имитаторы и интриганы. Если и верхи, и большинство низов устраивает жизнь в «королевство кривых зеркал», то кому нужны в такой обстановке правдивые

вредны, когда по ним принимаются решения, противоречащие субъективным (обыденным) представлениям о справедливости, принятым в организации (или в определенной социальной группе). Тонкость в понимании этого случая заключается вот в чем. Тут речь не идет о перевернутых шкалах тестовых баллов (когда лучшим выставляются худшие баллы), а тут речь идет о том, что решения на основе этих баллов не соответствуют представлениям сотрудников о справедливости. Например, в организации вдруг возобладали представления о справедливости в логике «уравнительной морали»: «Справедливость – это когда все члены команды получают равное вознаграждение». В этом случае перспективы успешного внедрения тестов весьма призрачны, ибо любой тест работает скорее на дифференциацию, чем на уравниловку. В общем случае можно сформулировать такое правило: дифференциация в решениях по результатам теста не должна резко противоречить представлениям людей о допустимой силе дифференциации. Наиболее типичная «имплицитная модель справедливости» формулируется, согласно современным социологическим исследованиям (Ролз, 1995), в виде «формулы пропорциональности»: люди должны получать вознаграждение, пропорциональное своему вкладу в общее дело. Если по результатам вполне качественного теста принимается такое решение, что премируются только «призеры» (тройка лучших), а все остальные оказываются «за бортом», то это будет, скорее всего, воспринято как нарушение принципа справедливости и, в свою очередь, усилит сопротивление тестированию на следующем цикле тестирования. Особый случай – конкурсное тестирование. Люди должны участвовать в нем, разделяя базовые принципы конкурса: награду получают не все, а только победители. Только создав в организации атмосферу понимания и принятия правил конкурсного тестирования, можно проводить тесты с резкой дифференциацией благ, выделяемых участникам по их результатам.⁷

4) Тесты вредны, когда большинство сотрудников в организации (или в группе тестируемых) знают (или догадываются), как именно можно обмануть тест, и есть люди, которые обязательно воспользуются этой возможностью. Таким образом, выражаясь

тесты – испытания, которые трудно обмануть? – Никому. Участь тестолога в этой обстановке – это распознавать демагогию и профанацию как разновидности социально-психологических механизмов сопротивления и учиться преодолевать это сопротивление.

⁷ Тут проницательный читатель может мне возразить: «Вы рассказываете о каких-то общих социологических и социально-психологических принципах, которые касаются любых управленческих решений. При чем тут тесты и тестология?» Я частично соглашусь с этим возражением, но... только частично. Дело ведь в том, что ошибки в принятии решения по итогам тестирования приписываются не людям, которые эти решения принимают, а... чаще всего самим тестовым методикам, самому тестовому подходу, ибо большинство воспринимает ситуацию тестирования не аналитически, а синтетически: «Все, что в результате тестирования ухудшает мое положение, есть свидетельство вреда от тестов». – Прим. автора

предельно бытовым языком, тут можно сказать так: «Уязвимые тестовые технологии вредны!». «Дырявые» в том смысле, что тестовая технология содержит лазейки, гарантирующие безнаказанную возможность для фальсификации. Под словом «уязвимые» мы имеем в виду уязвимость, которая сознательно используется фальсификаторами и профанаторами. Но нередко уязвимость возникает произвольно – вследствие халатности. Тестовые технологии по уровню своей защищенности должны соответствовать профессионально-этической и собственно профессиональной подготовленности пользователей-тестологов. Если в методических указаниях, адресованных пользователю, черным по белому написано: «Вы должны запрещать списывание», но на самом деле проводящий НЕ выполняет это требование методических указаний и смотрит на ситуацию массового списывания (со шпаргалок), спустя рукава, то в этом вреде – в появлении ложных, завышенных результатов – виноват не столько сам тест (набор заданий в узком смысле), но социально-информационная технология его проведения, не адекватная данным организационным условиям. Кто должен был хранить в тайне «ключи к тесту»? Заинтересован ли этот персонал (имеющий доступ к ключам) в том, чтобы добросовестно выполнять свои обязанности? Этими вопросами должен задаваться всякий, кто стремится внедрить тесты не формально, а с целью добиться реального полезного результата. На моих глазах сплошь и рядом получает доступ к секретной тестовой информации⁸ самый разный персонал, нередко очень низко оплачиваемый (например, администраторы и операторы внутренней компьютерной сети в вузах РФ оплачиваются очень плохо). Что в итоге происходит? Происходят массовые подтасовки, и достигается обратный эффект – более высокие результаты показывают не достойные, а как раз менее достойные люди – более наглые, едва ли не самые некомпетентные, сплоченные в «паразитарные группировки» – как раз те, которым просто «нечего терять» (ибо при честном выполнении теста им «ничего не светит»).

5) Тесты могут быть реально вредоносной технологией из-за того, что они, будучи внедрены как единственная оценочная технология, измеряют что-то менее важное, но действительно не измеряют и не могут измерить что-то более важное в человеке. Как уже говорилось выше, тесты рассматриваются как вредные с позиции тех весьма влиятельных людей в организации, которые применяли оценочные процедуры до внедрения тестов. Нередко эти люди сопротивляются внедрению тестов в логике «луддитов, ломающих станки», ибо видят в этих инструментах конкурирующую силу, заменяющую их труд в организации. Но... не надо рассматривать критику тестов с позиции этих людей только как проявление защитных эгоистических реакций. В этой

⁸ Секретной информацией следует считать: задания теста, правильные ответы, то есть «ключевые ответы», правила шкалирования результатов (перевода сырых баллов в стандартизированные шкалы), тестовые нормы и правила «отсечения» (категоризации результатов).

критике нередко есть рациональное зерно и большая ценность. Рассмотрим такой упрощенный пример. В организации решили внедрить тесты, в которых подавляющее число тестовых заданий (вопросов) проверяет то, как сотрудник помнит какие-то мелочи из инструктивных и регламентирующих документов. И при этом решили... отменить обычный человеческий контроль за тем, насколько работник в ходе живого контакта с наблюдателем (аналог устного экзамена) может просто показать (даже не рассказать), в каком порядке он производит те или иные производственные операции, решая определенную производственную задачу. То есть, подменили контроль за более существенным качеством (компетенцией) контролем за гораздо менее существенным, но легко формализуемым качеством (память на письменные документы). Обоснованным ли в данном случае будет протест специалистов по традиционным методам оценки? Да, совершенно обоснованным. Общее правило таково: тесты почти никогда не могут охватить все существенные качества работника, поэтому их надо применять в сочетании с традиционными методами, а не вместо них. Предлагаю читателю оглядеться «вокруг себя» и посмотреть, часто ли наши практики соблюдают это простое правило? Не соблюдают? Тогда чему удивляться, что внедрение тестов приводит скорее к вредному эффекту, чем к полезному.

б) Тесты оказываются вредными, если они приводят к неоправданному росту затрат при незначительном повышении точности оценки (а иногда даже понижении точности!). Если при внедрении тестов кто-то корыстно рассчитал, что тем самым «освоит определенный объем финансирования», то, скорее всего, эффект от тестов будет антиэкономическим (корыстное управление по затратному принципу вовсе не изжило себя вместе с социализмом, а по-прежнему процветает в больших и бюрократизированных организациях и суперкомпаниях-монополистах). Необходимо сравнивать затраты и полезный эффект. Не исключено, что традиционные оценочные технологии будут все-таки рентабельнее – с учетом того, что они не требуют особых дополнительных вложений. Но тонкости расчета рентабельности тестов – за пределами настоящего введения. Они будут рассмотрены в главе 6.

Завершая это введение, мне бы хотелось подчеркнуть следующее. Нет ни одного узкого специалиста, включенного в процесс тестирования, который бы обращал равное внимание, а главное был одинаково компетентен решать задачи по обеспечению указанных здесь критериев 1 – 7. Обычно в литературе по тестологии в центре внимания 2 фигуры:

- психометрист-тестолог (это очень часто, прежде всего, специалист по прикладной математической статистике),

- автор тестовых заданий (это очень часто специалист в какой-то предметной области, а при психологическом тестировании это психолог).

Но едва ли не большая часть указанных критериев не обеспечивается подготовкой психометриста и автора тестовых заданий. В процессе создания и внедрения тестовой системы требуется особый тип компетентности и особая ролевая позиция. Это позиция организационного социального психолога, способного чутко и оперативно отслеживать динамику фактических установок людей в отношении программы тестирования в данной конкретной социальной группе (организации), а также умело влиять на эту динамику в позитивном направлении – выстраивать систему конструктивного делового взаимодействия между разными участниками процесса, включая софт-инженеров и собственно администраторов-управленцев в данной организации. Поэтому, по убеждению автора, уже прошло время тестологии как науки об отдельных тестах, выхваченных из контекста их применения в конкретных социально-организационных обстоятельствах. Пришло время тестологии как науки о создании систем комплексного взаимодействия специалистов разного профиля, внедряющих тестирование в организации не столько как технологию формального ранжирования или селекции кадров (учащихся), сколько решающих сложную задачу инновационного управления и принимающих во внимание все идеологические, политические и эмоционально-психологические факторы происходящих организационных изменений. Тестирование затрагивает глубинные отношения внутренней конкуренции между людьми в организации (социальной группе), так что существует высокий риск при неграмотном внедрении обострить эту конкуренцию так, что она может принять деструктивные формы. В обычной тестологической литературе, включая даже самые современные работы, основное внимание традиционно уделяется более узкой психометрической тематике (см. например, Фер, Бакарак, 2010, Крокер, Алгина, 2010), а вопросы, от которых зависит судьба внедрения тестовых технологий, рассматриваются весьма конспективно и поверхностно в небольших главах и параграфах, связанных с профессионально-этическими нормами «правильного тестирования».

В понимании автора развитие практической тестологии должно происходить в ближайшее время в очень тесной методологической и концептуальной связке с развитием психосоциологической теории конкуренции (Шмелев, 1997; Шмелев, 2012б). Завершим данное введение следующим рабочим определением:

Практическая тестология – это прикладная междисциплинарная наука, которая базируется на системном анализе и учете всех существенных факторов и ресурсов, обеспечивающих внедрение системы тестирования как организационного инструмента, призванного обеспечить продуктивную конкуренцию между людьми.

