

ГОСУДАРСТВЕННЫЕ И МЕЖДУНАРОДНЫЕ ПОДХОДЫ К СТАНДАРТАМ КАЧЕСТВА В ТЕСТИРОВАНИИ: ПОЗИЦИЯ ВЕЛИКОБРИТАНИИ

П. Линдлей

Описываются разработка, применение и контролирование существующей в Великобритании официальной процедуры рецензирования и регистрации тестов, принятой Британским психологическим обществом, ее развитие от формата печатных изданий к принятому сегодня формату электронного рецензирования. Описано развитие структуры рецензии и ее отход от первоначального британского формата в соответствии со стандартами, разработанными действующим Комитетом по тестам и тестированию Европейской Федерации психологических ассоциаций (EFPA). Отражается текущее положение в области рецензирования и регистрации тестов, планы и перспективы развития рецензирования и регистрации тестов.

Ключевые слова: психодиагностика, психологическое тестирование, рецензирование тестов, сертификация тестов.

История возникновения британской модели рецензирования тестов

Британская модель рецензирования тестов была разработана Дейвом Бертрамом, Пэт Линдлей и Джулией Фостер в 1980-х годах для проекта правительства Великобритании по развитию служб занятости. В это время уровень безработицы был высок, и правительство спонсировало службы занятости, которые оценивали людей на способности, склонности и профессиональные предпочтения с целью максимизации их потенциала по обучению и трудоустройству. Издатели и распространители психометрически апробированных тестов продвигали на рынке огромное количество методик, призванных помочь службам занятости в достижении своих целей. Правительство объявило тендер на обзор тестов и других методик, который мог бы помочь службам занятости в выборе тестов, наиболее подходящих для их практических нужд.

Выигравшая тендер компания (Newland Park Associates Ltd) в 1990 году выпустила «Обзор стандартизованных тестов, применяющихся для оценки в сфере профессиональной подготовки». В этом же году службой занятости Великобритании было опубликовано 63 рецензии. После публикации и распространения по службам занятости в Шотландии, Англии и Уэльсе правительственный проект прекратился, а первоначальный набор рецензий и сама модель рецензирования была принята Обществом (BPS Books) в 1991 году.

В период с 1992 по 2001 гг. Обществом издавались разные по объему тиражи обзоров и ежегодников, главным редактором которых сначала был Дейв Бертрам (до 1998 года), затем – Пэт Линдлей. Еще ряд коллег выступали в качестве редакторов-консультантов. Линда Маршалл, а впоследствии и другие коллеги, были редакторами повторных обновленных изданий. Издаваемые обзоры при этом постепенно отходили от исключительно акцента на тестах, применяемых в профессиональной подготовке, вбирая в себя более широкий спектр тестов достижений и личностных опросников, применяемых в сфере организационной психологии в целом.

Цель рецензирования

Цель рецензирования тестов заключалась в том, чтобы получить независимую экспертную оценку тестов и снабдить квалифицированных пользователей независимой информацией, которой можно было доверять. Акцент ставился (и ставится) на написание таких рецензий, которые были бы информативными и полезными для практического использования тестовых инструментов. Они не только академичны, но и могут служить руководством для пользователя. С этой целью в каждой рецензии тест оценивался по определенному набору качеств (например, оформление, информативность, психометрические характеристики и т. д.) с использованием 6-балльной шкалы «звезд», в которой 0 «звезд» означает,

что искомое качество в тесте не представлено, а 5 «звезд» означает, что это качество проявлено превосходно. Эта система приводила к тому, что оценка содержания теста была интуитивно понятной для пользователя. Кроме того, рецензенты и редакторы писали к каждому из тестов оценочные комментарии. Одним из последствий использования такого четкого набора ясных критериев явилось то, что некоторые издатели тестов начали использовать эти критерии как шаблон для создания руководств к новым или обновленным психологическим тестам.

Развитие европейских критериев рецензирования тестов в 2002 году

Общество использовало британскую модель в качестве своего вклада в европейский проект, направленный на выработку общих критериев рецензирования тестов. Действующий комитет по тестам и тестированию EFPA выработал Европейский стандарт, который был одобрен Генеральной ассамблеей EFPA в 2002 году. Этот стандарт широко доступен и бесплатен, и его публикация резко повысила прозрачность процесса рецензирования, что означало, что издатели, авторы, распространители и рецензенты теперь могли пользоваться единым четким набором критериев оценки тестов и другого психодиагностического инструментария.

Под руководством Пэт Линдлей Общество в 2002 году приняло Европейские критерии рецензирования. В то же самое время издательство BPS Books было продано издательству Blackwell's Publishers, и Общество через свой Руководящий комитет по тестовым стандартам приняло решение публиковать рецензии не в бумажном формате, а онлайн. Это привело к следующему:

- переход от старых бумажных рецензий и обзоров к генерированию отчетов на основе баз данных, располагающихся онлайн;
- пересмотр и обновление предшествующих рецензий с целью приведения их к европейскому формату и включению в единую базу данных;
- повторная публикация (в электронной форме) рецензий и обзоров в новом формате, соответствующем европейским стандартам.

Переход от бумажных публикаций к электронным позволил сделать процесс рецензирования и обновления тестов непрерывным, решил проблему отсрочек в рецензировании новых тестов и сделал возможным не-

прерывное обновление цен и фактической информации.

Сама система рецензирования также получила ряд преимуществ. Приведение британских рецензируемых тестов к критериям EFPA дало редакторам рецензий опыт использования этих критериев и сделало переход от организационных тестов к образовательным более плавным; подготовка новых рецензентов также стала прозрачной и понятной, что привело к эффективному развитию системы регистрации тестов.

Рецензирование и сертификация

В настоящий момент Обществом для обеспечения высокого качества тестов используются две процедуры.

Рецензирование тестов: оценка теста, осуществляющаяся с применением критериев EFPA, результатом которой является публикация рецензии, призванной информировать конечных пользователей теста. Для издателей тестов эта процедура бесплатна.

Сертификация тестов: процесс, в котором издатель может заказать оценку своих тестов на предмет того, насколько они соответствуют минимальным стандартам качества для стандартизованных инструментов, определенным в терминах критериев EFPA. Эта процедура оплачивается издателем.

И рецензии, и сертифицированные тесты доступны на сайте Центра психометрики и тестирования Британского психологического общества: www.psychtesting.org.uk

Процедура рецензирования тестов

Издатели предоставляют тест для рецензирования либо, если есть потребность, редакторская коллегия запрашивает у издателя тест. В обоих случаях предоставление теста на рецензирование для издателя является добровольным.

Издателю необходимо предоставить четыре полных набора материалов, которые будут продаваться вместе с тестом или другим психодиагностическим инструментом. Редакторская коллегия назначает двух рецензентов и редактора-консультанта. Оба рецензента работают независимо друг от друга и не контактируют с издателем теста. Все рецензенты подписывают декларацию о заинтересованности, в которой они дают исчерпывающую информацию о личных связях и контактах с издателями теста. Редакторы подписывают эту форму во время своего назначения на ту или иную рецензию.

Оба рецензента готовят рецензию независимо друг от друга с использованием специальных форм, которые для каждой секции рецензии содержат соответствующие примечания и руководства. Эти формы отсылаются редактору-консультанту и одновременно главному редактору. Процесс подготовки рецензии отслеживается онлайн. Редактор-консультант объединяет и редактирует рецензии (также онлайн), после чего главный редактор завершает рецензию и отправляет ее анонимно издателю / распространителю для проверки фактической информации. Если вносятся какие-либо существенные правки, перед публикацией рецензии главный редактор согласовывает их с редактором-консультантом. Такая система «сдержек и противовесов» позволяет устранить субъективизм и предвзятость.

Как только все правки внесены, рецензия публикуется и становится доступной онлайн по адресу www.psychtesting.org.uk. Имена рецензентов и редакторов указываются в опубликованной версии рецензии. Каждый год у издателей запрашивают информацию по поводу обновлений в ценах или условиях поставки, а также информацию по поводу значительных изменений или о новых версиях теста. Соответствующие обновления в рецензию вносятся редактором-консультантом и главным редактором.

Процедура сертификации теста

Тесты, соответствующие минимальным стандартам качества EFPA (3 «звезды» или выше) в отношении валидности, надежности, норм и другой информации, необходимой для осмысленной интерпретации тестовых баллов, могут быть сертифицированы. Сделать это можно двумя способами.

Если тест уже прошел рецензирование и соответствует необходимым критериям, тогда он подлежит сертификации при условии оплаты со стороны издателя.

Если тест еще не рецензировался, издатель может назначить дипломированного психолога для публикации в руководстве к тесту доказательств того, что тест соответствует критериям качества. Заявка рассматривается членом редакторской коллегии, и если тест действительно соответствует критериям качества, то он признается пригодным для сертификации. В качестве условия сертификации издатель должен предоставить тест также и для рецензирования.

После сертификации издатель может размещать на тесте логотип «Сертифицирован РТС», что означает, что тест был рассмотрен Обществом и соответствует минимальным стандартам качества EFPA. Процесс сертификации длится 5 лет. На сегодняшний день сертифицировано 67 тестов. Еще множество тестов находятся в очереди на рассмотрение.

Проблемы, возникающие из-за использования европейских стандартов

В целом стандарты EFPA показали высокую эффективность и хорошо «работали» для тестов, разрабатываемых в Великобритании, а использование этих стандартов повысило качество процессов рецензирования и сертификации тестов. Тем не менее в рецензировании тестов существует несколько проблем, которые не разрешаются в стандартах.

По традиции в Великобритании всегда положительно относились к рецензированию тестов из других стран и культур, и именно в этой области возникают основные проблемы:

- недостаток сопоставимости между рецензиями на «один и тот же» тест, рецензируемый в Великобритании и в Швеции;
- проблемы рецензирования тестов, которые разрабатывались не в Великобритании;
- ожидания издателей тестов в тех случаях, когда рецензия на тест была получена в одной из европейских стран и ее оказывается достаточно для сертификации теста на территории Великобритании.

Сопоставимость рецензий

Проблема сопоставимости рецензий из разных стран впервые возникла тогда, когда одни и те же тесты, рецензируемые в Швеции и в Великобритании, получили неодинаковые оценки по некоторым ключевым показателям норм, валидности и надежности.

Второй пласт проблем выявился, когда на рецензирование в Великобритании начали предоставляться тесты из других стран. Некоторые из этих проблем касались того, как обходиться с наличием международных норм при отсутствии в нормах британского компонента; наличием международных норм с небольшой британской выборкой при отсутствии сугубо британских норм; тестами, стандартизованными на определенной выборке и затем переведенными на английский язык без стандартизации переводного варианта, однако с наличием британских норм; тестами с хо-

рошими психометрическими данными в изначальной стандартизации, которые проигрывают в британской рецензии, поскольку в них отсутствуют британские нормы.

Перевод и адаптация тестов

С течением времени становилось ясно, что все эти проблемы и дискуссии возникают из-за того, что в стандартах недостаточно представлен компонент, связанный с переводом и адаптацией тестов. Впервые на эту проблему пролило свет сопоставление шведских рецензий с некоторыми европейскими. Каждый из рецензируемых тестов был переведен на шведский язык. Различия в некоторых рецензиях могли быть объяснены не столько проблемами перевода как таковыми, сколько недостаточной стандартизацией шведскоязычных версий тестов.

Подобные проблемы появлялись и тогда, когда некоторые тесты, разработанные за пределами Великобритании и получившие там удовлетворительную рецензию, предоставлялись для рецензирования или сертификации в Великобритании. Например, «хороший» тест с хорошими нормами из некоей страны может иметь неподходящие «хорошие» нормы для населения Великобритании; либо же тест с международными нормами может проигрывать за счет диспропорциональной подвыборки британцев, приводящей к недостаточной обоснованности именно британского компонента тестовых норм.

Решение

Редакционный комитет Общества проанализировал стандарты EFPA и сделал вывод о том, что они целесообразны, но недостаточно освещают **вопрос об адаптации тестов**. Опираясь на работы Ван де Вийвера и Пуртинга (van de Vijver & Poortinga, 2005) была проанализирована проблема эквивалентности рецензий, адаптированных / переведенных на другой язык тестов.

Нами был разработан набор вопросов и ключевых позиций адаптации теста, который призван согласовать подходы разных редакторских коллективов к рецензированию тестов из разных культурных контекстов. Этот список теперь доступен на сайте EFPA по адресу <http://www.efpa.eu/professional-development/tests-and-testing>.

Руководство адаптированного теста должно быть проанализировано по следующим позициям.

Разработка теста через перевод / адаптацию

1. Присутствует ли в руководстве информация, доказывающая, что процесс перевода / адаптации включал в себя:

- работу специалистов, для которых новый язык (на который переводился тест) является родным;
- перекрестное рецензирование теста экспертами как по языку, так и по содержанию;
- обратный перевод с нового языка на язык оригинала;
- учет культурных и языковых различий.

Если информация по некоторым или всем этим позициям отсутствует, необходимо дать соответствующие комментарии.

2. Исследованы ли базовые психометрические качества переводной версии теста:

- характеристики пунктов;
- надежность шкал.

Если этого не сделано (или отсутствует в руководстве), необходимо дать соответствующие комментарии.

Нормы

1. В тех случаях, когда предоставляются местные нормы, рецензирование может проводиться в соответствии со стандартами EFPA.

В тех случаях, когда предоставляются «внешние» (неместные) нормы – достаточно ли убедительны доказательства эквивалентности обеих версий теста и обоих выборок, чтобы считать использование этого теста обоснованным? Обычно такого рода доказательства предполагают исследования, демонстрирующие скалярную эквивалентность двух разноязычных версий теста.

Если этого не сделано (или отсутствует в руководстве), необходимо дать соответствующие комментарии.

2. В тех случаях, когда предоставляются международные нормы в целом, из-за их повышенной неоднородности требования к этим нормам должны быть выше, чем типичные требования для «местных» норм.

По этому поводу необходимо дать соответствующие комментарии.

3. Международные нормы могут быть наиболее подходящими для использования в международном контексте (например, при сравнении людей, которые выполняли тест на разных языках), однако при определении качества этих норм необходимо принимать во внимание аспекты, перечисленные ниже. В

общем случае использование международных норм требует доказательства по меньшей мере эквивалентности измерения между тестами на исходном и на переводном языке.

Характеристика выборки

- Уравновешенность источников выборки (например, выборка, на 95 % состоящая из немцев, на 2 % из итальянцев и на 3 % из англичан, не может считаться по-настоящему международной выборкой). Выборка должна быть уравновешена для наиболее адекватного отражения соотношения частей популяции.

- Эквивалентность фоновых характеристик (профессиональная принадлежность, образование, ситуация тестирования и т. д.) различных частей общей выборки. Выборки стандартизации, в которых эти характеристики не могут быть сопоставлены, не могут считаться достаточными.

Характеристика теста

- В тех тестах, где используется меньше вербального материала, в меньшей степени будет проявляться эффект перевода. Это относится к тестам достижений и в некоторой степени к тестам абстрактного и схематического мышления.

- Эквивалентность версий теста, используемых с разными языковыми группами респондентов.

Должны присутствовать доказательства того, что все языковые версии были должным образом переведены/адаптированы.

- Есть ли доказательства того, что отдельные группы респондентов заполняли тест не на исходном языке?

Сходство баллов в различных выборках респондентов

- Необходимо, чтобы были приведены свидетельства о сходстве паттернов средних баллов по шкалам для подвыборок респондентов из разных стран. В тех случаях, где обнаруживаются значительные различия, они должны быть объяснены, а также должны быть разъяснены последствия этих различий для практики применения и интерпретации результатов теста. Например, если испанская выборка в среднем набрала по шкале балл значительно выше, чем голландская выборка, существует ли объяснение этой разницы и каким конкретно образом в этих условиях необходимо сравнивать представителей обеих стран по результатам тестирования?

Руководство к обобщению норм за пределы тех групп, которые включены в международную выборку стандартизации

- Например, если выборка стандартизации состоит на 20 % из немцев, 20 % французов, 20 % итальянцев, 20 % британцев и 20 % голландцев, использование этой выборки в качестве нормативной для кандидатов из Бельгии или Швейцарии может быть оправданным, а для кандидатов из Китая – нет.

Эквивалентность / Валидность / Надежность

В руководстве должны быть представлены свидетельства эквивалентности разноязычных версий теста. Без этого невозможно применять психометрические характеристики одной версии теста к другой версии (проводимой в другой стране или на другом языке). Примеры свидетельств эквивалентности тестов:

- инвариантность в структуре конструктора – например, через сравнение факторных структур или сравнение корреляций с некоторыми стандартными переменными;

- сходство в характеристиках критериальной валидности – например, в двух языковых группах тест коррелирует с одинаковыми компетенциями;

- сходство в паттернах нагрузки отдельных пунктов на шкалу: например, пункты коррелируют с другими шкалами с одинаковым паттерном в различных подвыборках; набор пунктов, имеющих наиболее высокие и наиболее низкие нагрузки на «свою» шкалу, совпадает в двух языковых группах;

- двуязычные респонденты получают сходные профили, заполняя тест на разных языках (надежность альтернативных форм).

Случаи, в которых генерализация коэффициентов валидности должна обосновываться более доказательно

- Когда перевод теста осуществляется в разных языковых группах (например, с индоевропейского языка на семитский), эквивалентность более уязвима по причине сильных различий в языках и культурах, следовательно, необходимы сильные доказательства эквивалентности разноязычных версий теста.

Случаи, в которых генерализация коэффициентов валидности может быть произведена теоретически (путем индуктивного вывода)

- В тех случаях, когда тест был переведен на множество языков, некоторые выводы

о генерализации коэффициентов валидности могут быть сделаны на основе свидетельств об инвариантности валидности в предыдущих переводных версиях теста. Например, если шведский тест уже был переведен на французский, немецкий и итальянский языки, и во всех этих случаях была показана эквивалентность разноязычных версий теста.

В целом при рассмотрении проблемы эквивалентности полезно исходить из классификации, предложенной Ван де Вийвером и Пуртингой

- Структурная / функциональная эквивалентность

Существуют доказательства того, что разноязычные версии теста измеряют один и тот же психологический конструкт. Обычно это демонстрируется путем сравнения паттернов корреляций между переменными в разных группах респондентов.

- Эквивалентность единиц измерения

Существуют доказательства того, что единицы измерения в двух группах совпадают, однако источники тестового балла неодинаковы (то есть индивидуальные различия в группе А сопоставимы с индивидуальными различиями в группе Б, но абсолютные сырые баллы в группе А и группе Б несопоставимы без определенной формы стандартизации баллов).

- Скалярная эквивалентность / Полная эквивалентность тестовых баллов

Единицы измерения совпадают, и источники тестовых баллов одинаковы (то есть сырые баллы имеют один и тот же содержательный смысл и для разных групп респондентов являются сопоставимыми).

Применимость рецензий

Третья проблема, с которой сталкиваются издатели тестов, которые прошли рецензирование по стандартам EFPA в другой стране. Зачастую такие издатели хотят, чтобы тест был автоматически сертифицирован и в Великобритании, поскольку «это европейские стандарты». Нами был разработан специальный бюллетень, в котором разъясняются проблемы

перевода и адаптации тестов и необходимость отдельного рецензирования в Великобритании.

Справочная информация по проблеме стандартов качества и тестирования¹

1. Bartram, D, Lindley, PA and Foster, JM (1990). A Review of Psychometric Tests for Assessment in Vocational Training. The Training Agency: Sheffield.

2. Van de Vijver, F.J.R., & Poortinga, Y.H. (2005). Conceptual and methodological issues in adapting tests. In Hambleton, R.K., Merenda, P.F., & Spielberger, C.D. Adapting Educational and Psychological Tests for Cross-cultural assessment. Mahwah, New Jersey: Lawrence Erlbaum Associates.

3. Bartram, D. (1998). The need for international guidelines on standards for test use: A review of European and International initiatives. *European Psychologist*, 3, 155–163.

4. Bartram, D. (2001). The development of international guidelines on test use: the International Test Commission Project. *International Journal of Testing*, 1, 33–53

5. Bartram, D, Lindley, PA and Foster, JM (1990). A Review of Psychometric Tests for Assessment in Vocational Training. The Training Agency: Sheffield.

6. Hambleton, R. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229–244.

7. Hambleton, R. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In, R.K. Hambleton, P. Merenda, & C. Spielberger (Eds.). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum.

8. Van de Vijver, F. & Hambleton, R. (1996). Translating tests: some practical guidelines. *European Psychologist*, 1, 89–99.

¹ Перечень процитированных в статье литературных источников представлен с сохранением оформления автора статьи.

Поступила в редакцию 27 апреля 2011 г.

Патрисия Линдлей. Доктор, председатель комитета по тестовым стандартам, Британское психологическое общество (BPS).

Patricia Lindley. BA, PhD, Independent Consultant in Occupational Psychology, Senior Fellow in Psychology, University of Hull (UK).