

ПЕД	
	измерения

ОБЕСПЕЧЕНИЕ ВАЛИДНОСТИ ТЕСТОВЫХ РЕЗУЛЬТАТОВ ПО РУССКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ

Наталья Андрюшина

МГУ им. М.В. Ломоносова, ЦМО

Российская система тестирования по русскому языку как иностранному (ТРКИ), созданная во второй половине 90-х гг. прошлого века, получила широкое признание как в стране, так и за её пределами. Выделенные в ней уровни владения русским языком — элементарный, базовый, первый, второй, третий и четвёртый — вполне соотносятся с уровнями общеевропейской системы владения иностранными языками (A1, A2, B1, B2, C1, C2). Конечно, соотношение выделяемых в обеих системах уровней нуждается в дальнейшем тщательном описании и анализе, однако не подлежит сомнению, что функционирование Российской системы тестирования способствует интеграции РКИ как самостоятельной прикладной науки в международное образовательное пространство.

Тем важнее обеспечить системе добротный фундамент, который позволит ей в дальнейшем успешно развиваться. В качестве такого фундамента в этой статье рассматривается проблема валидности.

Валидность — один из основных критериев качества тестовых результатов. Несмотря на это, на конференциях, где обсуждались различные аспекты, связанные с тестированием по РКИ, валидность неизменно оставалась за рамками внимания докладчиков. Между тем, после того как уляжется эйфория от признания самого факта создания Российской системы тестирования по РКИ, неизбежно встанет вопрос о проработанности основных теоретических понятий, о качестве заданий и не в последнюю очередь — о валидности тестовых результатов. В связи с этим необходимо уже сейчас начать работу по определению уровня валидности результатов, получаемых в ТРКИ.

Под валидностью (в широком смысле слова) большинство тестологов понимает «пригодность тестовых результатов для той цели, ради чего проводилось тестирование».¹

Иными словами, тестовые результаты считаются валидными, если тест способен измерять подготовленность испытуемых именно по той

1

Аванесов В.С.
Композиция тестовых заданий. М., 2002.

учебной дисциплине, для которой он разработан. Валидность показывает, что именно измеряет тест и насколько хорошо он это делает. Понятие валидности обычно относится к тестовым результатам в целом. В тестах по РКИ оно может использоваться и по отношению к субтестам.

Правомерно говорить и о валидности системы тестирования в целом. Анализ системы с этой точки зрения был бы очень полезен, но это многоаспектная проблема, требующая отдельного объёмного исследования.

В педагогических измерениях различается несколько видов обоснования валидности результатов. Наиболее актуальными на данном этапе развития системы тестирования представляются внешняя (*face validity*) валидность результатов, в существенной степени зависящая от содержания теста и соответствие критериям здравого смысла. Внешняя валидность представляет собой оценку эксперта или экспертов на основании просмотра результатов теста в целом, без его тщательного анализа и детального сопоставления со спецификацией и стандартом. Такую валидность иногда называют также эмпирической.

Содержательная валидность результатов (*content validity*) определяется отношением содержания теста к совокупности подлежащих проверке знаний и умений. Тестовые результаты обла-

дают высокой валидностью, если тест проверяет всё, что предполагалось проверяющими. При этом проверяющие не обязательно должны являться и разработчиками теста. Тестовые результаты считаются невалидными, или недостаточно валидными, если часть заданий, или субтестов, измеряет не то, что декларировано разработчиками.

По мере развития системы тестирования и накопления статистического материала, неизбежно придётся определить и так называемую концептуальную (*construct*) валидность тестовых результатов.

В настоящее время о валидности упоминается в лучшем случае при сдаче тестовых результатов на экспертную оценку. Заключение экспертов о валидности результатов по тем или иным тестам в России в основном можно отнести к констатации очевидной валидности, ибо для определения содержательной валидности требуется анализ, базирующийся не только на предлагаемой разработчиками спецификации тестов, но и на статистических данных, которые, как правило, на экспертизу не предоставляются.

В тестах с критериально ориентированной интерпретацией результатов важнейшее условие валидности — полнота отражения в тесте содержания проверяемого уровня владения языком. Соответственно, чтобы обеспе-

Методология

ПЕД
измерения

чить приемлемый коэффициент валидности, необходимо максимально точное описание языкового материала каждого уровня, ведь для качественного измерения надо правильно представлять объект измерения.

Решать столь сложную задачу следует поэтапно, идя от вопросов распределения языкового материала и речевых умений в целом по уровням, к анализу содержания субтестов, а затем и отдельных заданий. Хотя эта мысль может показаться странной, необходимо пересмотреть объём языкового материала на различных уровнях, ибо никем не оспаривается факт, что в его распределении наблюдается ярко выраженная диспропорция.

Возможно, это болезнь роста, поскольку в уровневых системах других стран и даже в «Общевропейских компетенциях»² имеется ряд нерешённых вопросов аналогичного плана. Например, в содержании дескрипторов допускаются слишком очевидные нарушения пропорции в распределении языкового материала по уровням. Так, при проверке навыков письменной речи в описании дескриптора «Заметки, записки, бланки» в этих компетенциях предполагается, что материал уровней В2, С1, С2 идентичен уровню В1 («Может написать записки с короткой важной информацией друзьям, ...»).

Невозможно представить себе, чтобы навыки письма иностранца, который владеет самыми простыми элементами делового стиля, не отличались от навыков иностранца, который находится на уровне профессионального владения языком. Вообще наличие ссылки на предыдущий уровень (а тем более на предыдущие уровни), как правило, означает, что в данном аспекте не выявлена специфика описываемого уровня. А если пропущено описание сразу нескольких дескрипторов, то возникает сомнение в правомерности выделения или данной шкалы, или данного уровня. Соответственно и результаты, полученные по тестам, сконструированным на основе подобного описания, не могут быть признаны валидными.

Учитывая изложенное, становится понятной необходимость решения задачи обоснованного описания объёма языкового материала по уровням владения иностранными языками, попутно решая вопросы логичного и последовательного представления созданной иерархии.

Итак, содержательная валидность тестовых результатов не может быть получена без пересмотра уровней владения русским языком как иностранным и научного обоснования их объёма и содержания этих уровней. Проблема пропорционального и обоснованного распределения

2

Общевропейские компетенции владения иностранным языком: изучение, обучение, оценка. Департамент по языковой политике, Страсбург, 2003, с. 85.

языкового материала по уровням, к сожалению, не единственная. Некоторые объекты контроля вообще не упоминаются в принятых на настоящий момент нормативных документах, которыми являются, например, Стандарты (Требования). В качестве иллюстрации можно привести отсутствие описания социокультурной компетенции. Конечно, такое описание — очень непростая задача, но перечислить хотя бы основные компоненты данной компетенции необходимо.

Нельзя не упомянуть о трудностях противоположного толка: при наличии в Стандартах (Требованиях) раздела «Фонетика» чёткие критерии оценки уровня владения русским произношением до сих пор не выявлены, не определён и объём проверки этого аспекта. Снижается внешняя валидность и в связи с непродуманным соотношением количества заданий, контролирующих различные умения. Такая ситуация возникает, когда разработчики поддаются соблазну включить в тест задания, которые легко создаются. Например, в субтесте «Грамматика. Лексика» 2 уровня умение различать виды глагола проверяется в 15 заданиях³. Столь очевидное количественное преобладание заданий одного типа ничем не может быть оправдано. Устранение упомянутых недостатков помо-

жет существенно повысить различающую способность тестовых заданий.

Следующим шагом к повышению валидности заданий ТРКИ должна стать работа по коррекции перечня объектов проверки. До сих пор не определён объём материала, подлежащий *обязательному контролю* при тестировании. Какую часть языкового материала уровня целесообразно предъявлять для контроля?

С одной стороны, очевидно, что желательно заложить проверку максимально большего объёма, тогда и валидность будет достаточно высокой.

С другой стороны, невозможно безгранично увеличивать объём проверки, поскольку тест не должен утомлять испытуемых (не будем лукавить, он должен быть и экономичным). Основанием отбора должна стать статистическая обработка результатов тестирования, вернее, анализ данных статистической обработки. Только таким способом можно отобрать релевантные языковые единицы и уменьшить объём контрольных материалов.

Наконец, необходимо проверить пригодность каждого задания теста. При рассмотрении содержания заданий с этой точки зрения возможны следующие случаи: а) задание имеет низкую дифференцирующую способность или не имеет её вовсе;

Методология

3

Аверьянова Г.Н. и др.
Типовые тесты по русскому языку как иностранному. II сертификационный уровень. Общее владение. М.—СПб.: изд. Златоуст, 1999. С. 36–37.

Здесь и далее приводятся примеры только из теста 2-го уровня, с целью сокращения объёма ссылок. Это не значит, что аналогичные примеры отсутствуют в других тестах.

ПЕД	
	измерения

б) задание дифференцирует испытуемых, и это связано с качественным подбором его содержания; в) задание дифференцирует испытуемых, но не по тому объекту контроля, что был заявлен разработчиками.

Для оценки дифференцирующей способности заданий необходимо сопоставить заявленные в спецификации или методических рекомендациях объекты контроля и непосредственное содержание задания. В субтестах по письму и говорению целесообразно рассмотреть и листы оценки (рейтерские таблицы).

А. Можно описать ряд теоретически возможных заданий, задуманных для проверки различных видов речевой деятельности и обладающих при этом низкой дифференцирующей способностью.

Однако более убедительным будет рассмотрение реально существующих заданий. Так, в субтесте «Письмо» 2 уровня в спецификации 1-го задания заявлено, что в нём проверяется способность иностранца репродуцировать письменный текст, демонстрируя умение «выделять и развёртывать информацию» в соответствии с заданием⁴. В формулировке же задания испытуемый должен *дать рекомендацию* на основе прочитанной рекламной информации, то есть задание проверяет также и умение продуцировать определённую интенцию, что заложено в оцен-

ку задания, но никак не обозначено в методических пояснениях к нему. Таким образом, объектом проверки данного задания стало умение, не вошедшее в спецификацию, что резко снижает его пригодность для разработки теста.

При внимательном рассмотрении тестовых заданий нельзя не заметить случаев, когда содержание отдельного субтеста плохо соотносится с содержанием спецификации, представленной в методических рекомендациях теста. Так, в субтесте «Грамматика. Лексика» 2 уровня при заявленной проверке умения правильно употреблять деепричастный оборот⁵ испытуемому предлагается для выбора 3 причастия и 1 деепричастие, т.е. данное задание проверяет способность иностранца различать причастия и деепричастия⁶. И внешняя, и содержательная валидность результатов при этом весьма невысока.

Б. Естественно, задания, обладающие внешней и содержательной валидностью, преобладают, иначе процесс тестирования был бы бессмысленным. Например, в субтесте «Письмо» 2 уровня в задании 3 проверяется умение иностранца продуцировать текст в виде неофициального письма-рекомендации, что чётко коррелирует с инструкцией и текстом задания.

В. Примером, когда наличие внешней валидности не подкреп-

4

Аверьянова Г.Н. и др.
Типовые тесты по русскому языку как иностранному. II сертификационный уровень. Общее владение. М. СПб.: изд. Златоуст, 1999. С. 62.

5

Там же. С. 92.

6

Там же. С. 41–42.

ляется содержательной валидностью, служат задания, помещённые в Типовые тесты второго, третьего уровней с целью проверки владения интонационными средствами русского языка, играющими, как известно, важную смысловозначительную роль и несущими высокую функциональную нагрузку. На каждом уровне владения русским языком иностранцы должны правильно воспроизводить и употреблять определённый набор интонационных конструкций (ИК). Поскольку интонационные средства многих языков существенно отличаются от интонационных средств русского языка, последние должны являться объектом серьёзной проверки (здесь «работает» принцип: проверяться должно то, что представляет трудность).

Одно из заданий в субтесте «Говорение» 2 уровня предлагает «воспроизвести реплики с интонацией, соответствующей намерению, которое предложено в задании».⁷ В частности, испытуемый должен прочитать фразу, выразив восхищение: «Как звучит рояль / Просто прелесть».

На первый взгляд, данное задание обладает дифференцирующей способностью, ибо по замыслу разработчиков служит

для проверки владения русскими интонационными конструкциями и испытуемые должны продемонстрировать указанное умение. По всей вероятности, данную фразу следовало прочитать с ИК-5, но основная масса испытуемых была незнакома со словом «прелесть», поэтому их усилия были направлены на то, чтобы правильно прочитать это слово. В результате 90% испытуемых читало слово с неправильным ударением и забывало о нужной интонации.

Столь же безуспешными были попытки выразительно прочитать и другие фразы задания. Таким образом, анализ содержания задания показал, что в данном случае проверяется, скорее всего, техника чтения. Следовательно, необходимо искать другие способы проверки владения интонационными средствами русского языка.

Приведённые примеры позволяют констатировать, что уровень валидности тестовых результатов в Российской системе тестирования по РКИ пока оставляет желать лучшего. А потому основная работа по повышению валидности ещё впереди. Хотелось бы надеяться, что представленный в статье анализ может улучшить ситуацию.

Методология

Методология

7

Аверьянова Г.Н. и др.
Типовые тесты по русскому языку как иностранному. II сертификационный уровень. Общее владение. М.-СПб.: изд. Златоуст, 1999.
С. 30.