

ПЕДАГОГИЧЕСКИЙ АНАЛИЗ ПОПУЛЯРНЫХ ТЕСТОВЫХ СИСТЕМ

// Школьные технологии. – 2006. - № 3. –
С. 126 -140.

Уривки

1. Можно ли доверять тестам?

Ответ на этот вопрос – задача этой лекции. И кратким ответом на него является следующий: Смотря каким... Вот давайте и посмотрим на существующие в настоящее время в образовательном мире тестовые системы и решим, каким доверять, а каким не доверять. Слово «тест» английского происхождения. Оно используется для обозначения всевозможных проб, испытаний и проверок, направленных на оценку состояния объекта или явления. В мире давно и успешно используются различные виды тестов. Почти все они строятся на базе четких научно-теоретических принципов и подробного описания свойств объектов и явлений, которые тестируются.

Вот уже более столетия как тесты вошли и в педагогический оборот. За это время многое сделано психологами и педагогами для того, чтобы понять их специфику и возможности для объективной оценки педагогических явлений. Из всего спектра этих явлений тестирование пока охватывает в основном испытание и оценку исходных (врождённых) интеллектуальных особенностей индивида и оценку качества усвоения им опыта (знаний и умений) конкретной деятельности.

Первое направление – тестирование интеллектуальных особенностей индивида – является в чистом виде предметом психологии. Начиная с Бине-Симона, французских психологов начала XX века, в этой области психологии потрудились целая плеяда замечательных психологов, во главе списка которых можно поставить известного английского психолога Айзенка. Вершиной их достижений явился чрезвычайно популярный в середине XX века и одновременно чрезвычайно противоречивый так называемый IQ-тест. Сегодня можно с уверенностью сказать, что цель, поставленная перед IQ-тестом: определить исходное состояние, врожденную силу человеческого интеллекта независимо от приобретенного опыта, к сожалению, все еще не достигнута. По этой причине IQ-тест в значительной мере потерял свою популярность и в организованной педагогической практике не используется. Других, более обоснованных и надежных интеллектуальных тестов пока не создано, хотя, судя по некоторым психологическим разработкам (Павлов И.П., Spranger E., Gardner H.), к этому открываются не плохие предпосылки. Сегодня в теории и практике обучения поле диагностики качеств личности занято почти исключительно тестами достижений.

Вот почему мы ограничиваем свой выбор для анализа только тестовыми системами, предлагающими тесты достижений, т.е., тестами, которые выявляют факт и степень усвоения учащимися конкретного опыта (учебного материала).

В Российской педагогической литературе понятие тест используется многими авторами в предельно узком смысле этого слова — только как контрольные задания, сопровождаемые несколькими вариантами ответа. Такое ограничение понятия *тест* существенно обедняет его действительное содержание. Любое контрольное задание, обеспеченное эталоном его исполнения, является тестом, не зависимо от формы его предъявления: с вариантами ответа или без них.

Знакомясь с сегодняшним рынком тестов, на первый взгляд может показаться, что существует множество тестовых систем и их анализ может превратиться в очень громоздкое предприятие. Но, если присмотреться к исходным посылкам, лежащим в основе создания и использования этих тестовых систем, то оказывается, что все они построены по одной и той же методической схеме, лишь незначительно варьируя в своей эмпирической части. Это особенно характерно для тестовых систем, используемых в США и России, которым мы и уделяем наше основное внимание. В данной лекции будут рассмотрены две наиболее широко представленные в образовании США тестовые системы: SAT и TOEFL.

Поскольку тестирование всегда осуществляется с помощью некоторого набора (батареи) тестов, рассмотрим теоретические вопросы их создания и применения как базу для анализа различных тестовых систем.

2. Тесты-лестницы, которым можно доверять

Термином «Тест-лестница» обозначают строго упорядоченную, соответственно описанным выше параметрам качества усвоения, последовательность из двух и более батарей тестов, позволяющих получить исчерпывающее представление о достижении учеником определённого мастерства. Под батареей тестов подразумевается набор тестов одного и того же уровня.

Общая высота, на которую ученик может подняться по лестнице тестов, и скорость перемещения по ней характеризуют качество усвоения учеником предмета и могут служить надёжными исходными ориентирами для выбора целей последующего обучения и распределения учащихся по группам с одинаковыми познавательными возможностями. Тест-лестница удобна также для итогового контроля усвоения, выявляя полную картину успехов учащихся, для этого тест-лестница должна состоять из батарей тестов высокой надёжности.

2.1. Надёжность батареи тестов

Возникает вопрос о выборе числа тестов в батарее тестов данного уровня для надёжного тестирования знаний учащихся. Под надёжностью тестирования понимается стабильность его результата при повторных испытаниях одного и того же испытуемого. Надёжность тестирования обозначим греческой буквой μ (мю). Надёжность батареи тестов выражается в процентах. Например, если надёжность $\mu = 40\%$, это означает, что в 40% повторных проб будет один и тот же результат, а в 60 % случаев — другие результаты. Конечно, такое тестирование ненадёжно и его можно применять только для ориентировочных срезов в текущей учебной работе. Никаких итоговых суждений по такой батарее тестов выносить нельзя. Минимальная надёж-

ность батареи тестов для итоговой характеристики качества знаний испытуемого должна быть не ниже $\mu = 75\%$.

Выбор надёжности тестирования зависит от характера диагностируемой деятельности: для летчика $\mu = 100\%$; для врача-терапевта минимум $\mu = 75\%$; для хирурга $\mu = 100\%$; для учителя $\mu = 75\%$, для ученика в школе традиционного обучения μ может не превышать и 50% и т.д.

При этом единицей для расчётов надёжности батареи тестов является существенная операция теста, а не сам тест, как ошибочно предполагают некоторые педагогические и психологические математики, разрабатывающие эту проблематику.

Надёжность батареи тестов зависит от числа существенных операций во всём тестовом наборе данного уровня. Для достижения $\mu = 75\%$ батарея тестов должна содержать не менее 80 существенных операций. Такое число операций сложно иметь в тестовой батарее для текущего контроля, но не сложно — для итоговой оценки. Нечего говорить, что надёжность теста-лестницы определяется надёжностью, входящих в нее тестовых батарей.

3. Методика педагогического анализа тестовых систем

Распространенный в педагогике метод анализа образовательных явлений состоит в написании рецензии, в которой рецензент высказывает свою точку зрения по поводу авторского творения. Рецензия — это повод для дискуссии, а не основание для решения, хотя во многих случаях она используется именно в этой, второй своей ипостаси. Ограничение педагогического анализа рецензией — это вынужденная позиция рецензента, поскольку педагогика все еще остается на консервативно-феноменологическом уровне своего развития и рецензенты привычно седлают своего умозрительного конька и бросаются в бой с беззащитным автором. Авторитарность все еще остается основным оружием рецензента и фатальной стрелой для «чего изволите?» автора.

В наших работах, подводящих итог полувековому поиску объективных показателей особенностей педагогических явлений, открывается возможность переходить в их анализе с феноменологических позиций на позиции качественной и количественной теории, позволяющих получать доказательные, а не умозрительные выводы.

В данном анализе тестовых систем мы вооружены следующими методиками анализа тестов достижения:

- 1) содержательной валидности теста;
- 2) функциональной валидности теста;
- 3) простоты теста;
- 4) надёжности теста;
- 5) корректности скоростного теста;
- б) шкалой оценки, сопряженной с процессом формирования знаний.

Содержательная валидность теста означает, что все задания теста могут быть разрешены на основе ранее изученного в предмете учебного материала. Что должно было быть изучено, о том и спрашивают — является афористическим правилом содержательной валидности теста. Однако если при создании теста не используется

формализованная методика подготовки учебного материала к построению на его основе тестов, отклонения от содержательной валидности становятся систематической ошибкой. Ниже, в анализе конкретных систем, эти отклонения будут наглядно продемонстрированы, но заключение о возможной содержательной инвалидности теста может быть вынесено немедленно, как только будет установлено, что в содержании предмета не выделены учебные элементы. Незаметно для себя составители тестов отклоняются от кажущейся им ограниченности содержания предмета и строят тесты на «догадку», которая кажется им такой «простой». Наличие графы учебного предмета с наглядным представлением структуры и номенклатуры учебных элементов в предмете позволяет практически полностью избежать ошибок в содержательной валидности теста, и это не умозрительный, а аналитический подход.

Функциональная валидность теста означает, что тесты построены с ориентировкой на некоторый, четко и точно сформулированный диагностируемый признак качества знаний по тестируемому предмету. Сопоставляя тест с диагностируемым признаком, на который он нацелен, приходят к выводу о функциональной валидности теста.

В теории уровней усвоения четко выделен диагностируемый признак *функциональной валидности* теста — это способ выполнения деятельности при решении задания теста: репродуктивно с подсказкой, репродуктивно по памяти, эвристически и творчески. Соответственно, любой тест, составленный для диагностики усвоения знаний и умений учащимися, может быть однозначно отнесен к одной из названных выше категорий деятельности и сделан вывод о валидности его использования относительно данной цели обучения. Здесь афоризм валидности может быть сформулирован, перефразируя излюбленное изречение диктаторов: «Цель определяет средства диагностики». Функциональная валидность теста не может быть установлена, если цель обучения не поставлена диагностично. К примеру, если цель обучения поставлена «научить учащихся читать литературные тексты на английском языке» - то можно предлагать любой тест, но ни его содержательную, ни функциональную валидность доказать нельзя. Перефразируем цель: «Научить учащихся самостоятельно читать литературные тексты на английском языке, составленные на основе базовой (800 единиц) лексики». При такой формулировке цели обучения, функционально валидным будет тест на узнавание по памяти английской лексики (второй уровень), не выходящей за пределы ограниченной 800-ми единицами лексики (содержательная валидность).

Простота теста — это чисто функциональное требование, предотвращающее ошибки оценивания из-за склонности составителей тестов усложнять их содержание путем наслоения в одном кадре нескольких тестов, да ещё и разного уровня. В этом случае природа ошибки, если она случается, часто не может быть объяснена.

Требование *надёжности теста* состоит в том, что оценивание должно производиться батареей тестов с надёжностью, допустимой для данной цели контроля качества обучения.

Иногда говорится о скоростных тестах - это те же тесты достижений с ограничением времени на тестирование. Ограничение времени превращает тест достижений в скоростной тест.

Выставление оценки качества знаний учащегося должно производиться в зависимости от достигнуто учащимся места на траектории процесса усвоения деятельности и по шкале, соответствующей природе оцениваемого признака. Мы везде используем 12 - балльную шкалу порядка, соответствующую процессу восхождения учащегося по уровням усвоения деятельности. Оценка по 12-балльной шкале однозначно показывает позицию учащегося и качество его знаний в полном диапазоне возможной человеческой деятельности: *от ученической до творческой*.

4. Анализ тестовой системы SAT

Тестовая система SAT является самой распространенной в мире: ее услугами пользуются более 200 стран в мире, а тираж ее тестов достигает 12 миллионов. Многие страны, не вошедшие в упомянутые 200, включая Россию, создают (копируют) свои тестовые системы под большим влиянием методики, разработанной для SAT.

Тест SAT I

Аббревиатура SAT 1 расшифровывается как Scholastic Aptitude Test, что в относительно свободном переводе означает «Тест учебных способностей». Авторы этого теста предполагают, что каждый человек в процессе обучения в зависимости от его врожденных способностей приобретает некоторые обобщенные навыки учебной работы (понимать прочитанное, находить решение проблем, излагать свои мысли и пр.). При этом прямо указывается, что эти навыки не обязательно усвоены в процессе обучения в школе, они могут быть приобретены самыми разными путями, которые предоставляются человеку его реальной жизнедеятельностью. С целью выявления таких обобщенных навыков и их уровня создается данный трехчасовой тест. В него включается материал из различных предметов, не обязательно изучаемых в школе, но который, по утверждению авторов теста, хорошо диагностирует речевые и математические способности учащихся и тем предсказывает их будущие успехи в вузовском обучении. Испытуемый в подавляющем числе предлагаемых тестом заданий должен догадываться, не опираясь на конкретные знания предмета, в чём состоит правильное решение или ответ. Обоснованием такого типа тестирования, считают его авторы, является тот факт, что в США много школ с низким уровнем образования. Талантливые дети из этих школ не могут соревноваться на приемных экзаменах в вузы по уровню усвоения конкретных знаний с выпускниками «хороших» городских школ. Но они, обладая хорошими учебными способностями, могут стать достойными студентами. Задачей теста SAT 1 является выявление таких способностей. Никакой психолого-педагогической теории способностей, на которую опирается построение данного теста, в литературе не приводится. Таким образом, данный тест, хотя и называется тестом достижений, но «подпольно» он переводится в категорию известных тестов интеллекта, отражающих умозрительную позицию их авторов. Утверждается, что SAT надежно предсказывает будущие успехи студента вуза на первом, критическом году обучения. Это утверждение авторов теста входит в катастрофическое противоречие со всей системой школьного образования: если можно

успешно учиться в вузе только за счет общих способностей, то для чего вообще необходимо все 12-летнее общее среднее образование? Тем не менее широко вещательные обещания теста привлекли внимание вузов, и тест одно время считался популярным мерилом подготовленности студентов к обучению в вузе. Каждый год более миллиона абитуриентов американских вузов соревновались по этому тесту за право обучаться в престижных вузах США. В то же время сами вузы не получали подтверждения предсказаний теста, а даже наоборот — успешность обучения принятых в вуз по этому тесту оставляла желать лучшего. Наиболее точное исследование этой проблемы было предпринято в Калифорнийском университете, где достоверно установлено, что предсказательная сила теста SAT 1 (21 %) не превышает таковой по обычному тесту достижений SAT II, диагностирующему качество усвоения школьниками учебных предметов. Калифорнийский университет выступил с инициативой отказаться от теста SAT 1 как вступительного испытания к поступлению в университет. За этим последовало значительное сокращение его использования в США. На первое место стал выдвигаться предметный тест SAT II, который дается всего на один час и используется вузами не столько как приемный критерий, а скорее, как материал для распределения студентов по направлениям подготовки.

Педагогический анализ тестов SAT I и II опирается на содержание многочисленных пособий, излагающих некоторый набор правил и рекомендаций построения этого типа тестов. Ознакомление с этими пособиями со всей наглядностью показывает, что тесты SAT являются примитивными творениями житейского эмпиризма, т.к. не имеют под собой никакой психолого-педагогической базы. Так, тест SAT 1 направлен в основном на лингвистические и математические способности испытуемых, но оставляет в стороне многие другие их способности, важные для обучения в вузе. Если учесть, кроме того, что оценивать одним и тем же тестом и артистов, и мыслителей не имеет никакого смысла, т.к. и те и другие по SAT 1 могут быть зачислены в один и тот же вуз, к которому у них диаметрально противоположные способности. Не это ли обстоятельство повлекло за собой недовольство вузов прогностическими возможностями теста. Ничего, кроме как качество разгадки самого теста испытуемым, о тесте сказать нельзя. В этой же роли могли бы выступить любые наборы загадок, которыми богат народный фольклор. Подобными загадками наполнены все три части теста: его лексическая, математическая и аналитическая часть”.

В *лексической части* проверяется владение учащимся английским языком, но каким английским? Не современным, общеупотребительным литературным английским, а специально подобранной смесью из редко употребляемых, архаичных и жаргонных слов, значение которых часто неизвестно взрослому и образованному носителю языка. В этом нам удалось убедиться в естественном мини-эксперименте. В пособии для учащихся, готовящихся к тесту, опубликован список из 100 «трудных слов», которые могут встретиться учащимся в тесте. Мы показали эти слова полудюжине образованных американцев. В среднем смысл 25% этих слов оказался для них неясным.

Неудивительно, что более половины выпускников американской старшей школы проваливают лексический тест. Это объясняется тем, что школьные предметы не содержат эту лексику, а следовательно, тест является содержательно инвалид-

ным. Авторы утверждают, что тест SAT 1 выявляет талантливых абитуриентов, но не объясняют, какой талант выявляется при угадывании значений ранее незнакомых слов родного языка. При полуторамиллионном словаре английского языка совсем не сложно найти две-три-четыре сотни редко употребляемых или вовсе вышедших из употребления слов, чтобы поставить в тупик рядового гражданина страны.

Математическая часть теста также изобилует загадками, видимо, для выявления математических талантов, но это делает математическую часть теста также со-держательно инвалидной.

В батарею SAT 1 включаются тесты логического мышления, которые также расцениваются авторами теста как диагностический инструмент для выявления талантливых абитуриентов. Не вдаваясь в психологические софизмы и спекуляции о том, что такое логическое мышление, авторы этого вида тестов формулируют многоходовые загадки, называя их решение признаком развитого логического мышления.

Вот типовая загадка:

Произошло восемь событий: А, Б, В, Г, Д, Е, Ж, З. Укажите их последовательность, если: Е является третьим; Ж не было четвертым; Е и Ж являются соседними в последовательности; два события произошли перед А; А совершилось перед Г.

Какое отношение эта и подобные завихренные загадки, предлагаемые тестом, имеют к нормальному человеческому логическому мышлению, необходимому для успешного обучения в вузе? Это ни компанией ETS, ни авторами многочисленных пособий по натаскиванию абитуриентов на преодоление теста не объясняется. Более того, наличие таких пособий и их рекламные утверждения об эффективности излагаемых в них тренингов для преодоления теста начисто перечеркивают весь тест. Действительно, если можно натаскать любого выпускника школы на успешное преодоление теста, то утверждение его создателей о том, что он выявляет талантливых, а не знающих, полностью опровергается, поскольку талант — это врождённое качество, которое ни подавить, ни создать на пустом месте невозможно. И наконец, последнее. Если можно натаскать на этот тест слабоуспевающих выпускников провинциальных, запущенных школ, то тем успешное на него натаскаются успешные выпускники из школ больших и развитых городов, и в итоге исходная посылка этого теста — уравнивать шансы и тех и других тестом, не зависящим от школьного образования, — терпит сокрушительное фиаско. Что и происходит сейчас на практике.

Перейдем теперь к анализу предметных тестов достижения SAT II.

Эти тесты, как утверждают их создатели, строятся на базе школьных учебных программ и соответствуют их предметному содержанию: математике, биологии, химии, английскому языку и другим предметам. О содержательной валидности этих тестов можно говорить с некоторой осторожностью и в определенной степени. Почему «осторожно и в определенной степени»? Потому что их содержательная валидность не доказывается создателями тестов, а поэтому велика вероятность ее нарушения.

Что же касается их функциональной валидности, то здесь, вне всяких сомнений, не было сделано никаких усилий для ее обеспечения. Из-за недиагностичности

целей обучения в американской общеобразовательной школе не представляется возможным соблюдать требование функциональной валидности теста. Что измеряет САТ II тест? Он измеряет степень усвоения знаний учащимися, но как?

В тестовой системе САТ тесты разделяются на «легкие», «средние» и «трудные», но никаких психолого-педагогических различительных признаков трудности или легкости тестов не приводится. О трудности-легкости судят по числу учащихся, преодолевших данный тест. Но это уже зависит не от теста, а от качества обучения: при изменении качества обучения трансформируется принадлежность тестов к категории «легкий-трудный».

К категории легких тестов относятся такие тесты, на которые большинство студентов отвечают правильно. К категории тестов средней трудности относят тесты, на которые только около половины учащихся отвечают правильно. И, наконец, к трудным относят тесты, на которые большинство студентов отвечают неправильно. Такая категоризация тестов не имеет ничего общего с корректной методикой измерения и оценки качества знаний учащихся, т.к. при соблюдении требований содержательной и функциональной валидности тестов, их соответствия целям обучения и успешном учебном процессе все тесты должны быть легкими и посильными для всех учащихся. В этом состоит задача разумного обучения. Если же половина или большинство учащихся не в состоянии преодолеть тест, то это не означает, что тест является слишком трудным и надо снижать его трудность. Это абсолютно точно указывает, что цель обучения ещё не достигнута и надо продолжать процесс обучения.

Тесты САТ не проверяют степень усвоения предмета, а ставят ловушки учащимся, из которых нет метода благополучного выхода на базе знания предмета.

Читатель еще больше удивится, если узнает, что в инструкции по оценке знаний учащихся говорится: «каждый правильный ответ оценивается одним баллом, независимо от того, легок ли тест или труден». Возникает комическое недоумение: к чему же тогда вообще «огород городить» из понятий «лёгкий — трудный», если преодоление трудных тестов не дает никаких преимуществ испытуемому?

В США и ряде стран мира, включая Россию, тестирование похожими тестоподобными материалами приняло огромный размах и превратилось в очень прибыльный бизнес.

Все зарубежные тестовые системы в настоящее время имеют один и тот же монопольный источник и построены по одному и тому же принципу: запутать учащегося, чтобы он не мог угадать правильный ответ. Но что же тогда учащемуся делать, если от него требуется угадывать, а не знать предмет? И выход остается один: в сторону предмет, в сторону систематическое изучение предмета и все силы и средства — на натаскивание на возможные ловушки и загадки одиозного вопросника под названием САТ или любой другой трудно расшифровываемой аббревиатуры. И возникает море коррумпированных возможностей для ловкачей и криминальных бизнесов, питающихся педагогическим невежеством деятелей образования.

В названных тестах также используются задания различной конструкции, которые на первый взгляд могут напоминать тесты-лестницы. На самом же деле они ничего общего с тестами-лестницами, в нашем смысле этого слова, не имеют, т.к. эти

тесты созданы безо всякой ориентировки на обоснованную систему иерархически построенных параметров качества обучения. Если даже по некоей случайности благодаря авторской интуиции предлагаемый набор тестов и содержит некоторую иллюзорную иерархию тестов, типа «лёгкий — трудный», то все равно эта иерархия не строится осознанно и не принимается в расчет при оценке работы испытуемого. Общая сумма баллов, которой оценивается качество исполнения теста, абсолютно некорректна (отсутствие аддитивности в наборе тестов) и абсолютно неинформативна. Такая оценка знаний испытуемого не только не отражает четко выраженного понятия «качество обучения», но и не дает достоверной информации о направлении совершенствования процесса обучения в случае неудовлетворительных показаний тестов.

5. Тестовая система TOEFL.

Эта система называется The Test of English as a Foreign Language, в аббревиатуре TOEFL.

5.2. Фактическая справка

Тестовая система TOEFL разработана той же фирмой ETS для оценки качества усвоения английского языка иностранцами, желающими учиться в американских высших учебных заведениях. Она используется по всему миру миллионами абитуриентов и приносит ее владельцам солидную прибыль, которая умножается примитивным манипулированием требованиями содержательной и функциональной валидности тестов, а также запутанной шкалой оценки.

Несмотря на то, что эта система тестирования используется во всем мире (180 стран) в течение более 30 лет, изложение ее педагогических основ автору этой лекции не удалось обнаружить ни в одной из известных публикаций по проблеме измерения качества знаний учащихся. Не обнаружены также мною отчёты о проверке этой системы в корректном педагогическом эксперименте. Мои запросы по этому поводу в ETS остались безответными. Имеющиеся публикации носят скорее информационный, чем исследовательский характер и не затрагивают педагогического фундамента системы. С тем же успехом, как авторы TOEFL представили свое твердение как тест проверки знаний английского языка иностранцами, можно было бы создать еще дюжину тестов, опирающихся на пресловутый здравый смысл, и тоже выйти с ними на тестовый рынок. Но бизнесмены от TOEFL были первыми.

5.3. TOEFL как он есть

Испытуемым предлагается четыре набора тестов:

- 1) Listening — слушание (аудирование);
- 2) Structure — структура (элементы грамматики);
- 3) Reading — чтение и понимание текста;
- 4) Writing — сочинение на свободную тему.

Дается следующее разъяснение смысла каждого из тестов:

«Тесты слушания проверяют вашу способность понимать английский язык, употребляемый в США и Канаде. Структурные тесты проверяют вашу способность распознавать правильность написания на стандартном английском. Тесты чтения

проверяют способность понимать тексты общего (не специального) содержания. Письменный тест проверяет вашу способность писать на стандартном литературном английском. Все четыре набора тестов, содержащих от 150 до 230 заданий, даются испытуемому на три часа с одним перерывом на 10 минут.

После выполнения теста учащемуся выставляется общая оценка, которая якобы характеризует его знание английского. По этой оценке учебное заведение решает, принимать этого абитуриента или не принимать. Никакого критерия для такого решения не существует: одни колледжи требуют наивысшего балла, а другие, не понимая конкретного смысла этого балла, могут принять абитуриента и без TOEFL с последующим изучением английского в самом колледже.

5.4. Конструкция тестов TOEPI

Тесты аудирования предлагаются в трех вариантах: в виде краткого диалога; развернутого диалога (несколько фраз) и короткой лекции.

В тестах чтения испытуемому предлагается для чтения пять коротких текстов по 120—250 слов каждый. После прочтения каждого текста испытуемый выполняет 12 тестов распознавания. Задача испытуемого усложняется тем, что в составе предлагаемых текстов содержится множество редких и специальных слов и учащемуся предлагается не только перевести текст, но и распознать скрытый в нем смысл.

В тестах-сочинениях необходимо показать владение языком для изложения своих мыслей. Необходимо в ограниченном по объему сочинении (200—300 слов) на свободную тему показать свою способность критически мыслить, оригинальность воображения и богатство словаря.

5.5. Педагогический анализ TOEFL.

Итак, система тестов TOEFL создана для оценки способности иностранных абитуриентов американских и канадских вузов читать и понимать стандартный устный и письменный английский язык, владеть его грамматикой и уметь им пользоваться для выражения своих мыслей.

Возникает вопрос: достаточно ли адекватна предлагаемая экзаменационная процедура цели экзамена?

Всякое измерение качества знаний учащихся должно удовлетворять следующим педагогическим требованиям:

- 1) быть содержательно валидным, т.е., осуществляться на ранее изученном множестве УЭ (чему учили, то и надо проверять);
- 2) быть функционально валидным, т.е., соответствовать цели изучения заданных УЭ;
- 3) быть простым, т.е., содержать одну задачу данного уровня;
- 4) быть надежным для точных суждений.

Прежде всего содержательно ли валидны, тесты TOEFL? Чтобы ответить на этот вопрос надо сопоставить содержание тестируемого знания с содержанием применяемых тестов. К сожалению, такое сопоставление сделать абсолютно невозможно, поскольку система TOEFL не ограничивает себя ни лексическим, ни грамматическим минимумом, на котором строятся тесты. Поэтому провозглашенная цель по-

нимать английский язык и уметь им пользоваться является безразмерной даже для носителя языка. Не случайно же американцы с молодых ногтей и до седых волос изучают свой язык, как это делают и русские, и немцы, и французы, и представители любой другой национальности. Любой национальный язык состоит из миллионов неслучайных лексических единиц, моделей, правил и исключений. Даже лучшие национальные писатели и поэты используют в своем творчестве ничтожную долю национального языкового богатства. Что же говорить об абитуриенте-иностранце, который готовится в специализированный вуз чужой страны, где даже самый речистый профессор использует в своем лексиконе пару-тройку тысяч стандартных лексических единиц? Изучению языка в его неизмеримом богатстве посвящают целую жизнь поколения учёных, и они уходят из жизни, так и не завершив изучение этой вечно развивающейся материи. Можно ли требовать от иностранного абитуриента безразмерного знания чужого языка? Безразмерная цель по объёму УЭ порождает и безразмерное содержание экзамена, и ни о какой содержательной валидности тестов системы TOEFL говорить не приходится. Экзамен вырождается в примитивную лотерею-ловушку с огромной вероятностью провала испытуемого в каждом тесте. Трудно себе представить, чтобы авторы TOEFL не были знакомы с требованием содержательной валидности теста, поскольку об этом говорится даже в самых примитивных и рецептурных пособиях по разработке тестов достижения. Напрашивается вопрос: не по коммерческим ли соображениям авторы этой тестовой системы преднамеренно и грубо нарушают это требование? Но тогда это уже не педагогическая, а моральная проблема, решение которой находится за пределами педагогики. Игнорируя требование содержательной валидности тестов, экзаменаторы лишают испытуемых законной возможности целенаправленно подготовиться к экзамену, им остается лишь надеяться на знаменитый русский «авось». Не поэтому ли даже хорошо подготовленные выпускники средней школы неизменно получают довольно низкие оценки по TOEFL?

Для исправления этой педагогической ошибки в тестовой системе TOEFL необходимо:

а) точно определить и задать тот лексический минимум, который используется преподавателями в типичном американском учебном заведении и авторами соответствующих учебников. Исследования показывают, что 3—4 тысячи лексических единиц в прикладных науках и огденовские 800 единиц общежитейской лексики являются хорошим базисом для развития языкового сознания учащегося и его дальнейшего языкового развития. Этот словарь-минимум по всем педагогическим принципам образования должен быть доступен учащимся для подготовки к экзамену;

б) точно также должны быть определены типовые грамматические формы, используемые в соответствующих науках, и модели языка, которые учащиеся должны усвоить.

Эти исходные данные, будучи доступными для учащихся, могут обеспечить содержательную валидность TOEFL.

Содержательная же инвалидность TOEFL усугубляется еще и тем, что его авторы, по тем же соображениям, что и в системе SAT, искусственно усложняют тесты путем включения в них каждый год новых специально подобранных 300 слов нестан-

дартной и редко используемой лексики. Как выше показано, это может поставить в тупик не только абитуриента, но и носителя языка. При этом авторами TOEFL неоднократно подчеркивается, что их тесты построены на базе стандартного английского, используемого в США и Канаде. Что это? Лицемерие или незнание собственного продукта?

Теперь рассмотрим, насколько функционально валидны тесты TOEFL.

Для этого произведем сопоставление деятельности, задаваемой тестами TOEFL, с реальной деятельностью студента в американском вузе. Эта деятельность состоит в слушании лекций, чтении учебной литературы, общении на занятиях и вне них, написании курсовых и дипломных работ. Выполняя все эти виды деятельности, студент не получает ни от кого ни помощи, ни подсказки, а действует на основе той информации, которая содержится у него в памяти. Это значит, что студент должен обладать способностью читать, понимать, общаться и писать на английском языке как минимум на втором уровне усвоения.

При этом в анализе структуры тестов TOEFL всплывает еще одна несуразность: при проверке понимания студентом английских текстов он переводит их с английского на английский, а на свой родной язык, на котором он может точно и полно показать свое понимание иностранного языка. Где гарантия, что при таком «дубль-тестировании» не накапливаются ошибки высказывания при двойном переводе: с английского на русский и с русского на английский?

Также весьма странным кажется условие выполнения теста: «Вы прослушаете диалог только один раз. Их не будут повторять вам не разрешается делать заметки». Оба ограничения противоречат естественной ситуации общения. В любой момент диалога не будет бестактным переспросить: «Простите, не понял». Запрещение повторить высказывание, которое должно быть понято испытуемым, во-первых, вносит ненужную дополнительную нервную напряженность в экзаменационную процедуру, а во-вторых, превращает тест достижения в тест памяти и внимания.

Запрещение же делать заметки и вовсе странно, поскольку собственные исследования работников корпорация TOEFL показывают, что, делая заметки, испытуемые получают «little benefit». Вот почему даже профессиональные переводчики делают заметки в процессе синхронного перевода. Почему же абитуриента лишать этого «little benefit»?

Анализ функциональной валидности тестов TOEFL начнем с тестов понимания устной речи. Ознакомление с этими тестами немедленно показывает, что все они выполнены на первом уровне усвоения в виде тестов различения. Следовательно, тесты на понимание устной речи функционально не валидны будущей деятельности абитуриентов, качество которой они призваны оценить. Они ориентируют преподавателей языка и студентов на заниженный уровень усвоения языка, чем создают для последних затрудненные условия адаптации в англо-говорящей среде и возможную перспективу академической неуспеваемости по этой причине.

В тестах распознавания правописания тесты подстановки сведены к первому уровню деятельности на различение правильного ответа из четырех предлагаемых вариантов.

Тесты по нахождению ошибки в предложении могут быть отождествлены с типовой задачей, т.е., валидны второму уровню усвоения только в том случае, если они потребуют от испытуемого не только распознать ошибку (первый уровень), но и исправить её.

Тесты на чтение и понимание прочитанного требуют от учащегося логической и семантической интерпретации текста, для выявления их глубинного смысла. Это делает эти тесты валидными третьему уровню усвоения, что является излишним усложнением тестов для абитуриентов, не нацеленных на филологические отделения университетов.

То же самое относится и к тестам-сочинениям. Сочинение в принципе не относится к сфере языковой деятельности. Это деятельность литературная, основными компонентами которой являются не лексика, а семантика, грамматика, логика, воображение и мышление. Все эти качества имеют отношение к личности писателя и его литературному мастерству, а не к его владению письменной речью. Об этих тестах, особенно сочинениях на свободную тему, выполняемых экспромтом в ходе экзамена, можно говорить, во-первых, как о содержательно инвалидных заданиях, и, во-вторых, как о заданиях четвертого уровня усвоения. В обоих случаях они неуместны для экзамена, который проверяет владение абитуриентом стандартным английским языком. Если абитуриенту в будущем и придётся писать какие-либо сочинения, то это всегда будут содержательно валидные задания второго уровня усвоения и никогда задания на свободную тему (четвертый уровень).

Много вопросов вызывают также критерии оценки качества знаний испытуемых. «Ураниловка» в оценке разноуровневых тестов и сумма баллов за их выполнение, во-первых, недопустима по математическим соображениям (отсутствие аддитивности), а во-вторых, не даёт объективной картины успеваемости и оснований для правильных суждений при приеме в вуз.

Немаловажно и то, что надёжность батарей тестов разного вида деятельности чрезвычайно низка (30—40 %), что также резко снижает достоверность суждений по результатам теста.

И конечно же полный волюнтаризм в назначении времени на выполнение батарей тестов разного уровня: 50 тестов первого уровня — 30 мин, (полминуты на тест); 40 тестов второго уровня — 25 мин. (тоже полминуты); 60 тестов 2—3 уровня — 55 мин. (почти минута на тест). Такая жёсткая регламентация для проверки знания иностранного языка ничем не оправдана: испытуемому может быть предоставлено столько времени, сколько ему необходимо, для завершения всего теста без спешки и нервного напряжения. Эта рекомендация поддерживается специальным исследованием фактора времени для качества выполнения теста.