



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. О Едином государственном экзамене речь пойдет в следующей лекции.

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*. Можно схематически представить структуру теста-лестницы в следующем виде (где Т — тестовая батарея):

$$T\alpha_1 \beta_1 \tau_{0.5} + T\alpha_2 \beta_1 \tau_{0.5} + T\alpha_3 \beta_1 \tau_{0.5} + T\alpha_1 \beta_2 \tau_{0.5}, \text{ etc.}$$

Здесь  $T\alpha_1 \beta_1 \tau_{0.5}$  означает, что это тестовая батарея *первого* уровня усвоения, выполненная на *первой* ступени абстракции, требующая автоматизации усвоенных действий с  $K\tau = 0,5$ . В тесте-лестнице параметр «уровень усвоения» ( $\alpha$ ) символизирует *общую высоту* лестницы, параметр «ступень абстракции» ( $\beta$ ) — *высоту* лестничной ступеньки, а параметр автоматизации ( $\tau$ ) символизирует *скорость* восхождения по лестнице. В приведенной символической формуле теста-лестницы ее максимальная высота равна  $\alpha_3$ , высота ступеньки в лестнице меняется с  $\beta_1$  на  $\beta_2$  и ос-

таётся неизменным требование скорости перемещения по ней.

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

Выбор надёжности тестирования зависит, как и в случае с коэффициентом автоматизации, 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  $\mu$  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 и определяется



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) *надежности теста;*
- 5) *корректности скоростного теста;*
- 6) *шкалой оценки, сопряженной с процессом формирования знаний.*

**Содержательная** валидность теста, как упоминалось выше, 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. Казалось бы, вполне простое и очевидное требование. 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 графа учебного предмета с наглядным представлением структуры и номенклатуры учебных элементов в предмете позволяет практически полностью избежать оши-

бок 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*». Конечно, этот афоризм уместен только в том случае, когда цель задана *диагностично*. 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 наслаивания в одном кадре нескольких тестов, да ещё и разного уровня. В этом случае природа ошибки, если она случается, часто не может быть объяснена.



по этому тесту оставляла желать лучшего. Наиболее точное исследование этой проблемы было предпринято в Калифорнийском университете, где достоверно установлено, что предсказательная сила теста SAT I (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 I 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 (Ebel, R.L.; Childs, R.A.), 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, которой бы постулировался некий процесс усвоения опыта человеком и следствием которого была бы строгая и однозначная методика тестирования состояния этого процесса. Так, тест SAT I 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 I 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. Понятно, что при отсутствии какой-либо членораздельной теории общих способностей и представления об их природе и процессе развития, невозможно определить ни содержательную, ни функциональную валидность предлагаемых SAT I тестов. Их создание и применение опирается на волюнтаристские «придумки» авторов, а результаты не сопоставимы ни с какими объективными характеристиками человеческого опыта, а поэтому 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 части.

В лексической части проверяется владение учащимся английским языком, но каким английским? Не современным, общепотребительным 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» (difficult words), которые могут встретиться учащимся в тесте. Мы показали эти слова полудюжине образованных американцев. В среднем смысл 25% этих слов оказался для них неясным. Когда же мы сделали выборку этих 25% неясных американцам слов:

Garrulous –	Assuage –
Buttress –	Desiccate –
Diatribе –	Guile –
Aberrant –	Prevaricate –
Propitiate –	Aver –
Culpability –	Turpitude –
Obviate –	Occlude –
Abscond –	Wheedle –
Harangue –	Insipid –
Inchoate –	Rarefy –

и стали показывать эти слова более широкой американской публике, то только «чемпионы» давали смутное объяснение смысла не более чем половине этих слов.

Неудивительно, что более половины выпускников американской старшей школы (high school) проваливают лексический тест. Это объясняется тем, что школьные предметы не содержат эту лексику, а следовательно, тест является *содержательно инвалидным*. Авторы утверждают, что тест SAT I выявляет талантливых абитуриентов, но не объясняют, какой талант выявляется при угадывании значений ранее незнакомых слов родного языка. При полуторамиллионном словаре английского словаря совсем не сложно найти две-три-четыре сотни редко употребляемых или вовсе вышедших из употребления слов, чтобы поставить в тупик рядового гражданина страны.



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 успешнее 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, но тестовый бизнес не сдаётся. Скомпрометированный тест SAT I переиздается под разными другими названиями-аббревиатурами (CAT, GRE, GMAT и др.), но педагогическая их несостоятельность от этого не меняется. Не снижаются и многомиллионные барыши таких компаний, как KAPLAN, BARONS и др., которые спекулируют на педагогической некомпетентности как родителей, которые платят за тестирование, так и вузовских и школьных администраторов, подвергающих миллионы выпускников школ ненужным и иезуитским испытаниям.

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 SAT II тест? Он измеряет степень усвоения знаний учащимися, но как?

В тестовой системе SAT 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.

Посмотрим, как критерии лёгкости и трудности тестов применяются в реальных тестах SAT. Вот два теста, представляющие собой примеры этих разновидностей. Для большей наглядности и точности они приводятся на языке оригинала:

**Easy** (легкий): *If 20% of p is 10, then 10% of p is:*  
A) 2; B) 4; C) 5; D) 8; E) 14.

**Hard** (трудный): *If 200% of 40% of X is equal to 40% of Y, then X is what percent of Y?*  
A) 10%; B) 20%; C) 30%; D) 50%; E) 80%.

В чем состоит различие между этими тестами? Оба они – тесты второго уровня, представленные в виде псевдотестов первого уровня. Оба содержат одно и то же число существенных операций (две: 1) усмотреть

различие; 2) вычислить разницу) и даже относятся к одному и тому же УЭ – вычисление процентов. Почему же первый тест отнесен к категории «лёгкий», а второй к категории «трудный»? Ответ на этот вопрос очевиден: учащиеся плохо подготовлены в элементарной арифметике, что к УЭ «Проценты» отношения не имеет. Следовательно, «трудный» тест содержательно инвалиден. Этот способ создания трудных тестов (нарушение валидности) сплошь и рядом используется авторами тестов SAT. Вот как создается «очень трудный» тест по математике (опять на языке оригинала для большей наглядности):

*Graham walks to school at an average speed of 3 miles an hour and jogged back along the same route at 5 miles an hour. If his total traveling time was 1 hour, what was the total number of miles in the round trip?*

a) 3; b) 3 1/8; c) 3 3/4; d) 4; e) 5.

Опять тест второго уровня, представленный в виде псевдо-теста первого уровня. Тест содержательно инвалиден, поскольку относится к предмету «Физика», а не «Математика». Чтобы решить этот тест, учащиеся должны знать формулу связи пути, скорости и времени в равномерном движении из курса физики, где эта формула изучается. А если они эту формулу еще не изучали или не усвоили должным образом? Тогда получается «очень трудный» тест по... математике, поскольку по предположению автора теста испытуемый должен «угадать» правильный ответ, а не вычислять его. Вот как по предположению автора должен протекать этот процесс угадывания (буквальный перевод из «Cracking the NEW SAT»):

*Джо (имеется в виду среднеуспевающий американский школьник) изберет ответ «d» на этот вопрос, потому что ответ 4 кажется ему правильным ответом. Джо думает (?), что тест требует от него вычисления средней скорости движения Грехам и 4 — это среднее между 3 и 5. Но Джо не догадывается (?), что он должен учитывать, что Грехам перемещался с разной скоростью в каждом направлении. Его путь в школу длился дольше, чем путь обратно домой. Ответ будет 4, только если Грехам шел полчаса в школу со скоростью 3 мили в час и бежал полчаса обратно домой со скоростью 5 миль в час, но это не так. Ответ «d» это ловушка: авторы тестов ETS включают их среди ответов, поскольку они знают, что такие ответы будут казаться правильными среднему студенту. Они включают*

*ловушку среди ответов, чтобы сделать тест трудным, а не лёгким. (Правильный ответ, между прочим, это «c»). Грехам тратит больше времени пешком, при 3 милях в час, чем бегом, при 5 милях в час, поэтому ответ должен быть несколько меньше, чем 4.*

Значит, данный «трудный» тест является трудным, поскольку учащийся должен догадываться, а не вычислять правильный ответ. Что же тогда проверяется этим тестом: знание предмета или изворотливость ума учащегося? Ведь с этой же логикой рассуждений, без вычислений он мог бы выбрать и ответ «b», который несущественно отличается от ответа «c».

Таким образом, тесты SAT не проверяют степень усвоения предмета, а ставят ловушки учащимся, из которых нет метода благополучного выхода на базе знания предмета.

Если показанный тест считается трудным, то как же выглядит легкий тест? В цитированном пособии показывается, как можно сделать этот тест легким. Вот как выглядит тот же тест, но в «легком» варианте:

*Graham walks to school at an average speed of 3 miles an hour and jogged back along the same route at 5 miles an hour. If his total traveling time was 1 hour, what was the total number of miles in the round trip?*

a) .01; b) .1; c) 3 3/4; d) 1,000; e) 10,000.

И далее, вполне серьёзно утверждается, что тест из трудного преобразуется в легкий «by writing different answer choices (путем изменения вариантов ответа)». Вот и вся педагогическая теория, стоящая за тестами SAT!

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. Целые корпорации производят тесты типа SAT, TOEFL, GRE, ими наводнен тестовый рынок и потребителю не оставляется выбора в море этой сомнительной продукции. Не отстают от своих американских учителей и российские

предприниматели – благо ни учителя, ни родители не понимают педагогической невежественности и психологического вреда, сопровождающих такое тестирование.

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 SAT 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. Так, собственно, и обстоит дело сегодня как в Американской, так и в российской школе: итоговые тесты показывают неудовлетворительное качество знаний учащихся, но никто ничего толкового не может сказать, в каком направлении совершенствовать образование, кроме как «*того же и побольше!*». Полученные показатели характеризуют только тест, а не качество образования. Педагоги не смогли сделать правильные выводы о направлении совершенствования

процесса обучения и пошли по проторенному, но заведомо бесперспективному пути: в США — долларовой впрыск; в России — «*увеличить, усилить, навалиться!*» То есть все та же «*Эх, дубинушка, ухнем!*». Информация, полученная с помощью волюнтаристски построенных итоговых тестов, оказалась столь же бесполезной, как если бы бюро прогнозов погоды объявило, что на завтра ожидается ухудшение погоды, но не объяснило, в чем это ухудшение состоит: будет дождь или усилится жара, усилится мороз или наступит оттепель с гололёдом. Также и тесты показали, что учащиеся неудовлетворительно выполнили итоговый **тест**, но какое это имеет отношение к качеству обучения, выяснить невозможно, т.к. тест и цель обучения между собой не сопряжены. Не случайно успехи учащегося по тестам SAT характеризуются абстрактной суммой баллов, не отражающей никакой известной шкалы оценки. *Это баллы за искусственно созданный тест, а не за реальное качество знаний*. Подтверждением этому выводу является статья в газете USA TODAY, в которой выражается общее недоумение резким падением оценок абитуриентов по новому тесту SAT, введенному College Board в прошлом (2005) году. И это недоумение понятно: не может же резко, в течение одного года измениться качество знаний учащихся, обучающихся в тех же самых школах и у тех же самых учителей! А тест показывает резкое ухудшение, и все недоумевают, почему бы это? Здесь уместна старая российская присказка: «А ларчик просто открывался!» И его открывалкой опять являются понятия *содержательной и функциональной* валидности теста. В новом тесте ретивые его составители еще дальше ушли от валидности тестирования и получили желаемый результат: учащиеся, натасканные на прежний тест, стали проваливать его новый вариант. Как ответят учителя на это событие? Точно так же, как они уже неоднократно отвечали на причуды прежних SAT-ов: они начнут «натаскивать» учащихся на этот новый тест, оставив в стороне систематическое обучение учащихся содержанию учебных предметов. Провал в успеваемости учащихся показывает, что они попросту еще не успели перестроиться под него. Это для них единственный выход: когда цель обучения не поставлена диагностично, тогда целью обучения становится диковинным образом составленный тест. К примеру, школьники Техаса пожаловались, что у них проводится до *двадцати* тестовых



Как пишут авторы книги, «*The purpose of the test is to evaluate the English proficiency of non-native speakers of English who wish to study in colleges and universities where English is the language of instruction*» (р. 3). На всякий случай перевожу: «Цель теста состоит в оценке качества знаний английского языка иностранцами, которые хотят учиться в колледжах и университетах, где обучение осуществляется на английском языке».

Соответственно этой цели 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.

В упомянутой книге 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 TOEFL

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

Например, краткий диалог (в переводе на русский):

А: *Вы звонили Пете?*

Б: *Я позвоню ему, как только вернусь домой.*

Текст этот читается только один раз и далее предлагаются один или несколько тестов первого уровня на английском языке!

Вопрос: *Что имеет в виду Б?*

Ответы для выбора:

- 1) *Б позвонит Пете до возвращения домой;*
- 2) *Б позвонит Пете по возвращении домой;*
- 3) *Б позвонит Пете домой;*
- 4) *Б позвонит Пете завтра.*

А вот тест слушания в оригинале.

Разговор двух студентов, который испытуемый слышит и наблюдает на экране телевизора:

**Female:** *That statistics exam was really difficult.*

**Male:** *I'll say! I studied for it, but I hardly knew any of the answers.*

Затем, опять же с экрана телевизора, даётся следующий тест первого уровня:

*What does the man mean?*

- 1) *Although he studied, he did not take the test.*
- 2) *He was sure about the answers on the test.*
- 3) *He hardly made any mistakes on the test.*
- 4) *Although he studied, he did not do well on the test.*

Четвёртый ответ предполагается правильным.

**Структурные тесты** предлагаются в виде псевдотестов подстановки и тестов распознавания ошибки. И те и другие выполняются в виде тестов первого уровня. Вот пример теста подстановки:

*An instant camera can produce a photograph almost immediately after the picture \_\_\_\_\_*

*had been taken*

*is taken*

*is taking*

*took.*

Тест распознавания ошибки требует от испытуемого найти ошибку в предложении, написанном на стандартном английском:

*Three of the most important qualities of a safe driver are experience, attitude, and to concentrate.*

В обоих типах тестов деятельность учащегося не превышает первого уровня мастерства,

поскольку в тестах подстановки правильный ответ содержится в предлагаемых вариантах и его надо только распознать. В тесте распознавания ошибки ошибка только распознается, но не исправляется.

**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?

Исходя из изложенной в лекциях № 2 и 3 теории педагогических измерений 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 сде-

лать абсолютно невозможно, поскольку система TOEFL не ограничивает себя ни лексическим, ни грамматическим минимумом, на котором строятся тесты. Поэтому провозглашенная цель *понимать 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. Любая 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 объему 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 (Ogden, С.К.) 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 нестандартной и редко используемой лексики. Как выше показано, это может поставить в тупик и носителя языка, а не только абитуриента. 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 про-

верке понимания студентом английских текстов он переводит их *с английского на английский*, а не 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, и 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 корпорации TOEFL (Hale and Courtney, 1991) 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

