

Дидактические тесты как средство измерения

Михайлычев Евгений Аркадьевич — зав. кафедрой общей педагогики Таганрогского государственного педагогического института, научный руководитель Центра педагогических измерений Ростовского областного института повышения квалификации и переподготовки работников образования, профессор, доктор педагогических наук.

В систему современных средств контроля знаний, умений и навыков учащихся многие десятилетия входят дидактические тесты, используемые широко в зарубежной образовательной практике и возрождаемые в нашей стране. Но научно обоснованными диагностическими методиками принято считать только стандартизованные тесты, прошедшие экспериментальную апробацию. Это отнюдь не те комплексы диагностических заданий, которые публикуются в наших педагогических журналах или отдельными выпусками продаются как готовый научно-методический продукт.

Процесс создания, апробации и стандартизации предметного дидактического теста является полномасштабным теоретико-эмпирическим исследованием. Тест должен иметь чётко определённое место в системе диагностического контроля и выполнять диагностические функции, вписываясь в существующую в учебном заведении образовательную практику, стимулируя её совершенствование через систему оперативной обратной связи.

В условиях перехода России на новые, ещё апробируемые стандарты образования особо актуальным становится научно обоснованное сопоставление результатов обучения. Для этого необходима стандартизация средств контроля, их сопоставление на основе методологии диагностики.

Тесты в системе диагностического контроля знаний

Известно, что понятийный аппарат педагогики весьма далёк от совершенства. Многие, даже давно известные понятия аморфны, имеют различные, часто несовпадающие контексты. Нередко встречается отождествление разных по сути понятий, некорректно используется терминология. Эта общая для педагогики слабость её как науки полностью отражается и в дидактической тестологии — как в формирующейся заново, так и в давно сложившейся зарубежной. Особенно «болезненным» для практического решения диагностических проблем является отождествление понятий *тест, тестовое задание, субтест и батарея тестов*. Всё это зачастую именуется тестом (хотя реально это разные виды педагогической продукции).

Под *тестом* мы понимаем только научно разработанную, профессионально составленную диагностическую или исследовательскую методику, но не самодельные разработки. **Тест обязательно должен успешно пройти апробацию и стандартизацию (строгое определение процедур тестирования, обработки и анализа результатов, выработку нормативов, проверку на надёжность и валидность).**

Всё остальное, внешне напоминающее тест, целесообразнее считать наборами тестовых заданий, диагностическая ценность которых неизвестна, пока разработчиками не представлены данные, её подтверждающие. Здесь напрашивается аналогия с котом в мешке или консервной банкой без этикетки. *Тестовое задание — это диагностическое задание в виде задачи или вопроса (вариации-совокупности предложений), с чёткой недвусмысленной инструкцией к выполнению и обязательно с эталоном ответа или алгоритмом требуемых действий со стороны испытуемого.*

Субтест — составная, относительно автономная структурная часть теста, содержащая определённым образом систематизированную совокупность диагностических заданий. В профессионально составленных тестах (и психодиагностических, и дидактических) в основе каждого субтеста имеется определённая **шкала оценивания** того, что измеряет этот субтест. Причём основания и критерии оценок, технология оценивания в каждом субтесте могут быть

свои. Ранее разработчики стремились унифицировать эти шкалы для более лёгкой обработки тестов вручную или с помощью счётной техники. Внедрение в диагностику компьютерной техники качественно меняет ситуацию, особенно при тестировании в дисплейном классе (если тест составлен «под компьютер», а не просто перенесён с бланка на монитор). Умело составленная программа в секунды обрабатывает самые сложные шкалы, и диагност получает точные результаты. Компьютерная обработка бланковых тестов (при профессионализме программиста) также даёт не только качественные результаты по всем субтестам (сколько бы их ни было в тесте) при самых различных формах предъявления заданий обследуемым, но и обобщённые результаты по разнообразным подгруппам учащихся — по мальчикам и девочкам, сильным и слабым ученикам разных классов и т.д. Чёткое выделение субтестов возможно по нескольким основаниям.

Батарея тестов — это не любая их подборка, а комплекс скоординированных между собой методик, используемый в ситуации, когда для прогноза необходима именно данная серия специально подобранных тестов. При использовании батареи тестов надо решить, как учитывать результаты отдельных тестов при выработке решения-диагноза в каждом индивидуальном случае. Основная сложность совмещения тестов в батарею заключается в том, что тесты, прошедшие апробацию и известные своими достоинствами, обычно составлялись не только в разное время (нередко с разрывом в десятилетия — из-за поразительной живучести некоторых из них), но и не по одной технологии. Не всегда эти технологии совместимы, ибо построены на разном понимании целей обучения, структуры личности и мотивов учения, ориентированы на различные диагностируемые базовые умения и навыки учебного труда, которые проявляются при решении тестовых заданий учащимися (например, в степени понимания текста инструкций, в планировании своей самостоятельной работы учеником).

Надо сказать, что и наши разработчики тестов (даже кустарных, самодельных) тоже используют разные подходы и технологии их составления и апробации. Совмещая московскую, ростовскую и марийскую методики в одну батарею, мы рискуем получить что-то вроде фантастического зверя Василиска — чудовища в виде дракона с петушиной головой, грудью и когтями орла, крыльями летучей мыши и змеиным хвостом. «Диагностический голод» приводит к тому, что всё издаваемое, более или менее похожее на тесты, нашими педагогами раскупается. Педагогическая пресса почти не отражает сложившуюся ситуацию, хотя очевидна необходимость строго научной экспертизы всего, что публикуется и поступает в продажу, что сбывается торговцами псевдонаучной макулатурой малокомпетентным в вопросах диагностики педагогам. Без обязательной **сертификации** тестовых материалов, используемых в учебном процессе, практически нет охраны прав педагога — потребителя диагностико-методической продукции и прав обучающихся на объективное оценивание их достижений от околонуучной халтуры, даже если она выпускается из лучших побуждений, «на безрыбье...»

За рубежом ситуация качественно иная. В США и других развитых странах «право на приобретение тестов обычно предоставляется лицам, имеющим определённую квалификацию, — к их числу относятся либо профессиональные психологи, либо специалисты, чьё образование может быть приравнено к психологическому». Как отмечает А. Анастаси, «некоторые издатели распределяют тесты в зависимости от уровня квалификации использующих их людей». При этом начальным уровнем являются тесты достижений в учебной деятельности как наиболее простые и менее способные навредить человеку*.

* Анастаси А. Психологическое тестирование. М.: Педагогика, 1982. С. 53.

Разумеется, такое положение было достигнуто не в одночасье — ему предшествовал вековой путь развития тестологии (и психологической, и педагогической). Причём первую половину пути этого (до 1936 года) Россия прошла вместе с другими цивилизованными странами в качестве одного из лидеров, нередко обгоняя зарубежных коллег. В чём же преуспели наши зарубежные коллеги в области педагогической и психологической диагностики, особенно в области стандартизации дидактических тестов?

В 1931 г. в монографии Л.Тэрстона «Надёжность и валидность тестов» были впервые системно сформулированы требования к стандартизации тестов (в основном психодиагностических). Применявшиеся в системе образования дидактические тесты до второй половины XX века чаще всего составлялись совместно с психологами либо самими психологами по методологиям, принятым в том или ином исследовательском центре психодиагностики. Сейчас дидакты и психодиагносты больше занимаются проблемами специфики **дидактической тестологии**. Кое-что в этой области уже достигнуто, несмотря на малочисленность профессиональных исследований.

Реальными достижениями дидактической тестологии в России можно считать:

- существенный перелом «антитестовой» и, по сути, антидиагностической психологии большинства учителей и педагогов системы профессионального и дополнительного образования, которым в курсах педагогики и её истории более полувека внушалось, что тесты— методологически порочный инструмент буржуазной науки, что у них тьма недостатков и главная их реальная функция— классовая селекция детей из трудящихся слоев для того, чтобы не допустить их к качественному образованию;

- высокий реальный интерес к тестам как к оперативному средству диагностики, переходящий порою в тестоманию. Это проявляется не только в массовом спросе на любую тестовую продукцию (чем пользуются недобросовестные издатели и разработчики), но и в стремлении самостоятельно разрабатывать и внедрять в учебный процесс тесты как средство контроля (не имея для этого достаточной методологической подготовки);

- интерес к методологии и технологии тестологии со стороны ряда учёных, активно начавших разработку и пропаганду методологических и научно-методических вопросов конструирования и апробации дидактических тестов (В.А. Аванесов, Н.Ф.Ефремова, А.И. Майоров, М.Б. Чельшкова, автор этой книги и ряд других). Появились первые диссертации по педагогическому тестированию и системам компьютеризированного тестового контроля (Л.А. Гиннис, Л.А.Сучкова), занимаются этой проблематикой способные аспиранты и соискатели — С.А. Софронцев, Н.А. Раннева и другие;

- проведение на федеральном и региональном уровнях семинаров и конференций по проблемам тестирования;

- создание и успешное функционирование в течение нескольких лет системы централизованного тестирования с профессионально прописанной технологией её организации, системы дистанционного тестирования «Телетестинг», проведение конкурсов на создание тестовых материалов и их использование в аттестационных технологиях;

- появление на рынке дидактической продукции всё больше тестовых материалов по самым различным учебным предметам (однако несертифицированных и без каких-либо достоверных данных об их успешной апробации как научно-методического продукта);

- оформление официальной системы требований к разработке и сертификации тестовых материалов, к их апробации, оформлению, экспертизе (приказ Министерства образования Российской Федерации от 17 апреля 2000 г. № 1122 «О сертификации качества педагогических тестовых материалов»).

- наконец, практическое признание самими учащимися тестов как средства контроля, более адекватного, чем традиционная учительская оценка.

Для разработки дидактической тестологии следует учитывать мнение опытных психодиагностов. В педагогике ещё не использован весь богатый психодиагностический опыт. Многие положения его (включая элементарные принципы и этические нормы диагноста) малоизвестны большинству педагогов массовой школы. И, к сожалению, очень часто нарушаются разработчиками тестов, порой ещё на стадии замысла методики, что приводит, как и игнорирование опыта сопредельной науки, к ошибкам.

На I Международном конгрессе по применению тестов в школе (Берлин, 1967) исходным теоретическим положением стратегии дидактической диагностики стало положение об отказе от традиционных тестов достижений, созданных на базе классической тестологии. Новая дидактическая тестология ориентирована на оценку достижения учеником дидактических

целей и качество его знаний — насколько хорошо он эти цели достиг. Традиционные тесты достижений не измеряли успеваемость школьника с учётом заданных ему целей обучения. Надо не столько определять место учащегося в его учебной группе, сколько «измерять в абсолютных величинах дистанцию, отделяющую ученика от поставленной цели...»*.

* Ингенкамп К. Педагогическая диагностика. М.: Педагогика, 1991. С. 111.

Следует оговориться — нет смысла отказываться от нормативных тестов. Во многих отношениях они не так уж и плохи, решают стоящие перед ними диагностические задачи. Тесты *учебных достижений* до середины XX века строились преимущественно как нормативные, показывающие место учащихся в учебной группе по нормам, выработанным на более или менее **репрезентативных выборках**. Это их сильная сторона в сочетании с данными надёжности и валидности, накопленными за десятилетия практики применения таких тестов. Нормативы для них разрабатывались по той же технологии, что и для большинства психодиагностических тестов, выявлявших индивидуальные особенности психических процессов, сенсомоторные реакции, относительно стабильные личностные качества.

Специфика формирующего личностные подструктуры педагогического процесса при этом зачастую либо игнорировалась (порой принципиально, с позиций господствовавшего в психотехнике классического бихевиоризма), либо просто недостаточно учитывалась в русле уже сформировавшихся традиций психотехнического измерения, стремившегося к абсолютной чистоте эксперимента. Но тестирование учебных достижений в реальном учебном процессе сопровождается множеством «внешних факторов» — от неподточенных вовремя карандашей, которыми надо отмечать варианты ответов, до врывающейся в аудиторию разгневанной мамы, стремящейся здесь и сейчас выяснить свои взаимоотношения с педагогом. Просчитать все составляющие сложных педагогических воздействий вряд ли возможно на уровне современных знаний о личности, педагогическом процессе и их измерении.

В 70-е годы происходит существенный сдвиг в рассмотрении целей и функций тестов учебных достижений — в сторону более обоснованного их теоретического построения. В тестах, ориентированных на цели обучения, основой должна быть идеальная норма, идентичная этой цели. Традиционные тесты могут дополнять эти сведения (относительное место ученика в учебной группе, а не достижение им поставленных целей обучения в полном объёме или частично). Идеально было бы получение одновременно в тесте и той, и другой информации. Г. Витцлак считает это вполне возможным. Необходимо, однако, преодолеть сложившийся отрыв тестов учебных достижений от учебного процесса — включить их в систему управления развитием учебного процесса, его корректировки как органическую составную часть. Желательно, чтобы результаты тестирования давали информацию и рекомендацию об оптимальном способе обучения учащихся при наличии тех или иных предпосылок*.

* Психодиагностика: теория и практика. М.: Прогресс, 1986.

Разумеется, это сделать непросто — необходим высокий профессионализм разработчиков и сочетание в одной команде создателей теста методологов-диагностов, методистов-предметников и психологов. А также проведение многомесячных методологических экспериментов и накопление многолетних данных. Потребуется большие трудозатраты, но стратегически они будут оправданы, если удастся в итоге создать высокопрофессиональную по исполнению, но доступную массовому педагогу (и, естественно, учащемуся) методику. Это будет оправдано многолетней широкой эксплуатацией теста, который с каждым последующим годом его практического использования будет надёжнее, валиднее и экономичнее. Особенно при условии создания банков диагностических данных.

Многие широко применяемые в педагогике тесты весьма долговечны. Так, тест способностей к обучению (SAT — Scholastic Amplitude Test) Совета колледжей — авторитетнейшей в США ассоциации — применяется с 1926 г., что позволяет сравнивать результаты учащихся разных лет. Апробированные профессиональные дидактические тесты не остаются при этом навеки законсервированными в первоначальном варианте, сохраняя своё ядро, идею, обычно

шкалы (иногда добавляя к ним новые). Станфордский тест достижений, получивший широчайшее использование в школах США (SAT), впервые был опубликован в 1923 г. и неоднократно подвергался переработкам. Одна из последних модификаций этого теста была в 1973 г.: шесть батарей охватывают классы с I по IX, то есть уровень базовой школы. Тестирование занимает 4–5 часов (это у привычных к тестам американских школьников) и осуществляется в несколько сеансов.

Модификации и переработки тестов связаны с экспансией метода тестирования на те контингенты обследуемых, для которых они ранее (в первом варианте) не предназначались. В Европе и США «взрослые» тесты адаптируются для учащихся различных возрастных групп. Кроме того, некоторые тесты специально адаптируются для учащихся с отклонениями в развитии. Так, очень распространённые, важные для отбора и оценки результатов тесты SCAT (тест способности к обучению в колледже), IRE (тест способности для выпускников университетов) имеют свои варианты со шрифтом Брайля для испытуемых с недостатками зрения.

* Анастаси А. Психологическое тестирование. Т.1. М.: Педагогика, 1982. С. 251.

Дидактические тесты при их переводе на другие языки подвергаются сложной переработке, зачастую приходится изменять и содержание инструкций, и терминологию (на привычную в той стране, на язык которой переводится тест). Иногда перевод оказывается бессмысленным без выработки новых **нормативов** теста, перепроверки его **надёжности, со-держательной и критериальной валидности**. И хотя, по мнению опытных специалистов, объём работ по переводу и адаптации теста почти равен трудозатратам на создание качественно нового теста, лучшие тесты и переводятся, и адаптируются. Это даёт возможность сопоставлять на межнациональном уровне результаты обучения, сравнивать сильные и слабые стороны доминирующих в той или иной стране дидактических концепций, организационных форм, методов и условий обучения.

Скорее всего, в обозримый период российские дидактические тесты вряд ли будут переводиться за пределами СНГ (в частности, из-за существенных различий в системах образования бывшего СССР и стран Запада). Но переводы на русский и другие языки **тестов обучаемости** и тестов, ориентированных на глобальные дидактические (предметные и межпредметные) цели, необходимы для сопоставления наших учебных достижений с тем, что есть за рубежом. В том числе с учебными достижениями русскоязычных соотечественников, оставшихся после развала СССР в странах СНГ, получающих там далеко не первосортное образование и стремящихся продолжить или завершить его в России.

Тестирование является только одним из видов контроля— эффективным, но имеющим свои ограничения. Это недостаточная способность тестов диагностировать в структуре профессиональных и предметных знаний и умений учащихся их творческие способности. Для этого требуются специальные исследовательские задания, не имеющие строгого алгоритма решения и тестами не являющиеся. Некоторые феномены процесса обучения лучше диагностируются не тестами, а программированным наблюдением (отношение к учению, добросовестность, упорство в учёбе и т.п.) или другими методами (групповая оценка личности, педагогический консилиум, социометрия — для выявления структуры реальных учебных возможностей, коммуникативной активности учащихся). Много в мотивации учения, в затруднениях учебной деятельности можно выявлять в диагностических беседах при индивидуальной работе педагога или школьного психолога с учащимся (или же при анкетировании). В технологическом аспекте тестирование рассматривается нами как совокупность взаимосвязанных последовательных процедур использования для педагогической диагностики дидактических (и по мере необходимости — психодиагностических) тестов. Обязательной составной частью этих процедур является анализ обработанных данных, формулировки педагогических диагнозов и предлагаемых коррекционных мер. В таком целостно-диагностическом понимании тестирование отличается от распространённой узкой, дилетантской его трактовки только как применения теста и обработки его данных.

Типология дидактических тестов и тестовых заданий

Проблема типологизации тестов и тестовых заданий имеет одновременно и теоретическое, и прикладное значение. В ней, как в зеркале, отражается всё многообразие тестов, применяемых для диагностики учебного процесса. Прежде всего необходимо отличать **дидактические** тесты от **психодиагностических** (выявляющих индивидуальные различия и личностные особенности). Дидактические тесты следует также отделять и от диагностических программ, в состав которых они нередко входят. В европейской (особенно немецкой) практике дидактические тесты даются в диагностических программах «на входе» и «на выходе», а между ними строится коррекционная работа педагога и психолога с учащимися.

Существуют разные подходы к типологизации дидактических тестов. Общими их чертами, на наш взгляд, является неполнота, односторонность оснований для типологизации. Можно выделить как бы два уровня типологизации дидактических тестов: по качеству их составления и по целевой направленности. Качество разработки теста зависит от опыта и профессионализма создателя.

Польский дидакт Ч. Куписевич отмечает деление тестов на профессиональные (стандартизированные) и преподавательские. «Первые требуют подготовки соответствующими специалистами выполнения многих условий, следующих из теории диагностики и статистики, а также продолжительных проверочных исследований, необходимых для целей стандартизации, вторые же могут быть составлены лицами, исследующими результаты учёбы, достигнутые на проведённых ими учебных занятиях, и хотя они также должны отвечать условиям правильной диагностики и статистики, но они не принимают характера стандартизированных тестов»*.

* Куписевич Ч. Основы общей дидактики. М.: Высшая школа, 1986. С. 37.

Профессиональные тесты обычно бывают **стандартизированными**, хотя это не одно и то же. Таким тест становится, только пройдя в процессе апробации (то есть настоящего научно-педагогического **методологического эксперимента**) все процедуры стандартизации. Стандартизацию можно проводить и с преподавательским тестом. Однако если он составлен неумело, разработчик обречён на бесконечные переделки. В отличие от профессионально составленных стандартизированных тестов те разработки, которые создаются самими учителями для текущего и итогового контроля в своём классе, обычно называют *учительскими, преподавательскими, дилетантскими (любительскими)* или **неформальными** тестами. Последнее название представляется более адекватным, так как все эти тесты не проходят обычно проверок на надёжность и валидность. При достаточном научно-методическом мастерстве составителя эти неформальные тесты могут быть содержательно валидны, особенно в аспекте **куррикулярной валидности**. Но могут быть построены и согласно субъективному, личностному видению педагогом ключевых целей и задач своего предмета.

По нашему мнению, следует заниматься переработкой неформальных тестов в профессионально составленные стандартизированные методики. Делать это следует не в одиночку, надо создавать творческие группы и авторские коллективы под руководством опытных психологов или педагогов-диагностов, владеющих методологией и психологией педагогического измерения.

В публикациях по проблемам тестирования обычно встречается не более 2–3 оснований для типологизации дидактических тестов (причём нередко авторы отождествляют тест как методику с тестовыми заданиями, что вносит дополнительную путаницу в непростой вопрос о типологии тестов). На основе анализа литературы нами разработана иерархическая типология дидактических тестов, отражающая их существенные характеристики по 9 основаниям. Эта типология может использоваться как при определении целей создания нового дидактического теста, так и при экспертизе имеющихся тестов или претендующих на этот статус наборов тестовых заданий.

На уровне типологии тестов подбор заданий, их структурная взаимосвязь и формулировка зависит от таких факторов, как:

1. Предметная область применения тестов:

а) **монопредметные** (пред-метные) тесты для проверки элементов или всего объёма знаний и умений из области одного учебного предмета: по теме, главе — в тесте текущего контроля; эта же группа тестов используется для рубежной проверки в рамках одного учебного предмета политематических знаний и умений;

б) **полипредметные** тесты по циклу изучаемых в учебном заведении дисциплин, куда как составляющий элемент может войти предметный в качестве субтеста или же в виде рассыпанных по всему полипредметному цикловому тесту заданий (имеющих, если тест составлен профессионалами высокого класса, свою шкалу);

в) **межпредметные** (интегративные по своей структуре) тесты, построенные для выявления «сквозных», межпредметных и надпредметных знаний и умений.

2. Общая ориентация замысла построения теста как нормативного или критериального. Они различаются по тому, как оценивается результат выполнения. **Нормативные тесты** по результатам подсчёта набранных баллов позволяют сравнить одну группу с другой, а внутри группы — обследуемых друг с другом. **Критериальные тесты** выявляют, в какой степени каждый учащийся владеет знаниями и умениями, необходимыми для решения каких-либо учебных (а чаще профессиональных) задач. В дидактических тестах критериальный подход является весьма значимым, так как позволяет глубже и разностороннее оценивать аспекты усвоения знаний и умений. Это не исключает построения субтестов или группы блочных заданий внутри теста на нормативной или смешанной основе (для проверки знания номенклатуры терминов, правил, ключевых фактов, дат, имён и пр.).

3. Дидактико-психологическая ориентация теста: какой тест разрабатывается (или как характеризуют тест, проходящий адаптацию, его авторы):

а) **тест достижений** для контроля знаний теории (простых, элементарных и сложных, неоднородных по структуре учебного материала);

б) **тест достижений** для контроля умений и навыков различной степени сложности по данному предмету;

в) **тест обучаемости** (диагностики реальных учебных возможностей по данному кругу предметных или цикловых знаний). В этом случае надо чётко определить, какая обучаемость диагностируется — общая или специальная (математическая, лингвистическая и т.п.).

От дидактико-психологической ориентации теста зависит характер включаемых в тест заданий как по содержательно-целевой направленности, так и по форме их предъявления.

4. Ориентация (на определённый этап контроля) дидактических тестов:

а) **тесты входного контроля** знаний и умений (при поступлении в вуз, колледж, начальную или элитарную школу, переходе в старшие классы и так далее);

б) **тесты текущего контроля** (тематического), составленные для проверки усвоения ключевых учебных элементов, которые могут выступать в качестве субтестов в тестах рубежного или итогового контроля;

в) **тесты рубежного контроля** результатов обучения за полугодие, семестр;

г) **тесты итогового контроля** (по курсу, изучаемому длительное время, по циклу учебных дисциплин или по итогам обучения на данной ступени образования (что надо оговаривать в каждом конкретном случае).

Специфика тестов на каждом этапе контроля проявляется в уровне фактологичности и обобщённости тестовых заданий, в масштабности диагностических целей заданий и субтестов.

5. Диагностический уровень тестирования (заложен в замысле теста):

а) **тесты симптоматической диагностики** выявляют самый «верхний слой» знаний и умений (знает — не знает, знаком — не знаком с формулой, правилом и т.д.);

б) **тесты этиологической диагностики** нацелены на выявление причинно-следственных связей (почему не знает именно этот раздел учебного материала, в какой степени пробелы учащегося зависят от других знаний — и от каких именно);

в) **тесты типологической диагностики** призваны определять, какие позитивные и

негативные тенденции в усвоении знаний, умений наметились у данного учащегося, к чему это может привести и что надо делать, чтобы смягчить негативное и усилить позитивное в учебной деятельности учащихся. Если тест ориентирован на этиологическую или типологическую диагностику, то надо ещё на стадии замысла планировать такое построение заданий, чтобы можно было по неудачным ответам выявлять типичные ошибки и затруднения в учении.

6. Доминирующая деятельность испытуемого при выполнении тестов — устные или письменные ответы. Либо же сенсомоторные действия — более или менее сложные — выполняются учащимися в **тестах действия**, особенно в аппаратурных (в профпедагогике) и компьютерных. А также и в очень важных для начальной школы тестах готовности детей к обучению в школе. Возможны и комбинированные варианты, хотя они труднее поддаются обработке (но используются для отбора поступающих в вузы США, как, например, тест TOEFL, проверяющий знания английского языка как иностранного). В тестах действия на учебном или рабочем оборудовании, в тестах на тренажёрах возможны задания на сложные многошаговые операции, особенно в системе профобучения при компьютерном тестировании.

7. Степень гомогенности (однородности) теста. Характеризует методику в двух аспектах:

а) критериальном: что проверяет тест — один критерий интегративного характера (например, количество выполняемых на должном уровне операций) или несколько — качество, количество, скорость, строгую последовательность, осознанность тех же операций;

б) формальном: однородность (гомогенность) или разнородность (гетерогенность) форм построения заданий.

8. Скоростной фактор: является (или должен быть) тест скоростным, с обязательным функционально необходимым фиксированием времени выполнения задания (каким? Как обоснованным?) или же нескоростным, где какие-то (какие?) затраты времени испытуемого при работе с тестом несущественны для оценки качества выполнения заданий.

9. Форма организации тестирования: массовая (аудиторная, общая для всех), индивидуальная или редко применяемая, но очень эффективная в отдельных случаях — бригадная (в условиях взаимодействия в группе обследуемых, совместно решающих какие-либо проблемы). Индивидуальное тестирование может проводиться при непосредственном личном контакте, лицом к лицу с диагностом или за компьютером, с индивидуализированной программой тестирования на ПЭВМ.

Используя предложенную схему, даже малоопытный разработчик теста (а таковых большинство) сможет чётко и обоснованно определить самостоятельно, какой тип теста нужен ему и его коллективу. Можно конкретно и диагностично сформулировать все основные цели и задачи создания теста и в соответствии с ними подбирать необходимые тестовые задания.

Следует особо остановиться на типологизации тестов по диагностируемому уровню усвоения знаний известного дидакта В.П. Беспалько*.

* Беспалько В.П. Слагаемые педагогической технологии. М.: Педагогика, 1989.

Она широко известна, применяется и сейчас многими разработчиками неформальных (преподавательских) тестов. Нами она будет рассмотрена ниже в контексте типологии тестовых заданий, поскольку уровни усвоения, выделенные В.П. Беспалько, характеризуют, на наш взгляд, не столько тест как целостную и структурно сложную диагностическую методику, сколько локальные, частнодидактические диагностические задачи.

Предложенная нами типология тестов, как показывает опыт, помогает разработчику теста (профессионального или неформального) чётко сформулировать свои конечные цели и осознать ограничения, связанные с замыслом его создания. Для этого вполне достаточно «пробежаться» по иерархии целей создания теста, чтобы увидеть, каким должен он быть в результате.

Типология тестовых заданий

Все разработчики тестов при определении цели и формы заданий пользуются той или иной их типологией. Как и в типологизации тестов, здесь существует множество различных классификаций. И точно так же ни одна из них не охватывает весь спектр видов заданий. Предлагаемая типология заданий теста позволяет многоаспектно охарактеризовать виды тестовых заданий по существенным основаниям, которые влияют, в конечном итоге, на содержание, структуру и внутреннюю композицию теста.

В основе любого тестового задания лежит **задача**, более или менее сложная, которая может быть по-разному сформулирована (в зависимости от замысла разработчика и его методического мастерства как тестолога). В современной дидактике существуют различные по степени полноты, охвата возможных ситуаций, классификации учебных задач и заданий. Наиболее развёрнутыми из них являются классификации учебных задач Г.А. Балла по восьми основаниям и классификация учебно-исследовательских заданий В.И. Андреева по шести основаниям*.

* Андреев В.И. Эвристическое программирование учебно-исследовательской деятельности. М.: Высшая школа, 1981; Балл Г.А. Теория учебных задач: психолого-педагогический аспект. М.: Педагогика, 1990.

Каждая задача, как известно, имеет свой предмет, условия, требования (задание в собственном смысле слова), логику и форму построения. Г.А. Балл выделяет следующие основные типы задач:

1. В зависимости от того, является ли исходный предмет задачи индивидуальным или родовым (относящимся к определённому классу предметов)— *индивидуальные и родовые*. Родовая задача требует нахождения алгоритма решения, а не только самого решения. Родовые задачи позволяют строить тесты на понимание закономерностей, правил выполнения определённого вида расчётов, грамматические правила и т.п., то есть затрагивающие относительно самостоятельный раздел знаний.

2. В зависимости от характера предмета решения выделяются задачи *материально-направленные* (решаются, чтобы обеспечить какое-то свойство реального материального предмета — химический состав, нахождение в каком-либо месте и т.п.) и *информационные* (предмет задачи — модель какой-либо системы, её описание, структура, изображение и т.п.).

3. В зависимости от отношений между предметом и требованием некоторые задачи могут классифицироваться как *принципиально неразрешимые* (построение вечного двигателя, оживление умерших). Такими они бывают, когда оказываются невозможными требуемое состояние предмета задачи либо переход к этому состоянию из заданного в условиях задачи исходного состояния предмета. Все прочие задачи *принципиально разрешимы*. В первом случае при построении тестов могут использоваться псевдозадачные формулировки, задачи с ложными данными, когда целью тестирования становится выявление умения учащихся (или педагога — при аттестации) различать реальные и нереальные условия, его знаний исходных принципов, закономерностей данной области знания. Например: в треугольнике ABC $A=82^\circ$, $B=64^\circ$, $C=44^\circ$. Вычислить внешние углы треугольника. (В евклидовой геометрии не существует треугольник с суммой углов, не составляющей 180° .)

4. В зависимости от субъективного уровня подготовленности решающего задачи подразделяются на *рутинные и нерутинные* (то есть, по сути, проблемные). В качестве нерутинных уже на первых этапах текущего контроля могут выступать тестовые задания, построенные в необычной, непривычной для учащегося форме (например, при проверке знания терминологии — задания в форме чайворда или мини-кроссворда, а также любой другой формы, с которыми учащиеся ранее не сталкивались). Освоение этих форм, привыкание к ним «переводит» тестовые задания такого рода в категорию рутинных для данного контингента учащихся. В параллельной учебной группе (классе), где с такими заданиями тесты не применялись, они будут до их внедрения и освоения учащимися рассматриваться как нерутинные.

5. В зависимости от чёткости построения задача считается *чёткой* (или хорошо определённой), если существует алгоритм для принятия заключения о её решении, и *нечёткой*

(плохо определённой), если алгоритм не существует или проблематичен. Нечёткость задач может использоваться для диагностики умения учащегося обнаружить эту нечёткость, а также для стимулирования самих учащихся к чёткой постановке задач (особенно в различных вариантах проблемного обучения).

6. Различаются также *внешние и внутренние* задачи в зависимости от того, как осмысливает обследуемый основные компоненты задачи. Почти любая задача по-разному может быть понята и доопределена (переопределена, «подменена») решающим в зависимости от его индивидуальных особенностей (знаний, отношения к задаче, мотивов, уровня и особенностей интеллектуальной деятельности). В жизни подобные ситуации складываются постоянно. Внутренняя задача может игнорировать часть условий внешней или домысливать те условия и требования, которые не были заданы и могут накладывать дополнительные ограничения на способы решения задачи. Приём постановки тестовых задач, которые провоцируют учащегося на подобные домысливания, широко используется в психодиагностике для определения консерватизма установок личности.

7. Задачи подразделяются также на *теоретические и практические*. В первых какое-либо изменение предмета задачи возможно только в результате воздействий решающего, а не объективной внешней материальной среды. К их решению всегда можно вернуться при неудаче. Большинство тестов содержит именно этот вид задач. К практическим относятся задачи, в ходе решения которых изменяются предмет и условия (например, выполнение трудовых операций, изменяющих исходный материал — заготовки, или состояние работающей аппаратуры и т.п.). Они могут выступать в качестве тестовых заданий при работе с различными приборами, установками, химреактивами и т.п., в системе производственного обучения, в имитационных упражнениях и анализе ситуации, на лабораторно-практических занятиях в школе.

8. В зависимости от особенностей исполнительской деятельности обследуемого при решении задания задачи подразделяются на шесть видов: *задачи исполнения, задачи преобразования, задачи восстановления, задачи построения, задачи использования процедуры, задачи использования имеющегося состояния**.

* Балл Г.А. Теория учебных задач: психолого-педагогический аспект. М.: Педагогика, 1990. С. 50–98.

Приведённая типология задач Г.А. Балла может быть дополнена рядом параметров (оснований) типологизации задач других исследователей. Значительный интерес представляет разработанная В.И. Андреевым классификация учебно-исследовательских задач*.

* Андреев В.И. Эвристическое программирование учебно-исследовательской деятельности. М.: Высшая школа, 1981. С. 89–96.

Не приводя её полностью, отметим, что она позволяет дополнить классификацию Г.А. Балла ещё тремя основаниями типологизации содержательной стороны задач: *логической схемой построения задачи, доминирующими приёмами требуемого для её решения метода научного познания и формой организации выполнения задания*.

9. В зависимости от логической схемы построения задания могут подразделяться на *индуктивно построенные*: в процессе выполнения учащиеся, опираясь на эмпирические знания и самостоятельные наблюдения, отбирают некие факты, а затем на основе их анализа и сравнения делают индуктивные обобщения; *дедуктивно построенные* — при их выполнении учащиеся, опираясь на соответствующие теории и законы, строят логические объяснения и доказательства вновь наблюдаемым (рассматриваемым) фактам и явлениям; смешанные варианты (индуктивно-дедуктивные или дедуктивно-индуктивные).

10. В зависимости от того, приёмы какого метода научного познания доминируют при выполнении задания, выделяются *логические, аналитические, экспериментальные, графические* и задания смешанного типа.

11. В качестве основания типологии тестов по особенностям их предмета и содержания можно привести распространённую в США и странах Западной Европы типологию тестов по

основным видам деятельности, выполняемым учащимися при их решении: на тесты действия, письменные и устные тесты. Доминирующий вид деятельности учащегося при решении теста в значительной степени определяет (уже на стадии замысла теста) многие характеристики и его содержания и формы построения тестовых заданий.

12. Массовая разработка дидактических тестов и за рубежом, и в СССР началась в 50–60-е годы. Нормативные тесты широко используются в различных вариантах программированного обучения. В российской литературе тесты чаще всего характеризуют по четырём уровням усвоения знаний, выделенным В.П. Беспалько*

* Беспалько В.П. Слагаемые педагогической технологии. М.: Педагогика, 1989.

и широко им пропагандируемым. Напомним эту типологию.

В тестах 1-го уровня в задании даются цель, ситуация и действия по её решению, а учащийся обязан дать заключение о соответствии (правильности) этих трёх компонентов в структуре задачи (тесты *узнавания, распознавания, классификации* ранее изученного материала). Иногда их называют *тестами на идентификацию*. Такие тестовые задания имеют подсказки в виде перечисленного набора вариантов ответов, из которого учащийся выбирает верный или наиболее полный (вариант — наоборот).

Тесты 2-го уровня характерны тем, что в задаче имеется цель и ситуация, а учащийся обязан применить ранее усвоенные действия по её решению. Действия его носят репродуктивный характер (работа по ранее усвоенному образцу, типа «делай, как я»). При выполнении этого задания учащийся самостоятельно воспроизводит в памяти и применяет информацию о ранее усвоенной ориентировочной основе выполнения данного действия (сама задача является типовой). Такие учебные действия называются *алгоритмизированными* уровнями. Всегда имеется эталон ответа, что сближает эти тестовые задания с 1-м уровнем (но этот эталон не предъявляется обследуемому). В тестовых заданиях 2-го уровня учащийся не пользуется подсказками при подстановке каких-либо компонентов (цифр, формул, графических изображений, слов) в ответе, используя собственную память. По своей форме это задания открытого типа, где надо вставить формулу, слово, фрагмент фразы или что-то написать, вычислить, нарисовать и т.д. В программированном обучении их принято было называть *заданиями с конструируемым ответом*.

Третий уровень усвоения знаний связан с тем, что в задаче задана только цель, но неясна ситуация, в которой цель может быть достигнута. Учащийся должен уточнить ситуацию и на основе её дополнения применить ранее усвоенные действия для решения такой *нетиповой* задачи. Он выполняет продуктивные действия *эвристического* типа: добывает новую (для него лично) информацию в ходе самостоятельного поиска и преобразования ранее известных ему способов и приёмов деятельности. Алгоритм этой деятельности заранее не задан (нет образца или жёстких правил), он формируется в ходе самой деятельности (разумеется, с опорой на все ранее изученные знания, приёмы и методы работы, которые учащийся приспособливает к условиям задачи). Результат такого решения может быть предсказан и описан лишь в самом общем виде. Поэтому построение тестов, нормативно-ориентированных для проверки знаний и умений 3-го уровня, вызывает очень большие затруднения.

Достижение *4-го уровня* связано с решением *творческих поисковых задач*, чаще всего они принципиально невозможны при дидактическом тестировании. Задания такого типа строятся как *критериально-ориентированные* и при их выполнении оценивается в первую очередь то, как учащийся владеет теми или иными способами, приёмами действия (например, написание сочинения на заданную тему или выполнение конкретного проекта, схемы или т.п. по общему методу).

Типология тестов В.П.Беспалько ещё в 60–70-х гг. критиковалась теоретиком программированного обучения Н.Ф. Талызиной. Она отмечала, что вряд ли во всех случаях процесса обучения репродуктивные действия учащегося сложнее, чем действия воспроизведения, а продуктивные действия эвристического характера сложнее репродуктивных. Скорее, это просто разные аспекты сложной и многогранной учебно-познавательной деятельности*.

* Талызина Н.Ф. Управление процессом усвоения знаний. М.: Изд-во МГУ, 1976.

Тем не менее, учитывая, что во многих тестах целые субтесты или блоки заданий строятся по этой логике, целесообразно включить уровень познавательной деятельности в классификатор заданий как самостоятельное (хотя и весьма условное) основание. Но не в типологию тестов, ибо профессионально составленный, стандартизированный дидактический тест по своей природе в подавляющем большинстве случаев гетерогенный (разнородный) по форме предъявления заданий и логике их построения.

Типология тестовых заданий по форме построения

В практике дидактической диагностики педагоги сталкиваются с некоторыми проблемами в связи с поиском разнообразия в **формах** разработки тестов. Однообразие тестовых заданий в значительной степени снижает активность интеллектуальной деятельности учащихся при тестировании, подталкивает на путь поиска «на удачу».

Наиболее развёрнутая классификация форм построения тестовых заданий предложена В.М. Распоповым. Он даёт полезные рекомендации по обучению самих учащихся составлению тестов заданий для программированного контроля.

По форме составления тест-заданий он выделяет следующие их виды:

а) в зависимости от *грамматического построения* — вопросы с набором ответов и предложения;

б) в зависимости от *взаимного расположения* — кадровые (отдельные вопросы или предложения с набором ответов) и блочные (группы вопросов или предложений с набором ответов);

в) в зависимости от *логических связей* тестовые задания бывают независимые и взаимозависимые (когда последующие задания требуют предварительного решения предыдущих);

г) в зависимости от *полноты содержания* выделяются задания:

с недостающей информацией (в вариантах ответов есть неправильные и только один правильный);

с избыточной информацией (все, кроме одного, варианты ответов правильные и предстоит отыскать единственный неправильный ответ);

с полной информацией (все ответы в принципе правильны, но один из них наиболее полный и его надо выбрать).

Комбинация возможных форм построения тестовых заданий даёт 24 варианта, то есть достаточно широкий диапазон для методического поиска педагога-разработчика.

Проблемы разработки стандартизированных дидактических тестов

Актуальность создания стандартизированных дидактических тестов в настоящее время самоочевидна. Решение этой задачи связано с определёнными проблемами.

Во-первых, стандартизированных дидактических тестов на государственном языке в России почти нет, в то время как в США, например, имеются прошедшие стандартизацию тесты практически по всем предметам и для всех уровней образования — от детских садов до выпускников университетов. Требование тщательной предварительной проверки дидактических методик давно известно психологам нашей страны. Чаще всего их формулировали ведущие учёные. Д.Б. Эльконин отмечал, что «ни одно дидактическое средство не может быть принято на вооружение для целей диагностики в отношении отдельного ребёнка, если оно не прошло проверки в сравнительном исследовании на массовом материале, если не доказана его дифференцирующая сила, его чувствительность к изменениям, связанным как с возрастным развитием ... так и с условиями обучения и воспитания. Только такие методические средства диагностики, которые обладают достаточно достоверными коэффициентами различий, могут включаться в качестве конкретных методик для целей диагностики индивидуального развития»*.

* Эльконин Д.Б. Избранные психологические труды. М.: Педагогика, 1989.

Требования были известны, но реально они почти не выполнялись. Ни в одном учебнике педагогики о надёжности и валидности методов педагогического исследования не говорится. Учителя с этими требованиями не знакомы. Поэтому, прежде чем разрабатывать стандартизированные дидактические тесты, надо подготовить их будущих авторов-педагогов к пониманию того, что создать тест — это провести настоящее научно-методическое исследование и научить это делать.

Во-вторых, недостаточно научно разработана методология и технология дидактического тестирования, т.е. прикладная дидактическая тестология. Здесь сказывается отставание российской педагогики и её теоретиков от реальной потребности практики. Отставание явно временное, потому что в отечественной педагогике много интереснейших разработок. Это в первую очередь огромный опыт программированного обучения, особо ценный для создания компьютерных дидактических тестов (в том числе в составе диагностико-коррекционных программ). Имеются ценные методологические разработки типологии учебных задач Г.А. Балла, самостоятельных заданий и работ, учащихся П.И. Пидкасистого, логической структуры учебного материала А.М. Сохора, исходные логические структуры обучения В.Т. Фоменко и др. Кроме того, чрезвычайно важен многолетний опыт зарубежной и советской экспериментальной психологии в разработке тестов для системы профессионального обучения и профотбора. Проверена почти в течение столетней практики технология разработки интеллектуальных тестов в психодиагностике.

Методология как ядро теории дидактической тестологии должна определить систему её специфических принципов и конкретизировать, заземлить на проблематике диагностического дидактического контроля общие принципы тестологии. Надо выяснить, в чём они конкретно проявляются (в каких специфических требованиях к разработчику) в типичных условиях разных подсистем образования — от дошкольной педагогики до послевузовского повышения квалификации. То есть составить алгоритмичную систему правил, следуя которым начинающий разработчик сможет сконструировать, апробировать и стандартизировать хороший дидактический тест без нерациональных потерь сил, времени, средств — как своих, так и испытуемых.

Практический опыт и анализ достижений современной психодиагностики позволяет сформулировать и систематизировать требования к последовательному циклу процедур методологии дидактического тестирования таким образом:

1. Определение в операционных понятиях диагностических целей и задач теста и его субтестов (по теме, разделу, учебной дисциплине, какому-либо умению).

2. Конструирование показателей теста и тестовых заданий (включая индикаторы и индексы), операционализация используемых в тестах (и при их обработке) понятий. Необходимо уточнить концепцию (модель) диагностируемого знания или качества личности, умения, навыка, определить в диагностичных формулировках требования к содержанию и уровню усвоения выявленных знаний (что надо знать наизусть, что уметь решать, что различать, какими приёмами владеть).

3. Спецификация создаваемого (или уже созданного и адаптируемого) теста. Следует заполнить таблицу, в которой указываются темы курса (их подразделы), учебные цели, выполнение которых диагностируется, намечается необходимое количество (и основные характеристики) тестовых заданий для проверки достижения каждой учебной цели (с учётом её «удельного веса», «значимости в общей структуре деятельности»).

4. Конструирование тестов в соответствии с заданной спецификацией: формулировка заданий, подбор тестовых задач различного типа и различной формы. Для облегчения этой работы могут использоваться разработанные нами схемы классификации тестов и тестовых заданий по содержанию, логике и форме построения. При экспертизе готового теста схема позволяет определить спектр приёмов, использованных при его конструировании, и разобратсья в том, что он собой на самом деле представляет, на что пригоден.

5. Предварительная апробация теста на репрезентативном контингенте, достаточно полно представляющем ту группу обследуемых, для которых тест предназначен.

6. Эмпирический анализ выполнения тестовых заданий: определение степени их сложности, селективности (возможности разделить сильных и слабых), качество подбора дистракторов, отвлекающих от эталона вариантов ответов. В тестологии принято считать, что минимально необходимая ориентировочная основа действий, диагностируемых ответом, достигается при коэффициенте усвоения знаний не ниже 0,7. Обучение считается успешным, когда учащийся на 95% вопросов профессионально составленного теста даёт правильные ответы. Коэффициент селективности заданий не должен быть меньше 0,30. Неудачным обычно считается дистрактор, который выбирает менее 5% опрошенных.

7. Определение надёжности теста как согласованности показателей, полученных теми же испытуемыми при ретесте— повторном тестировании или применении эквивалентных, дублирующих форм теста. Определяется также надёжность теста на скорость выполнения.

8. Определение валидности теста (насколько тест измеряет именно то, что запланировано измерять его разработчиками). Содержательная валидность определяет соответствие тестовых заданий содержанию изучаемого материала, конструктивная — соответствие индикаторов теста компонентам структуры измеряемого знания, качества, критериальная — сравнение результатов тестирования с внешними, значимыми критериями (например, знания определённых математических формул наизусть и успешности решения конкретных профессиональных задач). Прогностическая валидность показывает точность прогноза успеваемости или успешности какой-либо деятельности на основе данных тестирования, педагогическая — возможность выработки определённых эффективных педагогических коррекционных мер при однотипных результатах тестирования, полученных разными обследуемыми.

9. Корректировка первоначального варианта тестового комплекса на основе результатов эмпирического анализа, оценки надёжности и валидности. Это повторное прохождение алгоритма разработки и апробации теста.

10. Определение на основе усовершенствованного варианта теста выработанных в процессе апробации тестовых норм и диагностического описания системы критериев и оценивания результатов. Публикация теста для массового использования.

И, наконец, существуют проблемы, связанные со стандартизацией теста. К сожалению, в России и СНГ отсутствуют единые требования к стандарту дидактического теста. В США разработка тестов — монополия психологов и Службы педагогического тестирования. Например, для поступающих в вузы Америки иностранных студентов — ETS (Educational Testing Service), имеющая отделения во всех регионах земного шара, в том числе в Москве. Американская психологическая ассоциация давно определила жёсткую систему требований к стандартизации тестов: если методика им не соответствует— ей закрыт путь на рынок тестов.

Пока что ни в России, ни в других странах СНГ таких чётко сформулированных, общепризнанных и обязательных требований к стандарту дидактических тестов нет. В результате анализа литературы и практики тестирования нами разработан пробный вариант стандарта дидактического теста, отражающий основные требования стандартизации как предметного, так и межпредметного теста.

Требования к стандарту дидактического теста

Предлагаемая система требований к стандарту дидактического теста представляет собой минимизированную схему стандартизации психодиагностических методик, принятую Американской психологической ассоциацией и адаптированную нами применительно к специфическим российским условиям и реальным возможностям массовой педагогической практики. Она включает:

1. Полное название теста, характеристику назначения теста и его рекомендуемой области применения, диагностически корректно сформулированные в операционных понятиях цели и задачи тестирования, описание предполагаемого контингента обследуемых, например,

«учащиеся средних образовательных школ», «студенты технологических колледжей, изучавшие курс математики в объёме не менее ... часов по типовой программе такого-то курса» (указать авторов программы и адрес).

2. Точные сведения об авторах-разработчиках (фамилии, место работы, учёные степени и звания, должность, год создания теста).

3. Характеристика репрезентативности выборки стандартизации — сколько человек входило в выборку, насколько она отражает (в %) генеральную совокупность (тех, кого должен диагностировать тест), в какой степени были представлены в выборке половозрастные группы и группы обследуемых с разным уровнем успеваемости и другими значимыми признаками.

4. Краткое описание, обоснование и объяснение логики построения системы основных нормативов, а также, желательно, полученных при апробации (адаптации переводного теста) результатов по разным выборкам, если выборок было несколько.

5. Сведения о надёжности теста (ретестовой, параллельных форм или полученной другим методом — с указанием метода получения, а при ретестовой надёжности — с указанием временных интервалов между обследованиями).

6. Сведения о валидности теста: как минимум о содержательной, критериальной, но, желательно, и о конкурентной, локальной, психологической и педагогической валидности (выражаются в коэффициентах с указанием вида валидности).

7. Инструкция пользователю по методике проведения процедуры тестирования с подробным описанием поведения в типичных ситуациях и всех необходимых процедур обработки данных.

8. Инструкция обследуемым, которая может быть в тексте бланкового варианта или в компьютерной программе либо прилагаться отдельно, вместе с тренинговыми примерами заданий.

9. Текст теста с бланком ответов (или программно-педагогические средства при компьютерном варианте).

Организационно-методическая группа проблем связана с необходимостью создания практического руководства по стандартизации тестов, с описанием процедур определения различных видов надёжности и валидности. Руководство должно быть написано на лексическом уровне, доступном рядовому педагогу, с конкретными и понятными примерами. Для написания таких руководств нужен единый (хотя бы минимальный) терминологический словарь педагогической диагностики и дидактического тестирования.

Сейчас используемая разными авторами терминология нередко бывает методологически некорректна (например, отождествляется тест как методика и тестовое задание, как её элемент; налицо путаница с видами валидности, с типологиями тестов и тестовых заданий).

Для решения этих проблем необходима определённая централизация усилий и управленческих органов образования и разработчиков тестов. Должны быть проведены методологические эксперименты по отработке в российских школьных условиях различных процедур стандартизации, сопоставлению их эффективности при проверке надёжности и валидности разных типов тестов различными методами (например, для определения оптимальных подходов к валидации письменных, компьютерных, аппаратных тестов).