

ББК Ч-426-28 М 14

Майоров Алексей Николаевич

Тесты школьных достижений: конструирование, проведение, использование.
Издание второе — СПб.: Образование и культура, 1997. — 304 с.

В книге сделана первая за многие годы попытка объединенных знаний по составлению и использованию тестов в педагогической практике. На основании собственного опыта автор рассматривает такие вопросы как составление заданий и тестов, их метрологическое обоснование, стандартизация и нормирование, оснащение тестов и их использование для аттестации учащихся. Книга проиллюстрирована многочисленными примерами, содержит задания и упражнения, это дает возможность ее использования в качестве учебного пособия.

В книге содержатся научно обоснованные рекомендации для практического составления тестовых заданий и тестов.

Книга адресована руководителям информационно-диагностических, аттестационных служб, педагогам, руководителям образовательных учреждений и их заместителям по научной и учебной работе, педагогическим работникам занимающимся аттестацией учащихся. Книга будет полезна преподавателям высших учебных заведений, аспирантам и студентам педагогических вузов.

ISBN 8-88857-021-4

©1996,97 Майоров А. Н.

©1996,97 «Образование и культура»

(лицензия ЛР № 071128 от 2.02.95)

©1996,97 CMS(оформление)

Только тогда психология сможет стать действительной и точной наукой, когда она будет иметь своей основой эксперимент и измерения.

Дж. Кеттелл

Определение целей тестирования

Определение ресурсных возможностей разработчиков

Отбор содержания учебного материала

Конструирование технологической матрицы

Составление тестовых заданий

Построение выборки для апробации заданий и тестов

Компоновка заданий для апробации

Апробации тестовых заданий

Определение и расчет показателей качества тестовых заданий

Отбраковка заданий и составление теста

Апробации теста

Определение и расчет показателей качества теста

Составление окончательного варианта теста

Стандартизация теста
Нормирование теста
Оснащение теста



Общая схема этапов составления тестов школьных достижений

Рецензенты:

Лебедев О. Е., доктор педагогических наук, заведующий кафедрой управления РГПУ им. А. И. Герцена, председатель Комитета по образованию Санкт-Петербурга.

Вершловский С. Г., доктор педагогических наук, заведующий лабораторией ИОВ РАО.

Сухобская Г. В., доктор психологических наук, заведующий лабораторией ИОВ РАО.

Башмаков М. И., доктор физ-мат. наук, ректор ИПО РАО, академик РАО.

Тряпицина А. П., доктор педагогических наук, заведующий кафедрой педагогики РГПУ им. А. И. Герцена, член-корреспондент РАО.

Редактор: Юрьева Н. Н.

Оглавление

Введение. Цели и задачи настоящей работы.	7
Глава 1. Исторический обзор применения тестов в мировой практике..	11
Глава 2. Достижения учащихся и способы их формализации.	35
2.1 Психолого- педагогический анализ сущности достижений учащихся.....	35
2.1.1 Проблема контроля и оценивания..	35
2.1 .2 Достижения учащихся как итог обучения.	37
2.1.3 Психологический аспект достижений..	45
2.1.4 Целеполагание как основа определения достижений..	49
2.2 Педагогические цели и их конкретизация.	55
2.3 Таксономия целей Б. Блума. Международные исследования..	59
2.4 Технологическая матрица.	62
2.5 Отечественные разработки уровней целей.	65
Глава 3. Тесты и их виды. Тесты достижений.	69
Глава 4. Составление тестовых задач.	80
4.1 Типы и формы заданий, основные требования.....	80
4.2 Задачи открытого типа: свободного изложения и дополнения.....	83
4.3 Задачи закрытого типа.	89

4.3.1 Задачи альтернативных ответов,	90
4.3.2 Задачи с множественным выбором.	97
4.3.3 Задачи соответствия.	105
4.4 Использование заданий психологических тестов на выявление структуры интеллекта для тестов достижений.	112
4.4.1 Аналогии	112
4.4.2 Задания исключения лишнего.	116
4.4.3 Последовательности.	118
4.5 Другие виды задач.	120
4.6 Составление задачи через выделение концепта.	123
4.7 Учет дополнительных обстоятельств при выборе формы заданий.	125
4.8 Различные формы представления заданий.	126
4.9 Практические (аппаратурные) задания.	134
4.10 Создание двуразмерных тестов и возможность создания трехразмерных тестов.	140
Глава 5. Работа с заданиями после СОСТАВЛЕНИЯ.	152
5.1 Объединение задач в тест.	152
5.2 Упорядочивание заданий для испытания теста.	153
5.3 Проверка трудности задач. Определение места задачи в тесте.	155
5.4 Случайное угадывание правильного ответа.	157
5.5 Дискриминативность,	159
5.6 Надежность и валидность тестов.	163
5.6.1 Надежность.	163
5.6.2 Валидность.	171
Глава 6. Нормирование тестов.	183
6.1 Выборка.	183
6.1.1 Варианты расчета объема выборки. Стандартизация выборки.	183
6.1.2 Оценочный подход к расчетам выборки с учетом стратификации.	193
6.2 Варианты нормирования тестов. Основания для сравнений в тестах достижений.	198
6.2.1. Ориентированные на индивидуальные нормы.	198
6.2.2. Критериально-ориентированные и ориентированные на статистические нормы.	199
6.2.3 Ориентация на критерий.	201
6.2.4. Прогнозирование критерия.	203
6.2.5. Ориентированные на межгрупповое сравнение.	204
6.2.6. Статистические нормы и нормирование.	205
Глава 7. УСЛОВИЯ и процедура проведения тестов.	225
7.1 Оснащение тестов.	225
7.2 Подготовка процедуры проведения тестов.	232
7.3 Сценарий проведения исследования.	234
7.4 Требования к личности экспериментатора.	236

7.5 Наблюдатели.....	237
Глава 8. Предъявление данных и их интерпретация.....	238
8.1 Формы представления полученных данных.....	238
8.2 Интерпретация полученных результатов.....	242
Приложения.	
1. Инструкция и анкета наблюдателя.....	255
2. Основные трудности внедрения тестовых методик в практику работы образовательных учреждений.	263
3. Руководство для исследователя.....	267
4. Сценарий проведения тестирования.....	276
5. Протокол проведения тестирования.....	280
6. Глоссарий оснащения тестов для проведения аттестации учащихся.....	282
7. Таксономия учебных целей в когнитивной и аффективной областях.....	288
Литература.....	293
Ответы на задания.....	298

Введение

Цели и задачи настоящей работы

Проблема получения качественной, объективной информации в области педагогической деятельности имеет чрезвычайно важное значение. Такая информация необходима учителю — для самооценки профессиональной деятельности и оценки достижений ученика с целью корректировки педагогических воздействий, и руководителю образовательного учреждения — для определения стратегии развития образовательного учреждения и оценки качества работы педагогов и корректировки многообразных управленческих решений. Испытывают острую необходимость в такой информации и руководители более высокого уровня. В общем, виде решение этой задачи не на интуитивно-бытовом, а на научном уровне заботит практически всех работников образования.

Конечно, моментально эту задачу не решить из-за отсутствия надежного инструментария, специалистов, достаточной материально-технической поддержки, но ее решение во многом зависит также от профессиональной культуры руководителей в работе с информацией не только в вербальной форме. Управленцу необходимы умения переработки информации, умения принимать решения на основе лично не ориентированных данных. Высокий уровень потребностей порождает предложение, Появившиеся в последнее время на рынке многочисленные тесты не выдерживают критики по

большинству показателей качества, но вместе с тем появились и достаточно качественные продукты, однако распознать их человеку, не обладающему достаточными познаниями довольно сложно.

Данная работа преследует три основные цели. Во-первых, ее задача минимум - дать педагогу в руки инструмент, необходимые знания, которые помогли бы отличить тест от контрольной работы или заданий, хороший тест от плохого, позволили бы достаточно свободно ориентироваться на уже сложившемся рынке тестов.

Во-вторых, познакомить педагогических работников с обоснованными и надежными процедурами проведения тестирования при решении стандартных и исследовательских педагогических задач, когда за счет применения обоснованной процедуры значительно вырастает качество результатов работы с существующими тестами или с использованием уже готовых тестовых заданий. И в-третьих, и это задача максимум, дать заинтересованным людям инструмент, который бы помог САМИМ СОЗДАВАТЬ тесты различной сложности, ориентированные на педагогическую деятельность.

Выполнение последней задачи, по нашему глубокому убеждению, слишком трудоемко для отдельного учителя или группы педагогов, однако для коллектива достаточно слаженного, многообразного задача создания тестов вполне посильна.

Данная книга касается в основном тестов достижений. Метод тестирования и сами тесты достижений, при значительном количестве преимуществ как перед психологическим тестированием, так и перед традиционными контрольными работами, как и всякий измерительный инструмент, имеют целый ряд ограничений и условий в использовании. Попытки использования тестов достижений в областях для этого не предназначенных заведомо обречены на неудачу.

Тесты достижений рассчитаны на определение, в первую очередь, учебных успехов (достижений) учеников в когнитивной, психомоторной и аффективной сферах, и предназначены для решения стандартных и исследовательских педагогических задач.

Однако тесты достижений не являются инструментом для определения всего спектра школьных достижений. Определение достижений, например, в социальной сфере, необходимо проводить не тестами достижений, а специальным инструментом на это рассчитанным и для этих целей разрабатываемым.

Одним из существенных ограничений применения тестирования школьных достижений являются ограничения, которые накладываются на ответы. В силу чего анализ способов решения задач, мыслительных операций, которые использует ученик при решении задач в большинстве случаев оказываются затруднены или невозможны. Существует и ряд других ограничений, о которых мы скажем особо.

Однако в рамках существующих ограничений на сегодняшний день метод тестирования является как наиболее мощным, надежным и объективным при

решении широкого спектра педагогических задач, так и наименее теоретически и практически разработанным в нашей стране средством.

Каждая их глав данной книги относительно самостоятельна и может быть использована вне связи с другими. Полезной для педагогов может оказаться глава 3, посвященная составлению заданий. Самостоятельно рассчитать выборку можно используя данные главы 6, причем для любого вида исследований, а не только для тестирования.

Еще одна особенность данной книги — минимизация используемого математического аппарата. Формулы приведены преимущественно в главе 5, в разделе, посвященном расчетам объемов выборки. Во всех остальных случаях мы старались использовать таблицы, или описание вычислений, проиллюстрированное примерами и графиками.

В книге приведен ряд заданий и вопросов, ответы на которые с комментариями помещены после приложений. Эти задания могут быть использованы как для самостоятельного выполнения, так и предложены группам педагогов в случае проведения занятий по тестологии.

Книга содержит также примеры и пояснения к ним, взятые из работ В. К. Гайды и В. П. Захарова, П. Клайга, примеры тестов взяты из книг Айзенка, Л. ф.

Бурлачука, разработанных ЕТ5 (Education testing service), в РАО по руководством Г. С. Ковалевой, в Санкт-Петербургском университете педагогического мастерства под руководством Н. И. Элиасберг и И. Ю.

Алексашиной, в Лаборатории изучения образовательных систем под руководством автора и некоторые другие. Мы ни в коей мере не претендуем на авторство этих материалов. Полный перечень источников приведен к концу данной книги,

Глава 1 написана совместно с педагогом Ивановой Еленой Валерьяновной, глава 3, часть «Психолого-педагогический анализ сущности достижений учащихся», совместно с Шурухтом Самуилом Марковичем и Юрьевой Натальей Николаевной.

Автор выражает благодарность и признательность Сахарчук Л. Б. за помощь при подготовке рукописи.

Российская педагогическая тестология только начинает возрождаться и данная книга одна из первых после большого перерыва, который длится с 30—х годов. Автор будет благодарен за все критические замечания и предложения, касающиеся данной работы и просит сообщать их по адресу; 191025, Санкт-Петербург, ул. Марата, д. 25, лаборатория ИОС, тел. (812) 219 41 37.

Глава 1, Исторический обзор применения тестов в мировой практике

Метод тестирования широко известен за рубежом. Однако в нашей стране в силу различных причин тесты разного назначения и качества появились не так давно. Прежде чем более подробно поговорить об истории возникновения и современном состоянии тестологии, коротко определим, что же такое 'т.

Слово «тест» (Р. Пенто и М. Гравитц) произошло из старо французского языка и имеет синоним «чашка» (лат. testa — ваза из глины). Этим термином

обозначали небольшие сосуды из обожженной глины, используемые алхимиками для проведения опытов, В русском языке в 19 веке термин существовал в двух значениях:

I — «испытательная присяга, английская религиозная клятва, которую каждый вступающий в общественную должность должен давать, чтобы доказать, что он не тайный ка-тмлик». Скорее всего с этим связан тот факт, что ряд других источников (7, 11) рассматривает происхождение этого слова от английского test — испытание, исследование.

II — Плоский плавильный сосуд для выделения олова из золота и серебра (Русский энциклопедический словарь, 1872) (11, с. 158).

Приведем несколько современных словарных определений тестов для психолого-педагогической области:

Тест — это объективное и стандартизированное измерение, легко поддающееся количественной оценке, статистической обработке и сравнительному анализу.

Тест — стандартизированные задания, по результатам выполнения которых судят о психофизиологических и личностных характеристиках, а также знаниях, умениях и навыках испытуемого.

Тест — это система заданий, позволяющих измерить уровень развития определенного психологического качества (свойства) личности.

Тест — это специфический инструмент, состоящий из совокупности заданий или вопросов и проводимый в стандартных условиях, позволяющий выявить типы поведения, уровень владения какими-либо видами деятельности и т. п.

Тест — стандартизированное, часто ограниченное во времени испытание, предназначенное для установления количественных и качественных индивидуально-психологических особенностей.

При кажущемся разнообразии эти определения близки между собой, наиболее существенным для нас представляется то, что тест в психолого-педагогическом понимании этого слова означает проверку, испытание, но это не простое установление факта наличия или отсутствия какого-либо качества или свойства. Тест не является аналогией экзаменационным вопросам, анкетам, головоломкам и пр. В его основе лежит специально подготовленный и испытанный набор заданий, позволяющих объективно и надежно оценить исследуемые качества и свойства на основе использования статистических методов.

Тест как метод изучения индивидуальных различий возник сравнительно недавно. Бременем его возникновения считается конец 19 — начало 20 века, когда для определения физических, физиологических и психических особенностей человека психологи попытались применить различные способы измерения,

Вместе с тем, нельзя сказать, что до этого времени различия между людьми не вызывали интереса у представителей различных отраслей науки и искусства.

Уже в глубокой древности решение проблемы выявления индивидуальных различий имело жизненно важное значение, Так, в древнем Китае за 2200 л. до н.э. практиковалось своеобразное испытание для проверки

способностей и отбора персонала для правительственных должностей. Через каждые 3 года все чиновники экзаменовались г лично у императора по 6 «искусствам» — музыке, стрельбе из лука, верховой езде, умению писать, считать, знанию ритуалов и церемоний, а также оценивалось поведение чиновника в быту. В период Ханьской династии с 202 г. до н.э. дополнительно были введены испытания на знание гражданского права, военного дела, финансов, сельского хозяйства, географии. И в таком виде система испытаний просуществовала еще несколько столетий (1, с. 14).

В древнем Вавилоне использовались испытания в школах подготовки писцов. Претендент на должность писца был обязан знать все 4 арифметических действия, уметь измерять поля, распределять рационы, делить имущество, владеть искусством пения и игры на музыкальных инструментах. Большое значение придавалось умению разбираться в тканях, металлах, растениях и др. (44, с. 23).

В древней Греции и Риме также придавали большое значение выявлению различий в поведении и характере людей. Уже в Афинском государстве пришли к пониманию того, что могущество государства зависит от способностей управляющих им людей (1, с. 15). В трудах Аристотеля, Гиппократ, Демокрита, Галена указываются многие ценные диагностические показатели разных особенностей личности. Например, популярное в течение нескольких веков сочинение Теофраста «Этические характеры» содержало яркие портреты отдельных типов характера с описанием того, как они проявляются (65, с. 261).

Попытки систематизации различий между людьми и создания на этой основе «диагностических методов» предпринимались и в дальнейшем.

Например, в физиогномике существует представление, в соответствии с которым в выражении лица проявляются некоторые особенности личности.

Существует даже теория, которая это поясняет — с помощью примитивных аналогий между животными и отдельными характерами (лицо, похожее на лису — лстивый, как лиса). Похожим образом френология исходит из примитивной теории прямой локализации душевных свойств человека в определенных частях мозга и отсюда доказывает возможность диагностики индивидуальных особенностей исходя из формы черепа (64, с. 10).

Однако все эти попытки вряд ли можно назвать тестами в современном понимании этого термина. Как указывал Дж. Кеттелл, труды которого способствовали стремительному развитию тестологии, «только тогда психология сможет стать действительной и точной наукой, когда она будет иметь своей основой эксперимент и измерения* (44, с. 27). Соответственно, основной характеристикой теста является наличие специфического измерительного инструментария, позволяющего объективно оценить особенности испытуемого.

Проникновению точных измерительных методов в такие «неточные» сферы, как психология и педагогика, способствовало быстрое развитие естествознания, возникновение и распространение теорий, объяснявших процесс развития человека как изменяющийся и усложняющийся во времени. Одним из первых

ученых, попытавшихся измерить различия между людьми в области элементарных психических процессов был англичанин Ф. Гальтон. Исследуя индивидуальные различия, Гальтон использовал целый набор методик: на определение зрительной, слуховой, тактильной чувствительности, на мускульную силу, время реакции и др. Во время проходившей в Лондоне в 1884 году Международной выставки медицинского оборудования, средств и методов охраны здоровья была устроена лаборатория, где посетители в возрасте от 5 до 80 лет могли проверить свои физические способности, физиологические возможности организма и психические свойства по 17 показателям: рост, вес, жизненная емкость легких, сила кисти и сила удара, различение цветов, острота зрения и др. Было обследовано более 9000 человек. Гальтон считал, что наиболее достоверные результаты получаются при обследовании большого количества испытуемых (1, с. 15). В трудах Гальтона прозвучала мысль о необходимости внешнего критерия, независимого от экспериментатора. Позволило бы «узнать, какой тест является наиболее информативным» (1, с. 19).

Важным вкладом Гальтона в развитие теории тестов стало определение трех основных принципов:

1. Применение серии одинаковых испытаний к большему, количеству испытуемых.
2. Статистическая обработка результатов.
3. Выделение эталонов оценки (61, с. 5).

Эти принципы используются и по сей день — на основе проведения серий испытаний получают различного вида нормы для оценки результатов тестирования, все современные тесты построены на основе статистической теории измерений, а идея эталона оценки лежит в основе определения тестов как стандартизированного инструмента.

Гальтон называл испытания, проводившиеся в его лаборатории, умственными тестами (от английского *Сех1* — испытание), Однако наибольшую популярность этот термин приобрел после выхода статьи Дж. Кетгелла «Умственные тесты и измерения», опубликованной в 1890 г. с послесловием Гальтона (65, с. 267).

Дж. Кеттелл, американский психолог, некоторое время работавший в лаборатории Гальтона, с энтузиазмом воспринял идею о способах определения индивидуально-психологических качеств и статистической обработки результатов.

Кетгелл поставил задачу описать образ цельной личности посредством возможно меньшего числа экспериментов. С этой целью он предложил в 1890 г. нескольким лабораториям произвести в одинаковых условиях 10 основных экспериментов (измерение силы рук посредством динамометра, скорость реакции на звук, скорость ассоциации при назывании 10 цветов и т. д.). На этой основе впоследствии им были разработаны наборы заданий, которые он называл «умственными тестами» и использовал для определения «интеллектуальной физиономии* индивида. Таких тестов Кеттеллом было разработано 50 (17, с. 7),

Однако все они позволяли оценить элементарные психические процессы, где индивидуальные различия сравнительно малы и не затрагивали высших психических функций, лежащих в основе интеллекта (63, с. 3).

Но все же нельзя недооценивать вклад Кеттелла а дальнейшее развитие идеи применения измерений и статистического анализа при определении индивидуальных различий. Горячий сторонник и пропагандист тестового метода, Дж. Кеттелл считал тест средством для проведения научного эксперимента с соответствующими требованиями к его чистоте. Такими требованиями он считал:

- одинаковость условий для всех испытуемых;
- ограничение времени тестирования приблизительно одним часом;
- отсутствие зрителей в лаборатории, где проводится эксперимент;
- оборудование должно быть хорошим и располагать людей к тестированию;
- одинаковые инструкции и четкое понимание испытуемыми, что нужно делать
- результаты тестирования подвергаются статистическому анализу, находят минимальный, максимальный и средний результат, рассчитывают среднее арифметическое и среднее отклонение (1, с. 19).

Все эти идеи, выдвинутые Кеттеллом, в настоящее время составляют основу для современной тестологии. Одинаковость условий для всех испытуемых, одинаковые инструкции и четкое их понимание испытуемыми — принципы положенные в основу стандартизации процедуры проведения тестирования; ограничение времени, в настоящее время, после дополнительных исследований, реализуется в зависимости от возраста испытуемых и особенностей применяемого инструментария; идеи статистической обработки результатов реализованы в достаточно сложных методах статистического анализа и моделирования.

По возвращении в США Кеттелл немедленно начал применять тесты, активно их пропагандировать. Не прошло и не скольких лет, как возникла необходимость организовать специальные координационные центры. И в 1895—96 гг. в США были созданы два национальных комитета, призванных объединить усилия тестологов и придать общее направление появившимся тестологическим работам. В конструировании тестов принимают активное участие видные психологи из разных стран: Бине (Франция), Торндайк, Йеркс, Гилберт (США), Мюнстерберг (Германия) и многие другие (65, с. 268).

Качественный скачок в развитии тестологии связан с деятельностью видною французского психолога А. Бине. А. Бине может считаться родоначальником современных тестов, предназначенных для диагностики уровня развития интеллекта. Началом послужила публикация в 1896 г. А. Бине и В. Анри статьи по вопросам индивидуальной психологии, где критически оценивались исследования Ф. Гальтона и Дж. Кеттелла по «измерению ума». Исследование элементарных психических процессов не давало возможности адекватно оценить развитие высших психических функций и интеллекта в целом. А. Бине и В. Анри подчеркивали необходимость найти объективную оценку для

определения уровня интеллекта и предлагали принципы построения нового теста.

Вскоре возникла реальная возможность применить научные разработки в педагогической практике. В 1904 г. А. Бине пошел в состав комиссии по созданию в Париже специальных школ для умственно неполноценных детей. Требовалось отделить детей, способных к учению, но ленивых и не желающих учиться, от страдающих прирожденными дефектами. А. Бимс и Т. Симон разработали серию заданий для детей от 3 до 11 лет, Первоначально серия состояла из 30 тестов-заданий, расположенных по мере возрастания трудности таким образом, что вероятность успешного выполнения была тем выше, чем старше были дети. Уровень трудности был определен в результате обследования 50 нормальных детей и незначительного числа слабоумных. За пределами определенного уровня сложности слабоумные дети не могли решить задачи. Тест рассматривался как вспомогательный инструмент изучения интеллекта, общий показатель интеллекта еще не рассчитывался. Фактически, применение этого теста было первой попыткой определить индивидуальные различия между детьми с помощью измерения их умственного развития. А. Бине и Т. Симон несколько раз пересматривали созданную ими шкалу. В 1908 г. осуществлена новая редакция, в ходе которой ими было изменено количество тестов, убраны неудачные задания, расширена выборка стандартизации. В этой редакции была поставлена принципиально новая задача — не только дифференциация слабоумных и нормальных детей, но и выделение разных уровней интеллектуального развития нормальных детей. Важным изменением было то, что впервые задания группировались по возрастным уровням, что позволило определить нормы для детей разных возрастов, и вводилось понятие умственного уровня (позднее замененного на умственный возраст, а еще позже на показатель умственного развития IQ) (11, с. 24).

Суть этого понятия заключалась в следующем. Если ребенок справлялся с заданиями, выполняемыми большинством его сверстников, он считался нормальным. Но в некоторых случаях задания оказывались для него либо слишком легкими, либо непосильными. Так, если ребенок 6 лет справился со всеми заданиями для пятилетних детей и детей более младшего возраста, с двумя — для шестилетних, и с одним — для семилетних детей, тогда его базовый умственный возраст равняется 5 годам (60 месяцам), к этому показателю прибавляется еще 6 баллов (по 2 балла за каждое задание более высоких возрастных уровней). Таким образом, умственный возраст ребенка составит 66 месяцев, а физический — 72 месяца (64, т. 1, с. 98).

Вариант 1911 года содержал шкалу, продолженную до возраста 15 лет, некоторое изменение заданий. Здесь была сделана попытка стандартизировать шкалу и определить ее валидность (17, с. 9).

Наибольшее распространение система Бине получила в Италии, Великобритании, США, В 1916 американский психолог Л. Термен пересмотрел шкалу Бине-Симона. Он попытался определить относительную позицию

испытуемого в своей возрастной категории, используя коэффициент интеллектуального развития.

Поскольку интеллектуальный коэффициент (Intellectual Quotient — IQ) является основным количественным показателем уровня интеллектуального развития, используемым во многих тестах, рассмотрим его несколько подробнее. Термин «интеллектуальный коэффициент» был введен в 1912 году В. Штерном. Первоначально он определялся как отношение умственного возраста (МА) к хронологическому (СА), умноженному на 100. Несовпадение возрастов считалось показателем либо умственной отсталости (если МА ниже СА), либо одаренности (СА выше МА) (65, с. 272). В приведенном выше примере умственный возраст ниже хронологического, что может свидетельствовать об умственной отсталости ребенка. В современных тестах используется стандартный интеллектуальный коэффициент, процедура определения которого предложена Д. Векслером. Д. Векслер использовал для определения 1^{\wedge} единицы стандартных отклонений, что показывало, в каком отношении находится результат данного обследуемого для он) возраста (11, с. 55). В главе посвященной стандартизации тестов мы вернемся к этому понятию.

Новая редакция Л. Термена получила известность как Станфордская шкала интеллекта Бине. В редакции Термена было увеличено количество тестов с 54 до 90, впервые выработаны инструкции для предъявления оценки отдельных субтестов. И в течение 20 лет эта редакция была наиболее популярной в США. Однако практика такого длительного использования шкалы показала некоторые ее недостатки, и в 1937 году (шли осуществлена вторая редакция, а затем, в 1960 г. появилась третья. Эта редакция Станфордской шкалы используется и в настоящее время.

Тест включает в себя 6 заданий для каждого возраста начиная с 2-х лет и один резервный тест на случай, если один из тестов не удастся провести. Ниже приводятся варианты тестовых заданий для шкалы Бине-Симона в редакции 1911 года и 3-й редакции Станфордской шкалы:

Возраст 6 лет:

1. Определить время дня
2. Назвать назначение нескольких предметов обихода,
3. Незаконченные рисунки
4. Понятие числа
5. Обратные аналогии
6. Лабиринт

1. Словарь
2. Различия
3. Срисовать ромб
4. Сосчитать 13 предметов
5. Сравнить с эстетической

точки зрения 2 лица (3 задания).

Резервный тест ответы на картинки

Возраст 11 лет:

1. Противостояние внушению при сравнении линий разной длины
2. Словесные абсурды — 4
3. За 3 мин. сказать 60 слов
4. Определение отвлеченных понятий
5. Восстановить порядок слов

Резервный тест: найти причину (3 задания)

1. Память на рисунки
2. Составление предложений из трех слов
3. Абстрактные слова — 2
4. Память на предложения — 2
5. Проблемные ситуации — 2
6. Сходства: три предмета

Возраст 14 лет:

1. Словарь
2. Тест на индукцию
3. Мышление
4. Найти выход из положения
5. Ориентировка: направление
6. Сходство антонимов

Резервный тест: Найти выход из положения (64, т. 2, с. 12)

Длительное время тесты развивались как инструмент индивидуальных измерений, Массовый характер тестирования вызвал необходимость перейти от индивидуальных тестов к групповым. В 1917—18 г.г. в США появились первые групповые тесты для нужд армии. Наибольшее применение нашли тесты А. Отиса. Он предложил два набора тестов (Альфа и Бета тесты) для владеющих английским языком и «немые» тесты для не владеющих или плохо владеющих языком. Основные принципы, использованные при составлении этих тестов, были систематизированы М. Бернштейном и впоследствии легли в основу всей методологии групповых тестов:

1. Принцип ограничения во времени — только 5% испытуемых могли окончить проработку всего теста, т. о. показатель развития прямо зависит от скорости выполнения заданий испытуемым.
2. Принцип детализированной инструкции — как в отношении проведения, так и в отношении подсчета,
3. Введение тестов с выборочным методом формирования ответа с указанием подчеркивать наугад в случае незнания или сомнения.
4. Подбор тестов после тщательной статистической обработки и экспериментальной проверки (63, с. 7).

В это же время широко развиваются методики обработки результатов тестирования, создания тестовых систем. Еще Гальтон воспользовался работами бельгийского статистика А. Кетле для измерения умственных способностей. Гальтон первый разработал метод статистического сравнения двух рядов переменных и ввел индекс совместного отношения, названный коэффициентом корреляции. Также Гальтону принадлежит заслуга изучения взаимосвязи переменных и построения линий регрессии одной переменной на другую. В 1896 г. К. Пирсон заложил основы теории корреляции. В 1904 г. Ч. Спирмен сделал следующий значительный шаг в использовании метода корреляции для психологии. Он показал, что корреляция между двумя переменными свидетельствует о наличии общего фактора (причины, определяющей величины этих переменных) и специфических факторов, присущих каждой переменной. Спирмен явился основателем двухфакторной теории.

Впоследствии Г. Томсон, а затем Л. Терстоун развили далее идею факторного анализа. Терстоун исходил из понимания интеллекта как целого, состоящего из нескольких первичных способностей (вербальной, перцептивной, способности к восприятию пространства, памяти, беглой речи), Все эти факторы равноправны, единого фактора интеллекта не существует. С помощью факторного анализа не просто устанавливается связь изменения одной переменной с изменением другой, а определяется мера этой связи и обнаруживаются основные причины, лежащие в основе указанных изменений. Достижения корреляционного и факторного анализа находят свое применение и в современных исследованиях (50, с. 425).

Дальнейшее развитие тестологии во многом определилось нуждами практики, Разрабатывались тесты для профессионального отбора в различных сферах производственной деятельности. Развивались новые направления. Например, создание тестов на основе использования проективного метода. Наиболее известными являются тест швейцарского психолога Г. Роршаха, предложившего использовать для глубинной диагностики личности адаптированную технику чернильных пятен, и тест Вартега — незаконченных рисунков. По ассоциациям, возникавшим у испытуемого при восприятии чернильных пятен в тесте Роршаха, можно было сделать вывод о его личностных особенностях. В настоящее время проективные методики изучения личности пользуются большой популярностью (64, с. 12).

Активно развивалось направление, связанное с созданием тестов изучения уровня интеллектуального развития. Наиболее известным тестом в нашей стране является тест, созданный Д. Векслером. Шкала Д. Векслера (WISC) была разработана в США для определения уровня интеллектуального развития детей 5—15 летнего возраста. В основе теста, как отмечает

Д. Векслер, лежат две основные идеи: рассмотрение интеллекта как неотъемлемой части целостной личности и сравнение результатов конкретного ребенка со средними результатами группы людей соответствующего возраста. Шкала Векслера состоит из 12 субтестов, которые делятся на 2 подгруппы: нацеленные на изучение вербального и невербального интеллекта.

К вербальным тестам относятся тесты:

1. осведомленность;
2. понятливость;
3. арифметический;
4. СХОДСТВО;
5. словарный;
6. счет (повторение цифр).

К невербальным (практическим) тестам:

1. недостающие детали;
2. последовательные картинки;
3. кубики Косса;
4. складывание фигур;
5. шифровка;
6. лабиринт.

Тесты «повторение цифр» и «шифровка» — дополнительные. Анализ полученных по всем субтестам результатов позволяет вывести суммарный интеллектуальный коэффициент (3, с. 5-6).

С начала 20 века определилось и педагогическое направление в развитии тестологии. Американец В. А. Макколл разделил тесты на педагогические (Education Test) и психологические — по определению уровня умственного развития (Intelligence Test). Основной задачей педагогических тестов являлось измерение успешности учащихся по тем или иным школьным дисциплинам за определенный период обучения, а также успешность применения определенных методов преподавания и организации (63, с. 4).

Макколл определил цель использования педагогических тестов — объединение в группы учащихся, усваивающих равный по объему материал и усваивающих его с одинаковой скоростью (63, с. 7). История использования педагогических тестов началась несколько ранее, чем психологических. Еще в 1864 г. в Англии Дж. Фишер создал «A scale book» — градуированную книгу, где были даны вопросы и варианты ответов для выбора. Книга содержала вопросы по правописанию, арифметике, чтению, грамматике, навигации (63, с. 8). Но начало систематических измерений в педагогике относится к 1894, когда появился первый педагогический тест доктора Раиса. Раис считается одним из первых создателей тестов, при помощи которых определялся уровень усвоения отдельных учебных дисциплин. В 1894—95 гг. им были разработаны 2 теста по орфографии, которые должны были измерить эффективность более длительного или менее длительного обучения навыку письма, а затем тест по арифметике.

Однако разработка первого педагогического теста принадлежит американскому психологу Э. Торндайку. Он считается основоположником педагогических измерений, Торндайку принадлежат первые настоящие педагогические тесты. Результатом большой исследовательской работы в области измерений и использования метода тестов в педагогике явилась книга «Введение в теорию психологии и социальных измерений» (1904). Работы Торндайка отличались теоретической и методической законченностью. Первый стандартизованный

педагогический тест, вышедший под руководством Торндайка, был тест Стоуна на решение арифметических задач, впервые снабженный «нормами».

В 1909—1914 г.г. выходят шкала оценки детского почерка Торндайка, шкала оценки сочинений на английском языке Хиллегаса, шкала правописания Бекингема, оценочные шкалы почерка и правописания Айреса, арифметические тесты С. А. Кортиса, определяющих знание 4 арифметических действий (63, с. 8).

В 1915 г. Иеркс создал свою серию тестов, главное отличие которой — изменение системы подсчета. Вместо возрастных долей, предложенных Бине, испытуемый получает за каждый правильно решенный тест известное количество баллов. Это повысило удобство в отношении проведения и подсчета результата теста. Количество полученных баллов переводилось по приложенным стандартам в коэффициент одаренности или о успешности (63, с. 5),

В Америке тесты нашли особо широкое распространение. 1915—30 г.г. — настоящий бум развития тестологии. Тесты успешности были разработаны по всем учебным дисциплинам. В 1918 г. А. Отис опубликовал серию групповых тестов для школы.

Группа американских психологов (Йеркс, Торндайк, Тер-мен, Уиппл, Хаггерти) работает над созданием «интеллектуальных» тестов для школ всей страны. Эти тесты назывались «национальные тесты». Национальные тесты были составлены в 2 сериях по 2 варианта в каждой серии. Совершенствуя методику, американские тестологи в этих тестах впервые перед каждым разделом теста дают примерные упражнения.

В Европе распространение тестов проходило далеко не так благополучно. Во Франции психологи Ване (1905) и братья Анфруа (1914) попытались создать тесты по проверке правописания. Однако широкого применения они не нашли. С 1921 г. эту работу возглавил Симон и регулярно стали создаваться французские тесты проверки счета, чтения, письма с использованием аналогичных американских. Однако их применение было не таким широким, как в США (63, с. 9).

Особым путем шло проникновение тестов в Россию. До 1917 года вопросам тестирования уделялось недостаточное внимание. Практическое значение тесты получили после 1925 года, когда была создана особая тестовая комиссия, Тестовая комиссия существовала при педагогическом отделе Института методов школьной работы. В ее задачи входила разработка стандартизированных тестов для советской школы. И уже весной 1926 года вышли такие тесты, созданные на основе американских. Были разработаны тесты по природоведению, обществоведению, счету, решению задач, знанию географической карты, на понимание чтения и правописание. К этим тестам прилагались инструкции и личная карточка для учета прогресса учащегося. Проблемой разработки тестов вплотную занимались видные российские психологи и педагоги: С. Г. Геллерштейн, П. П. Блонский, А. П. Болтунов, М. С. Бернштейн, А. М. Шуберт, Г. И. Залкинд и др. В 1927 вышла книга С. М. Василейского «Введение в теорию и технику психологического, педоло-

гического и психотехнического исследования», где излагались теоретические и практические подходы к созданию и использованию тестов. Можно сказать, что это был первый учебник, где отражена методика составления анкет и тестов, статистические методы обработки результатов (включая корреляцию), методы изучения различных профессий, вопросы проведения эксперимента. Идеи, изложенные С. М. Василейским о методе, методике и технике анкетного опроса, обработке анкетной информации не потеряли ценности и в наши дни. В 20-е годы Центральная педологическая лаборатория МОНО под руководством педолога Е. Гурьянова разработала и выпустила в свет несколько тестов:

1. шкала для измерения умственного развития детей;
2. тесты для учета навыков в чтении, счете и письме;
3. тесты коллективного испытания умственной одаренности.

Под руководством П. П. Еланского в педагогическом кабинете Академии коммунистического воспитания осуществлялась проверка шкалы Бине-Симона, создание школьных тестов и построение собственных стандартов. В Ленинградском научно-педагогическом институте им. А. И. Герцена и Детском обследовательском институте им. А. С. Грибоедова разрабатывались тесты на одаренность и тесты школьной успешности для массовых обследований детей нормальных школ.

На состоявшейся в апреле 1927 года 1 Всесоюзной конференции педологов активно обсуждались вопросы тестирования. В своем выступлении П. П. Блонский отметил, что «огромная заслуга тестов состоит в том, что они позволяют обычный ответ учителя „Так мне кажется" заменить словами „я это знаю или я это не знаю"» (63, с. 11—12). §'

Однако вскоре положение кардинально изменилось. Вышедшее в 1936 году постановление ЦК ВКП(б) «О педологических извращениях в системе Наркомпроса» негативно сказалось на развитии педологии в целом и тестологии в частности. Метод тестов был признан буржуазным орудием дискриминации учащихся и «изгнан» из советской школы. И в дальнейшем в течение многих лет тестирование официально не признавалось объективным методом и не находило практического применения. Только с шестидесятых годов стали осуществляться серьезные попытки использования зарубежных тестов, например, теста Д. Векслера, адаптированного А. Панасюком в 1968 г. Впоследствии сильнее развивалось диагностическое направление, где тесты являлись составной частью. Однако тесты школьных достижений до настоящего времени не нашли должного применения.

Многие из созданных за рубежом тестов популярны и используются до сих пор. Так, долгое время использовались практические тесты, разработанные Ч. Расселом: 48 расположенных в порядке возрастания сложности уроков по математике были отпечатаны на специальных карточках. Дополнительно давались тесты, помогающие установить уровень освоения материала. Каждый учащийся мог практиковаться в выполнении тестов по мере необходимости и самостоятельно проверить ответ.

Турецко-английский тест В. В. Леммона. В начале теста — ключ из 20 турецких слов, напечатанных английскими буквами, и их английские эквиваленты. Тест состоит из 100 турецких слов. Таким образом, каждое из 20 слов ключа повторяется 5 раз в произвольном порядке. Предлагалось около каждого турецкого слова написать его английский эквивалент, при этом в случае необходимости можно использовать ключ. Давалось 2 попытки по 10 минут. Оценка — по общему количеству правильных эквивалентов. Недостатком теста являлось то, что он основывался на механическом запоминании слов без их последующего закрепления в активной речевой ситуации.

В 1929 г. американский психолог В. А. Хенмон издал книгу, где опубликовал тесты на определение успешности овладения иностранным языком. Он определил основные пункты, которые необходимо изучать и разработал соответствующие тесты:

1. Тест для проверки знания слов, где измерялось знание слов и их значений на различных этапах обучения и при различных условиях и методах обучения (A vocabulary test).
2. Тест для проверки умения понимать прочитанное (уметь читать и понимать письменную речь). Он состоял из 2-х частей: на понимание предложений и на понимание абзацев (A silent reading or comprehension test) ;
3. Тест для проверки умения переводить с иностранного языка на родной (A translation-into-English test);
4. Тест для проверки умения переводить с родного языка на иностранный (A translation test);
5. Тест для проверки умения писать сочинения на иностранном языке (A written composition test);
6. Грамматический тест на знание синтаксиса и морфологии (A grammar test);
7. Тест для проверки понимания устной речи на слух — восприятие устной речи. (An oral comprehension test);
8. Тест на правила произношения звуков и слов (A pronunciation test)
9. Тест для проверки умения говорить, вести беседу, делать сообщения (An oral composition test).

8 и 9 тесты представляют трудность при групповом проведении и письменном контроле и оценке (63, с. 22).

Большое распространение получили так называемые педагогические прогностические тесты. Использование прогностических тестов предполагает возможность определения способностей, необходимых для изучения какой-либо учебной дисциплины, предсказания успешности в определенной области. Процедура заключается в следующем. Учащийся выполняет несколько тестов в начале года. Результаты их анализируются и оставляются до конца года. В конце года проводится повторное ц: тестирование тестами достижений. Выявляется корреляция по §• каждому тесту, Тот тест, где выявляется наилучшая корреляция- 3 Нин (соответствие), обладает наиболее высокой прогностической ценностью (63, с. 25). Примером прогностического теста может служить тест, созданный Дж. Кэрролом в соавторстве со Ст. Сэпоном в

1959 г — Modern Language Aptitude Test (MLAT). Тест создан с целью определения возможной степени успеха в усвоении учащимися иностранного языка: прогнозирования успеха в научении говорить и понимать иностранный язык, а также читать, писать и переводить на иностранный язык. Возможно использование MLAT как для современных, так и для древних языков. Тест явился результатом 5 лет исследований, в течение которых было обследовано 5 тыс. человек. Результат проведения теста должен был показать, насколько Хорошо по отношению к другим данный индивид может изучить иностранный язык за определенный промежуток времени и при определенных условиях, как быстро испытуемый может приобрести основы знания иностранного языка, которые позволяют ему научиться читать, говорить и т. п. Тест включал 5 частей и проводился за 50—60 минут.

Разработкой и проверкой тестов занимаются специальные государственные службы. Еще в 1900 году в США был создан Совет по вступительным экзаменам. В 1926 Совет колледжей принял тест SAT, разрабатывал тесты для квалификационной и профессиональной оценки деятельности педагога. С 1947 года существует Служба тестирования (Educational Testing Service), которая считается наиболее представительным научно-исследовательским центром. К 1961 году только в США были созданы 2126 стандартизированных тестов (61, с. 8). Здесь существует множество тестовых издательств, общая продукция которых представляет несколько тысяч различных тестов. Естественно, они различаются по качеству и по спросу на них. Изданные тесты ежегодно аннотируются и наиболее значительные комментируются в ежегоднике “Mental Measurement Yearbook” (64, с. 20). Среди создаваемых методик численное превосходство имеют опросники и исполнительные тесты. Для исследовательской работы и психодиагностической практики характерно применение новейших методов статистики. Американские авторы тестов обычно применяют так называемую эмпирическую стратегию, т. е. создают как можно больший набор тестовых задач без какой-либо системы или внутренней логики, а после применения этой экспериментальной формы на большом количестве испытуемых результаты подвергаются корреляционному и факторному анализу.

Американские тесты для нужд просвещения активно используются практически во всех странах Европы. Только во Франции осуществляется попытка создания оригинальных тестовых методик, основанных на концепции Р. Заззо (64, с. 20).

Интеграционные процессы, характеризующие современную цивилизацию, дали толчок развитию новых направлений педагогического тестирования. Оценка качества образования приобретает все большую значимость по мере увеличения количества стран, осознающих преимущества сопоставления собственных достижений с международным контекстом. В связи с этим наибольшую актуальность приобретают международные сравнительные исследования в области школьной успеваемости. Проведение международных исследований предоставляет странам возможность всесторонней оценки эффективности функционирования своих образовательных систем. Междуна-

родное исследование имеет не только политическую и экономическую ценность, но и, в первую очередь, педагогическую. В современных условиях стремительного нарастания информации педагогам необходимо изучать опыт других стран. Поиску новых, более эффективных систем преподавания и должны способствовать международные сравнительные исследования. С 1959 года основные крупномасштабные международные исследования проводятся под руководством неправительственной организации «Международная ассоциация по оценке школьной успеваемости» (IEA). Учреждения IEA созданы в более чем 45 странах мира и включают министерства образования стран, педагогические факультеты университетов и научно-исследовательские институты.

Первое исследование проводилось в 1959—66 гг. и затрагивало вопросы успеваемости по математике в начальной и средней школе. В этом исследовании участвовало 13 стран. В 70-е годы проводилось исследование успеваемости по естествознанию, литературе, гражданскому воспитанию, усвоению английского и французского языков как иностранных. В 1981 году осуществлено 2-е международное исследование по обучению математике, в котором участвовало уже 20 стран. А также в 80-е годы проводилось изучение учебной среды в рамках класса, письменных сочинений и 2-е исследование успеваемости по естественным наукам.

В настоящее время осуществляется ряд проектов: по дошкольному воспитанию (15 стран), компьютерному обучению (20 стран), исследованию навыков чтения (31 страна). С 1994 года началось 3-е исследование в области математики и естествознания. Это исследование представляет собой поперечный и продольный срезы охваченного им контингента учащихся разных возрастов и из различных классов с целью определения уровня обучаемости по предметам. В конечном итоге результаты исследования должны послужить основой для выработки нового понимания этих дисциплин.

Исследования IEA строятся в определенной последовательности и базируются на объяснительных моделях, основанных на учебных программах. В тестировании обычно принимают участие следующие возрастные категории учеников: 9—10, 13—14 и 17—19 лет. Предполагается обследование тысяч учащихся в каждой стране с помощью академических тестов, составляемых на основе учебных программ и других школьных и внешкольных занятий.

Используемые тесты успеваемости направлены на измерение достижений учащихся, которые можно оценить в баллах, скорректировать и сравнить с аналогичными данными по стране, а так же соотнести с другими факторами, влияющими на успеваемость (например, с содержанием обучения, технической оснащенностью школы, полом учащихся, местностью проживания и пр.). Цель такого исследования — создание достаточно гибкой модели, позволяющей измерить комплексное воздействие школы на процесс овладения знаниями. Исследование, как правило, включает последовательное осуществление 6 основных этапов:

- разработка вопросов политики исследования;
- определение методологии и составление плана исследования;

- разработка инструментария;
- проведение выборки участников исследования;
- сбор и обработка данных;
- анализ результатов и составление отчета.

Например, для исследования по математике был разработан перечень основных понятий, включавший 133 графы по 5 категориям (арифметика, алгебра, геометрия, статистика, измерения). Каждый параметр проведения был разделен на 4 уровня: счет, понимание, применение и анализ. Страны-участницы представили по этим параметрам сведения, отражавшие то значение, которое придается этому параметру в национальных школьных программах, А также подготовили задания, которые считали необходимыми для проверки каждого параметра.

Уже первичный анализ показал различия в программах. Так, например, для некоторых стран оказались неактуальными такие понятия как конечное множество и экспонента целого числа.

Полученная информация после проведенного анализа составила основу для разработки тестов, в которые вошли наиболее оптимальные задания, предложенные странами-участницами (обычно из нескольких тысяч заданий отбирается около 200).

Результаты исследований позволяют собрать не только статистический материал, но и всесторонне проанализировать функционирование системы на различных уровнях. На макро-уровне системы образования в программах отражаются намерения — официальная цель, идеи и традиции образования в стране. Набор желаемых результатов вместе с материалом курса, официальными программами и учебниками составляет содержание образования, которое реализуют учителя в конкретных школах или классах (уровень школы), И, наконец, содержание образования, представленное в виде знаний, навыков и установок, которые усваивают учащиеся и которые оцениваются посредством тестов:

Уровни	Предварительная	Контекст учебной программы	Содержание образования
система	характеристика системы	институциональные структуры	планируемое
Школа или класс	характеристика общины, школы, учителя	учебные программы и условия обучения	реализуемое
учащийся	специфика среды обитания учащегося	поведение учащегося	усвоенное

Такой анализ позволяет ответить на многие вопросы. Например, насколько реализуемый вариант содержания образования совпадает с планируемым, т. е. работают ли учителя по программам; соответствует ли понимание учителями

планируемого содержания намерениям на уровне системы; изучаются все темы программы, или некоторые опускаются и т. д.

Помимо IEA существуют другие организации, занимающиеся проблемами международного тестирования. Например, Институт образования ЮНЕСКО (Гамбург) и Международный центр педагогических исследований в Париже проводят совместно с IEA исследование навыков чтения в 35 странах (1990 г.). Тест предполагал изучение 3 аспектов: чтение описательного повествовательного текста и восприятие графики (диаграмм, карт, докладов, требовавших от учащихся умения восстанавливать важные фрагменты информации). Так, например, по результатам изучения восприятия графики наилучшие результаты были у учащихся Финляндии, Швеции, Швейцарии, Гонконга, Франции, Венгрии, Словении. Россия в этих исследованиях не участвовала. Подчеркнем еще раз, что определения лучшей или худшей страны в изучаемой области не является целью тестирования. Задачей исследования становится поиск наиболее эффективных условий обучения.

И, завершая обзор становления и развития метода теста, хочется подчеркнуть важность и педагогическую ценность этого метода, дающего возможность статистически точно анализировать процесс получения образования, искоренять недочеты и видеть дальнейшие перспективы его развития. Критическое отношение к тестированию, понимание его закономерностей позволит педагогу адекватно использовать тесты для улучшения педагогического процесса.

Глава 2. Достижения учащихся и способы их формализации

2.1. Психолого-педагогический анализ сущности достижений учащихся.

2.1.1. Проблема контроля и оценивания.

Школьные достижения представляют собой систему определенным образом связанных элементов. К таким элементам относятся: освоение учащимися основных понятий конкретных учебных дисциплин, сформированность умственных и практических умений и навыков и т. д. Оценивание уровня развития каждого из них делает необходимым при конструировании тестов использование различных видов тестовых заданий.

В теоретической и практической педагогике важное место всегда принадлежало проблеме контроля, учета и оценки достижений, учащихся в учебном процессе, ибо они являются необходимым компонентом последнего. Актуальность данной проблемы сохраняется в силу ее многогранности и сложности разрешения.

В соответствии с педагогической теорией функциями контроля, оценки и учета являются информирование о результатах учебной работы и отражение динамики развития учащихся. Кроме того, они сами имеют воспитывающее и обучающее значение. Контроль представляет собой реализацию принципа обратной связи, без него невозможно управление. В процессе контроля происходит постоянное сравнение заданной программы учебной деятельности с фактическим выполнением.

В традиционной педагогике выделяются следующие задачи контроля:

1) Установить готовность учащегося к восприятию и усвоению новых знаний (восстановить внутри предметные и межпредметные связи).

2) Получить информацию о характере самостоятельной работы учащихся в процессе обучения.

3) Выявить трудности и ошибки, причины их возникновения.

4) Определить эффективность организации, методов, средств обучения.

5) Выявить степень правильности, объем, глубину знаний, умений учащихся.

Процесс оценивания должен сопровождаться конкретным разбором положительных и отрицательных сторон и недостатков в знаниях, умениях учащегося. При этом необходимо учитывать:

— объем знаний,

— степень понимания,

— прочность,

— умение применять знания,

— качество изложения (устного и письменного),

— количество и характер допущенных ошибок,

— своевременность выполнения,

— умение пользоваться учебными пособиями.

Данная система оценивания предусматривает использование оценочных суждений после выставления балльной оценки. При этом предполагается оценивание не только наличия определенного количества знаний, но и качества учебного труда.

Несовершенство данной системы оценивания достижений учащихся, прежде всего, связано с целеполаганием, с некоей односторонностью в определении целей учебного процесса. Кроме того, с закрепившимся в педагогической практике, в силу его простоты и доступности в использовании, оцениванием при помощи балльной отметки, адекватность и объективность которой не раз подвергалась серьезной критике в педагогических исследованиях.

Использование только балльной отметки как средства оценивания достижений учащихся при отсутствии надежного инструментария, позволяющего §(оценить достижения по всем заявленным параметрам, и, кроме того, наборе параметров, не отражающем результативность процесса обучения в целом, не может обеспечить необходимую эффективность процесса контроля и оценивания достижений учащихся. Так, например, в ряду последних исследований данной проблемы, можно назвать исследование, проведенное Е. Б. Шияновой на базе экспериментальной школы № 82 РАО (Московская область), которое показало, что оценивание посредством балльной отметки не реализует в должной мере функцию инструмента воздействия на ученика. По полученным в исследовании данным, уверены, что оценка не соответствует их знаниям 27% учащихся, считают отметку заниженной 24%, считают, что лучше всего об уровне своих знаний информированы они сами 64% учащихся, вообще не обращают внимания на отметку 70% учащихся.

2.1.2. Достижения учащихся как итог обучения.

В связи с появлением большого количества ОУ нового типа и множественности образовательных маршрутов, педагогическая практика потребовала более

динамичной, объективной и всесторонней системы оценивания достижений учащихся в ходе учебного процесса, которая реализовывала бы все присущие ему функции, в том числе функцию стимулирования учебной деятельности, ибо, как подчеркивает д.п.н. Лебедев О. Е. (47), проблема стимулирования учебных достижений учащихся весьма актуальна для современной школы. Причем подобная постановка проблемы крайне актуальна еще и по следующей причине: рассматривая проблему достижений, педагоги-исследователи по идеологическим соображениям не атрибутировали ее как собственно проблему достижений учащихся. Речь шла о результативности, эффективности процесса обучения, которые связывались в большей мере с деятельностью учителя. Постановка проблемы в нынешней ее формулировке стала возможной только на современном этапе, т. к. детоцентристский подход к процессам обучения и воспитания больше не сдерживается официальной идеологией. Проблемам достижений учащихся, их сущности, критериям, по которым их можно было бы оценить, разработке инструментария, который обеспечил бы объективность оценки и др. посвящены многие педагогические и психологические исследования. Над этой проблемой в частности работали В. М. Блинов (10), В. П. Беспалько (8), Б. Битинас (9), Н. Д. Левитов (34), Н. А. Менчинская (15), ? И. Калмыкова (25) и др.

В своем монографическом труде, посвященном проблеме эффективности обучения, В. М. Блинов (10) подвергает критике балльную оценку как показатель успеваемости, ибо она базируется на субъективном представлении, и потому надежность и объективность ее весьма невысоки. Существенный недостаток работ, посвященных совершенствованию существующей системы оценивания результативности педагогического процесса, он видит в том, что в центре внимания оказывается, главным образом, деятельность учения, которая является лишь одним из компонентов обучения, и потому оценка свойств учения недостаточна для суждения об обучении в целом. В. М. Блинов предлагает для характеристики учебного процесса использовать две переменные величины — обучаемость и обученность, где обучаемость независимая переменная, а обученность — зависимая. Диагностика достижений учащихся представляет собой в данном случае фиксацию уровня обученности, как комплекса ответных реакций на обучающее воздействие. В ходе исследования были выявлены зависимости между обучаемостью и обученностью и между изменением организации учебной деятельности и ее эффективностью; предложен способ количественной оценки уровня обученности. Предложенный автором способ оценивания достижений учащихся безусловно придает оценке большую объективность, поскольку увеличивает количество параметров, по которым происходит оценивание, и в процессе оценивания реализуется индивидуальный подход, так как оно осуществляется от прежнего уровня достижений каждого учащегося. Однако следует отметить, что обученность как результат процесса обучения понимается о здесь в узко дидактическом смысле, то есть не является показателем, позволяющим оценить результативность целостного процесса обучения, а представляет собой один из показателей, характеризующий

главным образом уровень освоения знаний и способов учебной деятельности и подлежащий оцениванию наряду с воспитанностью и образованностью. Данная проблема исследовалась сотрудниками НИИ содержания и методов обучения (Р. Ф. Кривошаповой (32), Э. А. Красновским, В. З. Резниковой), которые в качестве критериев эффективности обучения рассматривают требования к знаниям, умениям и навыкам учащихся, «по соответствию которым молено было бы оценивать результаты учебной подготовки», т. е. достижения учащихся. Требования к знаниям рассматриваются как форма конкретизации целей обучения, на которую учитель самым непосредственным образом ориентируется в своей повседневной работе. От логики описания требований и единства их понимания зависит возможность разработки объективных критериев и норм оценок. В ходе исследования были получены многообразные данные о содержании требований, предъявляемых к знаниям и умениям учащихся, выявлены компоненты содержания обучения, чаще всего контролируемые в процессе обучения. Результаты исследования свидетельствуют о широком распространении в школе проверки усвоения фактического материала учащимися. Контроль за усвоением понимания сущности фактов, взаимосвязи между ними, уровня овладения законами и закономерностями невелик. Еще реже контролируются умения применять теоретические знания на практике; умения же, являющиеся основой умственного развития, в школьной практике вообще специально не контролируются. Таким образом, данное исследование показало, что требования, предъявляемые к результатам процесса обучения в школьной практике, явно недостаточны, и что в учебном процессе оцениванию подлежат не все компоненты содержания обучения, чаще всего контроль достижений учащихся сводится к контролю за усвоением учащимися определенной суммы знаний.

Учитывая значимость всестороннего контроля и оценки достижений учащихся в учебном процессе, в педагогической теории многими исследователями разрабатывались различные подходы к анализу и оцениванию эффективности процесса обучения с учетом не только качества усвоения знаний, но и формирования умений и навыков учащихся.

Одним из направлений исследования данной проблемы является выделение в качестве достижений учащихся системы умений, овладение которыми позволяло бы считать процесс обучения эффективным. Примером реализации данной исследовательской задачи является предложенная группой исследователей: Краевским В. В., Высоцкой С. И., Шубинским В. С. (30), — система умений, которые вытекают из структуры деятельности обучения и отражают ее процессы, происходящие в каждом звене:

- а) находить или выбирать учебное действие в качестве ведущего, определять его конкретную значимость;
- б) воспроизводить и применять по образцу имеющиеся знания, умения, навыки;
- в) строить план предстоящего действия, представлять или уметь выбирать предполагаемые средства и способы его выполнения;

- г) предвидеть возможные результаты предстоящего действия, его свойства и качества;
- д) реализовывать план, найденные средства и способы в процессе взаимодействия с объектом действия;
- е) находить объект действия, правильно оценивать его свойства и качества, определять способы взаимодействия с ним;
- ж) контролировать и доказывать правильность процесса выполнения действия, его промежуточных и конечных результатов, соотносить их с соответствующим планом, выявлять допущенные ошибки и неточности;
- з) устанавливать объективную и субъективную, индивидуальную и социальную значимость выполненного действия, оценивать его результаты.

Предложенный перечень умений носит достаточно общий характер. Подобного рода или более конкретизированные перечни умений, овладение которыми позволяет судить о качестве процесса обучения, предлагались в качестве критериев его эффективности различными исследователями многократно. Но данный подход представляет интерес преимущественно для педагогической науки, ибо систематический контроль и оценивание формирования выделенных здесь умений представляется возможным лишь в условиях научного эксперимента, в условиях же школьной практики он представляется достаточно затруднительным.

Интересные возможности в разработке критериев эффективности процесса обучения открывает использование некоторых положений Лернера И. Я. (35), из которых автор считает необходимым исходить при отборе содержания образования в современной школе. Описывая процедуру отбора содержания образования, автор выходит на результат, который должен быть получен в процессе освоения этого содержания. А именно: освоение учащимися в оптимальном соотношении видов деятельности, извлекаемых из социального опыта, функций, к выполнению которых надо готовить учащихся, системы ценностей, которую предстоит освоить.

Интересный подход к рассмотрению проблемы достижений учащихся предлагает О. Е. Лебедев (47). Он исходит из того, что уровень достижений каждого учащегося складывается из четырех компонентов:

1. уровня академической успеваемости с учетом государственного стандарта образования и профиля или специализации конкретного учебно-воспитательного учреждения;
2. фонда коммуникативных умений;
3. творческой познавательной активности ребенка на протяжении всего курса обучения;
4. личных социальных достижений внутри школы и вне ее. Таким образом, в педагогической теории можно выделить несколько подходов к анализу проблемы достижений учащихся в процессе обучения.

Первый из них исповедует традиционное понимание достижений учащихся как возрастание объема усвоенных в процессе обучения знаний, умений и навыков, уровень усвоения которых оценивается при помощи балльной оценки (отметки). Оценивание, как показывают экспериментальные исследования, реально

осуществляется не по всем параметрам, которые декларируется как подлежащие оценке, в силу ограниченности возможностей оценивания с помощью балльной оценки и отсутствия надежного инструмента оценивания таких сложных образований как, например, умения умственной деятельности, личностные характеристики и пр.

Попытки усовершенствовать существующую систему оценивания достижений учащихся грешат некоторой односторонностью. А именно, рассматривается, как правило, не результативность всего процесса обучения в целом, что предполагает комплексную оценку достижений учащихся в ходе него, а какой-либо из его аспектов. Например, делаются попытки усовершенствовать существующий способ оценивания достижений учащихся при помощи балльной оценки за счет расширения возможностей самой оценки (отметки). При таком подходе свидетельством наличия достижений учащихся в учебном процессе остается рост количественных показателей и речь идет только о совершенствовании инструмента измерения последних.

Некоторыми исследователями делается попытка выведения некоторых качественных показателей. Однако, использование их в педагогической практике оказывается достаточно сложным в силу недостаточной разработанности механизма их применения.

Наиболее оптимальным представляется подход к рассмотрению данной проблемы, сторонники которого предлагают учитывать в процессе оценивания достижения учащихся в комплексе:

- а) академическую успеваемость (с учетом образовательного стандарта);
- б) овладение фондом коммуникативных умений, приобретенных в ходе учебного процесса (личностные приобретения);
- в) сформированность мотивов учебной деятельности;
- г) личностные социальные приобретения;
- д) систему ценностей.

Реализация данного подхода предполагает разработку соответствующего инструментария, который позволит обеспечить надежность и достоверность результатов оценивания.

Представляет интерес подход к разрешению проблемы отслеживания достижений учащихся, предложенный для системы TIMSS.

С целью обеспечения большей объективности при оценивании достижений учащихся TIMSS предлагает учитывать ряд факторов, влияющих на мотивацию учебной деятельности. Выделенные факторы подразделены на три группы.

1) Социально-экономические:

- а) демографические характеристики,
- б) экономическое состояние семьи, социально-экономический статус,
- в) культурное состояние семьи.

2) Социально-психологические:

- а) обстановка в классе, влияние учителя и сверстников,
- б) внешкольная активность.

3) Индивидуальные:

- а) образовательный статус учащегося,

б) оценка учащимся собственных достижений и мнение о факторах успеха. Учет перечисленных факторов в процессе организации совместной деятельности учащихся и педагога усиливает мотивацию учебной деятельности учащихся, которая в значительной мере определяет уровень их достижений. Интересен и концептуальный подход TIMSS к контролю усвоения учащимися содержания образования. Оценка достижений учащегося осуществляется с учетом всех заявленных социальных и социально-психологических факторов действующих в системе. Кроме того, оценка уровня усвоения содержания производится не только на основании соответствия его существующему стандарту, а с учетом особенностей конкретного образовательного учреждения. Интересно, что одним из компонентов содержания образования является и поведение учащегося.

В предлагаемой системе оценивания достижений учащихся прослеживается четкая детоцентристская ориентация, в русле которой разрабатывают проблему оценивания достижений учащихся и некоторые отечественные авторы — сторонники комплексного подхода.

С целью выяснения уровня осмысления проблемы достижений учащихся в Санкт-Петербурге ведущими методистами города была проведена деловая игра. Расшифровка записей с последующим контент-анализом показали, что методисты-практики выделяют три группы достижений учащихся:

- 1) система знаний умений и навыков, формируемых в рамках конкретного учебного курса;
- 2) система знаний умений и навыков, соответствующих стандарту образования;
- 3) так называемые отсроченные достижения, которые предполагают оценку:
 - а) качества мировоззренческих новообразований,
 - б) степень социализации личности.

Аспекты проблемы, касающиеся оценивания уровня развития интеллектуально-познавательной сферы, в процессе обсуждения затрагивались мало. В качестве средства стимуляции учебной деятельности с целью достижения ее максимальной эффективности участниками штурма было названо обеспечение положительного эмоционального состояния учащихся в процессе обучения за счет использования соответствующих педагогических технологий и оптимальной организации процесса контроля и оценивания результатов учебной деятельности (экзаменов, зачетов, собеседований).

Таким образом, результаты игры еще раз показали, что методисты-практики склонны анализировать различные аспекты данной проблемы, проблема достижений учащихся, организации контроля и оценивания их в целом не получила еще достаточного осмысления.

2.1.3. Психологический аспект достижений.

Разрешение проблемы определения сущности достижений учащихся, разработки адекватных способов их оценивания требует использования наряду с педагогическим и психологическим знаниями. В психологии проблема результативности процесса обучения рассматривается исходя из того, что сам он представляет собой процесс стимуляции и управления внешней и внутренней деятельностью учащегося.

Интерес представляет оценивание достижений учащихся в рамках концепции развивающего обучения Л. В. Занкова (22, 23), где предусматривается максимальное содействие умственному развитию учащихся. Умственное развитие характеризуется и тем, что отражается в сознании, и еще больше тем, как происходит это отражение. По мнению Л. В. Занкова, решающим в плане умственного развития является объединение в определенную функциональную систему разнохарактерных по своей природе способов действия. О прогрессе в умственном развитии можно говорить тогда, когда происходит объединение в одну систему, в единую аналитико-синтетическую деятельность этих разнохарактерных способов умственной деятельности. В связи с этим встает вопрос о содержательных критериях умственного развития. Перечень таких самых общих критериев дает Н. Д. Левитов (34):

- 1) самостоятельность мышления,
- 2) быстрота и прочность усвоения,
- 3) быстрота умственной ориентировки (находчивость) при решении нестандартных задач.
- 4) глубокое проникновение в сущность явлений (умение отличать существенное от несущественного),
- 5) критичность ума (отсутствие склонности к предвзятым, необоснованным суждениям).

Д. Б. Эльконин предлагает рассматривать в качестве основного критерия умственного развития наличие правильно организованной структуры учебной деятельности (сформированная учебная деятельность) с ее компонентами:

- а) постановкой задачи,
- б) выбором средств,
- в) самоконтролем,
- г) самопроверкой,
- е) правильным соотношением предметных и символических планов в учебной деятельности.

Н. А. Менчинская (15) рассматривает между ними такие особенности мыслительной деятельности как:

- а) быстрота усвоения,
- б) гибкость мыслительного процесса (легкость или трудность перестройки работы, приспособление к изменяющимся условиям задач),
- в) тесная связь или разрозненность наглядных и отвлеченных компонентов,
- г) различный уровень аналитико-синтетической деятельности.

Е. Н. Кабанова-Меллер основным критерием умственного развития считает широкий и активный перечень приемов умственной деятельности, сформированных па одном объекте на другой объект.

3. И. Калмыкова предлагает свою систему критериев, включающую:

- а) темп продвижения, который определяем и однотипных упражнений, необходимых для формирования обобщения (при этом быстрота работы и быстрота обобщения — разные вещи),

б) экономичность мышления: учащиеся с более высоким уровнем умственного развития обладают большим фондом знаний и способов оперирования ими, поэтому их путь к решению проблемы отличается краткостью и рациональностью.

Областью интересов названных выше исследователей является умственное развитие учащихся. Именно эта сфера — сфера умственной деятельности — чаще всего оценивается как показатель эффективности процесса обучения. Однако это не единственная цель, а следовательно, и не единственный результат процесса обучения. В процессе обучения оказывается влияние на психическое развитие ребенка в целом и потому контроль и оценивание должны по возможности касаться всех сфер, характеризующих психическое развитие ребенка, к которым Л. С. Выготский (16) относит:

а) познавательную сферу (восприятие, память, воображение, мышление),
б) сферу личностных отношений (эмоции, потребности, мотивы, ценностные ориентации и т. д.),

в) систему практических и умственных действий, обеспечивающих возможность практической и умственной деятельности.

Таким образом, основанием для оценивания уровня достижений ребенка является диагностика его психического развития, которая позволяет определить реальный уровень развития ребенка и зону его ближайшего развития.

Основная трудность в реализации контроля и оценивания психического развития учащихся в процессе обучения состоит в отсутствии надежного инструмента его измерения.

Сторонники комплексного подхода к контролю и оцениванию результатов процесса обучения именно в этом направлении осуществляют свой научно-педагогический поиск. / Таким образом, в проблеме достижений учащихся, их контроля и оценивания условно можно выделить два аспекта:

1. дидактический,
2. психолого-педагогический.

Первый из них подразумевает разработку определенного образовательного стандарта и оценивание достижений учащихся по степени соответствия их этому стандарту.

Второй — предполагает разработку механизма контроля и оценивания в процессе обучения динамики развития учащегося. Данный аспект проблемы может быть разрешен только совместными усилиями педагогики и психологии. Использование психологического знания позволит разработать надежный и доступный для применения его педагогами-практиками инструментарий, с помощью которого можно было бы адекватно оценивать уровень развития учащихся на каждом этапе процесса обучения.

Показатели, характеризующие достижения учащихся в рамках обоих аспектов рассматриваемой проблемы — это показатели личностных приобретений конкретного учащегося, показатели активного освоения и присвоения им интеллектуально-социального пространства. Достижения учащихся находятся в зависимости от степени интериоризации каждого из участников педагогического процесса. Разработка инструментария, доступного для

использования участниками учебного процесса, простимулирует ускорение и расширение процесса интериоризации, что, в свою очередь, обеспечит контроль и стимулирование достижений учащихся в процессе обучения.

Очевидно, что проблема результативности педагогического процесса постоянно находится в центре внимания как теоретиков, так и практиков, работающих в сфере образования. Актуальность ее объясняется наличием определенных несовершенств в "системе контроля и оценивания достижений учащихся в процессе обучения и сложности поиска путей повышения ее эффективности. Процесс обучения складывается из деятельностей учения и преподавания. Основными показателями эффективности и той и другой являются, прежде всего, достижения обучающихся, которые и являются объектом нашего интереса. В связи с этим возникает необходимость ответить на ряд вопросов. А именно: что можно считать достижениями учащихся? Чем они определяются? Каким образом их можно обнаружить и зафиксировать? Какие цели преследуются в процессе оценивания достижений учащихся?

2.1.4. Целеполагание как основа определения достижений.

Анализ проблемы результативности учебного процесса следует начинать, обратившись к проблеме целеполагания, ибо результатом любой деятельности является реализация поставленных в ходе нее целей. А достижения участников процесса могут быть оценены по степени их соответствия поставленным целям. Общеизвестно, что в традиционной педагогической теории и практике в процессе обучения реализуются три функции: образовательная, воспитательная и развивающая. Следовательно, в ходе него решаются задачи образования, воспитания и развития. Однако следует отметить, что существенным недостатком существующей системы обучения является то, что некоторые из провозглашаемых целей и задач процесса обучения только декларируются. Специально предусмотренные средства для их реализации, контроля и оценивания результатов предусматриваются лишь для решения задач образования (понимаемого как овладение суммой знаний, умений и навыков учебной деятельности).

То есть, фактически в современной педагогической практике и качестве достижений учащихся оценивается уровень овладения ими некоей, соответствующей определенному стандарту, суммой знаний, умений и навыков учебной деятельности. Ранее контролю подлежала также степень сформированности мировоззрения, его содержательной стороны, которая определялась лишь лояльностью по отношению к господствующей идеологии, то есть также соответствием определенному стандарту. Существовал еще один параметр, подлежавший оценке, которым оценивалось поведение учащегося — уровень его дисциплинированности.

И это все, чего мог добиться и что могло быть поставлено в заслугу учащемуся. Ограниченность и недостаточная эффективность такой системы контроля и оценивания результатов процесса обучения не раз подтверждались педагогической наукой.

Можно предположить, что причиной отсутствия в педагогической практике системы контроля и оценивания реализации всех поставленных целей была идеологическая установка (ориентация на формирование дисциплинированной, исполнительской, но безынициативной, инертной личности, что определялось социальным заказом и обеспечивалось использованием соответствующих педагогических средств, хотя цели провозглашались совершенно иные и оценивание степени их реализации было нецелесообразным). Однако, следует признать весьма серьезной причиной этого саму сложность разработки адекватной системы контроля и оценивания результатов процесса обучения, ибо оценке подлежат личностные приобретения учащихся, отслеживание которых на современном уровне человекознания представляет значительную проблему.

Так, например, контроль и оценивание степени реализации воспитательных целей характеризовались не систематичностью, случайностью, отсутствием инструментария, обеспечивающего объективность оценки.

Еще одной причиной, кроме вышеназванных, является существующее в традиционной педагогике деление педагогического процесса на обучение и воспитание, что еще в какой-то мере допустимо для научного анализа, но совершенно неприемлемо для педагогической практики, ибо единый педагогический процесс при этом искусственно разделяется на две составляющие, каждой из которых приписываются собственные цели, средства, результаты, назначаются лица, ответственные за реализацию этих целей.

Вследствие этого участники педагогического процесса, отвечающие за образование, неосознанно (неформально) освобождаются от ответственности за воспитательные результаты своей деятельности, так как на своем уровне они их не оценивают.

Выделение воспитания и обучения в отдельные процессы создает сложности с целеполаганием и осуществлением контроля и оценивания сформированности некоторых личностных формирований, которые являются продуктами в равной мере и того и другого процессов. Примером такого рода формирований могут быть, в частности, мировоззрение, фонд коммуникативных умений и пр.

Если реализация воспитательных целей, хотя и неполноценно, но в какой-то мере контролируется и оценивается, то реализация целей развития практически совсем не контролируется, что подтверждается педагогическими исследованиями. В настоящее время отсутствует какой-либо надежный инструментарий, с помощью которого можно было бы оценивать и в процессе обучения степень реализации целей психического развития и формирования личностных образований учащихся.

Теоретические работы, посвященные проблеме умственного развития учащихся (Л. В. Занкова (22,23), Н. А. Менчинской, В. В. Давыдова, Л. Б. Эльконина и др.), интересны и, безусловно заслуживают внимания. Однако на практике проблема оценивания уровня развития учащихся остается неразрешенной.

Причиной этого, кроме прочих, может быть сложность использования в педагогической практике инструментария, при помощи которого в ходе научных исследований оцениваются достижения учащихся, характеризующие

уровень их развития. Однако и здесь речь не идет о контроле и оценивании достижений личности в целом, а рассматривается только один из аспектов данной проблемы.

Таким образом можно констатировать, что в современной педагогической практике контроль и оценивание достижений учащихся осуществляются несколько односторонне, то есть достижения оцениваются только по конечному результату, а точнее по соответствию результатов деятельности учащегося по овладению определенной суммой знаний, умений и навыков учебной деятельности существующему стандарту. Собственно деятельность учащегося, его энергетические затраты, уровень его собственного развития, на который ему позволили подняться, его усилия, характер его личностных приобретений и т. д. остаются неоцененными.

В связи с этим приобретает актуальность вопрос о целях самого процесса контроля и оценивания. Недостаточная эффективность существующей системы контроля и оценивания достижений учащихся связана в значительной мере с тем, что в ходе них не реализуется цель — формирование у учащихся положительной мотивации к учебной деятельности. Такой вывод подтверждается экспериментальными исследованиями, проведенными Е. Б. Шияновой.

Исследования показали, что учащиеся и учителя совершенно по-разному оценивают уровень знаний учащихся. Такую неадекватность в оценке знаний учителем и учеником исследователи объясняют различием критериев, на которые они полагались. Среди учащихся, не согласных с отметкой, большинство составляли те, кто ориентировался в своей работе не на результат, а на затраченный труд или на способы выполнения заданий. Следовательно, для того, чтобы в процессе оценивания достижений учащихся добиться формирования положительной мотивации к учебной деятельности, необходимо отслеживать индивидуальное продвижение учащегося в процессе освоения знаний, умений и навыков, развития психических процессов, формирования личностных образований.

Таким образом, оценивание достижений учащихся в процессе обучения целесообразно осуществлять по двум основаниям:

— по соответствию результатов учебной деятельности определенному образовательному стандарту,

— по личному продвижению учащегося по лестнице достижений в процессе освоения знаний, умений развития психических процессов, формирования ценностных ориентацией, личностных качеств.

При оценивании достижений, выделенных по второму основанию, следует иметь ввиду, что организация контроля и оценивания их зависит от характера оцениваемых личностных приобретений, то есть от того, что именно оценивается: знания, умения или личностные качества, время формирования которых весьма различно.

Представляется целесообразным выстроить в соответствии с иерархией целей процесса обучения иерархию достижений учащихся. Система контроля и оценивания этих достижений определяется общими и частными

педагогическими целями и задачами. Педагогические цели и задачи определяют деятельность учащегося и характер контроля и оценивания ее результатов в процессе выполнения одного задания, в процессе изучения одной темы, в процессе освоения курса или в процессе учебной деятельности учащихся в рамках учебного года. Это деление достаточно условно, ибо более общие цели могут исключать в себя более частные.

В соответствии с этими целями и достижения учащихся могут быть структурированы. То есть существуют результаты, которые могут оцениваться по выполнению учебного задания или после освоения учебного курса и т. д. Могут быть результаты, например, продвижение в развитии тех или иных психических процессов, или, тем более, сформированность личностных образований, оценивание которых возможно по истечении достаточно длительного срока, зависящего от психофизиологических механизмов формирования оцениваемого образования.

Наиболее общей целью педагогического процесса, признанной в педагогической теории и практике, является формирование личности учащегося, реализация которой предполагает формулировку и реализацию более конкретных целей и задач.

Для того чтобы цели педагогического процесса были реализованы, необходимо предусматривать средства их реализации и обеспечивать функционирование механизмов контроля и оценивания результатов.

Разработка инструментария, с помощью которого можно было бы адекватно оценить результаты процесса обучения в педагогической науке, осуществляется постоянно. Недостатком многих таких разработок является то что в них анализируются различные аспекты проблемы, а не проблема в целом где в качестве объекта анализа должна быть представлена целостная система контроля и оценивания достижения учащихся позволяющая дать адекватную оценку совокупности достижений учащихся в процессе обучения.

Решение данной проблемы предполагает описание параметров деятельности учащегося, достаточно полно ею характеризующих, формирование и развитие которых подлежит оцениванию с учетом возможности их оценивания в условиях учебного процесса. Кроме того, необходима разработка критериев, по которым будет осуществляться оценивание методик, оценивания по выделенным критериям, система контроля и учета должны быть созданы условия для систематического отслеживания индивидуальных достижений каждого учащегося.

Под достижением учащегося понимается таким образом продвижение учащегося в процессе учебной деятельности, понимаемое и как процесс (движение от своего прежнего уровня к новому уровню овладения чем-либо) и как результат (соответствие достигнутого учащимся уровня овладения чем-либо определенному образовательному стандарту).

При таком понимании сущности достижений учащихся, контроль и оценивание их при соответствующей их организации позволит решить несколько задач: во-первых информировать самих учащихся об их успехах и неудачах в процессе обучения, что поможет обеспечить формирование у них положительной

мотивации к учебной деятельности; во-вторых: информировать родителей об успехах и неудачах их детей с целью оказания им консультационной помощи; в-третьих: даст возможность учителям получать достаточно полную информацию о результатах учебной деятельности учащихся что позволит учесть их индивидуальные особенности в процессе обучения и обеспечить условия для их более полной самореализации.

Для оценивания достижений учащихся, исходя из представленного выше понимания их сущности, в качестве одного из средств, позволяющего оценивать их с учетом как показателей роста, так и показателей развития, может быть предложено использование элективных тестов, то есть тестов с выбором ответов. Надежность и достоверность измерения и оценки достижений учащихся в процессе обучения с помощью элективных тестов подтверждены экспериментальными исследованиями (М. Ф. Колпаков, Г. "Н. Александров, В. И. Огорелков (45)).

2.2. Педагогические цели и их конкретизация.

Существенную роль при отборе содержания и конструировании тестов достижений играет процедура целеполагания в педагогическом процессе. Конструирование надежного, обеспечивающего достоверность результатов инструментария для оценивания его эффективности, прежде всего, предполагает наличие обоснованной, иерархически выстроенной системы целей.

Проблема целеполагания в системе образования достаточно широко освещалась в зарубежной педагогической литературе. Попытки ее осмысления имеют место и у отечественных авторов. Наиболее полный анализ постановки и разрешения длиной проблемы в различных зарубежных образовательных системах представлен в работе М. В. Кларина (28).

Мы же только назовем здесь области этих разработок и остановимся наиболее подробно на таксономии целей И. Блума.

На основании проведенного анализа, М. В. Кларин выделяет несколько основных способов определения целей обучения. Проанализируем каждый из них с точки зрения возможности использования сформулированных в соответствии с ним целей в качестве основания для конструирования тестов достижений, позволяющих оценить степень их реализации в учебном процессе.

1. Определение целей обучения через содержание.

Выделение целей различного уровня, ориентирующих на освоение учащимися содержания того или иного учебного материала, позволяет уточнить конкретные предметные области, разделы или темы, содержание которых необходимо усвоить. Грамотное построение целей обучения, ориентирующих на освоение содержания учебного материала, представляется достаточно важным для конструирования тестов достижений, поскольку практически все тесты достижений строятся на основании учебного материала и измеряют, в частности, уровень знаний в той или иной предметной области. Таким образом, цели, реализуемые в процессе отбора содержания для тестовых заданий, определяются целями, которые реализуются в процессе отбора содержания учебных программ и курсов.

2. Определение целей через деятельность учителя.

Постановка целей применительно к организации деятельности учителя не представляет для нас значительного интереса, поскольку их реализация предполагает выбор различных педагогических средств, методов, форм организации учебного процесса и т. д., что характеризует деятельность учителя, а не достижения учащихся, и не может служить основанием для разработки тестовых заданий, так как один и тот же уровень достижений учащихся может быть обеспечен при использовании различных педагогических средств.

3. Определение целей применительно к процессам интеллектуального, эмоционального, личностного и т. п. развития ученика.

Примерами такого рода целей могут быть: «формирование умения анализировать», «развитие познавательной самостоятельности» и т. д. Степень реализации данных целей может быть оценена при помощи тестов достижений при условии их конкретизации и детализации.

4. Определение целей применительно к учебной деятельности учащихся, в принципе, может быть дано через описание действий учащихся, и является наиболее приемлемыми при конструировании одной из шкал тестов.

Технология уточненного описания конкретных целей

Приведенные здесь примеры целиком заимствованы нами из книги М. В. Кларина. По нашему мнению они как нельзя лучше иллюстрируют подготовительную работу, необходимую для выделения целей тестирования в сфере образования, в чем-то с выводами автора можно не согласиться, например, с утверждением о невозможности непосредственного измерения достижимости некоторых целей более общего характера, не разложенных до результатов деятельности. Однако эти примеры показывают возможные пути конкретизации целей.

Общий прием конкретизации целей, по мнению М. В. Кларина — использовать в их описании глаголы, указывающие на определенное действие. Перечень таких глаголов приведен в приложении 6. Так, например, цель «изучить использование символических обозначений на погодной карте» может быть развернута в перечень возможных учебных результатов.

Ученик:

- 1) воспроизводит по памяти символы, употребляемые на карте погоды;
- 2) опознает символы на карте;
- 3) читает карту, используя символы;
- 4) составляет карту, пользуясь символами;
- 5) по заданной карте, пользуясь символами, дает прогноз погоды.

На этом примере, приведенном Н. Гронлундом, видно, что общая цель, с одной стороны, может быть сведена к простому результату низкого познавательного уровня (например, варианты 1 и 2), а с другой стороны — развернута в широкий перечень учебных результатов разного уровня.

Приведем еще один пример:

Общая цель «Ученик использует навыки критического мышления при чтении» может быть конкретизирована:

- 1) проводит различия между фактическими сведениями и оценочными суждениями
 - 2) проводит различия между фактами и предположениями;
 - 3) выделяет причинно-следственные связи;
 - 4) выделяет ошибки в рассуждениях;
 - 5) отличает существенные доводы от не относящихся к делу;
 - 6) проводит разграничения между обоснованными и необоснованными оценками;
 - 7) формулирует на основе текста обоснованные заключения;
 - 8) указывает на предпосылки, обосновывающие справедливость выводов.
- Этот пример конкретизации цели не дает ее однозначного, абсолютно воспроизводимого разложения на «наблюдаемые действия»; каждый учитель будет судить о наличии перечисленных признаков, исходя из собственного опыта и культуры мышления.

Степень конкретизации целей — детализация

Приведем пример трех ступеней конкретизации цели — «понимать значение письменного текста» (Гронлунд):

1. Помнить значение письменного текста:

1.1. Выделить все сведения, содержащиеся в тексте в явном виде:

1.1.1. Подчеркнуть конкретные детали в тексте (например, имена, даты, события и т. д.);

1.1.2. Выбрать высказывания, наиболее полно передающие смысл текста.

1.2. Выделить главную мысль текста:

1.2.1. Подчеркнуть предложение, выражающее основную цель;

1.2.2. Выбрать заголовок для текста.

1.3. Подытожить мысли, содержащиеся в тексте:

1.3.1. Написать конспективное изложение текста.

1.4. Вывести из содержания текста те идеи и соотношения, которые не раскрыты в нем в явном виде:

1.4.1. Назвать мысли, действия, события, которые не названы, но предполагаются в содержании текста;

1.4.2. Перечислить обозначенные в тексте действия или о событиях в их наиболее вероятной последовательности;

1.4.3. Подобрать наиболее вероятные последствия описанных действий или событий;

1.4.4. Объяснить, что связывает и объединяет между собой явления, предметы, представления в данном тексте.

2.3. Таксономия целей Б. Блума.

Развитие педагогической мысли на Западе шло долгое время достаточно изолированно, поэтому в настоящее время российские педагоги только открывают для себя значительный пласт разработок, которые там были выполнены. Анализируя открывшиеся для массового читателя разработки западных ученых можно говорить о параллельном, вместо изолированного развития отечественной педагогики.

Группа американских педагогов и психологов, под руководством Б. Блума (28), провела многолетнее исследование с тем, чтобы разработать общие способы формулировки и упорядочения педагогических целей. В 1956 г. издана первая часть «Таксономии», содержащая описание целей в познавательной (когнитивной) области. Эта система целей, названная таксономией Б. Блума, получила широкую международную известность. Ее используют при планировании обучения и оценке его результатов, она служит надежным инструментом при опытной проверке новых курсов.

Впоследствии была создана вторая часть «Таксономии» — и аффективной и психомоторной области. В настоящее время по данным Ковалевой разработаны около 9 различных таксономии учебных целей.

Две первые части таксономии Блума приведены в приложении 8. Мы уже касались проблем построения соподчиненных уровней обучения. Приведем описание целей, данное в работе Кларина.

1. Когнитивная (познавательная) область. Сюда входят цели от запоминания и воспроизведения изученного материала до решения проблем, в ходе которого необходимо переосмысливать имеющиеся знания, строить их новые сочетания с предварительно изученными идеями, методами, процедурами (способами действий), включая создание нового. По данным экспертных оценок, а также опросов учителей и анализа литературы, проведенных Б. Блумом и его сотрудниками, к познавательной сфере относится большинство целей обучения, выдвигаемых в программах, учебниках, в повседневной практике учителей.

2. Аффективная (эмоционально-ценностная) область. К ней относятся цели формирования эмоционально-личностного отношения к явлениям окружающего мира, начиная от простого восприятия, интереса, готовности реагировать до усвоения ценностных ориентации и отношений, их активного проявления. В эту сферу попадают такие цели, как формирование интересов и склонностей, переживание тех или иных чувств, формирование отношения, его осознание и проявление в деятельности.

3. Психомоторная область. Сюда попадают цели, которые связаны с формированием тех или иных видов двигательной (моторной), манипулятивной деятельности, нервно-мышечной координации. К этой области относится сравнительно небольшая доля из общей совокупности целей обучения. Среди них навыки письма, речевые навыки, а также цели, выдвигаемые в рамках физического воспитания, трудового обучения.

Авторы международных сравнительных исследований при создании тестов достижений используют содержательно-деятельностные матрицы, основанные на таксономии Блума, использование таксономии в целостном виде для целей тестирования как правило не проводится. Таксономия служит основой для разработки своих, более локальных целей или дополняется другими, в зависимости от целей предполагаемого тестирования.

Например, в международном исследовании IAEP-II, проведенном американским центром педагогического тестирования ETS в 20 странах мира в

1991 г. в котором участвовала и наша страна, для сравнительной оценки естественнонаучной подготовки школьников использовалась двухмерная система заданий. В соответствии с данной системой каждое задание теста было предназначено для проверки овладения учащимися определенными умениями, характеризующими отдельные компоненты познавательной деятельности на материале различных разделов естествознания. При выделении целей были использованы три позиции:

- знание (воспроизведение),
- применение,
- интеграция, в которую входили элементы понимания и анализа.

В исследовании TIMSS 1995 г. кроме предметной (содержательной) области выделены когнитивная и аффективная области тестирования, при этом к первой относятся следующие виды деятельности:

- знания,
- использование стандартных процедур,
- исследование и решение проблем,
- математические рассуждения,
- пропорциональность,
- коммуникативные умения. Ко второй:
- ожидаемые отношения,
- выбор профессии,
- участие в учебном процессе,
- интересы, свойства ума.

Как можно видеть из этих примеров, таксономия Б. Блума и ее развитие служит основой для разработки современных тестов школьных достижений, однако часть ее, реально используемая, определяется целями тестирования.

2.4. Целеполагание как основа определения достижений. Технологическая матрица.

В общем виде для обеспечения адекватности теста содержанию образования необходимо привлечение ряда экспертов, которые должны указать, что именно должны знать дети того или иного возраста по тому или иному предмету, взвесить важность в тесте одного аспекта некоторого предмета по отношению к другим аспектам данного предмета или смежным.

Должны быть сформулированы цели и задачи учебных программ или курсов так, чтобы просматривалась релевантность заданий по отношению к этим целям. То есть, необходимо определить содержательные и целевые требования к заданиям — сформулировать цели исследования.

Для педагогов первый этап этой работы не должен представлять значительных трудностей, поскольку они сами могут выступать в роли эксперта по проблемам содержания образования, однако основываться только на собственном мнении далеко не достаточно и при дальнейшей работе с заданиями привлечение экспертов становится необходимым.

Определив цели и задачи проверки, необходимо разработать структуру и основное содержание теста. Для этого нужно знать, какие группы учащихся будут участвовать в проверке, будет ли она текущей или итоговой. Если

проверка итоговая, то что она оценивает: достижения учащихся по отдельной теме, части курса или по всему курсу.

Как мы уже говорили, определение содержания для заданий теста может быть проведено через составление спецификации теста. Основную часть спецификации теста представляет собой содержательно-деятельностная матрица. Этот метод широко используется в мировой практике. Его описание предоставлено Г. С. Ковалевой.

Например, в международном исследовании IAEP-II для сравнительной оценки естественнонаучной подготовки школьников использовалась двухмерная система заданий, представленная в таблице 1.

В соответствии с данной системой каждое задание теста было предназначено для проверки овладения учащимися определенными умениями, характеризующими отдельные компоненты познавательной деятельности (воспроизведение, применение и интеграцию знаний) на материале различных разделов естествознания (биологии, физики и химии, наук о Земле и астрономии, методологии науки).

В соответствии с данной системой каждое задание теста было предназначено для проверки овладения учащимися определенными умениями, характеризующими отдельные компоненты познавательной деятельности (воспроизведение, применение и интеграцию знаний) на материале различных разделов естествознания (биологии, физики и химии, наук о Земле и астрономии, методологии науки).

Как видно из приведенной таблицы, при определении спецификации теста выясняются не только структура и содержание проверки, но и процентное соотношение различных вопросов в тесте. Причем вес каждой темы в тесте определяется важностью данной темы и количеством времени, отведенном на ее изучение. Такой подход дает возможность оценить, все ли основные вопросы программы включены в тест и в каком количестве они представлены.

Естественно, что невозможно охватить в одном тесте весь спектр учебных достижений школьников, но желательно, чтобы банк заданий отражал основное содержание проверяемого материала на разных уровнях трудности.

Таблица 1.

	Биология	Физика и химия	Науки о Земле и астр.	Методология науки	Процентное соотношение в тесте
Воспроизведение знаний					40
Применение знаний					35
Интеграция знаний					25
Процентное соотношение в тесте	35	35	15	15	100

Содержательно-деятельностная матрица спецификации теста по естествознанию (IAEP-II)

По сути технологическая матрица представляет собой содержательно-деятельностную модель теста. Содержание горизонтальной строки матрицы как мы говорили не должно представлять трудностей для педагогов. Здесь, как правило, определяются предметы, разделы, учебные темы, разделы учебных тем. Выбор того или иного материала напрямую зависит от целей тестирования. Сложнее дело обстоит с вертикальной составляющей, в западных и международных тестах она стоит как правило на основе той или иной таксономии, для отечественных тестов, в силу причин о которых мы скажем позже, это практически невозможно. В самом простом случае достаточно только одной шкалы, для более сложных тестов в качестве второй составляющей могут быть использованы:

- уровни овладения учебным материалом
- специальные или общешкольные умения и навыки
- уровень развития психических познавательных процессов и т.д. в зависимости от целей тестирования.

Принципиальной здесь представляется необходимость взаимно однозначного соответствия конкретного тестового задания и уровня (свойства, умения) графы матрицы, к которой отнесено это задание.

Более подробно к этому вопросу мы еще вернемся.

Кроме матрицы в спецификацию теста входят также определение типа заданий, которые будут использоваться для оценки достижений учащихся в соответствии с намеченными целями, времени на выполнение теста, особенностей проведения тестирования, которые могут повлиять на характеристики теста.

Для тестов, ориентированных на критерии, отбор содержания теста является самым важным этапом его создания, т. к. для принятия решения о достижении данной цели обучения, например, стандарта, необходимо достаточно точно и полно описать содержание стандарта и выразить его совокупностью заданий, которая была бы представительной для этой цели. Поэтому главной проблемой в разработке тестов, используемых для оценки достижения образовательных стандартов, является соотношение содержания стандарта и содержания теста. Следует отметить, что таблица спецификации теста также может быть использована как основа для анализа и представления результатов тестирования.

2,5. Отечественные разработки уровней целей.

Проблема целеполагания занимала и сейчас занимает значительное место в педагогической науке нашей страны, после фиксации нетехнологичности (Беспалько В.П. и др.) существующей системы целеполагания было сделано несколько практических шагов по их конкретизации. Были описаны требования и характеристики технологично поставленных целей (67), проведены попытки описания таких целей, однако эти широкого применения, в основном в силу слабой экспериментальной поддержки.

Попытки построения иерархично упорядоченных целей в когнитивной области делались и в Советском Союзе, однако эти работы не получили широкого распространения и практического применения. Как правило, выделенные уровни носили умозрительный, конструктивный характер. Экспериментального доказательства их реального существования не проводилось.

В качестве примера можно привести несколько уровней, выделенных разными авторами.

В работах Конфидератова Н. Я., Симонова В. П. (57) выделены уровни:

I уровень — различение — ученик может из предложенных примеров выбрать нужный;

II уровень — запоминание — учащийся может воспроизвести закон, формулу, понятие;

III уровень — понимание — ученик может объяснить закон, понятие, дать примеры его применения;

IV уровень — простейшие умения и навыки — ученик может применить закон, понятие на практике, решает задачи, делает выводы из их решения;

V уровень — перенос — подразумевает возможность применения знаний, полученных по данному предмету в какой-либо смежной области.

В. Г. Королева выделяет 4 уровня:

1. репродуктивное самостоятельное воспроизведение,.
2. репродуктивное алгоритмическое действие,
3. продуктивное эвристическое действие (или прикладной

уровень),

4. продуктивное творческое действие (креативный уровень).

В. П. Беспалько выделяет свои уровни усвоения:

1. ученический (деятельность по узнаванию),
2. алгоритмический (решение типовых задач),
3. эвристический (выбор действия),
4. творческий (поиск действия).

В. Н. Максимова выделяет следующие уровни:

1. узнавание,
2. запоминание,
3. понимание,
4. применение:
 - 4.1. тематическое обобщение,
 - 4.2. предметное обобщение,
 - 4.3. межпредметное обобщение.

Даже беглого взгляда достаточно для того, чтобы констатировать факт отсутствия в отечественной педагогике единого подхода к этой проблеме.

В настоящее время в Санкт-Петербурге, под руководством О. Е. Лебедева (68) проводятся широкомасштабные работы по обоснованию и экспериментальному подтверждению уровней обученности (достижений), которые можно представить следующим образом:

1. грамотность — уровень, характеризующийся способностью использовать основные способы познавательной деятельности,
2. функциональная грамотность — характеризуется способностью решать стандартные жизненные задачи в различных Сферах жизнедеятельности,
3. информированность — характеризуется усвоением определенного объема знаний и способностью репродуцировать их.

4. компетентность:

— общекультурная — уровень, достаточный для творчески-деятельностной самореализации личности, ориентирующейся в ценностях широкого культурного пространства, а также способность личности оценивать границы собственной компетентности,

— допрофессиональная — уровень, достаточный для осознанного выбора профессии и успешного обучения в соответствующем высшем учебном заведении,

— методологическая — уровень, достаточный для самостоятельного исследовательского подхода к решению широкого спектра задач, как теоретического, так и прикладного характера.

Данные уровни достаточно привлекательны, однако их использование невозможно в силу неформализованности и расплывчатости. Возможно, что после завершения этих работ мы наконец-то получим полноценный инструмент для разработки тестов, созданный с учетом особенностей российской школы.

В настоящее время возможно зафиксировать два факта: во-первых, использование зарубежных разработок уровней обученности ограничено целым рядом обстоятельств и в настоящее время практически невозможно; во-

вторых, отечественные разработки в данной области не соответствуют реальным требованиям, которые предъявляются к ним тестологами.

Глава 3. Тесты и их виды. Тесты достижений

Предметом нашего интереса являются тесты достижений, иначе — педагогические тесты или тесты успешности. Основными отличиями тестов от прочих способов выявления индивидуальных различий, таких как наблюдение, анализ документов или продуктов деятельности и др. являются их формализованность и более высокая степень объективности проводимого с их помощью оценивания, которая обеспечивается за счет стандартизации вопросов и ответов, особой процедуры проведения тестирования и способов обработки результатов; подтверждения качества теста: определения надежности, валидности, дискриминативности (сущность данных процедур подробно освещена в главах 3 и 4) и использования математического статистического инструментария для оценки полученных результатов. Практически все современные тесты создаются на основе математической теории оценки качества заданий и параметров личности.

Различие между тестами психологическими и тестами достижений, с одной стороны, и тестами достижений и традиционными контрольными работами (при многообразии видов контрольных работ), с другой, достаточно условно и зачастую разделить их сложно. Принципиальным для нас представляется то, что использование тестов достижений направлено на выявление степени овладения учащимися конкретными знаниями, умениями, навыками, измерение уровня развития которых может осуществляться на различных этапах процесса обучения, в отличие от наиболее близких им по структуре тестов интеллекта, которые предназначены для выявления уровня овладения испытуемыми разнообразными знаниями и умениями, осваиваемыми ими в процессе обретения жизненного опыта в целом. Измерение уровня достижений учащихся на различных этапах процесса обучения с помощью тестов достижений позволяет определить степень эффективности тех или иных специальных программ обучения, качественный уровень профессиональной или любой другой подготовки.

Тесты достижений как инструмент оценивания имеют значительные отличия от контрольных работ. Во-первых, тесты — значительно более качественный и объективный способ оценивания и, во-вторых, что наиболее важно и принципиально, показатели тестов ориентированы на измерение степени, определение уровня усвоения ключевых понятий, тем и разделов учебной программы, умений, навыков и пр., а не на констатацию наличия у учащихся определенной совокупности формально усвоенных знаний.

Стандартизированная форма оценки, используемая в тестах достижений, позволяет соотнести уровень достижений учащегося по предмету в целом и по отдельным его разделам со средним уровнем достижений учащихся в классе и уровнями достижений каждого из них. Анализ сущности понятий «достижения» и «достижения учащихся» в современной их трактовке дается в главе 2.

Для того, чтобы дать наиболее полную характеристику интересующего нас вида тестов и определить их место, значение» взаимосвязи в общей системе тестовых методик, мы считаем необходимым дать представление о существующих на данный момент видах тестов. Тесты могут быть классифицированы по следующим основаниям с выделением соответствующих видов;

1. По процедуре могут быть выделены стандартизированные и нестандартизированные тесты,

Стандартизация психологами понимается в двух аспектах:

— стандартизация процедуры и условий проведения тестирования, способов обработки и интерпретации результатов, которые должны привести к созданию равных условий для испытуемых и минимизировать случайные ошибки и погрешности как на этапе проведения, так и на этапе обработки результатов и интерпретации данных;

— стандартизация результатов, то есть получение нормы, шкалы оценки, которая служит основанием для определения о уровня овладения тем, что выясняет данный тест, при этом не важно, какого рода нормы получаются и какие шкалы используются.

Многие специалисты-тестологи вообще не считают возможным рассматривать так называемые нестандартизированные тесты как вид тестов, так как на их взгляд, стандартизованность является необходимой особенностью, определяющей сущность теста, как диагностической методики, его характеристикой.

Для нас важно разделить эти два понятия. Поэтому тесты нестандартизированные по результатам мы будем называть ненормированными (см. классификацию 10), а тесты нестандартизированные по процедуре — нестандартизированными.

Это важно, поскольку в педагогике можно выделить ряд задач, которые могут быть решены ненормированными тестами. Однако при этом необходимо учитывать несколько существенных обстоятельств. Сфера применения ненормированного инструмента ограничена: с точки зрения субъекта — разработчиком теста, а объекта — той группой учеников, с которыми он в состоянии работать. Круг задач для ненормированных тестов — частные, специальные методические и узкометодические. Например, оценка качества усвоения материала той или иной темы, выявление усвоенных понятий, определений данной темы и т. д. Вторая область применения ненормированных тестов — сравнительные исследования групп учащихся в тех же узких областях. Однако при сравнительных исследованиях необходимость получения оценок и их сравнения приводит к необходимости нормирования в той или иной мере. Требование стандартизации для тестов школьных достижений остается обязательным.

В дальнейшем мы будем говорить преимущественно о стандартизированных тестах. При этом необходимо сделать еще одно замечание: правила и

требования к составлению заданий и их качеству остаются едиными для любых, в том числе и ненормированных и нестандартизированных тестов. Конечно, надежные, метрологически обоснованные результаты дает применение стандартизированных тестов. Результаты, полученные с помощью нестандартизированных тестов, малонадежны и требуют подтверждения.

2. Классификация тестов по назначению:

- общедиagnostические (тесты личности по типу вопросников Кеттелла или Айзенка, тесты общего интеллекта Векслера, Бине-Симона в редакции Термена и Меррилла и др.);
- профессиональной пригодности (батареи общих классификационных тестов — общий армейский классификационный тест, классификационные тесты способностей Фланагана и др.);
- специальных способностей (технических, музыкальных, тесты для пилотов, радиооператоров и т. д.);
- достижений (например, произношения, качества написания письменных сочинений и т.п, т. е. тесты, предназначенные для оценивания результатов, достигнутых учащимися в процессе обучения).

3. По средствам, используемым в процессе тестирования:

- бланковые (тесты «бумага и карандаш», в которых используются тестовые тетради или бланки, в которых испытуемые отмечают или вписывают правильные ответы — вопросники личности, тест механической понятливости Беннета, большинство тестов школьных достижений и др.);
- предметные (в которых необходимо манипулировать материальными объектами, результативность выполнения этих тестов зависит от скорости и правильности выполнения заданий — кубики Косса, тест Стенквиста на сборку конструкций и узлов деталей, тест «Сложение фигур» из набора Векслера и др.);
- аппаратные (тесты с использованием устройств для изучения особенностей внимания, восприятия, памяти и мышления — «детектор лжи», аппарат для определения энцефалограмм мозга и др.);
- практические (появившиеся относительно недавно эти тесты схожи с известными у нас лабораторными работами — по химии, физике, биологии и пр., — однако они снабжены соответствующими инструкциями и имеют тестовое оснащение).

4. По количеству одновременно обследуемых людей:

- индивидуальные (тесты Векслера, тест Люшера и др.);
- групповые (большинство стандартизированных вопросников, оценивающих основные характеристики личности, набор тестов Фланагана, большинство тестов школьных достижений и т. д.).

5. По форме ответа тесты делятся на устные и письменные.

6. По ведущей ориентации:

— тесты скорости, содержащие простые задачи, время решения которых ограничено настолько, что ни один испытуемый не успевает решить все задачи в заданное время (Кольца Ландольта, Бурдона, «шифровка» из набора Векслера);

— тесты мощности или результативности, включающие трудные задачи, время решения которых либо вовсе не ограничено, либо мягко лимитировано. Оценке подлежит успешность и способ решения задачи. Примером такого рода тестовых заданий могут быть задания для письменных итоговых экзаменов за курс школы;

— смешанные тесты, которые объединяют в себе черты двух вышеперечисленных. В таких тестах представлены задачи различного уровня сложности от самых простых до очень сложных. Время испытания в данном случае ограничено, но достаточное для решения предлагаемых задач большинством обследуемых. Оценкой в данном случае служат как скорость выполнения заданий (количество выполненных заданий), так и правильность решения. Эти тесты наиболее часто применяются на практике и именно к ним относится большинство тестов школьных достижений.

7. По степени однородности задач:

— гомогенные, имеющие как правило одну шкалу, которые позволяют оценить одно свойство или качество личности и включают задачи, сходные по характеру, но различающиеся конкретным содержанием;

— гетерогенные (многомерные), имеющие несколько шкал, которые позволяют оценить разнообразные характеристики личности и включают задания, отличающиеся как по характеру, так и по содержанию. К этим заданиям относятся современные тесты школьных достижений.

8. По характеру действий:

— вербальные (связанные с необходимостью произведения умственных действий — словесно-логические тесты, вопросники на проверку знаний, установление закономерностей и пр.);

— невербальные (связанные с практическим манипулированием предметами — карточками, блоками, деталями).

9. По направленности, т. е. по тому, что именно предполагается изучать с помощью данного теста:

— тесты интеллекта, выявляющие особенности последнего;

— личностные тесты (иногда называемые тестами темперамента), с помощью которых изучаются особенности личности испытуемого, включая помимо его интеллекта тесты мотивов (то же, что и тесты динамики), предназначенные для диагностики мотивационно-потребностной сферы личности, позволяющие определить на что направлена активность индивидуума и каким образом осуществляется им саморегуляция поведения; тесты настроений и состояний, направленные на изучение временных состояний, таких, например, как эмоции;

— тесты достижений и другие.

10. По виду нормирования:

— ориентированные на статистические нормы — тесты, основанием для сравнения в которых служат соответствующим образом обоснованные статистически полученные значения выполнения данного теста репрезентативной выборкой испытуемых;

— критериально ориентированные — тесты предназначенные для определения уровня индивидуальных достижений испытуемого относительно некоторого заданного критерия, существующего в реальной практике и заранее известного: уровня знаний, умений, навыков, необходимых для выполнения определенного вида деятельности. Критерий может быть определен на основании экспертной оценки (например, критерий школьной успешности может быть определен путем опроса педагогов, работающих в данном классе или с данным ребенком) либо практической деятельности испытуемых (критерий школьной успешности может быть определен по оценкам за четверть или год);

— прогностические, ориентированные на успешность дальнейшей деятельности;

— ненормированные.

11. По характеру ответов на вопросы:

— открытого типа (со свободными ответами — когда испытуемому необходимо самостоятельно дописать слово, словосочетание, предложение, знак, формулу и т. д.)

— закрытого типа (с предписанными ответами — когда испытуемому необходимо выбрать из предложенных вариантов ответов тот или иной вариант).

Кроме этого, выделяют следующие возможные варианты ответов, используемые в тестах школьных достижений:

открытого типа:

— задания дополнения, (с ограничениями на ответы, с заданными ограничениями) — испытуемый должен сформулировать ответы с учетом предусмотренных в задании ограничений;

— свободного изложения, (свободного конструирования, без заданных ограничений) — испытуемый должен самостоятельно сформулировать ответы, ибо никакие ограничения на них в задании не накладываются.

закрытого типа:

— альтернативных ответов — испытуемый должен ответить да или нет;

— соответствия (на восстановление соответствия) — испытуемому предлагается восстановить соответствие элементов двух списков;

— множественного выбора (ответы с вариантами выбора) — испытуемому необходимо выбрать, как правило, один правильный ответ из приведенного списка возможных ответов;

— исключения лишнего (устранения лишнего элемента, «встретил лишнее — убери»);

— аналогии — испытуемый должен выделить отношение аналогии между парами элементов (слов, свойств, качеств и т. д.);

— последовательности (на завершение последовательности) — испытуемый должен завершить некоторую последовательность элементов.

Особенности различных заданий и требований к ним будут раскрыты в разделе, посвященном их составлению.

Кроме классификации по приведенным выше одиннадцати основаниям, можно выделить, так называемые, «объективные» тесты — тесты, объективность оценки результатов которых обуславливается тем обстоятельством, что в процессе обработки результатов тестирования не предусматривается использование их субъективных толкований тестирующим, к этой группе тестов относятся тесты школьных достижений. Проективные тесты — совокупность методик, разработанных в рамках проективного подхода в психологии, характеризующихся неопределенностью, неоднозначностью используемых в ходе тестирования стимулов (стимулы тестов рассматриваются как экран, на который тестируемый проецирует свойственные ему потребности, внутренние конфликты, способы защиты своего «Я»), что допускает чрезвычайно большое разнообразие ответов и проявление определенной субъективности при их толковании тестирующим; индивидуально ориентированные тесты, специально создаваемые или являющиеся модификацией уже известных, но предназначенные для обследования определенного испытуемого.

Тесты достижений, как и другие тесты, можно классифицировать по большинству оснований приведенных выше. Тесты достижений среди прочих психодиагностических методик составляют одну из самых многочисленных групп. В свою очередь среди них выделяют:

— широкоориентированные, позволяющие оценить эффективность процесса обучения по степени реализации одной из его основных целей, то есть степени освоения учащимися системы знаний, умений и навыков в ходе учебного процесса;

— узкоориентированные, направленные на выявление достижений учащихся в процессе освоения отдельных предметов, отдельных тем и т. д.

Таким образом, мы познакомились с главными основаниями классификации тестов и увидели место тестов школьных достижений среди них. Данная часть работы представляет интерес с точки зрения оценки спецификации предлагаемых на рынке тестов.

Задание 1.

Инструкция: Выпишите на отдельном листе варианты ответов «да» или «нет» для каждого номера вопроса.

Вопрос: вам необходимо проверить уровень усвоения материала учащимися по теме «Треугольники». Аналогичную работу должны провести и ваши коллеги в других школах, чтобы сравнить с вашими результатами. Какие из приведенных классификационных характеристик тестов вас устраивают?

1. стандартизированные да/нет
2. нестандарттизированные да/нет

3. специальных способностей да/нет
4. достижений да/нет
5. аппаратурные да/нет
6. бланковые да/нет
7. индивидуальные да/нет
8. тесты скорости да/нет
9. смешанные тесты да/нет
10. гетерогенные
11. нормированные

Задание 2.

Инструкция: выпишите на отдельный лист букву, соответствующую правильному ответу.

Задача: вам необходимо сравнить подготовленность учащихся по химии в конце первого полугодия с целью выявления трех лучших учеников, которые будут награждены. Вам предложили на выбор 5 наборов тестов. Кроме этого, тесты имеют разную цену. Какой из них целесообразно приобрести?

- А. Тест 1 стоимость 150 тыс. руб. — стандартизированный с ориентацией на норму.
- В. Тест 2 стоимость 120 тыс. руб. — критериально-ориентированный.
- С. Тест 3 стоимость 40 тыс. руб. — нестандартизированный.
- Д. Тест 4 стоимость 80 тыс. руб. — прогностический.
- Е. Тест 5 стоимость 60 тыс. руб. — ненормированный.

Глава 4. Составление тестовых задач

4.1. Типы и формы заданий. Основные требования к тестовым заданиям.

Прежде чем переходить к проблеме составления тестовых задач, попытаемся их классифицировать. Частично мы это сделали во второй главе. В отличие от самих тестов, классификация заданий несколько проще. Задания могут быть разделены по форме вариантов ответов на два больших типа — с закрытыми и открытыми ответами.

Задания с открытыми ответами в свою очередь можно разделить на две группы: задачи свободного изложения (свободного конструирования) и задачи дополнения.

Задания с закрытыми ответами значительно разнообразнее. В этой книге мы рассмотрим следующие их виды: с альтернативным выбором ответов (альтернативные задания, АО), с множественным выбором, задачи соответствия и некоторые другие. К отдельной разновидности относятся задания психологических тестов на выявление структуры интеллекта, которые возможно использовать в тестах достижений: аналогии, задания на исключение лишнего и последовательности — они тоже относятся к заданиям закрытого типа.

Большая часть многообразия существующих заданий может быть сведена к данным видам или их сочетанию. Например, задачи с переструктурированием

данных могут рассматриваться как совокупность задач последовательности и соответствия, задания на нахождение ошибок — частный случай заданий на исключение лишнего и так далее. Отдельный тип заданий представляют собой практические задания. 3:

Выбор типа и вида задачи определяется прежде всего целью- §< ми, в соответствии с которыми проводится тестирование, характером материала, усвоение которого необходимо выявить, возрастными особенностями детей. Немаловажное значение при этом играет финансовое, кадровое и ресурсное обеспечение, а также запас времени, которым располагает разработчик.

В общем виде тестовое задание состоит из четырех частей:

- 1. инструкции;**
- 2. текста задания (вопроса);**
- 3. вариантов ответов;**
- 4. правильного ответа.**

1. Инструкция содержит указания на то, что испытуемый должен сделать, каким образом выполнять задание, где и как делать пометки и записи.

Например:

— отметьте знаком X те варианты ответов, которые Вы считаете правильными...;

— решите предложенную задачу, ответ запишите в рамку, которая находится ниже задания, для промежуточных вычислений используйте место слева от вопроса...;

— в третьем столбце, над строчками впишите цифры, соответствующие понятиям, обозначенным буквами в этой же строке...;

— используя калькулятор, проведите вычисления, ответ запишите в бланке на строке 4... и т. д.

В реальных тестах дается, как правило, одна инструкция для группы однотипных заданий, которая помещается в начале данной группы заданий.

Для проверки того, как испытуемые поняли инструкцию, ее снабжают одним или несколькими примерами, которые разбираются вместе с проводящим тестирование.

2. Текст задания или вопроса представляет собой содержательное наполнение задания.

3. Варианты ответов не используются для заданий открытого типа, где испытуемый самостоятельно формулирует и отражает ответ,

4. Правильный ответ — обязательный атрибут любого тестового задания — без него задание, за исключением пожалуй самых тривиальных, теряет смысл, поскольку не может быть точно оценено с учетом авторского замысла.

Перечисленные четыре составные части тестового задания являются минимально необходимыми для составления тестов. Кроме этого, составителей тестовых заданий просят указать возраст, на который рассчитано это задание, тему (предмет или предметную область), предполагаемую статистическую сложность, уровень, который соответствует данному заданию, или умения, которые оно выясняет, данные составителя, возможные варианты невербальной поддержки, возможно, и некоторые другие сведения, содержание

которых определяется, как правило, целями, с которыми создается данный инструмент. Как правило, для составителей заданий готовятся специальные бланки, в которых требуемая информация формализуется.

**Основное требование к тестовым заданиям:
Тестовое задание должно иметь однозначный
правильный ответ.**

Познакомимся подробнее с заданиями, которые используются в тестах школьных достижений.

4.2. Задачи открытого типа: задачи свободного изложения и задачи дополнения.

Задания этого типа дают наиболее разностороннюю информацию.

К ним относятся задания двух видов:

- свободного изложения или свободного конструирования. Они предполагают свободные ответы испытуемых по сути задания. На ответы не накладываются ограничения. Однако формулировки заданий должны обеспечивать наличие только одного правильного ответа. В этом состоит основная трудность формулирования заданий этого вида;
- дополнения (другое название задачи с ограничением на ответы). В этих заданиях испытуемые также должны самостоятельно давать ответы на вопросы, однако их возможности ограничены. Ограничения обеспечивают объективность оценивания результата выполнения задания, а формулировка задания должна дать один возможный вариант ответа.

Отличительной особенностью заданий дополнения и свободного изложения является то, что в заданиях дополнения требуется вместо каждого многоточия вписать только одно слово, символ, знак и т. д., то есть один пропуск — одно слово, в то время как в задании свободного изложения вместо многоточия можно вписать словосочетание, фразу, предложение или даже несколько предложений.

Инструкция для заданий свободного изложения может быть, например, такой: закончите предложение, впишите вместо многоточия правильный ответ; дополните определение, записывая ответ в бланке, и т. д. Инструкция для заданий дополнения может выглядеть так: вместо многоточия впишите нужное слово, выпишите на бланк слова которые пропущены тексте, вместо многоточия впишите нужный символ и т. п.

Задачи дополнения относятся к наиболее ранним видам задач, применяемых при составлении тестов. При решении таких задач от испытуемого требуется, чтобы он с помощью нескольких слов, одного ключевого слова, символа или рисунка ответил на вопрос, закончил утверждение и т. п.

Пример 1. (17) В данном примере материалом для тестирования служат знания о появлении кибернетики. В 1948 г. Н. Винер опубликовал свой труд «Кибернетика или управление и связь в животном и машине», положив начало новой науки — кибернетики.

Возможно поставить вопрос:

1. Какое событие произошло в 1948г.?

Однако событий в 1948 произошло так много, что точного ответа на этот вопрос не существует, поэтому такой вопрос нельзя использовать в качестве теста, поскольку он не имеет однозначного ответа.

2. Чем известен И. Винер?

История знает несколько Н. Винеров, живших в разное время и известных в различных сферах деятельности. И опять вопрос не имеет однозначного ответа.

Какими же могут быть вопросы к предложенному примеру? Определяющим здесь является то, что мы хотим проверить, и вопросы, соответственно, могут быть разными:

1. Кто является основоположником кибернетики — науки об управлении в животном и машине?

2. В каком году Н. Винер опубликовал свой труд «Кибернетика или управление и связь в животном и. машине»?

3. Как определил Н. Винер новую науку — кибернетику?

4. Какой труд опубликовал И. Винер в 1948 г.

5. Как назвал Н. Винер открытую им науку об управлении и связи, в животном и машине?

Варианты вопросов дополнения могут выглядеть следующим образом:

1—а. Основоположником кибернетики — науки об управлении в животном и машине является ...

2—а. Н. Винер опубликовал свой труд «Кибернетика или управление и связь в животном и машине»- в ... году?

3-а, 4—а. Сформулировать вопрос с дополнением полностью аналогичный для вариантов. «Как сформулировал И, Винер новую науку — кибернетику?» и «Какой труд опубликовал Н. Винер в 1948 г.?» представляется достаточно затруднительным б силу большого количества слов в определении, возможными вариантами, здесь могут

Кибернетика. — это наука о ... и Или Кибернетика — наука об управлении и связи в ... и Естественно, что два этих вопроса применять в одном тесте нельзя, поскольку один раскрывает другой, 5~а. Н. Винер открытую им науку об управлении и связи в животном и машине назвал...

На каждый из поставленных вопросов существует только один однозначный ответ, следовательно эти вопросы могут быть использованы при составлении тестов.

Пример 2. (6б) *Какое слово должно быть следующим в стихотворении, озаглавленном «Необразованный» :*

Жизнь его схожа с Прокрустовым ложем:

День за днем ее он терял

За каждую книгу, что не ... (прочитал)

В данном случае составить вариант ответа- свободного изложения невозможно. Заметим, что без заглавия стихотворения дать определенный вариант ответа становится затруднительно. Для ответа на этот вопрос требуется способность .как подбирать, так и рифмовать слова.

Пример 3. *В данном примере материалом для тестирования являются знания о геометрической зависимости в прямоугольном треугольнике: квадрат гипотенузы равен сумме квадратов катетов.*

Задача свободного изложения: Чему равна сумма квадратов катетов Прямоугольного треугольника?. В силу однозначного определения существует только один ответ, что обеспечит, высокую надежность по тесту.

Задана свободного изложения: Какая связь между гипотенузой и катетом прямоугольного треугольника? Как и в первом случае, возможен однозначный ответ.

В случае, подобном этому, использование задания на дополнение не приведет к улучшению — высокая специфичность правильного ответа заставляет включать в текст задания его определенную часть. Задание дополнения: «Сумма квадратов катетов равна прямоугольного треугольника» или «Если дана сумма квадратов катетов, то гипотенузу прямоугольного треугольника, можно найти как». Необходимые здесь слова не столь очевидны, но их количество может привести к путанице для испытуемых с нетвердыми знаниями

Положительными сторонами хорошо составленных заданий дополнения и свободного изложения являются:

1. краткость и однозначность ответов;
2. необходимость воспроизведения ответа по памяти;
3. отсутствие необходимости искать несколько вариантов ответа;
4. простота формулировки вопросов; - . -(.
5. простота проверки.

Вместе с тем имеются и недостатки, которые необходимо учитывать при конструировании вопросов подобного типа:

1. простое воспроизведение по памяти недостаточно для получения полной информации о знаниях и умениях учащихся;
2. трудности формулировки заданий — невозможность иметь один ответ на многие вопросы;

3. трудность в формулировке вопросов, требующих сопоставления, сравнения, анализа;

4. невозможность использовать технические средства для обработки.

Для того, чтобы предотвратить возможные недочеты при формулировании вопросов дополнения, необходимо учитывать ряд правил, предложенных Б. С. Аванесовым (1);

— использовать не более трех пропусков подряд, лучше 1—2;

— дополнять нужно наиболее важное — то, знание чего нужно проверить;

— дополнения лучше ставить в конце предложения;

— вопрос должен быть сформулирован четко;

— таким же четким и однозначным должен быть ответ;

— следует чаще пользоваться количественными величинами.

Трудность в применении этого вида задач заключается в формализации ответов: их неоднозначность затрудняет стандартизацию. Громоздкость процедуры и большие затраты времени на обследование тоже не в пользу этого метода.

Задания данного типа рассматриваются как; дополнительные методы представления заданий в тестовой форме. Однако какая из форм заданий лучше, определяется специфичностью проверяемой информации. Если она очень специфична, а в педагогической практике это не редкость, то эффективными будут задания свободной изложения, если она определена не столь четко, то лучше использовать задания дополнения.

Клайн отмечает, что задания свободного изложения и задания дополнения больше нравятся испытуемым, чем другие формы заданий, что в свою очередь приводит к снижению утомляемости и дает возможность составлять более длинные тесты, а это приводит в конечном итоге к большей надежности.

Основной трудностью при составлении заданий открытого типа является соблюдение основного требования к тестовым заданиям — наличия однозначного правильного ответа. Существует несколько приемов, которые позволяют формализовать ответ, сделать его однозначным.

Пример 4.

Вопрос: Биосфера это

Ответ: часть геологической оболочки земли, населенная живыми организмами.

Это пример неудачно сформулированной ответа. Задание с таким вариантом ответа невозможно обработать, ведь вероятность точного, дословного совпадения варианта ответа с представленным вариантом чрезвычайно мала.

Ответы «часть геологической оболочки земли» или «часть оболочки земли, населенная живыми организмами» могут считаться частично правильными, что нарушает основное - требование к тестовым заданиям. В данном случае возможно четыре варианта выхода из создавшегося положения: 1.

Попытаться переформулировать вопрос таким образом, чтобы ответ стал однозначным, например:

Вопрос: Часть геологической оболочки земли, населенная живыми организмами называется

2. Выделите ключевые, наиболее существенные, важные слова или словосочетания, в этом случае задание будет выглядеть так: Вопрос: Биосфера это

Ответ: ответ считается правильным, если содержит в себе два слова: геологической и организмами.

3. Возможен и более сложный способ формализации ответов с усложненным шкалированием: Вопрос: Биосфера это

Ответ: ответ на данный вопрос оценивается двумя баллами, если ответ содержит слова геологической и организмами, если ответ содержит одно из этих слов, то одним баллом, во всех остальных случаях задание считается невыполненным.

4. Данное задание может быть преобразовано в другую форму — например, в задание альтернативных ответов, которое мы рассмотрим ниже.

Задание 3.

Преобразуйте предложенное задание в тестовые задания дополнения и свободного изложения

Вопрос: Безотходная технология – это ...

Ответ: Комплексное использование сырья и создание производств с замкнутым циклом без сброса сточных вод и без выброса в атмосферу вредных веществ.

Задание 4.

Найдите недостатки в приведенном задании и предложите пути их устранения.

Инструкция: Вместо многоточий впишите пропущенные слова:

Вопрос: Имя ... — самостоятельная часть речи, которая обозначает.

Ответ: существительное, предмет, кто, что.

При выборе формы задания бывает полезно на одном и том же материале сформулировать задания всех возможных видов, а затем выбрать наилучшие.

Познакомимся теперь с другими формами заданий.

4.3. Задачи закрытого типа.

Наибольшее распространение применительно к тестам школьных достижений находят задания закрытого типа. Остановимся на них более подробно, Задания закрытого типа — тип тестовых заданий, предусматривающих различные варианты ответа на поставленный вопрос. Из ряда предлагаемых выбирается один или несколько правильных ответов. Это задачи с предписанными ответами, что предполагает наличие ряда предварительно разработанных вариантов ответа на заданный вопрос. Существуют несколько видов таких задач:

4.3.1. Альтернативный тип ответов (АО).

К каждой задаче альтернативных ответов дается только два варианта ответов. Испытуемый должен выбрать один из них — да-нет, правильно-неправильно и пр.

Инструкции для заданий альтернативных ответов могут быть такими: отметьте любым знаком вариант ответа да или нет, который вы считаете правильным,...; отмечайте знаком X номера только тех вариантов ответов, на которые вы даете утвердительные ответы...; в бланке ответов отмечайте в соответствующей графе 0 — если вы даете на вопрос ответ нет и 1, если вы даете ответ да.

Альтернативный тип ответов (АО) является самым простым, но не самым распространенным при составлении тестов. Это связано, в основном, со специфичностью того материала, которому в большей степени соответствует эта форм заданий. Задания альтернативных ответов применяются для оценки одного элемента знаний. Использование заданий альтернативных ответов в виде отдельного вопроса, одиночно приводит как правило к тривиальному тестированию и используются достаточно редко. Наиболее эффективно использование заданий этого вида в серии, когда для одного элемента знания задается несколько вопросов. В такой форме задания альтернативных ответов в большей степени подходят для выявления уровня овладения сложными определениями, знания достаточно сложных графиков, диаграмм, схем и т. д. Приведем несколько примеров:

Пример 1. *Сумма квадратов катетов равна квадрату гипотенузы прямоугольного треугольника: да — нет.*

Совершенно непригодное задание ввиду его тривиальности, которая вытекает из точного соответствия формулировке учебника.

Пример 2. *Багдад является столицей Ирана: да — нет. В данном случае задание может использоваться для формулировки вопроса АО, поскольку оно выявляет один элемент знания. Однако лучше будет вариант вопроса задачи дополнения: Столицей Ирана является Эта формулировка сводит на нет возможность угадывания. Забегая вперед, скажем, что можно сформулировать вопрос с множественным выбором: Столицей Ирана является:*

- а) Кабул*
- б) Багдад*
- в) Эр Рияд*
- г) Исфахан,*

Пример 3. *(66) Туберкулез заразное заболевание ~ верно Задание в принципе подходящее для альтернативной формы, поскольку для его выполнения необходим только один фрагмент знаний. Форма со свободным ответом связана с субъективностью суждения и поэтому ненадежна. Однако это задание может быть представлено в форме задания с вариантом выбора: Туберкулез — это*
А) заразное

- б) инфекционное
- в) заразное и инфекционное
- г) незаразное
- д) наследственное заболевание

Такая форма представления так же хороша, как и альтернативная, однако при этом сокращается вероятность случайного угадывания и е выявляются знания о формах распространения туберкулеза.

Приведенные примеры показывают неудачные случаи применения заданий с альтернативными ответами и возможность их замены на другие. Однако существует целый ряд заданий, когда альтернативные ответы применять наиболее целесообразно. Это относится к крупным определениям, сложным процессам, графикам, диаграммам, таблицам, тем элементам знания, которые могут быть структурированы или разбиты на более мелкие части.

Приведем примеры, когда элементы знания адекватны этой форме задания. Особенно это важно для тестов школьных достижений как в части усвоения знаний, так и умения анализировать данные, работать с разной формой представления результатов (графики, диаграммы, таблицы и т. д.).

Пример 4. Вернемся к примеру 4 заданий открытой типа.

Вопрос: В определение биосферы входят следующие утверждения....

- | | |
|--|--------|
| 1. ... это часть геологической оболочки, земли | да/нет |
| 2. ... это часть биологической оболочки земли | да/нет |
| 3. ... это часть гидрологической оболочки, земли | да/нет |
| 4. ... это часть, где присутствует воздух | да/нет |
| 5. ... это часть, населенная организмами | да/нет |
| 6. ... это часть, где возможна жизнь | да/нет |

Ответ да-нет-нет-нет-да-нет.

Пример 5. В этом, вопросе нас интересует процесс гидролиза.

Сформулируем вопрос: «В процессе гидролиза соль взаимодействует с водой. Так ли это?». Эта задача не подходит для формулировки открытого вопроса, поскольку процесс гидролиза со)

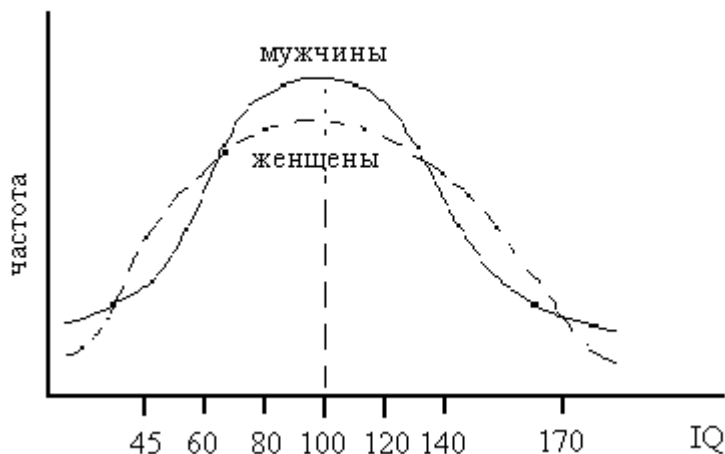
рядом явлений — изменением окраски индикатора, выделением поглощением тепла и т. д. формулировка вопросе

лиза называется взаимодействие соли и воды: да/нет» не подходит б силу своей тривиальности. Однако вопрос может быть значительно улучшен, оставаясь в рамках задач альтернативных ответов:

Гидролизом называется процесс, при котором:

- 1. соль разлагается с помощью электрического тока
 - 3. изменяется окраска индикатора
 - 4. соль кристаллизуется
 - 5. соль взаимодействует с водой да/нет
- Попытка переформулировать эту задачу в вариант с выбором вариантов ответов не приведет к улучшению, поскольку придется указывать по два варианта правильных ответов или формулировать несколько заданий.*

Пример 6. (66).



Распределение уровня интеллекта в большой выборке мужчин и женщин

Исходя из информации представленной на этом являются следующие утверждения истинными

- | | |
|---|--------|
| а) Больше мужчин, чем женщин низкий $IQ < 45$ | да/нет |
| б) Больше мужчин, чем женщин высокий $IQ > 140$ | да/нет |
| в) Распределение уровня интеллекта среди мужчин приближается к бимодальному | да/нет |
| г) Распределение уровня интеллекта среди женщин приближается к нормальному | да/нет |
| д) Среднее значение IQ для мужчин и женщин примерно одинаково | да/нет |
| е) Больше женщин, чем мужчин имеют IQ от 120 до 140 | да/нет |
| ж) Больше мужчин, чем женщин имеют IQ от 50 до 60 | да/нет |

Эти альтернативные задания в наибольшей степени соответствуют задаче выявления того, в какой степени испытуемый понимает данные. Кроме того это задание содержит проверку форм нормального и бимодального распределения, умений работать с графиками, навыки приближенного вычисления. Любая другая форма представления заданий будет гораздо более громоздкой и менее удобной.

В. К. Гайда и Б. П. Захаров (17) считают, что задачи АО целесообразно применять в тех случаях, когда:

1. тест должен быть разработан быстро;
2. проведение теста не должно занимать много времени;
3. тест не предполагает изучения творческих возможностей испытуемого;
4. есть возможность разработать большое число задач для теста;
5. задачи просты по содержанию и легко доступны для понимания испытуемых;
6. случайные решения не играют значительной роли, или могут быть учтены.

Отметим основные трудности при составлении заданий с альтернативными ответами.

Большим недостатком АО является высокая вероятность случайных правильных решений. Это снижает надежность теста. Преодолеть этот недостаток можно с помощью увеличения размера теста или увеличения количества заданий в серии.

Довольно трудно формулировать задания, на которые можно дать однозначный ответ да или нет. Недостатки в формулировках могут привести к тому, что наиболее подготовленные испытуемые увидят противоречие в формулировках и дадут неправильный ответ, или не ответят на задание.

При составлении вопросов альтернативных ответов следует особенно осторожно использовать слова «иногда», «часто», «всегда», «все», «никогда», которые сами по себе содержат неопределенность и могут пониматься субъективно, что может приводить к ошибочным ответам.

Инструкция и обработка результатов просты, но для большинства задач использование альтернативных ответов приводит к утрате специфики задачи, к ее упрощению.

Для получения объективных результатов также следует соблюдать ряд правил, предложенных Ж. Гильбертом:

- вопрос должен содержать одну законченную мысль;
- в вопросе следует избегать слов, дающих возможность учащимся догадаться о правильном ответе (например, слова «иногда», «обычно» подталкивают к ответу);
- вопрос должен быть четко сформулирован, избегая слов «большой», «небольшой», «малый», «много», «мало», «меньше», «больше* и т. д.;
- чаще использовать количественные термины;
- избегать вводных фраз или предложений, мало связанных с основной мыслью;
- не следует прибегать к пространным выражениям, т. к. они приводят к правильному ответу, даже если учащийся его не знает;
- следует избегать модализованных вопросов (типа «не считаете ли вы, что...»), особенно использовать их многократно в одном предложении;
- число ответов «да» и «нет» в тесте должно быть приблизительно равным, что исключает тенденцию отвечать одинаково на все вопросы;
- не следует задавать вопросы с подвохом (с одной стороны, скорее всего в заблуждение будут введены наиболее способные или осведомленные учащиеся, которые знают достаточно для того, чтобы попасться в ловушку, с другой, это противоречит цели — определение уровня знаний и понимания);
- необходимо избегать двусмысленных утверждений.

Задание 5

Найдите 4 существенных недостатка в приведенном задании альтернативных ответов. Предложите способы устранения недостатков.

Инструкция: выберите правильный ответ.

Вопрос: Как известно, в 1703 году Петром I был заложен город Санкт-Петербург. Одобряли или не одобряли бояре, жившие в то время, это событие да/нет

Правильный ответ: нет

Задание 6. /

Найдите 2 существенных недостатка в приведенном задании альтернативных ответов.

Инструкция: Обведи кружочком тот вариант ответа да или нет, который ты считаешь правильным.

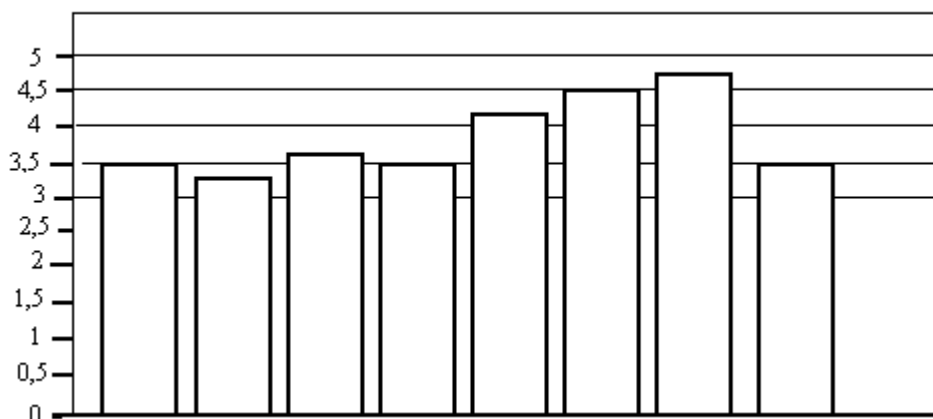
Вопрос: Имя числительное - самостоятельная часть речи, которая обозначает...

- | | |
|--|---------------|
| 1. число | <i>да/нет</i> |
| 2. количество предметов | <i>да/нет</i> |
| 3. порядок предметов при счете | <i>да/нет</i> |
| 4. отвечает на вопросы сколько? Который? | <i>да/нет</i> |

Правильный ответ: да — да — да — да

Задание 7.

На диаграмме представлена зависимость успеваемости учащихся выпускных классов (в средних оценках) в гимназии № 48 от года выпуска. До 1990 г. это была школа № 99, в 1994 году на экзаменах в данном образовательном учреждении присутствовали работники РОНО. Сформулируйте 4—5 вопросов АО на понимание этой диаграммы.



4.3.2. Задачи с множественным выбором.

Это основной вид заданий, применяемый в тестах достижений. Задачи с множественным выбором предполагают наличие вариативности в выборе. Эти задачи состоят из двух частей — формулировки задания и вариантов ответов. Испытуемый должен выбрать один из предложенных вариантов, среди которых

чаще всего только один правильный. Однако задачи формулируются так, чтобы в них было не менее 3 правдоподобных, похожих на правильные ответы. Именно этим обеспечивается независимость результата от случайного выбора. Инструкции для данного вида задач могут быть следующими; обведите кружком в бланке ответов букву, соответствующую варианту правильного ответа..., отметьте любым знаком номер варианта, ответа, который Вы считаете правильным..., из пяти предложенных вариантов ответа заберите один и подчеркните его..., выберите правильные ответы из предложенных вариантов и подчеркните их, в каждом вопросе может быть один или два правильных ответа....

Поиск правдоподобных однородных ответов составляет основную трудность для разработчиков заданий множественного выбора. Решить эту проблему возможно только на основе анализа результатов решения задачи.

Приведем несколько примеров, связанных с формулировкой заданий с выбором варианта ответов:

Пример 1. *К какой группе веществ относится серная кислота?*

- б) сильный электролит*
- в) слабый электролит*
- г) электролит средней силы*
- д) восстановитель*

В данном примере можно говорить о двух неудачах в подборе вариантов ответов. Во-первых, первые четыре варианта представляют собой разновидности одной группы веществ — электролитов, а последний определяет вещества по другому основанию, что по сути является подсказкой. Во-вторых, расположение в задании характеристик силы электролита хаотично, что приводит к потере времени на чтение задания. Более удачным был бы вариант ответов:

- а) не электролит*
- б) слабый электролит*
- в) электролит средней силы*
- г) сильный электролит*

Пример 2. *В. А. Сологуб вспоминал: «Часто приходит мне на ум Нева, дремлющая в огненных отливах солнечного заката. Еще чаще вижу ее сизо-серую как сталь... И с ужасом вспоминаю я, как однажды река перестала быть рекою и обратилась в море бушующее, разъяренное, смывающее Петербург с лица земли. Когда это произошло, в каком году?*

- а) 1715 г.*
- б) 7 ноября 1824 г.*
- в) 1703 г.*

В таком задании несколько недостатков. Задание содержит лишние детали, представлено в слишком длинном виде, при этом из него совершенно очевидно, что речь идет о наводнении. Второе — неудачна формулировка вопроса к

заданию, которая содержит в себе по сути два вопроса. Задание содержит в себе две подсказки: первая заключается в наличии среди вариантов ответа 1703 года — года основания Санкт-Петербурга. Вторая подсказка, более существенная, заключена в подобранных вариантах ответа, когда один из них приводится с датой, а остальные нет. Еще один недостаток этих вариантов ответа — значительный разброс в датах. Попытки переформулировать задание не могут иметь успеха, поскольку не понятен замысел автора. Можно рассматривать данное задание как выясняющее знание времени самую страшного наводнения, однако само по себе такое знание не представляется существенным.

Примеры 3—5 взяты из книги П. Клайна (66).

Пример 3. *У речной излучины на поляне появилась горилла-самка и внезапно издала громкий вопль. & этом предложении, дополнением является:*

- а) горилла*
- б) полоне*
- в) вопль*
- г) самка*

Это простое, понятное задание, в котором вся информация содержится в основной части, возможен только один правильный ответ, все варианты ответов являются существительными, то есть тем видом слов, которые могут быть дополнениями, что не позволяет испытуемым, знающим, что дополнениями не могут быть глаголы или наречия, сократить количество вариантов выбора.

Пример 4. *Внезапно громкий вопль был издан гориллой-самкой, которая появилась у речной излучины на поляне. В этом предложении подлежащим является:*

- А) горилла*
- в) вопль*
- б) излучины.*
- г) поляне '*
- д) самкой*

Этот пример аналогично прост, и вся информация, содержится в основной части.. Он умышленно связан с заданием 3, поскольку те испытуемые, знания которых о дополнении и подлежащем лишь приблизительны, и которые полагают, что подлежащее — это тот, кто «делает», не смогут выполнить данное задание. Аналогично, тот факт, что в предложении с почти тем же самым смыслом теперь подлежащее и дополнение поменялись местами, будет составлять трудность для тех, чьи познания в грамматике неполны. Заметим опять, что все варианты выбора могут быть подлежащим.

Пример 5. Я переживаю, когда думаю о тебе, переходящим через дорогу самостоятельно в это время дня. В этом предложении есть ошибка.

Неправильным, словом является:

- а) я
- б) переходящим
- в) дня
- г) самостоятельно

Это пример плохого задания. Понятно, что вся информация содержится в основной части задания и существует однозначный правильный ответ. Более того, содержание задания не является тривиальным. Однако в предложении есть только одна ошибка, так что испытуемые могут прийти к правильному ответу, исключив варианты ответов, которые явно не являются неправильными. Это задание должно быть каким-либо образом перефразировано, например: «Если в этом предложении есть грамматическая ошибка, то неправильным словом является: а) я, б) переходящим, в) дня, г) тебе, д) ни оно из этих слов, если в предложении ошибки нет». Такая формулировка устранил проблему, но она несколько громоздка. Вероятно, наилучшим решением было бы использовать другую форму задания, такую как: «Какое из следующих предложений грамматически правильно: а) Я переживаю, когда думаю о тебе, переходящим

г) Тесты школьных достижений: конструирование, проведение, использование через дорогу самостоятельно в это время дня. б) Я переживаю, когда думаю о тебе, переходящем через дорогу самостоятельно в это

Этот пример хорошо иллюстрирует ограничения формы заданий со многими вариантами выбора. При определенных условиях, когда задание слишком простым. Следовательно, более предпочтительно использование разных форм заданий. В этом случае трудно сформулировать общие правила. Здесь бея надежда на сообразительность разработчиков тестов. В данном примере показан ход рассуждений, который следует использовать разработчикам тестов при подготовке заданий.

Пример 6. Катер плывет по реке, скорость течения равна a , скорость катера в стоячей воде равна b . Какая из следующих формул выражает время, которое затрачивает катер на то, чтобы спуститься вниз по течению на 30 км, а потом вернуться обратно? (Время стоянки не учитывать)

а) $\frac{2b-a}{60}$

в) $\frac{30}{a} + \frac{30}{b}$

б) $\frac{30}{a+b} + \frac{30}{b-a}$

г) $\frac{30}{b-a} - \frac{30}{b+a}$

Это сложное задание, в котором необходимо знание формул и умение совершать действия с алгебраическими выражениями в уме. Однако той

сложности состоит смысл данной задачи. Сделать эр, не совершив определенных действий, практически, невозможно.

Пример 7. *Форма и цвет тела у богомола, палочника, окраска красного клопа обыкновенного, упрощенное тело ската — это примеры;*

- а) ароморфоза*
- б) идиоадаптации*
- в) дегенерации*
- г) атавизма*

В этом задании форма адекватна проверяемому материалу. В нем представлены четыре понятия, не зная которые невозможно найти правильный ответ. Используя, задания АО, нам пришлось бы формулировать четыре отдельных задания, чтобы провести ту же работу. Кроме этого, задание исключает возможность угадывания правильного ответа.

Как уже отмечалось, достаточную сложность представляет подбор вариантов правдоподобных неправильных ответов. К сожалению, как показывает опыт применения тестирования, ошибки связанные с этим достаточно широко распространены. Рассмотрим несколько примеров:

Пример 8. *Мы будем изучать латинский язык потому, что... А ... на нем говорят во многих странах мира*

В ... хотим лучше понимать родной язык, так как в нем много слов заимствованных из латыни

С ... хотим лучше понимать историю и культуру древнего мира В данном случае вообще отсутствует правильный ответ. Налицо путаница между тестовыми заданиями и вопросами на выявление мотивации. При этом он плох и как вопрос на выявление мотивации, поскольку не перекрывает возможного минимального поля вариантов ответов и не содержит открытой части.

Пример 9. *Империя это ...*

- А. Большое государство*
- В. Государство, управляемое императором*
- С. Государство, состоящее из метрополии и колоний .*
- Д. Сильное государство*
- Е. Недемократическое государство.*

Правильный ответ С.

С точки зрения материала, изучаемого в курсе истории, вариант ответа С представляется правильным, однако среди вариантов неправильных ответов существует ответ В, который, строго говоря (согласно словарным определениям), тоже является верным. Засчитывая только ответ С составитель получает искаженные данные о знаниях детей по вопросу.

Еще один возможный источник ошибок связан с желанием автора упростить понятия, перевести их на более понятный «детский» язык. Связано это с желанием разнообразить задания, сделать их менее утомительными. Однако такие задания могут применяться только после жесткой проверки на валидность. Приведем пример.

Пример 10. *Если твой сосед по парте пытается занять своими вещами не только свою половину парты, но и залезть на твою, то это называется... Предполагается, что правильным ответом будет слово «экспансия».*

Преимущества заданий с несколькими вариантами выбора:

- задания могут быть надежны, поскольку отсутствуют факторы, связанные с субъективными оценками, которые снижают надежность;
- задания этого типа легко обрабатываются и тестирование быстро проводится, особенно при проведении бланковым методом;
- простой алгоритм заполнения снижает количество случайных ошибок и описок;
- эти задания позволяют охватить большие области исследования, что для тестов достижений особенно важно;
- возможна машинная обработка ответов;
- низка вероятность угадывания правильных ответов;
- возможно получение точной оценки содержательности теста, что особенно важно для определения соответствия теста целям исследования.

Для правильного составления задач с множественным выбором важно соблюдать следующие требования:

- лучше использовать один вариант правильного ответа и, если инструкция требует выбрать правильный ответ, то таковым должен быть только один, в противном случае в инструкции необходимо указать, что правильных ответов несколько;
- все варианты ответов должны быть грамматически согласованы с основной частью задания;
- вопросы должны содержать только одну мысль или утверждение;
- неправильные ответы должны быть разумны, умело подобраны, не должно быть явных неточностей;
- как можно реже использовать отрицание в основной части;
- с одной стороны, использование отрицания приводит к противоречиям при чтении задания, с другой, отрицательные знания не так видны как позитивные (знание того, что дельфин не рыба, не рептилия, не птица не является необходимым для определения что же это такое), если все таки в вопросе содержится отрицание, его следует подчеркнуть, чтобы привлечь внимание;
- вопрос не должен содержать лишних деталей;
- ответ на поставленный вопрос не должен зависеть от предыдущих ответов;
- место правильного ответа должно быть определено таким образом, чтобы оно не повторялось от вопроса к вопросу, без закономерностей, в случайном порядке;

— правильные и неправильные ответы должны быть однозначны по содержанию, структуре и общему количеству слов;

— если ставится вопрос количественного характера, то ответы к нему должны располагаться упорядоченно от меньшего к большему или наоборот;

— лучше не использовать варианты ответов «ни один из перечисленных» и «все перечисленные». Применение первого целесообразно, когда существует недвусмысленный правильный ответ. Второй приводит к допустимости подбора вариантов ответов с низкой дискриминативностью, поскольку разработчик знает, что все ответы правильные.

И еще одно правило, позволяющее сделать процесс тестирования для учащегося достаточно быстрым и доступным: лучше использовать длинный вопрос и короткий ответ. В противоположной ситуации на прочтение ответов уходит больше времени и больше сил тратится на анализ высказываний, что противоречит поставленной в данном случае цели — выявлению усвоенных учащимся знаний.

14.3.3. Задачи соответствия (Задания на восстановление соответствия).

Задачи соответствия (восстановления соответствия) — задания, где необходимо найти или приравнять части, элементы, ^ понятия конструкциям, фигурам, утверждениям; восстановить соответствие между элементами двух списков. Данный вид заданий имеет достаточно много модификаций, от которых зависят инструкции. Наиболее распространенной формой ответа на данный вопрос, которая реально применяется педагогами и отражается в инструкции, — это вариант с использованием стрелочек: нарисуйте стрелочки от элементов первого списка ко второму..., соедините стрелками соответствующие понятия и т. д. Сам по себе способ с использованием стрелочек вполне правомерен, однако он имеет два существенных недостатка: сложность проверки, особенно когда необходимо проверить большое количество работ и, второй, привыкнув к способу выполнения этих заданий стрелочками, ученики, встретив в дальнейшем классическую форму задания, не будут к ней готовы, воспримут ее как неизвестную, что может снизить их результаты.

Наиболее классической формой представления заданий и соответствующей инструкции для выполнения заданий может быть следующая: на месте для ответа впишите цифры ряда 1, соответствующие ряду 2.

Ряд 1	Ряд 2	Место для ответов
1.	A	_____
2.	B	_____
3.	C	_____
4.	D	_____
5.		
6.		

Другие виды инструкций к этим заданиям мы рассмотрим в примерах.

Если мы хотим выявить знания испытуемого об авторах книг, один список должен содержать фамилии авторов, а другой — заглавия книг, и между ними должно быть установлено правильное соответствие.

Эта форма заданий может применяться в разнообразных предметных областях. Практически в каждом школьном предмете существует широкая возможность их использования. В области биологии, например, может быть использован список растений и список соответствующих соцветий, в русском языке — слова и соответствующие им части речи или орфограммы, в истории — даты и события и т. д.

Как и ранее, будут приведены примеры заданий шесте с комментариями, чтобы выявить сильные и слабые стороны этих заданий и проиллюстрировать простые указания по созданию эффективных заданий,

Задачи соответствия требуют подбора подходящего ответа. Обычно задание соответствия состоит из трех столбцов: в первом, под заглавными буквами (или цифрами) — вопросы, утверждения, факты, понятия и т. д., во втором — пронумерованный (или, в случае если первый список был пронумерован, то под буквами) список утверждений или слов, которые надо поставить в соответствие, и, наконец, третья графа — вариант ответа. Для каждого пронумерованного слова или утверждения следует отобрать один признак под заглавной буквой, наиболее тесно связанный с ним.

Пример 1. Кто написал?

<i>произведение</i>	<i>автор</i>	<i>цифра</i>
1. Обломов	<i>а). Толстой</i>	
2. Капитанская дочка	<i>б). Островский</i>	
<i>Севастопольские рассказы</i>	<i>в) Пушкин</i>	
<i>Лес</i>	<i>г). Чехов</i>	
<i>Крыжовник</i>	<i>д). Гончаров</i>	

..... __ этого примера, задания на восстановление соответствия лучшим образом подходят для выявления знания фактической информации. Можно было бы и просто спросить: «Кто написал „Севастопольские рассказы“?» Однако данное задание на восстановление соответствия — это простой метод задавать вопросы, который позволяет легко подсчитывать оценки. Это не очень удачный пример по двум причинам. Во-первых, неудачна формулировка вопроса: она слишком кратка, вызывает вопросы и требует пояснений. Во-вторых, каждому произведению поставлен в соответствие

только один автор, что увеличивает вероятность догадки. Приведем другой пример, построенный на аналогичном материале. ...

Пример 2. (66).

Укажите буквой, какому автору принадлежат книги, указанные в следующем списке (оставьте незаполненной графу напротив книги, автор которой не указан в левом списке):

авторы	произведения	буква
А) Диккенс	Ярмарка тщеславия	
Б) Скотт	Уэверли	
В) Смоллет	История двух городов	
Г) Теккерей	Утерянный рай	
	Гамфлри Клинкер	
	Памела	
	Лунный камень	
	Черный карлик	

Это задание лучше предыдущего: список, из которого выбираются ответы, длиннее, чем первый список, в противном случае, как мы уже говорили, случайное угадывание станет более вероятным. Так, если в первом примере испытуемый знает четыре из пяти ответов в одинаковых по длине списках, но пятый будет неизбежно правильным. Кроме предъявления списков неравной длины можно указывать в задании, что некоторым элементам нет никаких соответствий, как в данном примере. Эти меры с очевидностью уменьшают вероятность случайного угадывания правильного ответа.

Можно было бы сконструировать и задание с несколькими вариантами выбора: Диккенс написал: а) «Ярмарку тщеславия», б) Уэверли», и т. д. Однако в нашем примере потребовалось бы отдельное задание на каждую книгу, чтобы протестировать тот же самый объем знаний. Таким, образом, для выявления подробной информации такого рода задания на восстановление соответствий являются более предпочтительными, чем задания с вариантами- выбора: они гораздо более компактны.

Можно было бы, конечно, использовать альтернативные задания: Диккенс написал: а) «Ярмарку тщеславия», Д/Н; б) «Уэверли», Д/Н; и т. д. Очевидно, что для данного материала альтернативное задание не столь эффективно, сколь задание на восстановление соответствия. Во-первых, потребуется четыре задания: по одному на каждого писателя. Во-вторых, нельзя использовать одни и те же примеры в каждом задании, потому что в противном случае ответы на первое задание будут влиять на другие. Следовательно, для данного типа фактической информации лучшей формой задания является задание на восстановление соответствия, за ним следует альтернативно! задание, а задание с несколькими вариантами выбора является наименее эффективным.

Однако не всегда задание на восстановление соответствий оказывается более совершенным, чем задание с альтернативными ответами. Давайте еще раз рассмотрим пример 5, который мы рассматривали при описании альтернативных заданий

Пример 3. *(пример 5 альтернативного задания (стр. 92).*

Гидролизом называется процесс, при котором:

- | | |
|---|--------|
| 1. соль разлагается с помощью электрического тока | да/нет |
| 2. соль окисляется | да/нет |
| 3. изменяется окраска индикатора | да/нет |
| 4. соль кристаллизуется | да/нет |
| 5. соль взаимодействует с водой | да/нет |

В данном примере задание на восстановление соответствий вообще не может быть сформулировано. Этот пример позволяет понять относительные достоинства альтернативных заданий и заданий на восстановление соответствий.

Заметим, что для тестов на проверку понимания данных нет простого очевидного способа сформулировать задание в виде восстановления соответствий. Могут быть сконструированы задания с вариантами выбора, однако они будут слишком громоздкими, а выбор действенных вариантов ответов будет трудным. Помимо этого, такие задания неизбежно будут иметь тенденцию к взаимозависимости, что уменьшит их эффективность. Метод восстановления соответствий может оказаться более предпочтительной чем метод альтернативных ответов для подробного тестирования некоторой учебной информации, при условии, что содержание материала позволит составить соответствующие списки.

Пример 4. *Перед вами фамилии людей, известных в различных областях человеческих знаний. Вам необходимо поставить нужные буквы в ваших бланках, соответствующие основной сфере деятельности того или иного лица.*

Э — экономика

П — политика
Н — наука
Л — литература
И — живопись, музыка

<i>Имя, фамилия</i>	<i>буква</i>
1. Лев Толстой	
2. Борис Ельцин	
3. Луи Армстронг	
4. Томас Эдисон	
5. Василий Леонтьев	
6. Петр Столыпин	
7. и т.д.	

С технической точки зрения задание не очень удачное, поскольку испытуемому приходится все время справляться о правильной букве в списке, однако оно может быть улучшена путем выведения подсказки на бланк. В этом случае задание будет сформулировано следующим образом:

Перед вами фамилии людей, известных 5 различных областях человеческих знаний. Вам необходимо обвести кружком нужные буквы на бланке, соответствующие основной сфере деятельности той или иной лица,

Э — экономика
П — политика
Н — наука
Л — литература
И — живопись, музыка

<i>Имя, фамилия</i>	<i>буква</i>
1. Лев Толстой	Э П Н Л И
2. Борис Ельцин	Э П Н Л И
3. Луи Армстронг	Э П Н Л И
4. Томас Эдисон	Э П Н Л И
5. Николай Бородин	Э П Н Л И
6. Василий Леонтьев	Э П Н Л И

Здесь инструкция требует отметить знаком (кружком, подчеркнуть, крестиком и пр.) соответствующую букву. Не следует применять в инструкции слово «зачеркнуть», поскольку оно носит негативный оттенок и может привести к путанице. Кроме этого не следует включать в список

фамилию Бородина, который известен и как ученый-химик и как композитор, что может вызвать вопросы или ошибки, у наиболее подготовленных учеников.

Следует еще раз напомнить об опасности, возникающей при использовании альтернативных заданий и заданий на восстановление соответствия в тестах достижений — опасности тривиального тестирования.

Хотя и важно знать, что овес имеет соцветие «метелка», а «воробьиные» — это отряд птиц, однако ответы на эти вопросы не предполагают понимания. Знание того, что Пушкин написал «Капитанскую дочку», представляется не столь важным по сравнению со способностью выполнить анализ этого произведения или выразить отношение к проблемам, в нем поднимаемым. Главные преимущества задач такого вида: возможность быстрой оценки знаний, умений и навыков в конкретной области знаний и компактность размещения задач в тесте.

Для того, чтобы задачи соответствия позволяли получить результат не зависящий от внешних причин, при их конструировании необходимо учитывать требования, вытекающие из особенностей восприятия:

- число входных данных одного списка не должно превышать 10; если их больше, лучше составить еще одну или несколько задач;
- все списки и вопросы должны помещаться на одной странице;
- если длина списков не совпадает, то об этом необходимо упомянуть в инструкциях и ключе;
- все ответы по конструкции должны быть по возможности однородны.

Задание 8.

Найдите 3 существенных недостатка в приведенном задании

Вопрос: Как известно сахар является одним из важнейших продуктов питания.

Какие из перечисленных стран являются самыми крупными экспортёрами сахара

- А) США Б) Китай В) Бразилия Г) Куба

Правильный ответ: В) Бразилия и Г) Куба.

Задание 9.

Найдите 2 существенных недостатка в приведенном задании.

В каком году основаны города Санкт-Петербург и Рязань

- А) 1095 и 1703 Б) 1078 и 1713 В) 1705 и 1805 Г) 1090 и 1701

Правильный ответ А.

4.4. Использование заданий психологических тестов на выявление структуры интеллекта для тестов достижений.

Согласно анализу психологов (Клайн (66), Кеттел) выявляются два фактора интеллекта: (здесь и далее использована терминология способностей, упрощенная П. Клайном) g_f — текучие способности — g_c — кристаллизованные способности. Текучие способности весьма сходны с показателем интеллекта Спирмена — это способность отыскивать связи и выявлять отношения, зависимости, на материале, который, как предполагается, минимизирует индивидуальные различия в образовании и культурном уровне. Кристаллизованные способности — это результат обучения, конденсат опыта, и они измеряются в зависимости от культуры, к которой принадлежит человек, уровня образования, активности, интересов личности. Поскольку предмет нашего изучения — тесты школьных достижений, нас будут интересовать кристаллизованные способности в силу того, что они определяют результат обучения, В этой связи хочется еще раз подчеркнуть, что форма и содержание задания зависят от целей тестирования и ведут к разным результатам.

4.4.1. Аналогии.

В общем, виде задание аналогии выглядит следующим образом: А так относится к В, как С относится к ... По форме это может быть задание как открытого типа — задание дополнения, так и закрытого. В случае применения открытой формы испытуемому предлагается самостоятельно вписать в предназначенное для этого место слово-ответ. В случае, если применяется закрытая форма вопроса, то испытуемому предлагается на выбор несколько вариантов ответов, так же как и в заданиях множественного выбора. Задания аналогии имеют сокращенную форму записи, которая применяется когда заданий несколько и они представлены серией и, следовательно, нет необходимости повторять инструкции для каждого задания: А:В=С:?. Инструкции для этих заданий зависят от того, какие ответы (закрытые или открытые) предусматриваются.

Аналогия — тип заданий, обычно присутствующий во всех тестах интеллекта. Он особенно может быть полезен по двум причинам. Во-первых, легко изменять уровень сложности таких заданий (это касается самих отношений аналогии, а не сложности соотносимых объектов), что делает задания подходящими для всех возрастных групп и уровней развития. Во-вторых, практически нет ограничений, относящихся к материалу заданий, использующих принцип аналогии. Это означает, что аналогия подходит для тестов, измеряющих как g_c , так и g_f .

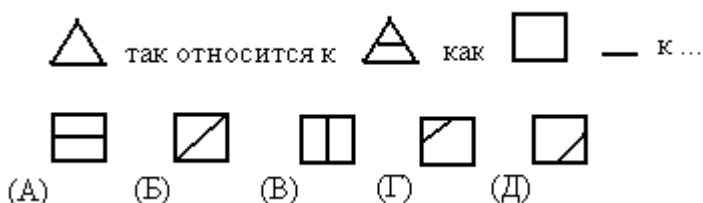
Одно из преимуществ аналогий как формы заданий — это то, что в них могут быть заложены разнообразнейшие виды отношений.

Для психологов аналогии привлекательны тем, что они могут быть оформлены в невербальном виде, что идеально при тестировании текучих способностей. В педагогике существует ряд задач, когда необходимо выявление текучих способностей, например, при отборе детей в классы разного профиля, внутри классной дифференциации. Однако этот класс педагогических задач не является предметом настоящей работы.

Психологи потратили довольно много усилий на решение задачи определения текущих способностей. Как; утверждает Кеттел, текущие способности лучше всего тестируются либо заданиями, для выполнения которых все представители некоторой культуры были обучены, либо заданиями, с которыми все испытуемые, независимо от образования и воспитания, в равной степени незнакомы.

П. Клайн предполагает, что задания на аналогии, приведенные ниже, — это задания, почти полностью связанные с фактором g_f . Наилучшими заданиями являются те, в которых используются буквы алфавита и цифры (при условии, что для логических рассуждений в случае использования цифр не будут требоваться математические способности). В последнем случае существует большое разнообразие абстрактных фигур и образов.

Пример 1.



Это типичное геометрическое, невербальное задание для измерения фактора g_f . Для выполнения этой простую задания не требуется почти никаких специальных знаний, при условии, что испытуемый с двумерными изображениями на бумаге.

Пример 2. Б так относится к Д, как К - к...

- а) Н,
- б) Р,
- в) Т,
- г) М,
- д) С.

Это простой пример, основанный на алфавитной последовательности.

Для нас более важно другое содержание заданий. Покажем, как в форму задания на аналогии можно включать разнообразные знания для выявления которых может требоваться дополнительная информированность испытуемого. Практически нет никаких ограничений на то, насколько сложными могут быть эти задания как по сложности используемых отношений, так и по информации.

Пример 3.

Быстрый относится к быстрой, как медленный — к ...

- а) медлительности, б) скорости, в) торопливости, г) неспешности, д) вялости.

Здесь отношение является весьма абстрактным — связь существительного с прилагательным. Очевидно, что выполнение задания зависит от наличия соответствующего словарного запаса. Такое задание является весьма существенным, поскольку вербальные логические рассуждения обычно являются лучшими индикаторами кристаллизованных способностей. Для обследования испытуемых, находившихся в равных возможностях для формирования богатую словарного запаса, подобные задания, хотя и не обязательно в аналогичном виде, наиболее значимы.

Пример 4.

Глагол так относится к слову спрягать, как существительное к слову...

а) изменять; б) образовывать; в) употреблять; г) склонять; д) писать.

Это простое задание, нагруженное фактором де. Для выполнения этого задания достаточно знаний по русскому языку.

Пример 5. Эпителий : ткань = аорта

а) сердце;

б) внутренний орган;

в) артерия;

г) вена;

д) кровь.

В данном примере приведена сокращенная форма представления заданий на аналогию, которая подходит для батареи заданий на аналогии и должна сопровождаться соответствующей инструкцией. Это задание конечно гораздо более сложное нежели предыдущее и лишь биологической информации недостаточно, чтобы выполнить это задание, поскольку все варианты ответов принадлежат к этому виду.

Пример 6. Птицы : воробьиные = млекопитающие :

а) земноводные; б) лошадь; в) теленок; г) насекомые; д) грызуны.

Данный пример мы привели, чтобы показать возможность формулировки задания аналогичную приведенному в разделе заданий на соответствие.

4.4.2. Задания на исключение лишнего (отношения и связей; «встретил лишнее — убери»).

Широко используемыми и результативными формами заданий являются задания на исключение лишнего. В таких заданиях испытуемому предъявляется список объектов: слов, фигур, чисел или чего-нибудь еще — всего, что только может придумать разработчик тестов; а испытуемый должен найти общие закономерности отношения между элементами списка, на основании которых и делать заключения об их подобии или различии. Очевидно, что при этом

необходимо выявление отношений и связей. Подобно аналогиям, трудность заданий можно легко регулировать как по уровню сложности закономерностей между элементами списка, так и по наличию специальных знаний, необходимых для выявления этих отношений.

Инструкция для этих заданий предполагает выделение каким-то образом лишнего элемента в списке, либо запись или отметку в бланке индекса, соответствующего лишнему элементу.

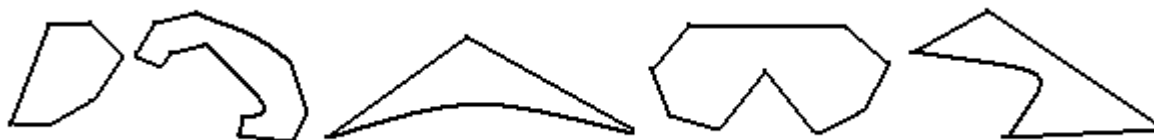
Пример 1.

24, 63, 10, 48, 35.

Это задание нагружено как g_c , так и g_f . Все числа, кроме одного, — это квадраты некоторых чисел минус один. При условии, что испытуемые знакомы с квадратными корнями и возведением в квадрат, выполнение задания зависит от того, будет ли испытуемым выявлена закономерность.

Пример 2. *Сокол, орел, стриж, петух, индюк. Это гораздо более коварное задание. Пример этого задания приведен здесь, чтобы проиллюстрировать логику рассуждения разработчика. Здесь лишним является слово «стриж», поскольку все другие элементы списка — это названия птиц, которые имеют еще одно значение и используются для характеристики людей. Обратите внимание, что если бы вместо слова индюк было использовано слово «гусь», то оно стало бы вторым правильным ответом, поскольку это была бы единственная водоплавающая птица. Слова «сова» и «курица» также не годятся, так как первое было бы названием единственной ночной птицы, а второе — единственным существительным женского рода. А заменив слово «индюк» на «жаворонок», получаем второй правильный ответ «петух», так как это будет единственная домашняя птица. При использовании этих слов задание будет простым — более зависимым от знаний. Хотя для выполнения задания со словом «индюк» и необходим хороший словарный запас, это задание становится более серьезным тестом фактора g_c .*

Пример 3.



Это простое задание, нагруженное фактором g_f . Четыре элемента имеют по крайней мере одну прямую линию. Будучи простым, оно требует от испытуемых нахождения общей закономерности для большинства элементов. Знания здесь не нужны. Это задание настолько очевидно, что оно годится только для детей начальной школы, и даже по отношению к ним оно может оказаться не дискриминативным.

Пример 4.

а) скорость; б) колебание; в) сила; г) вес; д) плотность.

Это задание по физике довольно простое, и для его выполнения необходимо знание основных понятий.

Пример 5.

а) аорта; б) вена; в) сердце; г) артерия; д) капилляр. Это пример задания на исключение лишнего, составленного на материале, аналогичном, заданию, данному в примере 5 аналогий (стр. 115).

Эти пять примеров демонстрируют, что исключение лишнего, как и аналогии, — это форма заданий, для выполнения которых требуются логические рассуждения; они необходимы в гестах интеллекта, могут применяться в тестах достижений и могут быть адаптированы к самому разнообразному материалу.

4.4.3. Последовательности.

Эти задания предполагают продолжение испытываемым какого-то ряда или нахождение предыдущего, среднего, последующего элемента или их сочетание. Инструкции для этих заданий сходны с инструкциями для заданий на исключение лишнего.

Пример 1.



Невербальный материал, как эти треугольники, полезен при конструировании последовательностей для тестирования фактора d_2 . Это простое задание. Однако можно сконструировать и значительно более трудные задания такого типа.

Пример 2.

12, 15, 17, 20, 22, ...

Это относительно простое задание, в котором члены последовательности увеличиваются на 3 и на 2 поочередно. Числа особенно > использовать при тестировании отношений, так как ими легко манипулировать, а для выявления отношений в данном примере необходимы лишь ограниченные математические познания. Трудность такого задания может быть увеличена за счет усложнения математического материала и тогда такого рода задания вполне можно использовать для проверки, не только d_6 , но и знаний по математике.

Пример 3.

Микроскопический, мельчайший, крошечный, ..., большой, крупный. Это пример вербальной последовательности, представляющий собой простой тест для измерения g_c . Понятно, что он также затрагивает вербальные способности. Здесь необходим большой выбор вариантов ответов; огромный, маленький, тяжелый, гигантский, чудовищный.

Пример 4.

..., пропан, пропанол, бутан, бутанол,

Пример более сложной последовательности, где необходимо найти как предыдущий, так и последующий элемент ряда, с достаточно простым наполнением знаниями по химии. Не обладая этими знаниями, справиться с заданием невозможно.

Форма записи этих заданий, приведенная в примерах, вполне удовлетворительна. Во избежание монотонности (с точки зрения испытуемого) могут быть использованы разные ее варианты. Задания последовательностей имеют несколько модификаций:

- 1) от испытуемых может требоваться нахождение некоторого элемента, не крайнего по расположению;
- 2) можно попросить испытуемых дополнить не следующий элемент, а через один или еще более дальний;
- 3) испытуемые могут переупорядочить элементы в соответствии с выявленной последовательностью.

По мнению П. Клайна, используя эти три типа заданий — аналогии, исключение лишнего и последовательности — можно создать большое количество тестов для измерения текучих и кристаллизованных способностей на разнообразном материале и разного уровня сложности.

Принципиальным для тестов достижений, если они не носят двуразмерный характер, становится составление заданий трех последних видов с достаточно простыми аналогиями и последовательностями, но при этом различным уровнем сложности включенного в них содержательного материала.

Мы рассмотрели три больших вида задач. Очевидно, нет необходимости говорить, что это не единственные типы заданий, которые могут и должны использоваться. Тем не менее эти формы детально изучены и поэтому наиболее пригодны для формулировки эффективных заданий. Рассмотрим кратко некоторые другие виды задач, отметив при этом отсутствие у них выделения единого типологического основания.

4.5. Другие виды задач.

Мы рассмотрели основные виды задач, которые применяются в тестах школьных достижений, Это так сказать *базовые* виды. Как мы уже отмечали, к этим основным видам может быть приведено большинство существующих задач. Для нас существенно то, что требования и особенности составления заданий перекрываются тем набором, который был нами рассмотрен. Приведем с краткими пояснениями другие виды заданий, которые могут встретиться в литературе, Подчеркнем еще раз, что все они могут быть сведены к уже рассмотренным нами.

Наибольшую сложность при создании тестов представляет выявление умения использовать полученные знания, уровня овладения действиями и операциями, характерными для данного предмета. Но изучать это необходимо, поскольку

именно то, как применяет учащийся знания, показывает реальный уровень овладения ими. Наиболее эффективными здесь могут быть следующие:

1. Задачи с переструктурированием (задачи на преобразование).

Это задания, требующие от учащегося анализа имеющихся данных и восстановления или изменения их последовательности в соответствии с поставленным условием. Например, предлагается определенная последовательность фаз деления клетки с некоторыми неточностями. Учащийся должен восстановить правильный порядок следования, или указать недостающие или лишние фазы.

Эти задачи предполагают перестановку частей с целью получения целого. Здесь требуется восстановить правильную или лучшую последовательность букв, слов или фигур, разобрать или собрать устройство или прибор и т.п.

Пример 1. *Укажите грамматически правильную последовательность слов, поставив цифры 1, 2, 3, 4, 5 в скобках:*

собака верная хозяина свою спасла
() () () () ()

Главным преимуществом таких задач является возможность регистрации хода решения задачи. Но вместе с тем достаточная трудность в стандартизации и невысокая экономичность делает такие задачи мало употребляемыми.

2. Задачи на нахождение ошибок.

Учащиеся анализируют схемы, планы, высказывания и определяют, какая схема наиболее точно отражает процесс, какое высказывание соответствует предложенному понятию и т. д. При этом в материале задания специально заложены ошибки, задача учеников их отыскать и объяснить.

3. Задачи на соотнесение.

В данном случае задачи на соотнесение будут включать большее количество правильных ответов, чем необходимо. При этом необходимо, чтобы ответов было приблизительно в 2 раза больше, чем заданий (но не более 12 перечислений).

4. Задачи на доказательство.

Задачи на доказательство нацелены на выявление понимания предложенного понятия и умения логически построить умозаключение.

5. Задачи по типу «открытой книги».

Предполагают использование справочной литературы для точного ответа.

6. Логические задачи.

Логические задачи могут быть представлены в виде подробного описания процесса или ситуации с пропусками. Пропусков может быть достаточно много, но без нарушения смыслового понимания текста.

7. Проблемные задачи.

Используются в нескольких вариантах: с недостатком данных, с избыточной или ненужной для решения исходной информацией; с противоречивыми сведениями в условии; допускающие вероятностные решения.

Заканчивая рассмотрение заданий, которые могут применяться в тестах достижений, можно сказать, что при конструировании тестов этого типа следует ориентироваться на цели создания тестов и разнообразие форм заданий.

4.6. Составление задачи через выделение концепта.

Любая тестовая методика включает совокупность тестовых задач. Тестовые задачи — это отдельные задания, пункты методики, служащие в совокупности для моделирования, проверки или анализа исследуемой деятельности.

Способ конструирования задания через выделение концепта описан у В. К. Гайды (17). Как уже отмечалось, отбор содержания для составления оригинальных задач осуществляется посредством анализа компонентов деятельности или компонентов учебного материала, представленного в виде теоретических схем или матриц. В результате дальнейшего анализа выделяется концепт задачи, очерчивающий ее содержательно, описывающий ее в свободной форме с помощью одного или нескольких утверждений. На основе этих утверждений формулируется вопрос или составляется задача, а также отмечается правильный и возможные ошибочные ответы или варианты ответов. Например, при конструировании теста на проверку знаний, умений и навыков в области механики не обойтись без концепта, описывающего «золотое» правило механики. Содержание может быть таким: «Установление связи между числом подвижных блоков и величиной прикладываемого к веревке усилия при подъеме груза; суть правила — выигрываем в силе, проигрываем в расстоянии; каждый подвижный блок позволяет примерно в два раза снизить прикладываемое к веревке усилие для подъема одного и того же груза. Вербально задача, разработанная на основе этого концепта, может быть сформулирована следующим образом: «При заделке вертикальных швов между панелями построенного здания для подъема рабочего на необходимую высоту применяется „люлька“. Она поднимается с помощью троса, который перекинут через неподвижный блок, закрепленный на балке. Балка жестко связана с крышей здания. Лебедка развивает усилие в полтора раза меньше, чем необходимо для подъема люльки с рабочим. Как выйти из положения и заделать вертикальный шов?»

Задача может быть представлена в виде двух рисунков: на одном — рабочий поднимает груз с помощью веревки, перетянутой через неподвижный блок; на другом — рабочий поднимает тот же груз, закрепленный на подвижном блоке с

помощью веревки, одним концом привязанной к потолку и пропущенной через подвижный и неподвижный блоки. Какому рабочему легче? Даны три варианта ответов: а) первому; б) второму; в) одинаково (17, с. 21).

Одним из источников при составлении концептов являются задачи и вопросы из тестов, сходных по характеру, в том числе из зарубежных. Однако при этом необходимо учитывать культурные различия, и разницу в системах школьного и профессионального образования. Поэтому сначала у заимствованной задачи определяется концепт, а затем на основе концепта составляется новая задача, не обязательно сходная с заимствованной по содержанию и форме.

Число концептов задач при конструировании теста по возможности должно быть большим (в 2-3 раза превосходить планируемое в конечной форме число задач). Опыт многочисленных создателей тестов показал, что это необходимо в силу следующих причин:

1. не все концепты можно преобразовать в адекватные по форме и содержанию задачи для теста;
2. не всегда к задаче можно подобрать «хорошие» неправильные ответы;
3. часть задач оказывается непригодной для использования в составе первичной формы теста в связи с неадекватностью результатов их решения.

Создание концепта является необязательным для однородных задач с общей инструкцией (числовые ряды, определение понятий и др.).

Рассмотрим несколько обстоятельств, которые желательно учитывать при составлении тестов достижений.

4.7. Учет дополнительных обстоятельств при выборе формы заданий.

1. Специфика однотипных заданий.

Если тест состоит из заданий только одного типа, возникает опасность влияния фактора, связанного с этой формой задания. Навык отвечать на задания определенного вида может привести к выявлению не какого-то факта или явления, а данного навыка, сделать вывод о характере обучения на основе специфики заданий. Возможное средство борьбы с этим явлением — разнообразие типов используемых заданий.

2. Специфика применения заданий тестов достижений в педагогике.

С некоторой степенью условности можно сказать, что психологи, применяя тестирование в своей практике, не учитывают, в отличие от педагогов, влияния, оказываемого самим процессом тестирования на обучение и воспитание. Для психолога тестирование — постановка диагноза, для педагога очень часто еще и коррекция, а иногда и обучение одновременно. Это обстоятельство накладывает отпечаток на тесты школьных достижений, причем, как на форму заданий — они должны быть значительно более разнообразными, так и на способ их представления, на чем мы остановимся несколько позже.

3. Утомление.

Источником ошибок при тестировании часто становится утомление, особенно там, где необходима концентрация, сосредоточение внимания. Разнообразие заданий, разнообразие форм представления материала в заданиях могут сделать тест менее монотонным и утомительным для испытуемых.

4. Соответствие формы заданий целям тестирования.

Как было показано в анализе различных типов заданий, каждое из них наиболее эффективно при определенных условиях. Основной подход к конструированию заданий сформулирован П. Клайном следующим образом:

- 1) стандартным заданием, которое подходит к большинству видов материала и вопросов, является задание с вариантами выбора;
- 2) для получения подробной фактической информации, связанной с содержанием заданий, эффективный и компактный тест обеспечивается типом заданий на восстановление соответствия и альтернативными заданиями;
- 3) для тестирования понимания материала, представленного в вопросе, особенно полезны альтернативные задания;
- 4) по причинам, обсуждавшимся в пунктах 1 и 2, имеет смысл разрабатывать тест с заданиями всех типов, хотя и нет необходимости иметь одинаковое количество заданий каждого типа.

4.8. Различные формы представления заданий в тесте.

Кроме рассмотренной нами классификации тестовых задач по типам и видам можно выделить еще и различные формы задач.

Рассматриваемые нами до сих пор задания носили в основном вербальный характер. Однако существует множество способов представления заданий в невербальной форме или же в вербально-невербальной. Причем, в таких формах могут быть представлены не только сами задания, но и варианты ответов. Вариативность форм представления заданий является, в первую очередь, мощным средством сделать задания разнообразными, значительно менее утомительными и даже интересными для испытуемых, особенно младшего возраста. Вторая цель разнообразного представления заданий заключается в возможности тестирования некоторых специальных умений, например, умения работать с каргой, понимать и читать схемы и графики и прочих.

Вопрос или задание могут быть сформулированы в виде фразы, текста, рисунка, схемы, цифры, символа, графика и пр. Как и в предыдущих разделах, приведем несколько примеров, на которых поясним применение разных вариантов представления заданий и вариантов ответов.

Пример 1. *Смотри пример 2 заданиям открытого типа (стр. 85).*

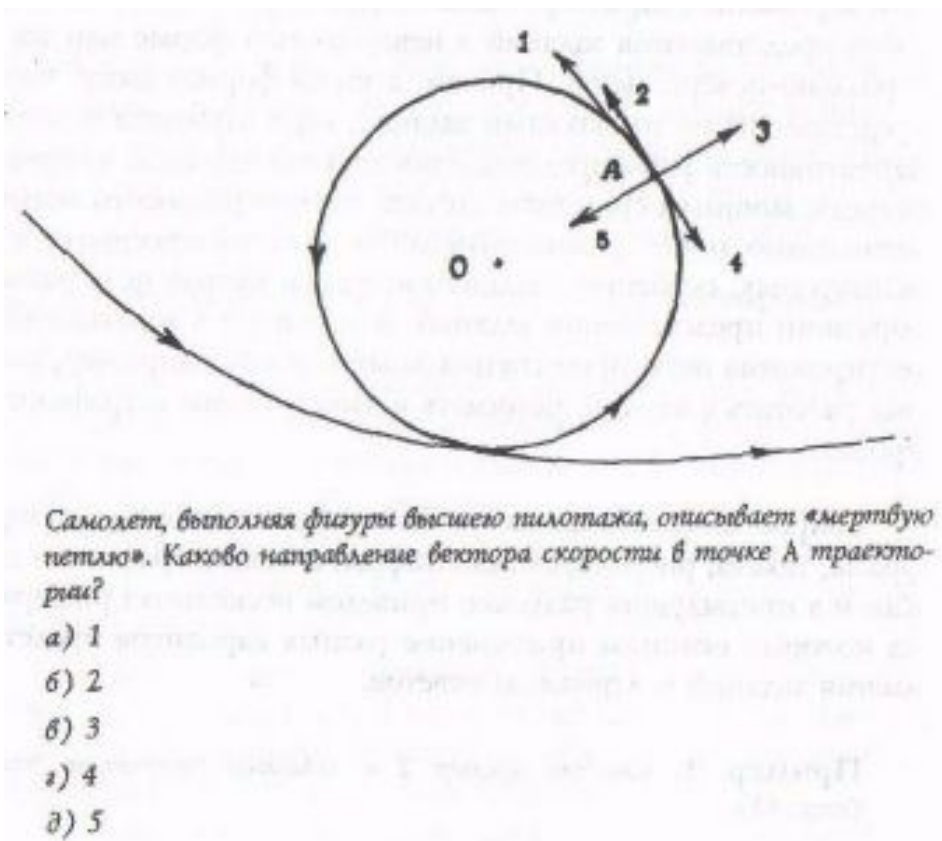
Приведем еще один пример, когда задание сформулировано в виде отрывка текста, в данном случае стихотворного:

Пример 2. В какой пьесе поют романс:

Не искушай меня без нужды
Возвратом нежности твоей
Разочарованному чужды
Все обольщенья прежних дней...

- а) «Гроза»
- б) «Вишневый сад»
- в) «Бесприданница»
- г) «Волки и овцы»

Пример 3. Задание представлено в виде рисунка



Пример 4. Задание представлено в виде схемы

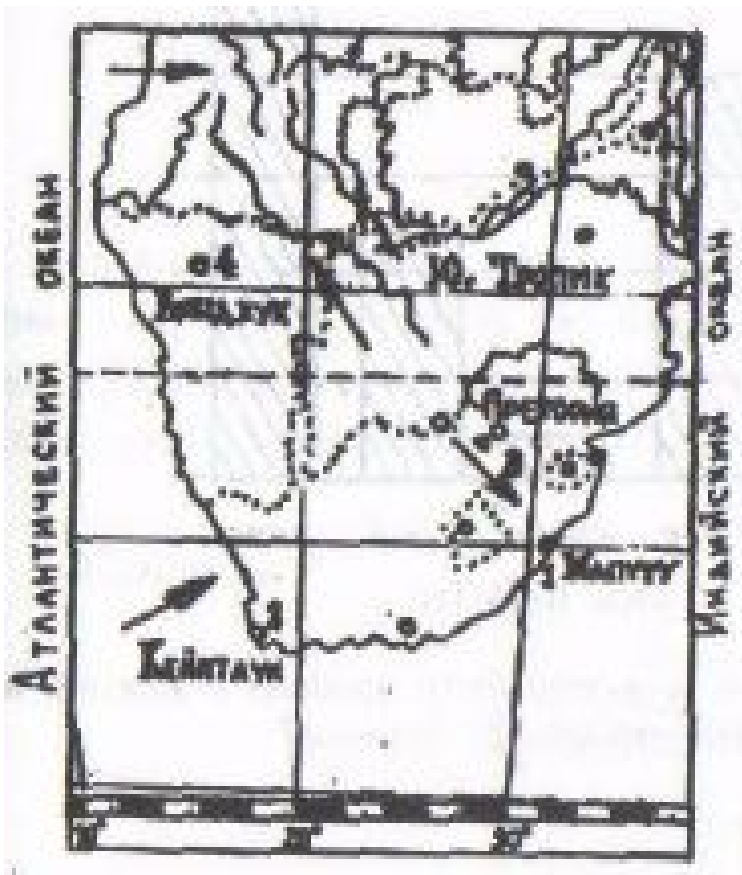


Рассмотрите схемы строения сердца позвоночных животных различных классов. Определите, какой буквой обозначено сердце земноводных (амфибий)

A B C

Пример 5. Задание представлено в виде графика. См. пример 5 альтернативных заданий (стр. 92).

Пример 6. Задание представлено в виде карты



По климатообразующим факторам, указанным на карте, определите, какой цифрой обозначен город, расположенный в средиземноморском типе климата.

- A) 1
- B) 2 .
- C) 3
- D) 4

Пример 7
Задание представлено в виде гистограммы



На рисунке показано, сколько билетов было продано в каждый из пяти дней недели. В какой день было продано 26 билетов

- а) в понедельник
- б) во вторник
- в) в среду
- г) в четверг

Пример 8. Задание представлено в виде таблицы: Ученых проводил опыты с магнитом и разными предметами, В результате он составил такую таблицу:

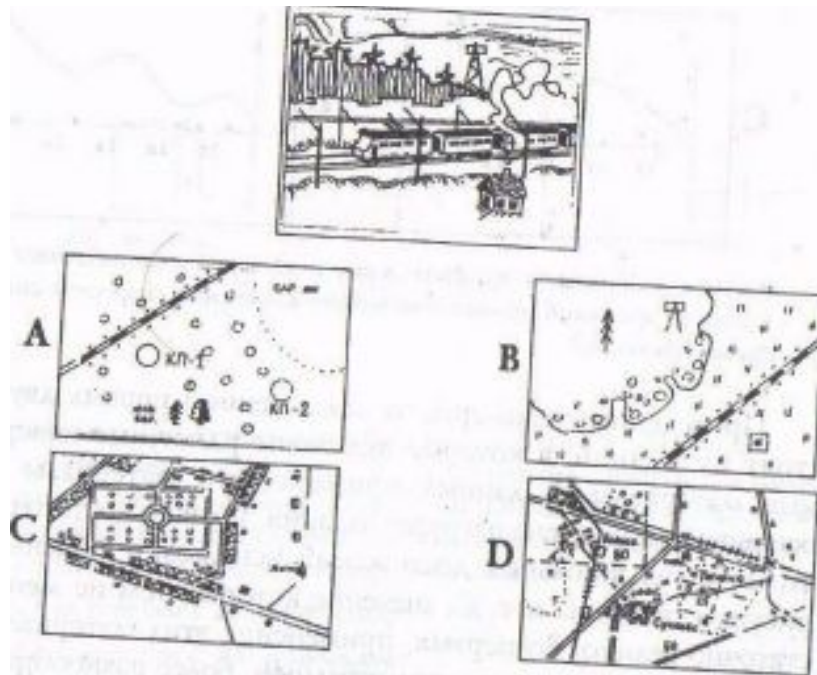
предмет	притягивает	не притягивает
Железный гвоздь	X	
Деревянная зубочистка		X
стеклянный шарик		X
Стальная ложка	X	
Монеты из серебра		X
Медная проволока		X

Выбери наиболее правильный вывод, который ученик мог бы сделать из полученных данных:

- а) магнит одни металлы притягивает, а другие не притягивает;
- б) магнит притягивает все металлические предметы и не притягивает неметаллические;
- в) магнит притягивает некоторые стеклянные предметы;
- г) магнит притягивает дерево.

Кроме заданий в различном виде могут быть представлены и варианты ответов. Большой трудности их использование не представляет, поскольку правила их отбора и составления совпадают с общими правилами составления заданий.

Пример 9.



Какой буквой обозначена часть топографической карты, отображающая местность на рисунке?

Пример 10.

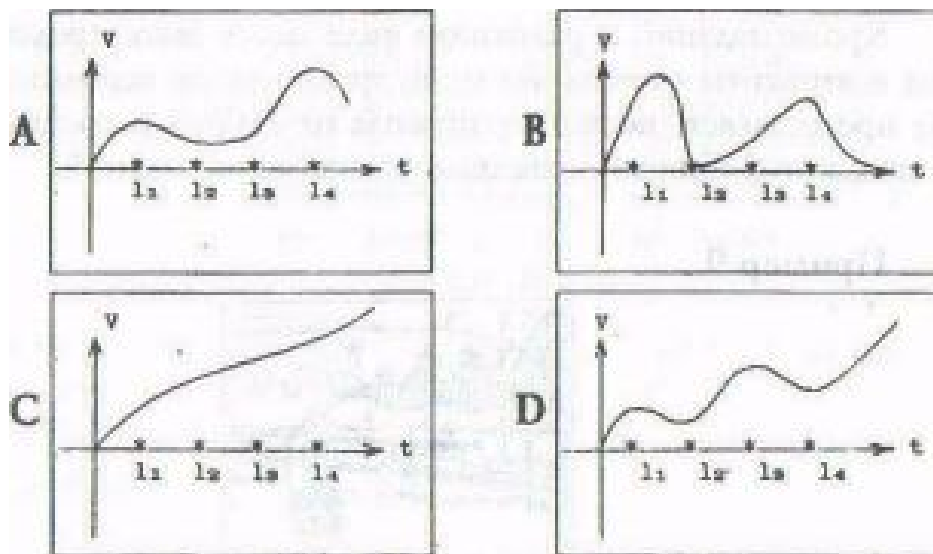
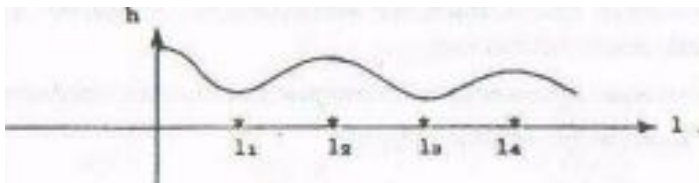
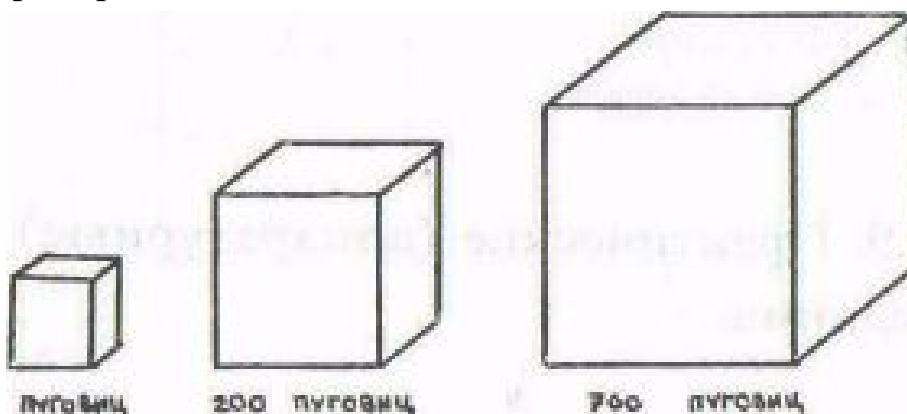


Рисунок изображает профиль юры, с которой скатываются сани. Какой из графиков точнее описывает изменение скорости саней во время движения?

Приведенные примеры, за исключением первых двух, состоят из заданий, в которые включены различные невербальные материалы. В данных примерах эти материалы несут основную смысловую нагрузку задания. Однако достаточно часто в тестах школьных достижений используют поясняющие рисунки, таблицы и т. д., значение которых тем не менее достаточно велико. Во-первых, применение этих материалов делает тестирование менее утомительным, более разнообразным, снижает количество случайных ошибок. Во-вторых, рисунки служат дополнительным мотивом при выполнении тестирования, вызывая интерес к заданию, особенно у детей младшего школьного возраста. В-третьих, эти рисунки дают возможность сгладить различия в уровне развития абстрактного мышления а у школьников начальных классов. Наличие вербальной и не- §' вербальной форм задания одновременно снимает эту разницу в развитии, которая у детей этого возраста довольно велика.

Приведем два примера, когда наличие рисунков не объясняется содержанием задачи.

Пример 11.



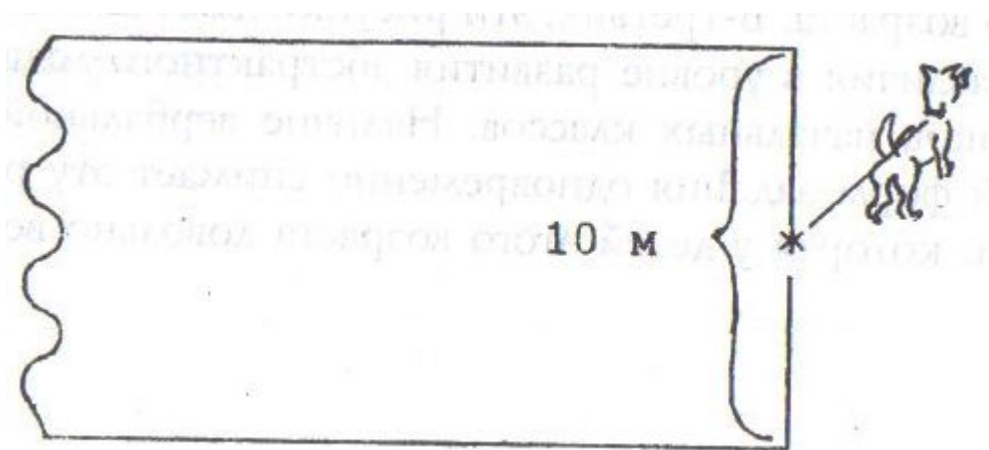
Имеется три коробки с пуговицами в первой — 25, во второй — 200, а в третьей — 700 пуговиц. В каждой коробке с пуговицами имеется по одной синей пуговице. Надо, не заглядывая в коробку, вынуть одну пуговицу. Из какой коробки надо вынимать пуговицу, чтобы возможность Вынуть синюю пуговицу была наибольшей? а.) из коробки с 25 пуговицами;

б) из коробки с 200 пуговицами;

в) из коробки с 700 пуговицами;

г) из любой из этих коробок.

Пример 12.



Собака привязана к середине стены гаража длиной 10 метров. Длина веревки 5 метров. Чему равна площадь участка, который она может контролировать?

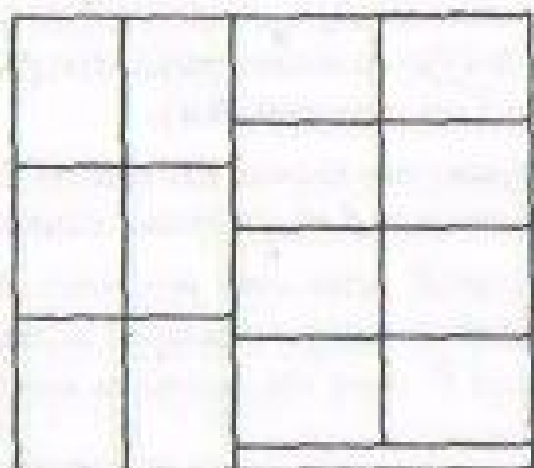
- а) 25π кв. метров;
- б) 15 кв. метров;
- в) 25 кв. метров;
- г) 12.5π кв. метров.

4.9. Практические (аппаратурные) задания.

Практические или аппаратурные задания стоят несколько отдельно в ряду заданий для тестов школьных достижений. Их психологические аналоги достаточно успешно применяются в практике, например, кубики Косса. В практике тестов достижений практические задания пока не нашли значительного применения ни у нас, ни за рубежом. Причин тому достаточно много: сложность группового проведения, слабая стандартизированность, значительные материальные и трудовые затраты, необходимость длительного подготовительного этапа к каждому тестированию. Однако есть у таких тестов неоспоримые преимущества, которые заключаются в большом объеме получаемой информации, возможности наблюдения и фиксации процесса работы и, наконец, что наиболее ценно для педагогов, возможность использования такого тестирования для практического обучения, о чем мы уже говорили.

Приведем несколько примеров таких тестов (задания рассчитаны на 9 или 13 лет) из опыта ETS с комментариями авторов.

Пример 1.



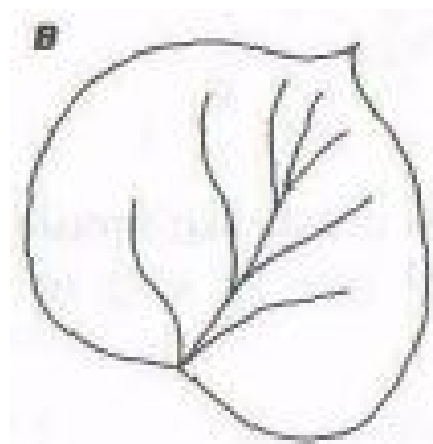
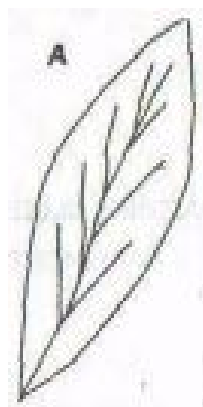
Название задания: БИЛЕТ.

Описание задачи. Определить, какое наибольшее число билетов (прямоугольников) может быть вырезано из листа бумаги. Приборы и материалы. Один билет (5x7 см) и чистый лист бумаги (24x21 см) (линии на рисунке показывают правильное решение). Указания. Пол сумел вырезать 12 билетов из такого листа. Жюли удалось вырезать 13, Найдите наибольшее количество билетов, которое можно получить из листа бумаги. Свое решение изобразите линиями на листе.

Правильный ответ. Правильный ответ засчитывается, если сказано 14 и линии проведены так, как показано на рисунке.

Комментарии, процент правильных ответов на это задание оказался довольно низким: самый высокий едва перевалил за 30. Еще меньше, оказались в силах нарисовать правильное решение.

Пример 2.



Название задания: ЛИСТЬЯ.

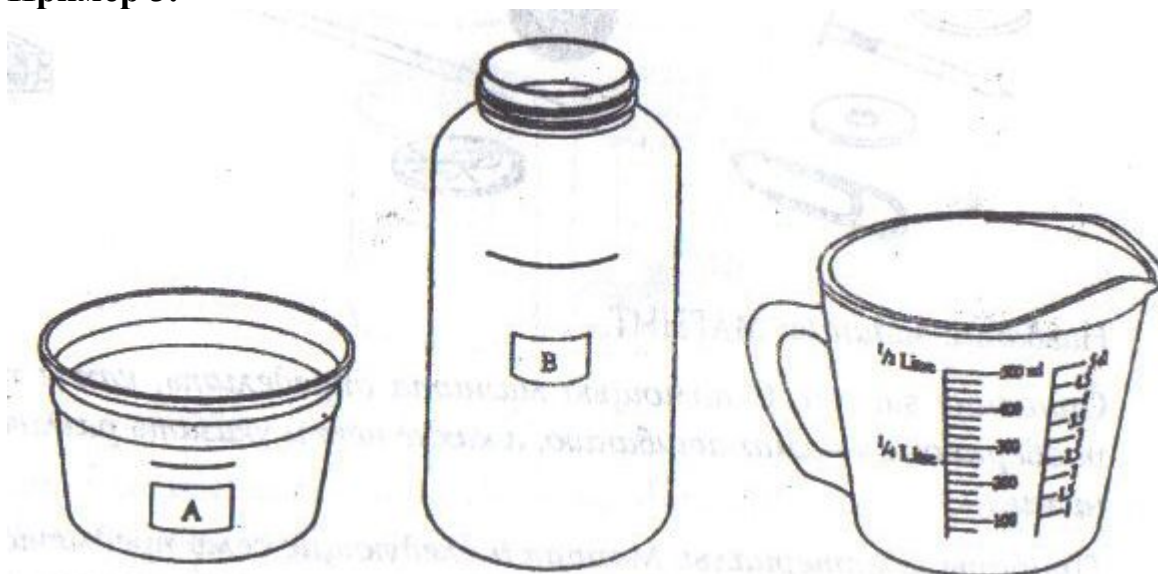
Описание задачи. Найти площадь двух негеометрических фигур с помощью измерительной решетки (миллиметровки). Приборы и материалы. Два изображения листа, площадью 21 кв.см и 48 кв.см, калька и решетка, размеченная в квадратных сантиметрах. Указания. Найдите площадь листьев, используя решетку. Вы можете пользоваться калькой, положив ее поверх решетки и отмечая на ней клеточки, которые попадают в площадь листа полностью или

Правильный ответ. Правильный ответ засчитывается в случае указания истинной площади листьев с точностью плюс-минус 4 кв.см.

Проблемы. При фотокопировании изображений листьев их размер немного изменялся. Поэтому, правильными пришлось считать ответы при больших отклонениях, чем вначале предполагалось. Ввиду неточностей, перевода студенты в Советском Союзе пропустили часть информации.

Комментарий. Процент правильных ответов выше для маленькою листа.

Пример 3.



Описание задачи. Определить объем двух пластиковых емкостей с помощью измерительной чашки.

Приборы и материалы. Большая емкость с водой, две меньших емкостей, помеченные буквами А и В, и измерительная чашка (500 мл с ирной деления разметки 25 мл). На емкости А и В нанесены черные линии на уровне 375 мл и 725 мл соответственно. Указания. Наполните емкости А и В до черной линии из большой емкости. С помощью измерительной чашки определите, сколько воды находится в каждой емкости, Ответ, дайте в миллилитрах. Правильный ответ: правильный ответ засчитывался при указании истинного объема плюс-минус 25 мл.

Проблемы. Проверка правильности нанесения разметки на емкостях А и В, а также на измерительной чашке выявила значительные различия. Для емкости А эти различия были достаточно малы для того, чтобы засчитывать правильный ответ в пределах ошибки плюс-минус 25 мл. Поскольку искажения для емкости В оказались больше, эти результаты были проигнорированы. Комментарии. В основном результаты были высокими (около 90%) за исключением Советского Союза и Тайваня, где они оказались несколько ниже. Было очень мало случаев, когда школьники не давали ответа вовсе (1% из всех стран и территорий), однако много раз вода проливалась!

Пример 4.



Название задания: МАГНИТ.

Описание задания. С помощью магнита определить, какие предметы подвергаются намагничиванию, а какие нет и указать различия между ними.

Приборы, и материалы. Магнит и следующие семь предметов: пластмассовая пуговица, железная или стальная шайба, стальная скрепка, железный гвоздь, стеклянный шарик, пластмассовая палочка и медная монета,

Указания. Проверьте, реагируют ли эти предметы на магнит, и разделите их на две группы. Составьте списки предметов, входящих в эти группы, и объясните что отличает эти предметы друг от друга. Правильный ответ.

Правильный ответ засчитывался при правильном разделении на группы.

Встречались четыре вида объяснений: по именам; одна из групп состоит из предметов, сделанных из железа или опали; одна из групп состоит из предметов, реагирующих на магнит; одна из групп состоит из предметов, сделанных из железа или стали и реагирующих на магнит.

Комментарии. Как правило, ученики хорошо справлялись с этим заданием, результат колеблется в пределах от 86% до 95%. По 10% студентов в каждой стране или территории дали неудовлетворительные объяснения.

Пропусков было мало, однако в Азалии 6% студентов пропустили какие-либо предметы.

Наиболее частым основанием для группирования оказалась намагничиваемость предметов (79%). Меньшее количество (4—30%) упоминали материал, и этот показатель сильно изменялся в зависимости от страны или территории.

Пример 5.



Название работы: ПЛАВАНИЕ.

Описание задачи. Выбрать правильные описания наблюдений над двумя идентичными наборами объектов.

Приборы и материалы, Два небольших стеклянных сосуда, помеченные буквами X и Y, содержащие прозрачные жидкости и похожие пластмассовые игрушки, одна из которых плавает (сосуд X), а другая утонула (сосуд Y).

Указания. Посмотрите внимательно на эти сосуды (вы можете, их поднимать). Пятеро студентов, посмотрев на эти сосуды, сделали следующие заключения. Какие из этих заключений описывают то, что студенты действительно видели? А. Я вижу игрушку, плавающую в сосуде X. В. Я вижу игрушку, плавающую в сосуде Y.

С. Я вижу, что игрушка в сосуде X сделана из другой пластмассы, нежели игрушка в сосуде Y.

Д. Я вижу сосуды, содержащие бесцветную жидкость и цветные игрушки,

Е. Я вижу, что игрушка в сосуде Y тяжелее, чем игрушка в сосуде X.

Правильный ответ. Правильный ответ засчитывался, если испытуемый обвел верные утверждения А и D и не обвел неверные В, С и Е.

Комментарии. Процент учеников, которые обвели правильные ответы и не обвели неправильные был невысок и колебался в пределах от 10 до 34 процентов, большая часть учеников указала ответы А и D как наблюдения.

Почти все дети определили, что В не является наблюдением в противовес А.

Большая часть учеников определила, что утверждение С о том; что игрушки сделаны из разного материала, ложно, возможно, что игрушки очень похожи.

Однако утверждение Е о различной массе игрушек показалось привлекательным многим, и они его обвели, несмотря на то, что масса, игрушек им была неизвестна.

4.10. Создание двуразмерных тестов и возможность создания трехразмерных тестов.

Перед тем, как обсуждать эту тему, необходимо рассмотреть принципиальную проблему, которая характерна для тестов» достижений, — проблему формы и содержания. В психологических тестах, близких к тестам достижений,

например, в тестах интеллекта, содержание состоит в закономерности, которую должны выявить испытуемые. Проблема создания тестов интеллекта, таким образом, заключается в конструировании такой формы задания, которая позволила бы выявить эту закономерность или зависимость.

В тестах достижений, как правило, основой является не форма, а содержание задания. Некоторые формы заданий, как мы уже видели, подходят к определенному содержанию материала лучше, чем другие. Существует ряд заданий, которые по форме невозможно совместить с тем или иным материалом. Следовательно, перед конструктором теста прежде всего стоит проблема создания адекватного целям исследования содержания заданий, а затем поиска формы, которая наилучшим образом соответствовала бы содержанию. При этом необходимо отметить, что часть форм заданий нейтральна; именно к таким относятся задания, которые мы рассмотрели первыми.

Из особенности тестов достижений, заключающейся в первичности содержания, вытекает очень важное следствие — возможность создания двух- и трехмерных тестов: по содержанию они, например, могут выявлять знания, по форме — составляющие интеллекта, а по способу представления материала — специальные или общешкольные умения.

Первый вариант двуразмерных тестов.

Двуразмерные тесты уже достаточно давно и успешно применяются как в мировой практике, так и в нашей стране. Опыт применения таких тестов можно разделить на две большие группы в зависимости от оснований двуразмерности, которые в них заложены. Первый вариант основательно разработан международными организациями по проверке подготовленности школьников и основан на результатах исследований американской организации Education testing service (ETS, Служба педагогического тестирования). Применение такого рода тестов проводится Международной ассоциацией по оценке результатов обучения при ЮНЕСКО — IEA (International Association For The Evaluation Of Educationall Achievement), TIMMS и других. В России применение этих тестов связано с работами в НИИ ОСО под руководством Г. С. Ковалевой. Размерность этих тестов определяется по двум основаниям: категориям познавательной деятельности и по знаниям и умениям детей по тому или иному разделу предмета. Приведем в качестве примера матрицы заданий для двух исследований естественнонаучного образования.

Как можно видеть, принятые основания для выделения двух шкал достаточно близки друг к другу, хотя и различаются по названиям.

В силу терминологической разницы, сложившейся в результате изолированного развития России и западных стран, сказать что либо о приведенных названиях шкал практически невозможно. Раскроем некоторые понятия, особенно по категориям познавательной деятельности, так как они определялись авторами исследования, одновременно обращая внимание на терминологические различия.

Воспроизведение знаний. «Задания данного типа проверяют уровень сформированности фактических знаний: умение

	Биология	Химия и физика	Наука о Земле и Астрономия	Методология науки
Воспроизведение знаний				
Применение знаний				
Интеграция знаний				

Модель теста ETS

воспроизводить факты, понятия, принципы и методы, используемые в данной предметной области, а также знание научной терминологии, умение распознавать основные идеи данной предметной области в другом контексте, умение передать информацию другими словами или в измененном виде» (28). Умение распознать идеи данной предметной области в другой отечественная психология относит к другому, более высокому уровню познавательной деятельности.

Пример 1. *Какое из следующих веществ наименее растворимо в воде?*

а) поваренная соль; б) сахар; в) спирт; г) бензин. Предметная область — химия.

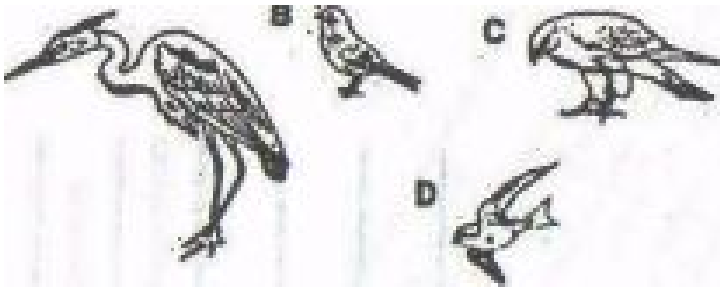
Уровень познавательной деятельности — воспроизведение знаний.

Это простое задание, относящееся одновременно к обозначенным двум шкалам.

	Биология	Физические науки	Науки о Земле	Естествознание, технологии, математика	Естествознание и другие дисциплины	Экологические проблемы	История естествознания и технологии
Использование теории, анализ и решение проблем							
Использование приборов и материалов, стандартных процедур							
Исследование мира							

Модель теста TIMMS

Пример 2. *Какая из изображенных на рисунке птиц может ходить в воде мелкого водоема и есть рыбу?*

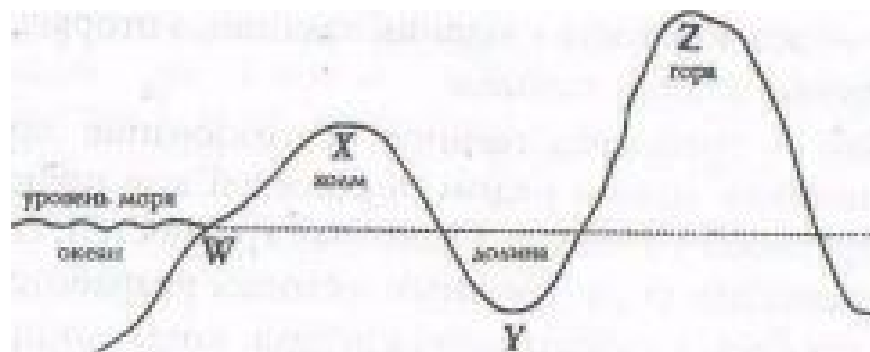


Предметная область — биология.

Уровень познавательной деятельности — воспроизведение знаний. Это задание не укладывается в рамки сложившихся у нас представлений о воспроизведении знаний, поскольку для ответа на этот вопрос испытуемому необходимо произвести некоторые интеллектуальные операции. К силу терминологической разницы, это задание большинство наших педологов отнесли бы к заданиям на применение знаний.

Применение знаний. «Данные задания проверяют умение применять на практике фактические знания для достижения поставленных целей в соответствии с определенными правилами, формулами или алгоритмами. Успешное овладение этими умениями предполагает применение знания научных фактов и принципов в знакомой и незнакомой ситуациях; интерпретацию информации, используя при этом основные научные идеи, способность устанавливать связи между наблюдаемыми явлениями, полученными данными и основными научными понятиями, фактами и принципами» (28).

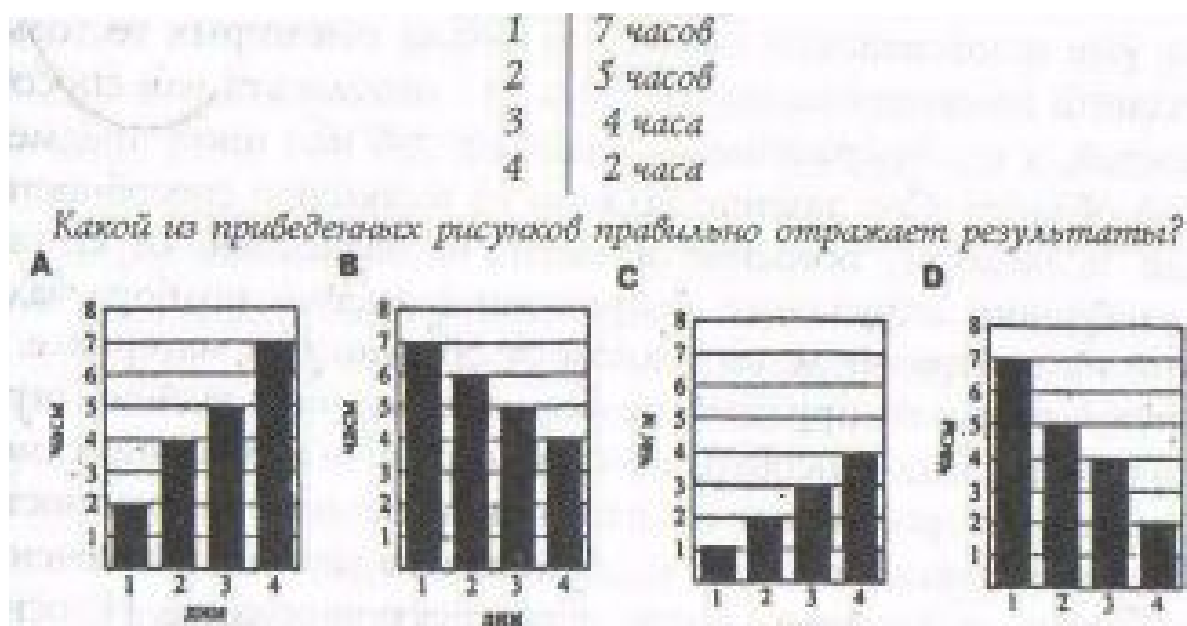
Пример 3. *В местностях W, X, Y, Z, изображенных на рисунке, были записаны одновременно значения атмосферного давления. Погода везде была одинаково ясная. В какой местности эти значения будут наибольшими?*



- А) W Б) X В) Y Г) Z

Предметная область — науки о земле. Уровень познавательной деятельности - применение знаний.

Пример 4. *Почтовые голуби — это птицы, которые могут найти дорогу домой, находясь на большом расстоянии от него. Человек выпустил, голубя в 30 км от дома. Он проделывал это в течение 4 дней подряд и записывал, за какое время голубь вернется домой. Он получил следующие результаты: •*



Предметная область — методология науки.

Уровень познавательной деятельности — применение знаний.

Завершая рассмотрение исследования, проведенного ETS, необходимо зафиксировать два принципиально важных факта.

Первый — возможность создания заданий, которые относятся одновременно к двум шкалам.

Второй — непосредственное использование зарубежного опыта связано с целым рядом трудностей как практического, так и теоретического плана. Основных трудностей, связанных с заимствованиями из зарубежных тестовых разработок, две. Во-первых, проблема адекватного перевода, когда по причине тех или иных переводческих ошибок, в русском тексте появляются слова, обозначающие понятия, уже имеющие устойчивые русские обозначения. Например, «варианс вместо «дисперсия», «общность» вместо «генеральная совокупность». Вторая и, на наш взгляд, основная трудность связана с терминологической разницей, вызванной изолированным развитием наук, как педагогики так и психологии, у нас и на Западе.

Второй вариант двуразмерных тестов.

Второй вариант составления двуразмерных тестов, их можно назвать психолого-педагогическими по своей идее, основан на уже отмечавшемся факте, что форма некоторых тестовых заданий позволяет выявлять структуру интеллекта или способностей, а их содержание — знания в той или иной предметной области. Составление заданий на выявление способностей для психологов, особенно в части, не зависящей от социо-культурного окружения, сопряжено с задачей подбора заданий на нейтральном, свободном от обученности материале. С другой стороны, применение форм целого ряда заданий ограничено в связи с необходимостью ограничить влияние самой формы, которое может вызывать определенные погрешности. В данном случае

речь идет о заданиях на аналогии, исключение лишнего, последовательности, переструктурирование. На основе идеи объединения этих двух шкал был создан школьный тест умственного развития (ШТУР), который получил достаточно широкое распространение,

Приведем несколько примеров из этого теста, которые могут служить образцом составления заданий.

Пример 5. *Вам предлагаются три слова. Между первым и вторым существует определенная связь. Между третьим и одним из пяти слов предложенных на выбор, существует аналогичная, та же самая связь. Это слово вам следует найти и подчеркнуть.*

слово : буква = предложение : ?

а) союз; б) фраза; в) слово; г) запятая; д) тетрадь.

Аналогии.

Предмет - русский язык.

Пример 6. *Вам даны пять слов. Четыре из них объединены одним общим признаком. Пятое слово к ним не подходит. Его надо найти и подчеркнуть. Лишним может быть только одно слово*

а) Канада; б) Бразилия; в) Вьетнам; г) Испания; д) Норвегия.

Классификация.

Предмет — география.

а) пословица; б) басня; Классификация. Предмет — литература.

г) былина; д) сказка.

Пример 7. *Это пример открытого вопроса.*

Вам предлагаются два слова. Нужно определить, что между ними общего. Старайтесь в каждом случае найти наиболее существенные признаки обоих слов. Напишите свой ответ рядом с предложенной парой слов.

Ответ....

метафора — аллегория

Обобщения.

Предмет — литература.

Ответ....

реформа — революция

Обобщения

Предмет — история

Таким образом возможно создание двуразмерных тестов на основе сочетания форма-содержание.

Возможность создания трехразмерных тестов.

Попытаемся представить одно и тоже задание с иллюстрациями в разных формах,

Пример 8. Дано числа учащихся в шести сельских школах: в первой — 50, во второй — 70, в третьей — 60, в четвертой — 30, в пятой — 50, в шестой — 110. На сколько меньше учеников в школе с наименьшим числом учащихся, чем в школе с наибольшим числом учащихся.

а) 10

б) 400

в) 550

г) 50

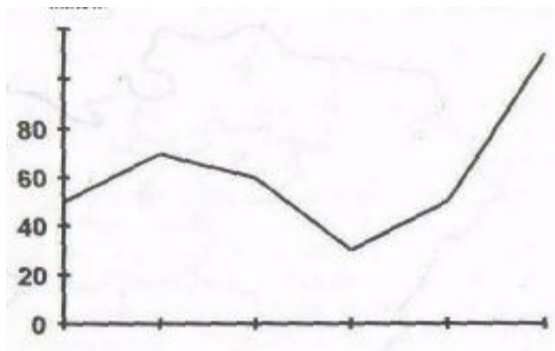
д) 80

Количество учащихся в сельских школах может быть представлено несколькими способами.

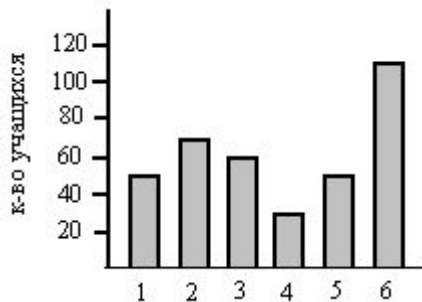
А) В виде таблицы

Школа	Количество учеников
1	50
2	70
3	60
4	30
5	50
6	110

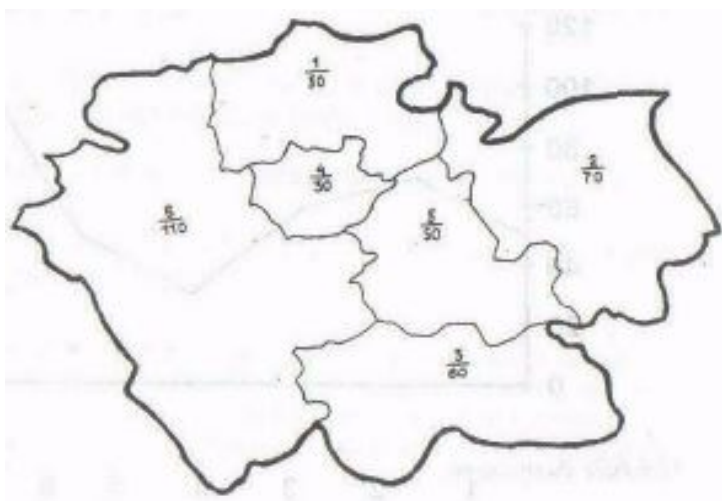
б) в виде графика:



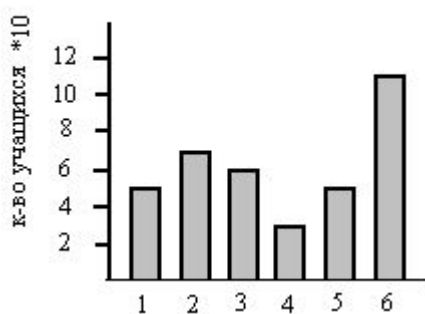
В) в виде диаграммы



г) в виде карты:



д) В виде диаграммы с масштабом (то же можно проделать, и с таблицей и графиком):



Очевидно, что разные формы представления данных в этой задаче повлияют на трудность ее выполнения. Таким образом, представляя информацию для задания в разных видах, не изменяя при этом ни форму задания (в нашем случае это задам, ответа), ни содержание задания (в данном случае это с больших и меньших чисел и нахождение разности между больше и меньшим), а выбирая лишь форму представления данных, мы можем существенно изменить задание, усложнить или упростить его

В данном случае, манипулируя представлением данных, мы имеем возможность проверить иных качеств подготовки учеников, которые не были заложены ни в форме самого задания, ни в его содержании. „ В данном примере мы имеем возможность выявить умения учеников работать с картой, таблицей, диаграммой, графиком.

Таким образом, мы имеем возможность говорить о создании трехразмерных тестов. По одной их шкале можно определить знания в той или иной предметной области — это определяется содержанием задания. По другой шкале можно определить способности и элементы структуры интеллекта — это является формой задания. И, наконец, по третьей инсайде можно определить школьные умения и навыки, которые Г определяются формой представления задания в тесте.

Кроме общих правил составления задания, для двумерных и трехразмерных тестов имеет большое значение количество заданий в тесте, способ и принципы формирования их в группы. Это и будет предметом рассмотрения в следующей главе.

Глава 5. Работа с заданиями после составления

5.1. Объединение задач в тест.

После того, как задачи будут подготовлены, следует проверить их объективность и корректность формулировки. Задачу или вопрос можно считать объективными или корректно сформулированными, если мнения ряда экспертов об их назначении, о правильности формулировки и пригодности вариантов ответов совпадают между собой, Число экспертов, необходимых для работы над тестами, колеблется у разных авторов от 3 до 8. Количество необходимых экспертов зависит от двух обстоятельств: целей создаваемого теста и качества экспертов Для тестов достижений локального применения достаточно и трех экспертов, для тестов, применение которых планируется в широких масштабах, количество экспертов не должно быть менее пяти. Вопрос оценки качества экспертов и процедура их отбора достаточно подробно описана в работах по технологиям проведения экспертных оценок, Экспертам предлагается по трех- или пятибалльной шкале оценить задачи с точки зрения соответствия целям тестирования (отвечает — частично отвечает — не отвечает), однозначности формулировки (однозначна — не совсем однозначна -неоднозначна), и пригодности вариантов ответов (подходят - частично подходят — не подходят). Если хотя бы треть экспертов оценивает задачу как частично соответствующую или не соответствующую тесту, ее исключают. Если процент такой оценки ниже, то задача пересматривается и корректируется.

5.2. Упорядочивание заданий для испытания теста.

Как было указано в первой главе, явные конструкции, обеспечивающие простоту понимания того, что испытуемые должны делать, способствуют повышению надежности и валидности теста. Есть несколько простых правил для упорядочения заданий при испытаниях теста, которые настолько очевидны, что заслуживают лишь кратких комментариев.

1. Составьте базу данных для заданий, в которой каждая задача представлена на отдельном экране, предусмотрев место для занесения экспертных оценок, времени, необходимого для ее выполнения, уровня сложности и прочих данных, которые могут ее охарактеризовать.

2. Проверьте содержание и формулировку задач во взаимосвязи друг с другом,

3. Располагайте задания каждого типа вместе, При этом инструкцию и пояснения к ним необходимо давать один раз для каждой группы заданий, Это дает возможность испытуемым приспособиться к данному типу заданий, а следовательно, достаточно понимания одной части инструкции для множества заданий.

4. Располагайте задания в порядке возрастания предполагаемой трудности. Это предотвратит случаи, когда слишком старательный испытуемый тратит все свое время (или слишком много времени) на задания, которые он не может решить и, таким образом, лишает себя возможности попытаться выполнить другие, по которым он мог бы получить баллы, а в результате все формы анализа заданий будут неточными. При апробации теста бывает полезно включить в инструкцию пункт (о том, что если испытуемому не удастся справиться с заданием, его необходимо пропустить, а после окончания работы над всем тестом вернуться к вызвавшему трудность заданию, если ', у испытуемого останется время.

5. С учетом (3) и (4), располагайте материал с разными формами представления данных с чередованием, насколько это возможно, Это снижает монотонность и связанное с ней утомление.

6. Не комплектуйте вместе такое количество заданий, что среднему испытуемому для их выполнения потребуется более получаса — для детей начальной школы; для старшеклассников — более часа. Столько длится период сосредоточения у детей. Если существует необходимость выполнения заданий большей продолжительности, технология проведения должна предусматривать перерыв. Необходимо отметить, что время появления утомления во многом зависит от мотивации (при этом слишком высокая и слишком низкая - мотивация быстрее вызывают утомление), разнообразия материалов тестирования, способа проведения, эмоциональной подготовленности учеников.

7. Гайда и Захаров рекомендуют дать испытуемым указание отмечать, какое задание теста они выполняют в различные моменты времени в процессе тестирования, для того чтобы убедиться, сколько времени требуется на выполнение заданий теста. Наш собственный опыт апробации тестов говорит о нерациональности этого пути; испытуемые отвлекаются или забывают отметить задание. Возможно, этот способ хорош при работе со взрослыми, При работе с детьми наиболее рационально оказалось поручить эту работу

испытателю, который может одновременно фиксировать время выполнения заданий у 5—6 испытуемых, отмечая его на специальном бланке.

8. При конструировании бланковых тестов лучше разместить задачи на листах брошюры так, чтобы они были пространственно разнесены и легко воспринимались. Задания и варианты ответов к ним должны располагаться на одной странице.

9. Важные части инструкции должны быть подчеркнуты или выделены особым шрифтом. Сделайте бланки ответов. Размножьте брошюры и бланки. Можно считать, что для проведения апробации все готово, за исключением одного — необходимо выбрать контингент, на котором будет испытан тест.

5.3. Проверка трудности задач. Определение места задачи в тесте.

Важным шагом в конструировании теста является проверка трудности предложенных задач. Для этого необходимо провести предварительное тестирование экспериментальной группы (выборки). Отбор участников экспериментальной группы может производиться различными путями. Наиболее доступным является отбор в случайном порядке по алфавиту, по таблице случайных чисел или другим способом. Вместе с тем состав выборки должен быть сбалансированным, т. е. тестируемые должны обладать разными возможностями решения задач. Количественный состав выборки рассмотрен отдельно в главе 6.

После того, как определен состав выборки, испытуемым предлагается решить составленные задачи. Полученные ответы анализируются с целью установления трудности, обоснованности и дискриминативности каждого вопроса, пригодности каждого варианта ответов. Результатом анализа должен явиться отбор и корректировка задач, а также их перераспределение внутри теста.

Трудность задачи является важнейшей характеристикой, определяющей место задачи в тесте. Трудность может быть субъективной и объективной.

Субъективная трудность задачи связана с индивидуально-психологическим барьером. В психологии величина этого барьера определяется различными факторами, в том числе:

1. условиями решения задачи (временем, отведенным на решение, понятностью инструкции и т. п.);
2. уровнем формирования необходимых для решения знаний, умений и навыков;
3. состоянием испытуемого и т. д.

Для снижения влияния перечисленных факторов при тестировании определяется стандартная форма процедуры проведения — этот вопрос мы рассмотрим несколько позже.

В большинстве случаев тесты достижений учитывают только правильность решения задач и меньше внимания уделяют способу решения, характеру затруднений, энергетическим затратам испытуемого. В связи с этим определяется объективная трудность задач (статистическая). Статистическая трудность определяется долями выборки, решивших и не решивших задачу. Например, если задачу решили только 20% участников тестирования, то ее

можно оценить как трудную для данной выборки, если 80% — как легкую. При этом значимым является только факт выполнения или не выполнения задания, причины неудач не рассматриваются.

Трудность задачи является относительной характеристикой, поскольку зависит от особенностей выборки и может меняться от выборки к выборке. Так, для выборки первоклассников задача «Сколько будет 7х8» является трудной, для старшеклассников — легкой. Поэтому указание уровня трудности задачи без указания контингента лиц, для которого эта трудность установлена, не имеет смысла.

Статистическая трудность позволяет определить место задачи в тесте. Так, если задачу решает большинство испытуемых, то ее, как легкую, помещают в начале; в том случае, когда с задачей справляется незначительный процент испытуемых, то ее, как трудную, помещают в конце теста. Самые легкие задачи (одну-две) выносят перед основными задачами теста и используют в качестве примеров. Итогом распределения задач по степени их трудности должна стать «лестница» усложняющихся задач, каждая ступень которой представлена процентом, испытуемых, решивших соответствующую задачу.

Подчеркнем, что в тестах достижений трудность задач лучше всего определять в условиях «мягкого*» лимита времени или совсем без его ограничения.

Фиксируется успешность и время решения.

Основным показателем трудности задач является индекс трудности, который рассчитывается по следующей формуле:

$$U = 100 \left(1 - \frac{n}{N}\right) \quad (\text{Гайда})$$

где U — индекс трудности в процентах, n — число испытуемых, правильно решивших задачу, N — общее число испытуемых.

Для определения индекса трудности в случае альтернативных заданий, где, как мы уже отмечали, максимальна вероятность угадывания правильного ответа, используется формула:

$$U = 100 \left(1 - \frac{n - \frac{Nn}{m-1}}{N}\right)$$

где N_n — число испытуемых, не решивших задачу, m — число вариантов ответа.

Задания, которые имеют значения показателя индекса трудности больше 80 или меньше 20, как правило в тест не включаются. Логика этого отсеечения достаточно проста — задания, с которыми справляются все или не справляется никто, являются низкоинформативными.

5.4. Случайное угадывание правильного ответа.

Существует еще одна проблема в конструировании тестов достижений, о которой мы уже упоминали — случайное угадывание правильных ответов.

Очевидно, что для объективных заданий случайное угадывание до некоторой степени исказит показатели по тесту. Если тесты предназначены для широкого использования, или используются для таких серьезных целей, как отбор, контрольные или итоговые работы, то очевидно, что испытуемые имеют сильные мотивы для угадывания ответов.

Случайное угадывание снижает надежность и валидность тестов, Проблема угадывания ответов стоит особенно остро для альтернативных заданий. Кроме того, если требуется высокая скорость выполнения тестов, то это подстегивает испытуемых к случайному угадыванию ответов, Еще хуже, однако, проведение тестов в такие сжатые сроки, что испытуемые не могут завершить их выполнения. Это привносит в тест внешнюю переменную скорости выполнения.

В случае с заданиями, содержащими много вариантов выбора, и достаточно большим количеством заданий в тесте, случайным угадыванием можно пренебречь. Однако здесь возникает еще одна проблема — оценки пропущенных заданий, на которые испытуемые не дали ответа вообще. Может сложиться ситуация, когда одни испытуемые вообще не заполнили часть заданий, а другие заполнили непосильные задания случайным образом. В таком случае следует, как советует П. Клайн:

1) Ввести в инструкции требование угадывать ответы. Это легко сделать, снабдив инструкцию подобным пояснением: «Не пропускайте ответы на вопросы. Если вы не знаете ответа или не уверены в его правильности, попытайтесь угадать правильное решение».

2) Перед подсчетом показателей по тесту ответьте на вес; пропущенные испытуемыми вопросы произвольным образом. Это будет эквивалентно случайному угадыванию.

Если будут выполнены действия (1) и (2), все показатели будут искажены случайным угадыванием. Следовательно, валидность теста не будет очень сильно искажена. К сожалению, даже при указании в инструкциях не пропускать ответов на вопросы, не все испытуемые подчиняются этому требованию, поэтому, если исследователь не прибегает к действию (2), следует вносить некоторую коррекцию на случайное угадывание ответов.

Формула коррекции угадывания.

П. Клайн приводит следующую формулу коррекции угадывания:

$$X_{\text{correct}} = \frac{X - W}{(n - 1)}$$

где: X_{correct} — показатель, скорректированный на угадывание;

X — количество правильных ответов «нет»;

n — количество вариантов выбора в заданиях,

Следует отметить четыре момента, касающиеся этой формулы.

1) Предполагается, что все неправильные ответы даны в результате угадывания, Это не всегда так, поскольку испытуемые могут часто иметь неверные сведения.

2) Предполагается, что при угадывании каждый вариант ответа может быть выбран с одинаковой вероятностью. Это не так, поскольку в заданиях с несколькими вариантами выбора испытуемый может правильно отбросить все варианты, кроме двух. Следовательно, для данного испытуемого коррекция на угадывание будет занижена (однако, см. пункт 4).

3) Для заданий с различным количеством вариантов ответов нужна различная коррекция угадывания.

4) Коррекция угадывания носит статистический характер, применима к так называемому среднему испытуемому. В индивидуальных случаях она будет неверной.

Из сказанного ясно, что коррекция на угадывание — это в лучшем случае лишь грубое оценивание.

5.5. Дискриминативность.

Дискриминативность задач определяется как способность отделять испытуемых с высоким общим баллом по тесту от тех, кто получил низкий балл, или испытуемых с высокой продуктивностью учебной деятельности от испытуемых с низкой продуктивностью. Дискриминативность обозначает различительную способность задачи. Для ее определения применяются коэффициент и индекс дискриминации.

Определение коэффициента и индекса дискриминации достаточно сложная, но необходимая процедура, поэтому остановимся на ней подробнее.

В. К. Гайда и Б. П. Захаров предлагают вычислять коэффициент дискриминации следующим образом. Любой ответ испытуемого, решающего задачу, можно оценить по двухбалльной шкале — «правильно» (1 балл), «неправильно» (0 баллов). Сумма баллов по всем задачам теста у испытуемого представляет собой первичную оценку его успешности в тесте.

Мера соответствия между успешностью решения одной задачи и всего теста по выборке испытуемых является показателем пригодности задачи для теста. Этот показатель и является коэффициентом дискриминации задачи и вычисляется по следующей формуле:

$$rd = \frac{x_n - \bar{x}}{\delta_x} \sqrt{\frac{n}{N_d - n}}$$

где \bar{x} — среднее арифметическое значение всех индивидуальных оценок по тесту, x_n — среднее арифметическое значение оценок по тесту у тех испытуемых, которые правильно решили задачу, δ_x — среднеквадратическое отклонение индивидуальных оценок по тесту для выборки, n — число испытуемых, правильно решивших задачу, N_d — общее число испытуемых, давших ответ на задачу,

Эта формула рекомендуется для расчета коэффициента дискриминации лишь в тех случаях, когда все испытуемые дали хоть какие-то ответы на все задачи теста. Случай, когда не все испытуемые дали ответ на поставленные задачи нами уже обсуждался. Расчет в этом случае производится следующим образом: в формуле используются показатели среднего арифметического значения

индивидуальных оценок только тех испытуемых, которые дали ответ на задачу, также рассматривается и среднеквадратическое отклонение.

Коэффициент дискриминации задачи может принимать значение от -1 до +1. Высокий и значимый положительный коэффициент является показателем того, что задача хорошо разделяет испытуемых с высокими и низкими оценками по тесту. Высокий, значимый отрицательный коэффициент свидетельствует о непригодности задачи для теста. Если значение коэффициента близко к 0, то задачи должны рассматриваться как не корректно сформулированные.

Определение дискриминативности по Фергюсону.

Значительно более простой способ расчета дискриминативности предложен и обоснован Фергюсоном. Его идея основана на том, что показатель дискриминативности связан по существу с ранжированием испытуемых. Для определения дискриминативности вводится отношение между показателями любых двух испытуемых. Эти показатели могут быть равны или отличаться. Общее количество возможных отношений между парами испытуемых, если испытуемых N , равно $n \cdot (n-1)$. Далее рассчитывается общее количество равенств среди этих пар и общее количество отношений различия между парами, которые равны между собой. Находя соотношение между этими величинами, получаем коэффициент Фергюсона — отношение между показателем дискриминативности, полученным для некоторого теста и максимальным значением дискриминативности, который он может обеспечить. Упрощенная формула для расчета выглядит так:

$$F = \frac{(n+1)(N^2 - f^2)}{nN^2}$$

где n — количество заданий,

N — количество испытуемых,

f — частота значения каждого показателя.

Если $F = 0$, все испытуемые получили одинаковые результаты, а при равномерном распределении $F = 1$.

Индекс дискриминации вычисляется по отношению к внешнему критерию — оценке продуктивности реальной деятельности испытуемых (эффективности обучения, академической успеваемости и т. д.). Проблему внешнего критерия мы уже обсуждали. Индекс вычисляется с применением метода крайних групп, т. е. при расчете учитываются результаты наиболее и наименее успешно справившихся с заданиями. Как правило, берут по 27% тех и других. Индекс дискриминации задачи вычисляется как разность долей испытуемых из высокопродуктивной и низкопродуктивной групп, правильно решивших ее, и обозначается D :

$$D = (N_{\text{рверх}}/N_{\text{верх}}) - (N_{\text{рниз}}/N_{\text{низ}})$$

где N_p — количество справившихся с задачей испытуемых,

N — общее количество испытуемых в крайних группах.

Полученные статистические данные позволяют определить трудность задачи и ее пригодность для использования в гесте.

Как один из возможных путей определения недостатков задач используется метод анализа предложенных заданий совместно с испытуемыми, вошедшими в выборку. В ходе этого анализа исследователь выясняет причины ошибок, допущенных в задачах, причины высокого процента неправильных ответов лиц из числа высокоуспешных и правильных — из низкоуспешных.

Типичными недостатками задач, оказавшихся непригодными, являются:

1. излишняя сложность, запутанность формулировки;
 2. неоднозначность условия;
 3. очевидность решения;
 4. зависимость результата от памяти или от других индивидуальных особенностей испытуемого, а не от уровня развития тех умений и навыков, для оценки которых разрабатывается тест (кроме заданий, где необходима именно работа памяти) ;
 5. абсурдность, нереальность вариантов ответов;
 6. появление двух и более правильных ответов, не оговоренное в условии.
- Только после учета всех указанных условий созданная задача может быть включена в тест.

5.6. Надельность и валидность тестов.

5.6.1. Надежность.

Надежность теста, является одним из критериев качества теста и показывает, насколько точно измеряет данный тест изучаемое явление, его «помехоустойчивость».

Надежность теста, как правило, определяется после окончания анализа задач и составления окончательной формы теста.

В психологии термин «надежность» применяется в двух значениях. Тест называется надежным, если он является внутренне согласованным. Также тест называется надежным, если он дает одни и те же результаты для каждого испытуемого при повторном тестировании, такая надежность называется, ретестовой.

Надежность характеризует точность теста как измерительного инструмента, устойчивость его к действию помех (состояния испытуемых, их отношения к процедуре тестирования и т. п.) - Качественный тест не может быть создан без тщательного изучения этого важного аспекта измерения. Использование ненадежных тестов, допуск большого количества ошибок в *каком* ответственном деле, каким является тестирование людей, может стать причиной педагогических и административных ошибок, последствия которых трудно исправить.

Ошибки при тестировании.

Как и любая человеческая деятельность, тестирование допускает наличие ошибок. При проведении тестирования возникают следующие виды ошибок: промахи, систематические и случайные ошибки.

Промахи возникают при грубых нарушениях процедуры тестирования. У опытных экспериментаторов промахи встречаются довольно редко. Они могут быть легко выявлены и устранены путем анализа резко отклоняющихся значений.

Систематические ошибки измерения возникают постоянно или закономерно меняются от измерения к измерению. В силу этих своих особенностей они могут быть предсказаны заранее, а в некоторых случаях и устранены.

Случайные ошибки имеют место, когда при последовательных измерениях постоянной характеристики получаются различные числовые оценки, т. е. при определении случайных ошибок предполагается, что измеряемая характеристика не изменяется во времени, а все отклонения обусловлены неточностью измерения.

Граница между случайными и систематическими ошибками достаточно условна. Суммарная ошибка любого теста состоит из ошибок двух этих типов, но в разном соотношении. Относительный вес каждого типа ошибок зависит от качества теста и условий проведения обследования. Результаты тестирования всегда содержат ошибки, как бы тщательно они ни проводились. Однако используя методы математической статистики, можно оценить величину суммарной ошибки тестирования и использовать ее для оценки надежности теста. Без статистической оценки надежности теста результаты измерения нельзя считать достоверными.

Ретестовая надежность.

Для математического определения ретестовой надежности теста необходимо применение используемого в психологических и педагогических исследованиях понятия дисперсии. Не вдаваясь в математические тонкости отметим, что дисперсия — показатель меры разброса величины около среднего значения, более подробно о дисперсии и способе ее расчета можно прочитать в главе 6. Надежность может быть определена как отношение дисперсии истинной компоненты к дисперсии тестирования, которая, в свою очередь, зависит от дисперсии ошибок. В своем аналитическом выражении надежность (или коэффициент надежности) может быть определен как квадрат коэффициента корреляции результатов тестирования с истинными результатами. Сам коэффициент корреляции в этом случае выступает в качестве индекса надежности.

Таким образом, под надежностью понимается степень сходства истинной компоненты с наблюдаемым при тестировании результатом. Что это означает? Каждый участник занимает определенное место на шкале результатов первичного тестирования. Эти данные принимаются за истинный показатель. Теоретически это место должно быть постоянно. Повторное выполнение теста теми же людьми в тех же условиях должно дать такое же распределение мест на шкале, как и в первый раз. В таком случае тест считается абсолютно надежным. Естественно, это идеальная ситуация, на практике и оценки и места испытуемых в результате повторного тестирования часто изменяются.

Получаемые отклонения повторных оценок являются основой для расчета коэффициента надежности.

Величина коэффициента надежности определяется величиной изменения оценок испытуемых на шкале в результате вторичного тестирования по сравнению с первичным. Следовательно, чем ближе оценки первичного и вторичного тестирования, тем выше точность измерения и, соответственно, надежность теста как измерительного инструмента. На практике используются три основных метода оценки надежности тестов:

1. повторное тестирование (ретестирование);
2. параллельное тестирование (тестирование параллельной формой теста);
3. расщепление (метод деления целого на части).

Рассмотрим каждый из них в отдельности.

Метод повторного тестирования (ретестирование) является основным при определении надежности психологических тестов. Однако применение его к гестам достижений ограничено. Этот метод предусматривает повторное тестирование через некоторый промежуток времени. Однако за это время дети успевают подрасти, узнать что-то новое, иногда забыть известное. То есть высокая динамика изменений объекта измерения ограничивает применение метода повторного тестирования для тестов школьных достижений.

Как мы уже отмечали, этот метод заключается в проведении через некоторый интервал времени повторного тестирования тех же испытуемых теми же заданиями. Надежность теста вычисляется как мера сохранения соотношений оценок испытуемых от первого ко второму испытанию, т. е. выявляется характер корреляции между результатами двух проверок.

Метод повторного тестирования обладает как достоинствами, так и недостатками. К числу достоинств относятся естественность и простота определения коэффициента надежности. Основным недостатком метода является неопределенность в выборе интервала между двумя тестированиями. Это связано с тем, что повторное тестирование отличается от первого.

Испытуемые уже знакомы с содержанием теста, они могут помнить свои первоначальные ответы и ориентироваться на них при повторном выполнении. Поэтому при повторном тестировании нередко наблюдается или «подгонка» под первоначальный результат, или демонстрация абсолютно новых решений как проявление испытуемым негативизма. В результате на месте старых возникают новые ошибки.

Для того, чтобы максимально снизить возможность возникновения новых ошибок, рекомендуется проводить ретестирование через достаточно длительные интервалы времени.

Разные исследователи рекомендуют проводить повторное тестирование в интервале от двух недель до 6 месяцев. Если такой возможности нет, то лучше отказаться от ретестирования и использовать другие методы определения надежности.

Существует несколько условий, влияющих на величину коэффициента надежности, при выявлении хотя бы одного из которых от метода ретестирования необходимо отказаться:

- запоминаемость содержания задач и способов их решения;
- интересность и оригинальность задач;
- небольшое количество задач;
- небольшое время между первым и вторым проведением теста.

Снижение коэффициента надежности при увеличении времени ретестирования говорит о функциональной неустойчивости теста.

Второй метод — метод определения надежности при тестировании параллельной формой теста — наиболее приемлем для тестов достижений.

Параллельными считаются такие тесты, которые измеряют одно и то же свойство с одинаковой ошибкой. Многие тесты имеют параллельные формы, направленные на измерение одного и того же свойства или качества, что для тестов школьных достижений полезно по целому ряду обстоятельств: при наличии параллельной формы упрощается процедура проведения, появляется возможность замены задания или групп заданий и т. д. Важнейшими признаками параллельности тестов являются следующие:

1. Одинаковая средняя величина результатов измерения. Это означает, что характеристика среднего уровня группы тестируемых не зависит от используемой части теста, и оценки по тесту зависят только от представленности изучаемого свойства у испытуемых.
2. Параллельные тесты имеют одинаковую дисперсию результатов измерения, т. е. позволяют получить одинаковую, стандартную оценку результатов измерения.
3. Истинная дисперсия (обусловленная только индивидуальными различиями) равна ковариации результатов тестирования параллельными формами (37).
4. Параллельные тесты имеют одинаковые коэффициенты надежности.

Хорошо составленные формы теста взаимозаменяют друг друга. Каждая форма должна обеспечивать испытуемому одно и то же место на шкале выраженности признака.

Важно помнить, что эквивалентность тестов, в первую очередь, определяется психологическим и педагогическим анализом, экспертными суждениями специалистов и только в дополнение к этому — статистическими критериями.

Параллельные формы тестирования рекомендуется применять в тех случаях, когда либо способ решения одной задачи • не может быть прямо использован для решения другой, либо влияние научения в процессе решения задач невелико, либо же задачи не уникальны. Тестирование параллельной формой может проводиться через минимальный интервал времени.

Процедура применения параллельных форм теста такова. Испытуемые делятся на две группы. Проводить это разделение можно либо случайным образом (по алфавиту, номеру, случайным числам и пр.), если число испытуемых достаточно велико, если же число испытуемых небольшое (менее 200), то лучшим способом деления будет деление в зависимости от успешности деятельности, которая определяется по внешнему критерию (например, отобрать из группы испытуемых тех, кто имеет школьную оценку 5 и эту группу поделить пополам случайным образом, затем отобрать группу, которая имеет оценку 4 и разделить пополам эту группу, таким же образом поступить с

остальными группами- Возможен и другой способ: всю группу ранжировать по внешнему критерию и уже по этому списку разделить группу пополам по четным и нечетным номерам). Затем одной группе предлагается форма А теста, а другой — форма В. Через несколько дней, но не больше чем через неделю, проводится второе тестирование, причем задания меняются по подгруппам. Результаты сравниваются с помощью расчета коэффициента корреляции. • Метод расщепления, или метод деления теста на части, возник как результат развития метода параллельного тестирования и основывается на предположении о параллельности не только форм теста, но и отдельных заданий внутри теста. Он применяется в тех случаях, когда повторное тестирование нежелательно (тесты достижений, некоторые интеллектуальные или творческие тесты) и существует только одна форма теста. Самый простой вариант деления теста — на две части. При этом распределение оценок за выполнение всего теста в целом должно быть близко распределению оценок при выполнении частей теста.

Важнейшим условием объективности этого метода является деление теста на равноценные части. Разделить тест на две части возможно несколькими способами: деление по принципу «чет-нечет»; деление задач теста на основе одинаковых коэффициентов сложности и дискриминации; деление задач в соответствии со временем решения каждой части (в тестах скорости). Однако только определения надежности теста не достаточно для того, чтобы признать тест готовым к употреблению.

Надежность по внутренней согласованности.

Окончательный вариант теста должен соответствовать требованию внутренней согласованности. Внутренняя согласованность — это характеристика теста, указывающая на степень однородности состава заданий с точки зрения измеряемого качества (т. е. все задания теста должны определять тестируемое качество). Для ее определения применяется процедура установления корреляций между результатами теста в целом и каждым отдельным заданием. Например, такая рассогласованность возможна, при использовании заданий на выявление структуры интеллекта для тестов достижений. Мы уже говорили, что такие задания нагружены двумя факторами. Например, в случае использования заданий аналогий, если создается тест на выявления владения какими-либо знаниями, а уровень аналогии, заложенный в эту форму заданий слишком сложен, то именно он будет определять успешность выполнения задания. То есть данное задание может быть не выполнено не в силу отсутствия того качества, которое тестируется (знание материала), а в силу неумения выявить аналогию.

Однако существует ряд сложностей, связанных с понятием этого вида надежности. Высокая внутренняя согласованность может быть противоположностью валидности, как утверждает Кеттелл. Поскольку валидность теста определяется как степень измерения того, для чего он предназначен, то достижение ее должно быть основной целью для разработки тестов. Надежность важна лишь постольку, поскольку она может обеспечивать высокую валидность. Мы не будем дальше обсуждать эту проблему, поскольку

она имеет большее практическое значение лишь для сложных психологических тестов. Подробнее с этим вопросом можно познакомиться в литературе (Н. ф. Бурлачук, П. Клаин и др.).

Мы приведем источники неудовлетворительной надежности тестов, кратко мы их уже касались.

1. Наличие в тесте заданий, которые допускают неоднозначное толкование результатов. Часто к таким заданиям относятся задания со свободными ответами.

2. Угадывание ответов.

3. Величина теста. Чем тест длиннее, тем он надежнее.

4. Инструкции и технология проведения. Мы еще коснемся этого вопроса. При помощи инструкций легко изменить трудность задания. Неоднозначные инструкции приводят к снижению надежности. Повышение надежности возможно двумя путями — ужесточением инструкций и повышением качества подготовки экспериментаторов.

5- Источники, связанные с испытуемыми. Усталость, скука, невнимательность (например, испытуемый может перевернуть две страницы, записать ответ не в той графе), жара или холод, самочувствие, различная мотивация и т. д.

Валидность и надежность — связанные понятия. В литературе мы находим различные примеры, иллюстрирующие эту связь. Вот один из них. Допустим, имеются два стрелка; А и В. Стрелок А выбивает 90 очков из 100, а стрелок В — только 70. Соответственно, надежность стрелка В — только 0.70. Однако стрелок А всегда стреляет по чужим мишеням, поэтому на соревнованиях его результаты не засчитываются, Второй стрелок всегда правильно выбивает мишени. Поэтому валидность стрелка А нулевая, а стрелка В — 0.70, т. е. численно равна надежности. Если стрелок А станет правильно выбирать мишени, его валидность тоже будет равна его надежности. Если же он будет продолжать иногда путать мишени, то часть результатов не будет зачтена и валидность стрелка А будет ниже надежности. В этом примере аналогом надежности является меткость стрелка, а аналогом валидности — точность стрельбы по строго определенной «своей» мишени. Поэтому надежный стрелок может стать валидным не только сменив мишень, но и сменив команду. В истории тестологии известны случаи, когда тест, признанный инвалидным для измерения одних свойств, признавался валидным по отношению к другим. ненадежный тест не может быть валидным, и, наоборот, валидный тест всегда надежен (37, с. 180)

5.6.2. Валидность.

Понятие валидность очень часто вызывает путаницу не только среди педагогов, но и среди психологов. Причины этой путаницы, которая возникла в силу исторических, лингвистических и прочих причин, кроются в особенностях того, что этим термином именуется, а именно несколько показателей, часть которых имеет между собой достаточно условную связь; то есть по одному показателю

валидности тест может быть вполне удовлетворительным, а по другому — негодным. При этом отсутствует возможность получения интегральной оценки в силу разницы тех показателей, которые называются валидностью.

В общем виде понятие валидности раскрывает, насколько полученные в результате тестирования результаты соответствуют объективной реальности, а поскольку отсутствует способ точного определения объективной реальности, то валидность — понятие относительное, поскольку тесты, как и любой измерительный инструмент, никогда не являются абсолютно эквивалентными реальному положению вещей. Соответственно, валидность устанавливается относительно характеристики, признака, величины, принимаемой (а не истинно являющейся) за объективную.

В современной тестологии выделяются следующие виды валидности:

1. очевидная валидность;
2. валидность по содержанию (содержательная);
3. конструктивная (концептуальная);
4. валидность по критерию (критериальная или эмпирическая валидность);
5. сравнительная или конкурентная валидность.

1. Очевидная валидность.

Очевидная валидность — это валидность с точки зрения испытуемого. Она определяет, насколько адекватно впечатление о предмете измерения, которое формируется у испытуемого при знакомстве с инструкцией и материалом теста, тому, что данный инструмент в действительности выявляет. Очевидная валидность (иногда называется доверительной валидностью) играет важную роль в процессе тестирования, поскольку именно она определяет отношение испытуемого к обследованию. Так, тест может не вызывать доверия, если в нем отсутствуют знакомые, традиционно используемые испытуемым термины, понятия, слова, определения. Известны случаи, когда отказ от выполнения теста обосновывался именно не соответствием используемой в тесте терминологии реально существующей.

Очевидная валидность может быть присуща тесту, а может и не быть, Это зависит от целей исследования. Для большинства тестов достижений очевидная валидность — позитивное либо нейтральное качество, поскольку нет достаточных оснований скрывать от испытуемых реальные цели тестирования. Для повышения очевидной валидности достаточно привлечение испытуемых для анализа теста и заданий: на основе их оценок можно внести некоторые коррективы.

2. Валидность по содержанию (содержательная).

Содержательная валидность устанавливается экспертами для деятельности близкой или совпадающей с реальной. Очевидная валидность может находиться в разных отношениях с содержательной: в одних случаях они совпадают, в других — внешняя валидность используется для маскировки содержательной валидности (чтобы скрыть истинную цель обследования для предупреждения возникновения установочного поведения при тестировании).

Определение содержательной валидности — основное для тестов достижений и тестов профессиональной успешности, когда должен быть точно определен материал, применяемый для тестирования, и когда существует достаточная ясность смысла измеряемого параметра. Например, если мы хотим определить умение сложения простых чисел, то задания $4+8=?$ или $42+16=?$ вполне подходящи, и трудно оспаривать их валидность.

Очевидно, что содержательная валидность будет полезна только тогда, когда могут быть определены специальные навыки и особенности поведения. Это можно сделать довольно элементарном уровне, при тестировании арифметических навыков (правил выполнения четырех арифметических операций, правил вычислений с 0 и т.п.), знаний в области искусства (знает ли испытуемый правила нотной записи, принципы архитектуры, и др.), а также знаний базовых элементов для большинства научных дисциплин, в которых накоплен багаж фактических данных. Однако, определить содержание теста таким образом для более высокого уровня знаний становится чрезвычайно трудно. Именно поэтому понятие «содержательная- валидность» наиболее уместно для тестов достижений.

В отличие от очевидной валидности, содержательная валидность всегда связана с оценкой отношения содержания тестового материала к истинной, а не мнимой цели измерения. Содержательная валидность определяется на основе экспертных методов. Эксперты оценивают задачи теста по шкале близости к реальным задачам, встающим перед человеком в процессе обучения или работать, и целям предполагаемого тестирования.

Первый этап экспертного оценивания может быть связан с отбором существенно важного материала на основе которого должны строиться задания. На первом этапе подбираются эксперты в данной области. Для тестов школьных достижений иногда необходимо привлечение экспертов нескольких специальностей: специалистов в области содержания предмета или науки и специалистов-методистов, особенно часто приходится это делать при разработке тестов по нетрадиционным предметам преподавания или области знания, где не удастся привлечь специалистов с достаточным уровнем подготовки по методике преподавания. Отбор экспертов — это отдельная задача, решение которой разработано соответствующей областью психологии и методологии исследования, и мы на ней останавливаться не будем. Однако необходимо отметить недопустимость отбора экспертов по должности, такой подход может сыграть очень злую шутку.

Окончательный вывод о содержательной валидности делается из обобщения суждений отдельных экспертов, которые они дают по реально составленным заданиям теста. В простейшем случае это может быть процент заданий, признанных группой экспертов соответствующими данному виду деятельности. П, Клайн предлагает следующую процедуру для определения содержательной валидности для тестов достижений:

- 1) Укажите точно категорию лиц, для которой предназначен тест.
- 2) Определите навыки, подлежащие тестированию; возможно, вам потребуется их проанализировать. Составьте список.

- 3) Передайте этот список экспертам в данной области (учителям и т. п.) для проверки нет ли упущений.
- 4) Преобразуйте этот список в перечень заданий, используя, когда это возможно, равное количество заданий на каждый навык.
- 5) Представьте эти задания экспертам для проверки.
- 6) Подвергните задания обычным процедурам конструирования тестов. В результате должен быть получен содержательно валидный тест.

3. Конструктивная (концептуальная) валидность.

Конструктивная валидность непосредственно для тестов достижения не используется. Этот вид валидности определяется в тех случаях, когда представление об измеряемом психическом феномене существует только в сознании исследователя. К такому роду конструктов (феноменов) относятся черты характера, темперамента, способности и интеллект. Разработчик теста может лишь строить гипотезу о существовании данного конструкта, о его формах и характере проявления. Устанавливается концептуальная валидность путем доказательства правильности теоретических концепций, положенных в основу теста, это особенно необходимо в тех случаях, когда результаты тестовых измерений используются не просто для предсказания поведения, а как основа для выводов о том, в какой степени испытуемые обладают некоторой характеристикой.

В. М. Мельников и Л. Т. Ямпольский предлагают проводить проверку концептуальной валидности через три основных этапа:

1. определение некоторой теоретической концепции, которая предположительно объясняет выполнение валидизируемого теста;
2. из теоретической концепции выводятся одна или несколько гипотез, связанных с тестом;
3. выдвинутые гипотезы подвергаются эмпирической проверке.

Если эмпирические данные подтверждают гипотезу, то тем самым подтверждается психологическая концепция, положенная в основу теста, и способность теста служить инструментом измерения данного конструкта. Чем убедительнее эмпирическое подтверждение, тем определеннее можно говорить о валидности теста по отношению к психологической концепции, положенной в его основу. Ошибки при определении валидности могут возникнуть как следствие неправильной теоретической концепции, положенной в основу теста, или отсутствия соответствия между тестом и теоретической концепцией, или ошибочного выдвижения гипотез,

4. Валидность по критерию (критериальная или эмпирическая валидность).

Основным видом валидности является валидность по критерию или эмпирическая валидность. Суть ее заключается в определении способности теста служить индикатором или предсказателем строго определенной психической особенности, формы поведения человека и др.

Валидизация теста по критерию состоит в сравнении результатов, полученных испытуемыми за решение теста с данными по критерию и вычислении коэффициента корреляции тестового результата с внешним критерием.

Например, школьный тест умственного развития (ШТУР) валидизировался на основе критерия школьной успеваемости — оценок детей по предметам в образовательном учреждении. В качестве критерия может выступать любой показатель, независимо и бесспорно измеряющий ту же психологическую характеристику, что и валидизируемый тест.

Вопрос о выборе критерия является в этом виде валидности основным. От выбора критерия зависят качественная и количественная оценка валидности. Критерий валидности теста — это признак его обоснованности, правомерности его применения как измерительного инструмента.

На практике в качестве критерия валидности обычно используются социально-демографические и биографические данные (стаж, образование, поступление или не поступление в учебное заведение, прием на работу или увольнение); производственные, научные, учебные показатели эффективности деятельности (число публикаций, оценки академической успеваемости); экспертные оценки выраженности черт личности, деловых качеств, успешности работы или учебы; результаты выполнения стандартизованных методик и др.

Помимо этих «жизненных*» критериев при составлении и апробации тестов используются разнообразные критерии, классифицировать которые можно по следующим основаниям:

— по объективности: субъективные и объективные;

— по метричности: качественные и количественные;

— по составу: одиночные и составные;

— по широте охвата явления: узкие и широкие. Важнейшей характеристикой критерия валидности считается его объективность. Объективность критерия валидности состоит в его независимости от субъективного мнения, оценок г. п. Для разных видов теста используются свои критерии объективности. Для тестов достижений наиболее актуальными являются более жесткие, однозначные, связанные с продуктивностью конкретной деятельности,

К субъективным критериям валидности относятся оценки, суждения, заключения об объекте (предмете, процессе, человеке), сделанные одним экспертом. В этом случае применяется метод индивидуальной оценки. Эксперт оценивает объект, используя предложенную составителем теста оценочную шкалу (5—, 7— или 9-балльную). Однако в тестологической практике этот метод встречается не очень часто. В серьезных исследованиях метод индивидуальной оценки использовать не рекомендуется. Объективность критерия валидности усиливается с увеличением числа экспертов. При этом выбор оценочной шкалы осуществляется по принципу привычности для эксперта. Так, учителю удобнее оценивать учащихся по привычной пятибалльной шкале.

Наибольшее распространение нашли такие способы экспертной оценки, как метод коллективной оценки, метод средневзвешенной оценки, метод

ранжирования и метод парного сравнения. Рассмотрим несколько подробнее эти способы.

Метод коллективной оценки предполагает участие нескольких экспертов (3—10), которые совместно оценивают объект с помощью определенной оценочной шкалы. Необходимым условием коллективной оценки является выработка общего, возможно, даже компромиссного мнения, общей оценки. Однако этот вид получения оценок ограничивается сложностью процедуры: необходимостью собрать в одном месте для длительной коллективной работы квалифицированных экспертов.

Чаще всего используется метод средневзвешенной оценки. В соответствии с этим методом каждый испытуемый оценивается по заранее установленной шкале несколькими экспертами. При этом оценки, данные одним экспертом, не известны другим. Затем оценки разных экспертов одного объекта усредняются. Такая усредненная оценка является более надежной, чем полученная другими способами.

Если выборка испытуемых меньше 30 человек, можно использовать метод ранжирования. Группе экспертов предлагается распределить (упорядочить) всех членов выборки в соответствии с заданным признаком. Усредненное ранговое место, может служить критерием для определения валидности теста. Если же в тестировании участвует более 30 человек, то для применения метода ранжирования можно распределить обследуемых на подгруппы из 10—15 человек по случайному признаку и упорядочить каждую группу. Дальнейшая обработка производится так же, как и в предыдущем случае.

Для повышения качества экспертного оценивания и при низкой квалификации экспертов используется метод парного сравнения. Работа экспертов состоит в попарной расстановке объектов по признакам сильнее/слабее, лучше/хуже и т. п. При этом каждый объект сравнивается с каждым по выбранному признаку. Этот метод более пригоден для малых групп оцениваемых объектов, так как уже для 30 объектов число сравниваемых одним экспертом пар составляет 435, что предъявляет особые требования к работоспособности экспертов и является недостаточно экономичным.

Автор при валидации тестов школьных достижений для учеников 10--го класса использовал метод парных сравнений, когда в качестве экспертов выступали педагоги, работающие с детьми данного класса. Им предлагалось оценить детей по четырем показателям: 1 — знания по предмету, 2 — школьные специальные навыки, 3 — психические познавательные процессы, 4 — уровень социализации.

При этом применялась шкала:

0 — качество выражено слабее

1 — качество выражено в равной степени

2 — качество выражено сильнее.

Был сконструирован бланк для оценок. Каждый эксперт заполнял по 4 бланка.

	1	2	3	4	...	Сумма
1. Кострова	X	2	1	1		4

2. Уварова	0	X	0	2		2
3. Кочкарева	1	2	X	2		5
4. Егорова	1	0	0	X		1
					X	

После заполнения просчитывалась сумма баллов и испытуемые ранжировались. После проведения тестирования просчитывался коэффициент ранговой корреляции, по которому и определялась валидность теста.

Попутно необходимо заметить, что учителя оказались слабодискриминативными экспертами. Корреляция их оценок по первым трем качествам оказалась чрезвычайно высокой от 0.79 до 0.92, что, в свою очередь, позволило при дальнейшей работе использовать только одну учительскую оценку. При этом встал вопрос поиска и отбора дополнительных критериев. В качестве объективных критериев используются результаты работы и их изменения во времени, наличие или отсутствие заболевания, признака и др. Качественные критерии дают представление о том, к какому классу (категории) относится данный объект или имеет ли он определенный признак, в этих случаях объекты располагаются в соответствии с законами номинативной шкалы.

Использование одиночных и составных критериев валидности определяется тем, сколько показателей принимается за основу. Одиночные критерии валидности представляют собой качественную специфическую характеристику, отражающую одно явление или признак. В реальной действительности характер явления обуславливается действием не одной, а нескольких причин. Поэтому для адекватной оценки явления, как правило, используются составные критерии валидности, предполагающие учет комплекса показателей. Природа изучаемого явления определяет и широту используемых критериев. Более узкие критерии применяются при обосновании тестов на измерение уровня развития конкретных знаний, умений и навыков. Их узость связана с узостью области объективной реальности, которую тест охватывает, с однородностью задач теста и конкретностью, однозначностью интерпретации результатов. Чем сложнее измеряемое явление, тем шире и обобщеннее должен быть критерий валидности. Такие критерии используют, например, в тестах интеллекта.

5. Сравнительная или конкурентная валидность.

Этот вид валидности рассчитывается по корреляции результатов выполнения, данного теста с результатами других тестов, которые доказательно

предназначены для измерения того же качества или свойства. Может показаться, что эта валидность не имеет большого смысла. Действительно, зачем создавать новый инструмент, когда уже есть готовый тест для определения какого-либо качества? На самом деле существуют два обстоятельства, при которых расчет этого вида валидности необходим. Первое — когда новый, создаваемый инструмент качественно отличается от существующего — более прост, технологичен для обработки и проведения и т. д. Второе — для нас более важное и применимое к тестам школьных достижений: при создании двух- и многомерных тестов на одну из шкал вполне могут оказаться готовые высококачественные тесты. Например, шкала интеллекта в тесте ШТУР проверялась на сравнительную валидность с тестом интеллекта Р. Амтхауера.

Для эффективного изучения конкурентной валидности существует несколько правил, представленных П. Клайном (приводятся с сокращением в части использования факторного анализа) .

1) Убедитесь, что выборка испытуемых отражает ту категорию лиц (популяцию), для которой данный тест предназначен, особенно по отношению к полу, возрасту, уровню образования и социальному положению.

2) Убедитесь, что выборки достаточно велики для получения статистически значимых корреляции, могущих быть затем использованными в факторном анализе. Минимальное количество испытуемых — 200.

3) Используйте настолько широкое разнообразие других тестов данной переменной, насколько возможно, чтобы убедиться, что корреляция получена благодаря близости групповых, а не специфических факторов.

Исследования конкурентной валидности, удовлетворяющие этим критериям, должны дать недвусмысленное свидетельство валидности, которое не может быть опровергнуто.

Глава 6. Нормирование тестов

6.1. Выборка.

1.1. Варианты расчета объемов выборки.

Корректное построение выборки — основа для стандартизации: все зависит от выборки заданий для теста и выборки репрезентативной выборочной совокупности. При формировании выборки следует учитывать две важные переменные: объем и репрезентативность. Выборка должна точно отражать категорию лиц, для которых предназначен тест (конечно, может быть несколько таких категорий и, следовательно, и несколько "выборок или же одна большая выборка), а также быть достаточной большой для обеспечения столь малой стандартной погрешности нормативных данных, чтобы ею можно было пренебречь. Ниже приведены расчеты объемов выборки для всех участников педагогического процесса, а не только учеников.

Объем выборки — число единиц наблюдения, включаемых в выборочную совокупность.

Что определяет объем выборки, и от чего он зависит? В общем виде объем выборки определяется следующими четырьмя обстоятельствами:

1. Задачи и условия проведения исследования.

1.1. Получение экспресс-информации, фундаментальное исследование, необходимость получения прогноза по результатам исследования требуют разной степени точности и определяют точность и доверительную вероятность результатов проводимого исследования.

1.2. Объем выборки зависит от разрешающей величины шкалы изучения (дихотомическая, процентная, семантический дифференциал) и требует учета при выборе способа расчета объема выборки.

1.3. От условий проведения зависит способ определения объема выборки, при этом возможны три случая:

— об измеряемых величинах есть статистическая информация по предыдущим исследованиям;

— будет проведено пробное исследование;

— о характеристиках статистических величин ничего не известно.

В последнем случае возможно получение оценок, необходимых статистических данных. Дисперсия может быть оценена по правилу «шести сигм» (4, с. 66):

$$\sigma = \frac{X_{\max} - X_{\min}}{6} = \frac{R}{6}$$

где X_{\max} — максимальное значение шкалы измерения,

X_{\min} — минимальное значение шкалы.

Остальные необходимые статистические данные могут быть либо заданы, либо оценены как максимально теоретически возможные. Таким образом задача расчета объема выборки в этом случае может быть сведена к задаче с известными статистическими данными.

2. Степень однородности генеральной совокупности.

Объем выборки определяется разбросом измеряемого показателя в генеральной совокупности, выражаемым дисперсией или стандартным отклонением.

3. Вероятность, с которой гарантируется достоверность результата.

Определяется условиями исследования. Обычно в практике исследований ее величина колеблется от 85% до 99%. Наиболее часто используется вероятность 95% (0,95).

4. Точность результатов, определяемая предельной ошибкой репрезентативности.

Предельная ошибка репрезентативности задается обычно в пределах от 0,01 до 0,10 с наиболее частым употреблением 5% (0,05).

Определяющим фактором является первое обстоятельство § (задачи исследования), но объем выборки зависит также от остальных трех обстоятельств (степень однородности генеральной совокупности, вероятность, с которой гарантируются результаты, предельная ошибка репрезентативности).

Дадим обозначения нескольким известным статистическим понятиям, без которых дальнейшее объяснение затруднено:

P — вероятность,

n — объем выборки,

Δ — предельная ошибка репрезентативности,
 σ^2 — дисперсия,
 σ — стандартное отклонение,
 N — объем генеральной совокупности,
 t — табулированная константа, смысл которой определим несколько позже (см. расчеты по формулам),
 $\alpha = 1 - P$ - уровень значимости.

Общий подход к расчету объема выборки.

Основная величина, от которой зависит величина объема (выборки) — дисперсия и ее поведение (изменение). При выборочном методе исследования дисперсия изучаемого признака складывается из двух составляющих — дисперсии генеральной совокупности (собственно мера рассеяния признака как такового) и дисперсии, вызванной неоднородностью генеральной совокупности. При малых объемах выборки обе составляющие вносят в нее свой вклад, при этом вторая составляющая может как увеличивать, так и уменьшать дисперсию. С увеличением выборки, составляющая, вызванная неоднородностью генеральной совокупности уменьшается, значение дисперсии при этом перестает изменяться, при этом дальнейшее увеличение

объема выборки нецелесообразно, поскольку ошибка выборки становится пренебрежительно мала,

Графически это может быть выражено следующим образом:

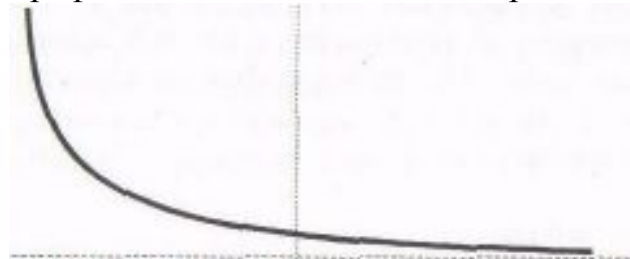


Рис. 1 . График зависимости дисперсии от объема выборки. N^{\wedge} — оптимальный объем выборки.

Способы расчета объема выборки.

ЗАМЕЧАНИЕ. Объем выборки, рассчитываемый в данной части, относится к:

1. чисто случайной выборке (безотносительно к способу построения выборки — по таблицам случайных чисел, периодическому и пр.);
2. неповторному случаю — один и тот же испытуемый в одном исследовании дважды не тестируется. Это обусловлено тремя обстоятельствами:
 - а) объем выборки для чисто случайного способа отбора не является максимальным, однако в практике применяется только этот способ;
 - б) объем выборки для доли и среднего рассчитывается одинаковым способом, считая признаки альтернативными;
 - в) расчеты более сложных выборок уменьшают объем выборки и их проведение подробно описано в литературе (53, 46, 38, 27).

Степень близости выборочной средней к генеральной средней при постоянной дисперсии при повторном отборе зависит только от объема выборки, а не от удельного веса выборочной совокупности в генеральной. Объем выборочной совокупности не зависит от объема генеральной (46, с. 252).

Объем выборочной совокупности рассчитывают четырьмя основными способами.

1. По изменению дисперсии.
2. По таблицам достаточно больших чисел и номограммам, больших чисел.
3. Эмпирически по среднему квадратичному отклонению.
4. По формулам математической статистики, Приведем варианты расчета объемов выборки всеми четырьмя способами из одинаковой генеральной совокупности.

1. По изменению дисперсии.

Расчет объема выборки по изменению дисперсии вытекает из общего подхода к расчету объема выборки. Применяется, когда известны значения статистик, характеризующих различные объемы совокупностей.

Для примера изучения образовательной системы Санкт-Петербурга взяты следующие совокупности: учителя, родители, учащиеся. Расчеты проведены для двух вопросов в каждой группе — один носит оценочный характер по отношению к образовательному заведению, второй — самооценка.

В таблице представлены простые статистики для разных выборок совокупностей для наиболее употребительной пятиразрядной шкалы.

Полужирным шрифтом в таблицах выделены значения дисперсии, статистически равнозначные, и обозначен момент с которого дальнейшее увеличение выборки не уменьшает дисперсию. Таким образом для данных вопросов объемы выборки могут быть определены в 150—350 единиц.

Глава 6. Нормирование тестов

Таблица 2.

категория	объем выборки	дисперсия вопрос 1	дисперсия вопрос 2
Учителя	100	0.35	0.54
	150	0.51	0.55
	200	0.44	0.67
	300	0.43	0.63
	400	0.44	0.62
	500	0.44	0.62
	600	0.43	0.62

Таблица 3.

категория	объем выборки	дисперсия вопрос 1	дисперсия вопрос 2
Ученики	50	1.37	1.86
	100	1.25	1.81
	200	1.03	1.89
	400	1.01	1.95
	500	1.00	1.93
	700	1.00	1.93

Таблица 4

Таблица 4

p	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02
0.90	67	83	105	138	187	270	422	751	1690
0.95	96	118	150	195	266	384	600	1067	
0.99	165	204	259	338	400	633	1036	1446	3501
0.997	220	271	344	449	611	880	1376		
0.999	270	334	422	552	751	1082		3007	

Достаточно большие числа наблюдений при допустимой ошибке (а) и доверительных вероятностях (P), рассчитанные без учета величины N — объем генеральной совокупности.

2. Расчет объема выборки по таблицам достаточно больших чисел и номограммам достаточно больших чисел.

Расчеты объема выборки по номограммам и таблицам дают примерно одинаковые результаты, но применяются при разных условиях. Таблицы достаточно больших чисел используют, если ничего нельзя сказать ни о средних показателях, ни о генеральной совокупности. При этом используются следующие вполне корректные (не уменьшающие объема выборки) допущения:

- объем генеральной совокупности не влияет на объем выборки;
- численность выборки зависит от P и величины вероятности, с которой делается заключение о достоверности выводов предельной ошибки репрезентативности и вероятности появления события,

Поскольку таблицы достаточно больших чисел построены с расчетом максимального значения признаков, то полученное число окажется несколько больше необходимого (46, с. 250). Одним из видов таблиц достаточно больших чисел является следующая таблица.

Чтобы найти по таблице достаточно большую численность выборки для обеспечения точности оценки $\alpha = 0,05$, ее надежности (доверительной вероятности) $P = 0,95$, на пересечении находим число 384, которое и является исходным достаточно большим числом.

Для случая изучения образовательных систем, наиболее употребителен максимальный объем выборки в 400—600 единиц.

Возможная ошибка тем меньше, чем больше численность выборки. Причем, чтобы уменьшить ошибку вдвое, численность выборки приходится увеличить в 4 раза. В таблице выделены значения выборок наиболее употребительных в практике социально-педагогических исследований на уровне города. По этой же таблице можно провести обратную работу — определение предельной допустимой ошибки и вероятности при известной выборке.

Если рассеяние признака известно на основании предыдущих исследований, то объем выборочной совокупности может быть определен по номограммам больших чисел представленных в работе (38, с. 247—275), при этом объемы выборки при тех же значениях P и α меньше, чем определенные по таблицам.

3. Расчет объема выборки эмпирическим способом.

Определение объема выборки по процентному выражению среднего квадратичного отклонения от величины измеряемого параметра. Основа этого способа представлена в работе (53 с. 210). Согласно эмпирическому закону объем выборки увеличивают до тех пор, пока среднее квадратичное отклонение, выраженное в процентах от величины средней, не станет меньше или равным 50%. Проведем расчет для вопроса 1 родителей.

максимальных достижений: x_i

выборка	отношение %%
150	52
300	49
400	46
500	44
700	45

Как видно из расчета, устойчивое значение отношения менее 50% соответствует выборке в 300—400 единиц. Аналогичные расчеты проведенные для учеников и учителей дают значения 200—300 и 300—400 соответственно,

4. Расчеты по статистическим формулам.

Напомним еще раз условия для которых проведены настоящие расчеты (см. замечание в начале этого раздела).

Различные источники дают одну формулу для расчета объема выборки для нашего случая:

где n — объем выборки.

Некоторые сложности появляются с определением величины t .

В работе (59) она определяется как нормированное отношение или стандартизованная разность и заданы ее значения:

$t = 2$ при $P = 0.95$, и

$t = 3$ при $P = 0.99$,

где P — вероятность (надежность заключения).

В работе (27) эта величина не вычисляется, а устанавливается равной 4.

В работе (53) t определена как критическая точка стандартизованного нормального распределения и составляет:

$t = 1,6449$, при $\alpha = 0.05$, и

$t = 2,3263$, при $\alpha = 0.01$,

Глава. 6. Нормирование тестов 192

где α — уровень значимости.

В работе (4) t определена как величина, связывающая среднюю (μ) и предельную ошибку выборки (Δ)

$$t = \frac{\Delta}{\mu}$$

Отсюда становится ясным статистический смысл этой величины — мера риска допущения ошибки связанной с выборкой — предельная ошибка выборки равна t -кратному числу предельных ошибок выборки. Табличные значения этой величины следующие:

$t = 1,96$, при $\alpha = 0.05$, и

$t = 2,58$, при $\alpha = 0.01$,

В этой же работе определены источники этих значений, вытекающие из значений функции Лапласа для нормального распределения. Кроме того, в этой работе для выяснения этой величины рекомендуется использовать эмпирическую формулу:

$$t = 3 + \frac{6}{n - 4}$$

Подробное рассмотрение этой величины вызвано ее значительным влиянием на получаемый в результате расчетов объем выборки. Разброс итоговых значений количества единиц наблюдения весьма значителен при констатировании остальных составляющих, так для случая $N = 29962$ — количество учителей в Санкт-Петербурге, при равенстве прочих переменных получаем следующие значения:

при $t = 2$, $n = 1198$, а

при $t = 3$, $n = 1752$, таким образом разброс составляет более 600 единиц, что для практики исследований весьма существенно,

Для дальнейших расчетов использовались данные значений t из работы (46), где, как нам кажется, они описаны наиболее подробно. Расчет объемов выборки по формулам для различных совокупностей образовательной системы города дает следующие результаты:

1. Для учителей города, при $N = 29962$, $n = 286$;
2. Для учащихся в школах города $N = 540481$, $n = 298$;
3. Для образовательных учреждений типа «школа» $N = 662$, $n = 78$;
4. Для классов $N = 2018$, $n = 184$.

Некоторые выводы.

Рассчитанные четырьмя различными способами объемы выборок дают примерно одинаковые результаты при сохранении показателей качества. (Достаточными объемами выборки можно признать значения в пределах от 150 до 600 единиц — это максимальное и минимальное значения для выборок с нашими граничными условиями. При этом необходимо заметить, что построение более сложных выборок приводит, как правило, к некоторому уменьшению количества измеряемых объектов.

В заключение необходимо сказать о стойком убеждении Среди педагогов и социологов в необходимости 5 или 10% Выборки для получения надежных результатов. Анализ литературных источников, посвященных расчетам объемов выборок, ни теоретически ни эмпирически не подтверждает такие объемы. Выявлен всего один источник, где рекомендуются подобные объемы: «... пока их (предприятий) немного, взять под наблюдение все... Затем, когда их станет очень много, обследовать 1/10 или 1/20 выборочным методом.» (В. И. Ленин, ПСС, т.35, с. 150-151).

6.1.2. Оценочный подход к расчету объема с учетом стратификации.

Как мы показали, для простого уменьшения значения стандартной погрешности вполне достаточна будет выборка из 300 — 500 испытуемых. Однако репрезентативность выборки с прямо зависит от ее объема. До сих пор мы рассматривали выборки без выделения страт. Для стратифицированных выборок существуют отдельные формулы расчета объемов (53, 46). Однако для практической работы более полезно будет рассмотреть вопрос оценки, а не расчета объема выборочной совокупности.

При получении нормы для общей популяции, например, детей школьного возраста, необходима выборка больше, чем выборка из столь ограниченной популяции, как укротители львов или факиры (56). До начала исследования нельзя сделать никакого утверждения относительно объема выборки безотносительно той популяции (категории лиц), из которой она подбирается. Здесь проясняется тот момент, что репрезентативность выборки является более важной, чем ее размер. Маленькая, но репрезентативная нормативная выборка будет предпочтительнее, чем большая, но непропорционально представленная.

Получение репрезентативной нормативной выборки.

Ясно, что наиболее неоднородной популяцией является генеральная популяция (все население), а все остальные являются ее подмножествами. По этой причине получить выборку заданного определенного качества из генеральной популяции — наиболее трудная задача, Однако в практике тестов школьных достижений этого и не нужно.

Правила для общих норм.

Принципиальным для оценки объема выборки является то, как будут использованы полученные данные. Если на их основе будут получаться нормы, то проведение расчетов объемов должно быть максимально строгим, если же предполагается критериальное оценивание, то оно требует меньшей точности и, соответственно, меньших затрат.

Для получения норм для генеральной популяции можно сформулировать несколько общих правил:

1) Выборка должна быть стратифицирована. То есть в выборке должны быть представители разных групп, реально существующих в генеральной совокупности, желательно в такой же пропорции.

Кроме этого страты должны быть выделены таким образом, чтобы дисперсия их ответов была максимальной. Обычно в психолого-социологических исследованиях адекватной признается стратификация на четыре уровня. Важными стратификационными переменными являются социальное положение, возраст и пол. Для тестов школьных достижений — возраст и пол.

2) В каждой подгруппе должно быть достаточное количество испытуемых, чтобы сформировать адекватную выборку. Выборка для всего населения означает, что с учетом всех возможных классификаций, необходимо очень большое количество испытуемых, например: 4 (социальное положение) x 2 (пол) x 5 (возраст) дает 40 категорий по 300 (среднее значение предварительных расчетов) испытуемых в каждой, что дает общее количество 12000 испытуемых. Следовательно, обеспечение адекватности норм для генеральной совокупности требует огромных затрат ресурсов. Нормы, полученные в менее масштабных исследованиях, также полезны, но должны использоваться с осторожностью.

По приведенной выше причине, а именно из-за необходимости привлечения огромных ресурсов, многие разработчики психологических тестов указывают более специфические групповые нормы, специально соответствующие характеру и практическому применению теста.

Формирование выборок для специфических групп.

Для получения адекватных норм для специфических групп необходимо стратифицировать выборку по основной переменной, влияющей на эти группы. На примере будет показано, как это делается в целом, хотя для выбора стратифицирующих переменных необходимы значительные исследования конкретной группы.

Учащиеся одной параллели.

Чтобы получить соответствующую выборку для учащихся одной параллели, необходимо обеспечить ее адекватность. Так, основанием для формирования выборки может быть подбор всех учащихся из шестых классов. Для школ некоторыми основными стратифицирующими, переменными являются:

1) Тип образовательного учреждения или реализуемой образовательной программы — школа, гимназия, лицей, авторская школа. Как показывает опыт, это разделение на категории можно отбросить, поскольку наиболее существенным для стратификации является тип образовательной программы, выделенный по следующим двум основаниям:

а) по уровню преподавания; коррекционная (компенсирующего обучения), базового уровня, повышенного уровня;

б) по профилю: физико-математический, естественнонаучный, гуманитарный, эстетический, военно-спортивный, профессионального обучения.

Таким образом, основанием отбора, если речь идет об учебных заведениях общего образования, должна быть образовательная программа, а единицей отбора — тот или иной класс, в котором данная программа реализуется.

2) Пол — мальчики, девочки, смешанная группа.

3) Географическое размещение — центр города, спальный район, поселок, сельская местность.

4) Статус школы в системе образования — начальная, неполного среднего, полного среднего, начального профессионального образования,

В зависимости от целей тестирования количество страт может быть уменьшено, например, если тестирование рассчитано только на городскую полную среднюю школу для детей обучающихся в гимназических и обычных классах, то по позиции «статус» стратификацию можно не учитывать, а по позиции :| «уровень преподавания» — уменьшить до двух. ; Таким образом для расчета норм по параллели классов может потребоваться $2 \text{ (уровень преподавания)} \times 2 \text{ (пол)} \times 2 \cdot [\text{(район города)} = 8 \times 300$, получим 2400.

Будет трудно отобрать выборку, точно сбалансированную I по всем переменным (поскольку на окраине города достаточно мало школ реализующих профильную программу обучения I на повышенном уровне), но, тем не менее, выборка, отражает пропорции каждой из школ в общей популяции, даже I было по две школы на категорию, должна была бы дать адекватные нормы.

П. Клайн предлагает следующие правила формирования с для специальных групп.

1) Найдите наиболее важные переменные, релевантные иных групп, и используйте их как основание для стратифицирования выборки.

2) Подберите настолько большую выборку, насколько возможно, как минимум из 300 испытуемых,

3) Помните, что маленькая выборка — это лучше, чем ничего. Если использовались небольшие выборки, активизируйте внимание пользователей теста на том, чтобы они использовали нормы, а если будут делать это, с предельной осторожностью.

Уже написано достаточно, чтобы стало ясно, что при установлении норм нет никаких теоретических проблем. Наоборот обычной трудностью является отсутствие ресурсов: времени, денег, испытуемых и помощников для проведения тестирования и обработки тестов, Однако, если тест предназначен для практических целей: отбора детей для продолжения образования, профотбора и профориентации, то естественно, что его нормы должны

удовлетворять высоким стандартам, описанным здесь. Существенно важно использование стратифицированных и больших выборок.

6.2. Варианты нормирования тестов.

Основания для сравнений в тестах достижений.

При разработке теста как инструмента измерения исследователь должен определить способ оценки полученных результатов, определить основание для сравнения. Прежде чем переходить к описанию процедуры нормирования тестов — основной формы сравнения психологических тестов — необходимо остановиться на возможных видах сравнений при проведении тестов достижений, которые могут быть не менее, а в некоторых случаях и более полезными, чем нормы, которые некоторые авторы рассматривают в качестве иной формы стандартизации.

Существуют четыре основных подхода к оценке результатов выполнения тестов достижений учащимися, которые в некоторой степени определяют особенности самих тестов. Выделяют тесты, ориентированные на:

- 1) индивидуальные нормы,
- 2) критерий (критериально-ориентированные),
- 3) статистические нормы,
- 4) межгрупповое сравнение.

6.2.1. Ориентированные на индивидуальные нормы конкретного ученика, реального уровня его развития в данный момент времени.

Результатом оценки в этом случае являются темп усвоения и объем усвоенного материала по сравнению с его начальным стартовым уровнем. На особенностях таких тестов мы еще не останавливались,¹ поэтому приведем пример. Как правило, такие тесты создаются на индивидуальных бланках, которые сохраняются все время обучения, а их анализ проводится после итогового тестирования. Бланки входного и выходного тестирования могут быть расположены на одном листе, а ученикам может быть предоставлена возможность откорректировать самостоятельно свои результаты входного тестирования.

Пример 1. Масса молекулы вещества:

<i>вход</i>		<i>выход</i>
<i>а</i>	<i>может быть измерена на чувствительных весах</i>	<i>а</i>
<i>б</i>	<i>может быть получена в результате эксперимента</i>	<i>б</i>
<i>в</i>	<i>может быть рассчитана по формуле</i>	<i>в</i>
<i>г</i>	<i>не может быть определена</i>	<i>г</i>

Привлекательность такого подхода заключается в педагогической направленности — по результатам может быть выявлено не только приращение в знаниях и навыках, но и способы деятельности, знания, полученные на бытовом уровне, особенности восприятия ребенком той или иной части

материала. А на основании полученных выводов может быть спроектирована программа индивидуальной коррекции.

В применении такого рода тестирования есть и существенные недостатки: большая трудоемкость, ограниченная сфера применения, низкая сравнимость результатов выполнения, сложность выделения предметной и возрастной детерминации. Однако для целого ряда задач обучения этот способ может оказаться наиболее адекватным.

6.2.2. Критериально-ориентированные и ориентированные на статистические нормы.

Критериально-ориентированные тесты позволяют оценить, насколько учащиеся достигли заданного уровня знаний, умений и отношений, определенного как критерий (проблему по-И и отбора критериев мы уже обсуждали). В данном случае оценка конкретного ученика не зависит от того, какие результаты получили другие ученики. Результат показывает, соответствует ли уровень достижений данного ученика социально-культурным нормам, требованиям стандарта или другим критериям. Шкала оценки, используемая при таком подходе, имеет только два уровня: достиг заданного уровня (критерия) или не достиг.

В тесте, ориентированном на статистические нормы, определяемые для данной совокупности учащихся, учебные достижения отдельного ученика интерпретируются в зависимости от достижений всей совокупности учащихся, среднего показателя — нормы. Происходит распределение учащихся по рангам. Независимо от того, какая шкала используется при этом (5-балльная, 100-балльная или др.), все эти шкалы не дают информации об овладении учащимися определенной системой знаний и умений или о достижении ими конкретных целей обучения.

На практике в последнее время наметилась тенденция смешения в одной проверке двух подходов, как ориентированного на норму, так и ориентированного на критерий. Чаше всего это смешение происходит при интерпретации результатов выполнения теста, когда результаты одновременно используются и для получения распределения учащихся, например, по проценту выполненных ими заданий, а также — для получения информации об уровне усвоения данной совокупностью учащихся изученного материала, представленного одним или группой заданий теста. Приведенная на рис. 2 гистограмма распределения учащихся по результатам выполнения некоторого гипотетического теста может помочь в понимании различий критериально-ориентированного тестирования и тестирования, ориентированного на норму,

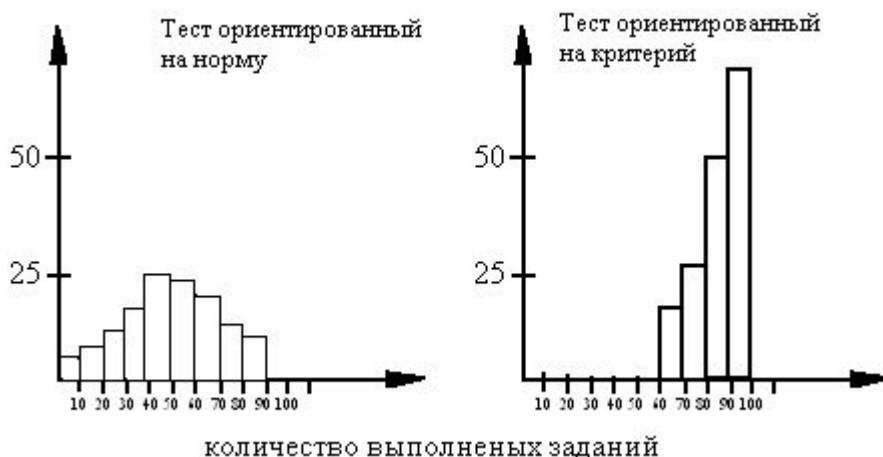


Рис. 2. Распределение учащихся по результатам выполнения теста.

6.2.3. Ориентация на критерий.

В психологии существует отдельное понятие «критериально ориентированный тест». Основным его отличием от тестов, ориентированных на нормы, считается то, что в нем указываются и определяются цели, достичь которые необходимо в процессе обучения, затем эти цели закладываются в тест, и измеряется степень достижения этих целей. Поэтому область применения этих тестов определена, с одной стороны, ясностью и диагностичностью целей обучения, которые для тестов достижений не всегда можно четко обозначить; с другой стороны, применение этих тестов возможно при максимальном показателе содержательной валидности.

Принципиального отличия тестов, ориентированных на нормы, и критериально ориентированных в части создания и процедур проверки качества, нет. Для критериально ориентированного теста не нужны нормы. Для критериально ориентированного теста важно, какую часть заданий испытуемый выполнил, а для теста, ориентированного на нормы, важна не сравнительная картина, а факт выполнения или невыполнения задания. Связь между этими тестами достаточно проста. Нормированный, содержательно валидизированный тест является по сути критериально ориентированным, а если критериально ориентированный тест стандартизировать, то он становится содержательно валидным, нормированным тестом. В любом случае, форма теста в этой части должна опираться, в первую очередь, на цели разработки теста.

Таким образом, как указывает П. Клайн, содержательный критерий применим к показателям теста, разработанного как содержательно валидный. Это замечание поясним на примере. Тест арифметических навыков может быть связан с проверкой овладения четырьмя арифметическими действиями. Задания будут состоять из ряда цифр, связанных арифметическими соотношениями, а испытуемые должны будут указать результат.

В таком тесте показатель 100% будет соответствовать полному овладению арифметическими действиями. В тестах такого типа необходимо иметь веское основание для того, чтобы по некоторому пороговому значению можно было бы принимать решение о переводе учащегося на следующий этап курса

обучения. Для этого требуется подтверждение, что ученики с показателями выше некоторого порогового являются успевающими или достаточно обученными, а те, у кого показатели ниже — неуспевающими.

Обычно, однако, такие пороговые значения задаются произвольно, на основании опыта преподавателей данной дисциплины. Для тестов такого типа эффективным будет всего лишь двухбалльный показатель: удовлетворительно и неудовлетворительно, да или нет.

Содержательный критерий полезен для тех тестов, в которых требуется проверка овладения некоторыми навыками или специальными знаниями. На самом низком уровне, в начальной школе, где важны базовые навыки, он может рассматриваться в качестве методики для конструирования тестов.

6.2.4. Прогнозирование критерия.

Вторым методом интерпретации показателей, отличным от использования норм и основанным на критериальном показателе, является построение серий таблиц ожиданий. К сожалению, автору неизвестны практические работы, которые проводились бы в нашей стране по построению таблиц ожиданий для тестов достижений. Поэтому представленные здесь материалы взяты целиком из литературы.

Таблица ожиданий — это таблица, в которой указывается вероятность того, что испытуемые, получившие некоторый показатель по тесту, достигнут в будущем некоторого критериального показателя. В этом случае критериальным показателем может быть дальнейшая успешность в обучении, результативность сдачи выпускных или вступительных экзаменов, в общем — уровень дальнейших достижений. Этот метод применим только тогда, когда возможно явное определение некоторого критериального показателя. Это чаще всего имеет место при тестировании результатов обучения, когда могут быть получены экзаменационные оценки и пень дальнейшего продвижения.

Аналогичные показатели, возможны в промышленности, где для профотбора используются экзамены и разного рода рейтинговые процедуры. В инженерной психологии могут использоваться предоставленные контролирующими лицами (экспертами) рейтинги успехов в, хотя надежность и валидность таких критериев может подвергаться сомнению.

Значения этих вероятностей должны определяться экспериментально. Это приводит к тем же самым проблемам формирования выборок, что присущи определению норм. В этом смысле значения вероятностей из таблиц ожиданий — это также нормативные данные. Вместо стандартных показателей для целевых групп или группового среднего и стандартного отклонения в них представлены вероятности достижения некоторого критериального показателя, Таблицы ожиданий иллюстрируют проблему индивидуальных и статистических показателей. Если конкретному показателю по тесту соответствует вероятность 0.43 успешной сдачи конкретного экзамена, то это означает, что 43% испытуемых с таким показателем по данному тесту успешно сдадут в будущем экзамен. Имеется в виду, что 43% испытуемых с таким

показателем смогут сдать экзамен — но какие 43%? Имея дело с отдельными детьми, трудно интерпретировать такие статистические предсказания. Однако, отсеив испытуемого с таким показателем будет вообще-то означать, что производящий отбор будет чаще прав, чем не прав. В этом смысле такие показатели полезны, но только в этом смысле. Не действенность статистических прогнозов в индивидуальных случаях имеет место для большинства нормативных исследований. Это очевидно и в случае с таблицами ожиданий, которые, казалось бы, представляют такие явные прогнозы.

6.2.5. Ориентированные на межгрупповое сравнение.

Этот вид сравнения результатов применяется при оценке результатов выполнения тестов группами учащихся. В данном случае не столь существенно, каким образом получены индивидуальные данные каждого ученика и в каких единицах они выражены. Межгрупповое сравнение позволяет определить, насколько лучше или хуже справились с работой ученики. Основная привлекательность такого метода — возможность использовать любые шкалы и единицы измерения, лишь бы они были единообразны. Во многих случаях такое сравнение дает количественную оценку (позволяет оценить «на сколько лучше»). Однако представить качественную картину результативности выполнения заданий такой способ сравнения не позволяет.

6.2.6. Статистические нормы и нормирование.

В начале нашей работы было показано, что одно из преимуществ, которыми обладают тесты по сравнению с другими методами измерений, — это то, что они нормированы. Следовательно, возможно сравнение показателя некоторого испытуемого с таковыми в генеральной совокупности или других релевантных группах, что в конечном счете дает возможность адекватной интерпретации полученного показателя. Из сказанного следует, что нормализация тестов наиболее важна в тех случаях, когда осуществляется явное или неявное сравнение показателей испытуемых, как, например, при профориентации или отборе в целях обучения. Нормы также могут быть полезны и в крупномасштабных скрининговых исследованиях. Тестовые нормы представляют собой установленные на базе репрезентативной выборки эмпирические усредненные количественные данные о результатах выполнения теста, полученные в стандартных условиях. Наличие норм позволяет определить характерные особенности шкал, установить место каждого испытуемого по отношению к другим для данного теста, установить статистически значимые сходства и различия между испытуемыми или группами испытуемых по одному и по разным тестам.

Виды норм.

По широте охвата можно выделить:

— универсальные нормы — устанавливаются для широкого контингента людей и лишь в малой степени зависят от действия каких-либо признаков;

- национальные нормы — применяются для представили конкретной народности или страны в целом и учитывают особенности культуры, норм и традиций обследуемых;
- региональные (локальные) нормы.

Помимо перечисленных можно выделить нормы для различных групп: профессиональные, армейские и другие групповые, в том числе школьные. Как правило, школьные нормы являются основными для тестов школьных достижений и разрабатываются для каждой школьной ступени на основе школьных программ.

При разработке норм необходимо учитывать следующее:

1. Нормы устанавливаются при разработке нового теста, адаптации или редактировании существующего, если он используется на выборке, отличающейся от стандартизированной по каким-либо критериям.
2. Введение нового типа нормировочного балла при разработке теста должно быть обосновано,
3. Стандартизированная выборка при разработке норм должна быть хорошо сбалансированной по составу и численности.
4. Все отклонения от процедуры нормирования тестовых результатов должны оговариваться в прилагаемых руководствах.

Обычно показатели некоторого индивидуума сравниваются с показателями релевантной нормативной группы посредством некоторого преобразования, которое выявляет статус этого индивидуума относительно данной группы. Для этого существуют различные методы преобразования исходных данных, наиболее употребительные из которых будут здесь описаны и оценены.

Наглядное представление нормирования. Нормальное распределение.

Итак, решая задачи сравнения в самом широком понимании этого слова, необходимо провести преобразование исходных данных в более удобный вид. Предварительно необходимо выяснить, чем же плохи первичные показатели и какие преобразования данных можно совершить. Учитывая наше обещание минимально использовать математический аппарат и вычисления, попытаемся показать наглядно формы представления данных. Однако без некоторых сведений классической теории ошибок нам не обойтись. Представим их здесь в вербальном виде и поясним на примере:

Пусть в результате проведенного тестирования получен некий ряд сырых показателей:

показатель	1	2	3	4	5	6	7	8	9	10
значение	3	5	6	2	9	4	8	5	5	2

$\bar{X}_{ср}$ — среднее значение показателя. Для его получения |димо сложить данные всех измерений и поделить их на количество измерений:

$$\frac{3+5+6+2+9+3+8+5+5+2}{10} = \frac{48}{10} = 4,8$$

σ — отклонение: показывает, насколько далеко данное значение показателя отстоит от среднего значения того же показателя. Для его расчета необходимо:

1. рассчитать среднее значение показателя;
2. вычесть из среднего значения показателя значение данного показателя.

Для показателя 1 отклонение будет равно:

$$4.8 - 3 = 1.8$$

для показателя 2 —

$$4.8 - 5 = -0.2$$

показатель	1	2	3	4	5	6	7	8	9	10
значение	3	5	6	2	9	4	8	5	5	2
отклонение	1,80	-0,2	-1,2	0,8	-4,2	1,8	-3,2	-0,2	-0,2	0,8

D — среднеквадратичное отклонение — мера разброса показателей относительно среднего значения. Для его расчета необходимо:

1. рассчитать среднее значение группы показателей;
2. вычесть из среднего значения показателя значение каждого показателя;
3. сложить полученные отклонения в абсолютном значении, не учитывая знак:
 $1.8 + 0.2 + 1.2 + 0.8 + 4.2 + 0.8 + 3.2 + 0.2 + 0.2 + 3.8 = 16.4$;

4. возвести сумму отклонений в квадрат:

$$16.4^2 = 268.96;$$

5. поделить полученное число на количество показателей:

$$\frac{268,96}{10} = 26,896 \approx 27$$

6. извлечь из полученной цифры квадратный корень:

$$\sqrt{27} = 5,2$$

α — частота признака или показателя: показывает, как часто встречается определенное значение того или иного показателя.

Для нашего примера частота, с какой встречается значение показателя 5, равна 3, частоты показателей 3 и 2 равны 2, для показателей 7 и 10 — равны 0, для остальных показателей частота равна 1, поскольку они встречаются по одному разу.

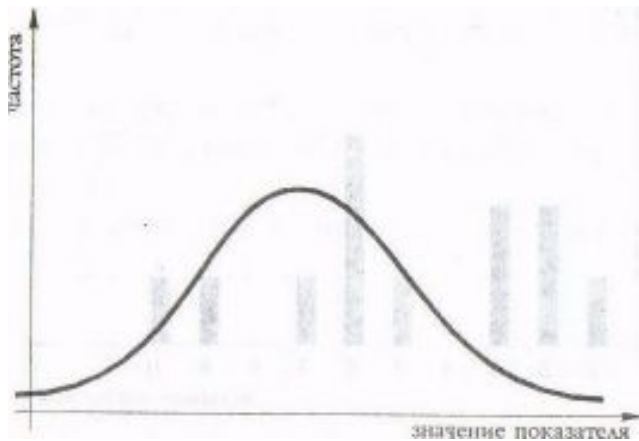
Распределение частот — это представление частот признаков в наглядном виде.

Для нашего примера распределение частот выглядит так, как на рис. 3:

Нормальное распределение.

Нормальное распределение — распределение частот, которое подчиняется определенному закону (закону нормального распределения), существование этого распределения обосновано эмпирически и математически. Именно по закону нормального распределения в большинстве случаев распределяются как чисто случайные величины, так и результаты выполнения тестов и их заданий. При этом необходимо сделать существенное замечание — количество значений показателя должно быть достаточно большим: не менее 30. Это утверждение

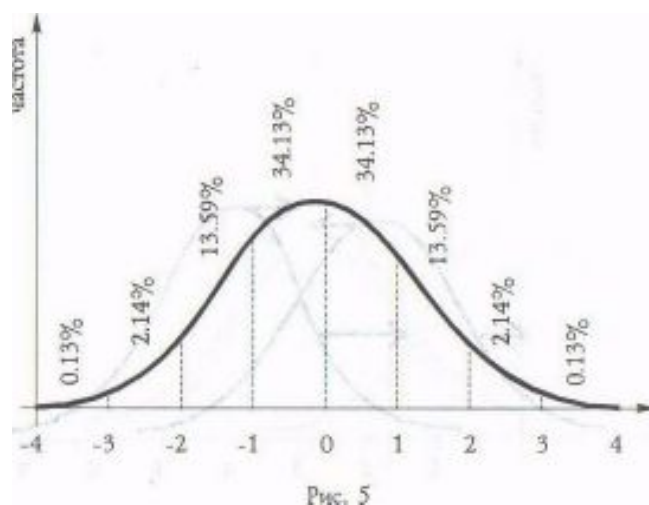
придется принять на веру или обратиться к литературе по математической статистике. Графики нормального распределения представлены ниже. Вообще, может быть бесконечное множество Графиков нормального распределения, но все они имеют одинаковый вид. Как можно видеть, график функции нормального распределения симметричный, с асимптотически приближающимися к 0 ветвями (рис. 4).



Принципиальное значение для дальнейших расчетов имеет определение площади под кривой нормального распределения. Определим на этом графике положение интересующих нас величин.

Среднее значение находится под максимумом пика кривой. Средне квадратичное отклонение — расстояние от среднего значения до точки А. Точка А выбрана таким образом, что площадь графика, ограниченная $X_{ср}$ кривой и ординатой точки А, составляет приблизительно 34.13% ограничения, а ограниченная расстояниями А с двух сторон — 68.26%.

Таким образом, мы определили необходимые для дальнейшего исследования понятия. Попробуем теперь выяснить, что же представляет собой нормирование тестов. Существует два этапа нормирования: центрирование и собственно нормирование.



Центрирование.

На рис. 6 изображены два графика распределения показателей, которые получены на одной шкале. Здесь мы можем увидеть один из недостатков первичных значений — они не со-1 впадают по шкале. Для того, чтобы сделать показатели сравнимыми, нам необходимо либо сдвинуть график А на три единицы вправо, либо — график Б на три единицы влево. На практике поступают несколько иначе: сдвигают оба графика в некоторую фиксированную точку, обычно это среднее значение выбранной шкалы измерения или нулевое значение. В данном случае для получения нормированных показателей к первичным достаточно добавить или вычесть некоторое фиксированное значение и нормы будут сформированы. Обычно для приведения данной кривой к нулевой необходимо вычитание из значений показателей $X_{ср}$. Действительно, если из значений показателей кривой А вычесть ее среднее значение 3, а кривой Б — 6, то обе кривые окажутся симметричными относительно нулевой шкалы.

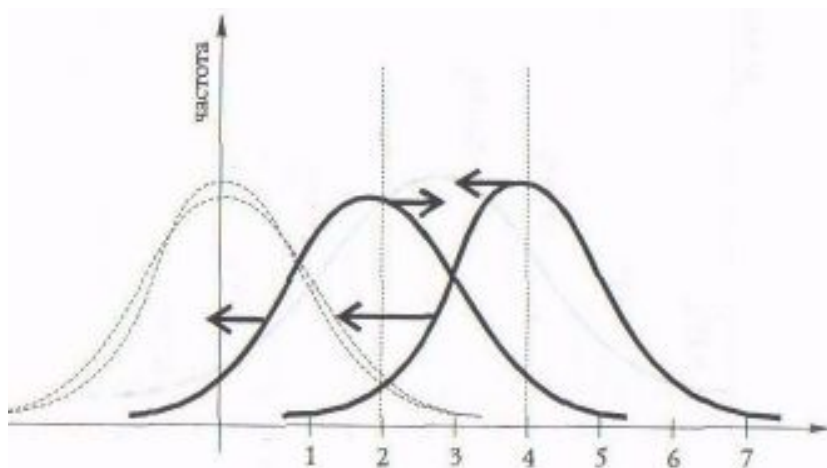


Рис.6

Нормирование.

Суть нормирования состоит в переходе к другому масштабу. Попытаемся проиллюстрировать эту операцию графическим примером. На рис. 7 представлены два графика, оба они уже приведены к единой шкале — их средние значения совмещены с нулевой отметкой. Однако форма этих графиков различна, и количественно частоты в точке А для этих графиков будут различны, что не дает возможности провести сравнения признаков для этих двух распределений. Для приведения их к нормальному виду нам следует провести еще одну операцию, а именно совместить в одной точке их среднее квадратичное отклонение. Для этого необходимо либо «сжать» один из графиков, либо расширить другой. Как и в случае центрирования, используют способ приведения графиков к нормальному стандартизированному виду.

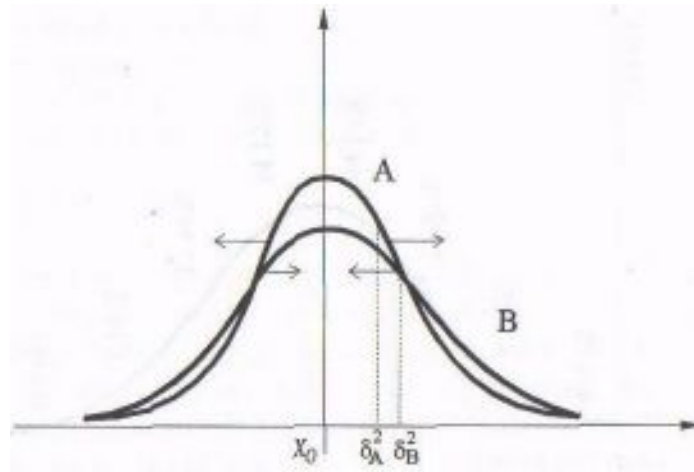
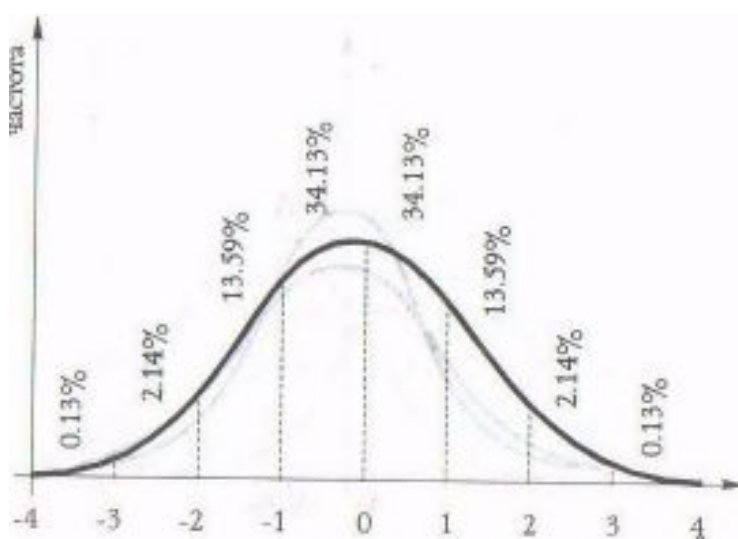
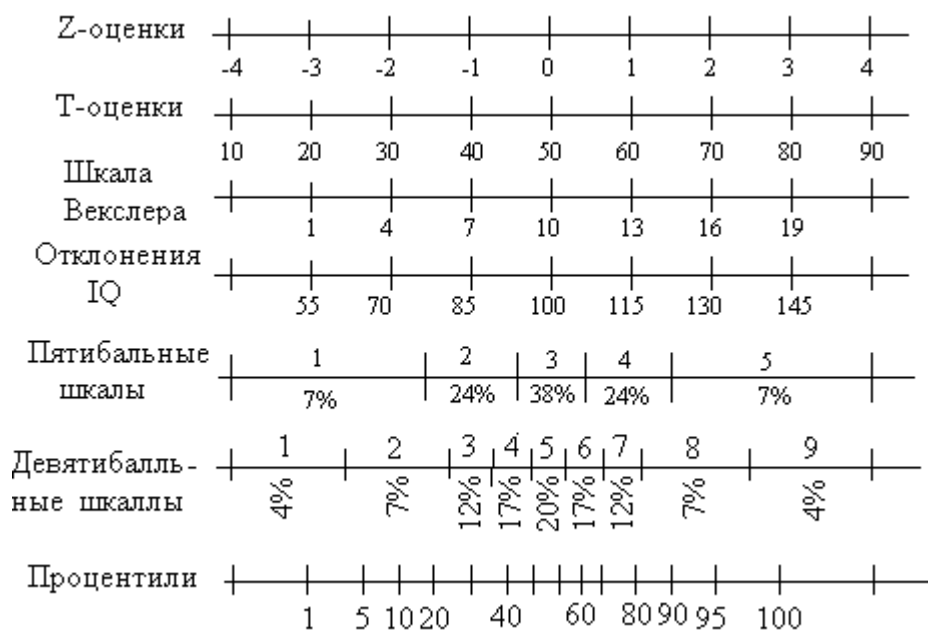


рис.7

На практике это означает деление показателей для стан-! на величину среднеквадратичного отклонения. В случае величина этого отклонения станет равна единице результаты исследований можно будет сравнивать. Таким образом, мы пришли к понятию стандартного показателя, который характеризуется средним значением 0, средним квадратичным отклонением 1. В общем виде кривых нормального распределения бесконечно много, также много может быть и стандартных кривых. Все зависит от того, каким образом мы определим среднее и среднеквадратичное отклонение, к которому необходимо привести наши первичные данные. Ниже приведены наиболее употребляемые стандартные значения (рис. 8), Эту же таблицу можно использовать для перевода одних шкал в другие. Первыми представлены Z-оценки. Далее T-оценки со средним 50 и отклонением 10. Остальные шкалы: шкала для широкого применяемого теста Векслера, шкала уровня умственного развития IQ, и, наконец, шкала процентилей.





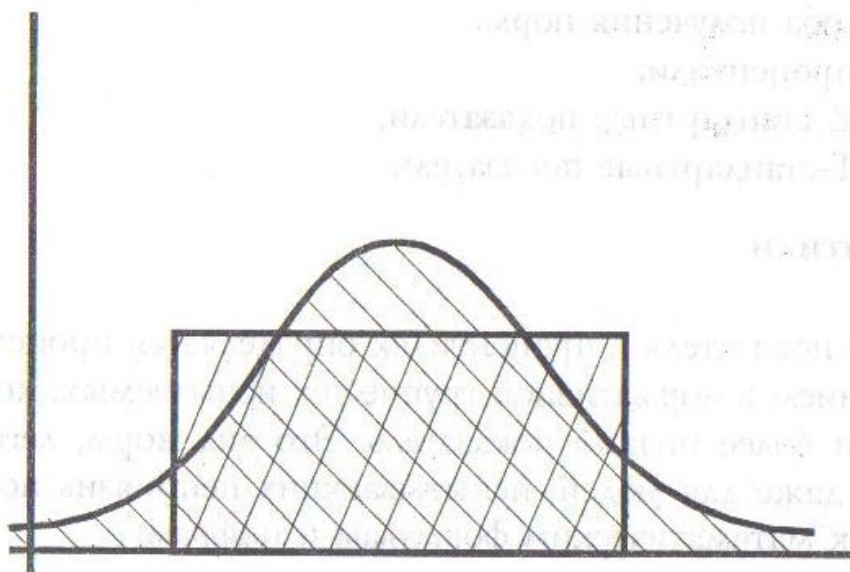
Не вдаваясь более в математические тонкости, приведем три способа получения норм:

- процентиля;
- 2-стандартные показатели
- T-стандартные показатели.

Процентили,

Ранг показателя в процентилях определяется процентным отношением в нормативной группе тех испытуемых, которые получили более низкий показатель. Это вид норм, легко понятный даже для людей, испытывающих неприязнь по отношению к математическим формулам и цифрам. Значение «двадцать пять процентилей» означает, что 25% из популяции имеют показатели ниже данного. Представление норм в процентилях просто и наглядно, однако процентиля имеют целый ряд недостатков. Процентили, являясь значениями порядковой шкалы, не могут быть использованы для последующего статистического анализа, как это может быть сделано для стандартизованных показателей. Во-вторых, так как распределение процентилей прямоугольное, тогда как распределение для большинства тестов приближается к нормальному, процентиля значительно искажают результаты. Большие отклонения от среднего значения сильно увеличиваются прцентилями, так же как и относительно небольшие отклонения на краях кривой. Этот факт можно проиллюстрировать измерением расстояний на плоской карте мира, когда измерение расстояний у полюса будет давать значительное искажение в силу невозможности адекватного отражения шара на плоскости. Для наглядности географы используют такие карты, однако для измерения расстояний они вряд ли пригодны. По этим же причинам (значительные искажения), использование процентилей крайне ограничено. В графическом виде взаимосвязь между представленными типами стандартизованных шкал можно представить следующим образом (см. рис. 9).

Процентили имеют прямоугольное распределение.



Стандартные показатели. Z-показатель.

Наилучший метод для представления норм — в виде стандартизованных показателей. Существует несколько типов стандартизованных показателей, которые, как видно из графика, и общем подобны.

Z -ПОКАЗАТЕЛЬ — показатель, имеющий, как мы выяснили, распределение со средним 0 и стандартным отклонением 1.

Как видно из графика, показатели Z имеют среднее значение 0, следовательно, принимают отрицательные или положительные значения. Чем больше показатель Z, тем дальше от среднего: положительные значения будут находиться выше, а отрицательные значения показателя Z — ниже среднего.

Например: проведя исследование, мы получили группу первичных показателей. Чтобы рассчитать стандартные Z-оценки необходимо:

1. Рассчитать среднее значение показателя, пусть оно равно $X_{ср.} = 60$;

Таблица 5.

Первичные показатели	Вычисление	Z
50	$(50 - 60)/10$	-1
60	$(60 - 60)/10$	0
75	$(75 - 60)/10$	1.5
30	$(30 - 60)/10$	-3

Z-показатели и первичные показатели

2. Рассчитать среднеквадратичное отклонение, пусть оно равно $D = 10$;

3. Взять значение какого-либо из показателей, например, $X = 50$, и перевести его в стандартную Z-оценку. Для этого провести центрирование (вычесть из значения показателя значение среднего; $50 - 60 = -10$) и нормирование (поделить полученное значение на среднеквадратичное отклонение; $-10/10 = -1$).

Таким образом, стандартная Z-оценка будет равна -1.

Преобразования в стандартный показатель, представлены в таблице, Так как преобразование первичных показателей в показатели Z является линейным, то распределение показателя 2 отражает распределение первичных показателей. Если распределение первичных показателей нормальное, то 2-показатели будут изменяться от +4 до -4. Приблизительно 68% этой кривой лежит в интервале между средним, плюс-минус одно значение стандартного отклонения.

Строго говоря, концы кривых нормального распределения асимптотически приближаются к нулю. Это означает, что будучи бесконечно близко от осей координат, концы кривых никогда нуля не достигают. Этот факт порождает много противоречий и неточностей среди авторов, освещающих данную тему. В качестве примера можно привести разницу даже в числовой интерпретации. П. Клайн утверждает, что максимальное значение Z -показателя — 3, а в книге В. К. Ранды утверждается, что 4.

При этом, пренебрегая абсолютно малыми значениями, и тот и другой авторы не совсем точны, поскольку ни в той ни в другой работе не указана точность, с какой проходит измерение. По своей сути точность в данном случае — это значение тех отрезков под кривой нормального распределения, которыми можно пренебречь. Так в первом рассматриваемом нами случае при шкале 4- 3 — -3 точность измерения составляет 99,72, а во втором — при шкале +4 — -4 — 99.99, то есть под кривой нормального распределения, ограниченной этими значениями, остается 99,72% и 99.99% площади поверхности соответственно. Показатель Z для 1 может непосредственно быть преобразован в проценты; то есть 16 процентов популяции имеют показатели выше, чем +1 или ниже -1. То же самое может быть сделано для любого показателя 2, определенного по таблицам нормального распределения.

У Z -показателей есть несколько существенных недостатков. Основной из них состоит в том, что их среднее значение равно 0, а наивысшее возможное значение равно 4 (при учете точности измерения), что при обсуждении результатов исследований с людьми несведущими вызывает необходимость дополнительных объяснений и в некоторой степени подрывает доверие к результатам исследований. К недостаткам Z -показателей могут быть отнесены наличие отрицательных значений, необходимость дробных вычислений, малая интервальность шкалы.

Это привело к тому, что разработчики тестов стали применять дальнейшие преобразования Z -показателей.

Т-показатель.

В общем виде стандартизованное преобразование Z -показателей состоит в линейном преобразовании вида $Y =$

Таблица 6.

Первичные показатели		Z		T
50	$(50 - 60)/10$	-1	$100 - 1.0 * 10$	90
60	$(60 - 60)/10$	0	$100 + 0.0 * 10$	100
75	$(75 - 60)/10$	1.5	$100 + 1.5 * 10$	115
30	$(30 - 60)/10$	-3	$100 - 3.0 * 10$	70

«Сырые» показатели, Z-показатели и T-показатели первичных показателей с $X_{ср.} = 60$ и $G = 10$. Преобразованный T-показатель с $X_{ср} = 100$ и $G = 10$.

$Ax + B$, где V — преобразованный показатель, B — среднее значение преобразованного распределения, A — стандартное отклонение, x — Z-показатель.

Пример такого преобразования приведен в таблице 5 (использованы данные из таблицы 4).

Шаги вычислений для показателей Z и T:

- 1) Вычислите среднее и стандартное отклонение показателей.
- 2) Представьте каждый показатель как отклонение от среднего: $X - X_{ср.}$
- 3) Поделите полученный результат на стандартное отклонение.
- 4) Вычислите произведение каждого Z-показателя на требуемое стандартное отклонение для преобразуемого распределения.
- 5) Сложите полученное значение, с требуемым средним; и получим необходимое значение распределения, в данном случае T.

При конструировании тестов обычно производится преобразование Z-показателей к распределению со средним значением $x' = 50$ и стандартным отклонением $G = 10$. В этом случае, если распределение приближается к нормальному, показатели будут изменяться от 90 до 10. Таким образом, если мы хотим получить для наших тестов нормы этого вида, то преобразуем наше множество нормативных показателей в показатели со средними, равными 50 и $G = 10$. Пользователи тестов, следовательно, могут взять «сырые» показатели своих испытуемых и выразить их в виде T-показателей.

T-показатели такого вида просты для понимания и в тех случаях, когда распределения показателей тестов аппроксимируют нормальное распределение. Такие показатели, конечно, пригодны для статистического анализа.

Еще одним преимуществом стандартизованных показателей является то, что стандартизованные показатели сравнимы; так, значение показателя 2, равное 1, представляет результат выполнения любого теста, находящийся на расстоянии, равном одному значению стандартного отклонения от среднего. Аналогично, если бы для батареи тестов использовался некоторый преобразованный стандартизованный показатель с одинаковыми значениями средних и стандартных отклонений, то можно было бы непосредственно сравнивать показатели, получаемые по любым тестам батареи.

С точки зрения П. Клайна, для тестов с распределением, если не нормальным, то, по крайней мере, симметричным, T-показатели со средним значением $x' = 50$ и со стандартным отклонением $G = 10$ являются точной значимой нормой.

Итак, мы привели примеры получения нормализованных показателей. Конечно, данным изложением мы только коснулись темы стандартизации, однако приведенных сведений вполне достаточно для того, чтобы самостоятельно рассчитать необходимые для школьных тестов достижения нормы. Подведем некоторые итоги.

Необработанные показатели имеют значение лишь при сравнении с показателями нормативных групп.

Значение норм зависит не только от способа расчета, но, в первую очередь, от качества формирования выборочных совокупностей. Нормы могут использоваться с какой-либо долей уверенности лишь тогда, когда нормативные выборки адекватны. Существуют различные методы для представления нормативных показателей, которые, в конечном итоге, определяются целями разрабатываемых тестов.

Процентили, хотя и просты для понимания, не пригодны для статистического анализа. Следовательно, для использования рекомендуются стандартные показатели, основанные на отклонении показателей от среднего.

Преобразованные стандартные показатели всегда сравнимы: одинаковые стандартные показатели находятся на одинаковом расстоянии от среднего, При определении норм для тестов должны выполняться следующие процедуры, благодаря которым тесты приобретают гораздо большее значение и реже приводят к заблуждениям:

1) Объем выборки, основание для ее стратифицирования (если таковое есть) и ее происхождение должны быть четко указаны.

2) Должен быть указан тип используемых норм.

3) Для каждой нормативной группы должны быть указаны «сырые» показатели по сравнению со стандартизованными показателями. Если необходимо, то рядом с «сырыми» показателями могут быть также указаны и процентили. Следствием применения норм и стандартизации является то, что некоторый показатель по отношению к показателям стандартной группы может интерпретироваться. Для изучения педагогических процессов во многих случаях достаточно определения критерия или первичных показателей, они более предпочтительны, ибо являются исходными данными.

Получение школьных оценок.

Мы рассмотрим два вида этих оценок. Школьный балл и оценки типа «зачтено-незачтено».

Получение оценки школьного балла.

Значительную трудность при переводе стандартных оценок в школьные вызывает то, что преобразование Z- и T-оценок в оценки школьного балла является нелинейным. Поэтому, для получения этих оценок предлагается следующий путь. Перевести Z-оценки в форму близкую к школьным — Ш-оценку (со средним 4, средним квадратичным отклонением 0.5 и шкалой от 2.5 до 5.5), а затем по приведенной ниже таблице дать школьную оценку. Для преобразования необходимо воспользоваться формулой:

$$Ш = 4 + \frac{Z}{2}$$

Пример 1. результаты выполнения трех субтестов тестов школьных достижений одним, из учеников пересчитаны в стандартные Z-оценки.

субтесты	Z-оценки	Ш-оценки
для математики	1.02	$Ш = 4 + 1.02/2 = 4.51$
для географии	-0.79	$Ш = 4 + (-0.79/2) = 3.61$
для истории	0.82	$Ш = 4 + 0.82/2 = 4.41$

Далее воспользуемся таблицей для перевода Ш-оценок

Ш-оценка	до 3	от 3 до 4	4 до 4.5	более 4.5
школьный балл	2	3	4	5

Таким образом, данному ученику могут быть даны следующие оценки: по математике — 5 (отлично), по географии — 3 (удовлетворительно), по истории — 4 (хорошо).

Стандартная Ш-оценка еще не является оценкой школьного балла, которая используется в школе. Для тех кто разрабатывает эту часть тестирования корректный перевод стандартных оценок в оценку школьную чрезвычайно важен и мы обращаем на это самое пристальное внимание. Дело в том что все стандартные оценки распределены нормально, а распределение школьных оценок не нормально, поэтому любые попытки прямого использования стандартных оценок для выставления школьных приводят к значительным искажениям. Для получения же таблицы перевода Ш-оценки в школьный балл можно предложить следующий путь. На репрезентативной выборке собрать данные об успеваемости учащихся. Разбить этих учащихся на четыре группы — имеющие оценки отлично, хорошо, удовлетворительно и неудовлетворительно. При этом первые три оценки рассчитывать усреднение по всем предметам, а оценку «неудовлетворительно» присваивать ученику, если он получил хотя бы одну такую оценку. После этого найти пропорциональные отношения этих групп учеников и на основе этой пропорции рассчитать значения коэффициентов для таблицы перевода. Именно таким способом была получена таблица, приведенная в примере 1.

Отметим, что для практики возможно создание таблицы, перевода непосредственно сырых оценок в оценку школьного балла. При разработке компьютерного варианта обработки результатов полезно иметь несколько стандартных оценок. Оценку школьного балла и оценку «зачтено-незачтено».

Выставление оценок «зачтено-незачтено».

В том случае, когда нет необходимости выставить школьную оценку, а необходимо оценить работу учащихся по двухбалльной шкале

«зачтено/незачтено», «сдал/не сдал», «поступил/не поступил» и т. д. удобнее пользоваться критериями значимости для стандартизированных оценок. Такие критерии могут быть двух видов: для относительной и абсолютной оценки. Этот критерий имеет числовое выражение и обозначает границу, результаты выше которой говорят о том, что данная работа выполнена, а ниже — соответственно о том, что она не выполнена.

Для абсолютных оценок проводится сравнение с нормами, - полученными при апробации данного теста, и для стандартизированных оценок они составляют минус одно значение среднеквадратичного отклонения:

показатели критериев значимости	
Для Z-оценок	-1
Для T-оценок	40
Для Ш-оценок	3.5

Для их использования необходимо провести процедуру обработки, и результаты сравнить с данными показателями критериев значимости, если полученные результаты их превосходят или равны, то дается позитивное заключение, если меньше — то негативное.

При использовании относительной шкалы все испытуемые ранжируются по результатам выполнения ими работ, и по необходимости выбирается определенное количество лучше (или хуже) справившихся с работой. Если данные расчеты проводятся только по одному из субтестов, то при этом возможно использование первичных оценок. Если испытания проводятся по двум или более субтестам, то необходимо вначале проведение процедуры получения Z-оценок.

Глава 7. УСЛОВИЯ и процедура проведения тестов

7.1. Оснащение тестов.

Подбор оснащения зависит от назначения тестов, от их вида и количества участников.

Индивидуальные тесты рекомендуется применять в следующих случаях:

1. большое количество заданий в тесте даны с открытыми вариантами ответов;
2. требуется повысить достоверность результатов;
3. требуется непрерывное наблюдение за работой испытуемых, фиксация промежуточных результатов, времени, или нужен тесный контакт с испытуемым с целью оптимизации его деятельности (например, корректировка промежуточных результатов);
4. испытуемые не умеют читать и писать, либо плохо или совсем не знают языка;
5. тестирование лиц с соматическими или нервно-психическими расстройствами или пониженным интеллектом.

Индивидуальные тесты требуют большего времени и более высокой квалификации экспериментатора. В этом смысле они менее экономичны по сравнению с групповыми. Индивидуальное тестирование играет особенно важную роль на начальном этапе многих психологических исследований. Большинство предметных и аппаратных тестов проводятся индивидуально. Групповое тестирование определяется, в первую очередь, запросами практики. Процедура их проведения значительно проще.

Опыт проведения тестов достижений среди детей России и стран СНГ позволяет говорить еще об одной особенности проведения тестов в нашей стране, особенно если исследование носит сравнительный характер. Эта особенность вытекает из исторически сложившегося отсутствия опыта заполнения тестов и ответов на вопросы, в которых нужно дать качественную оценку какому-либо явлению или процессу. В результате, количество случайных ошибок может существенно повлиять на выполнение тестирования. Отсюда необходима особая тщательность при написании инструкций для детей и экспериментаторов, их качественная подготовка, а в некоторых случаях бывает целесообразно проведение тренировочного задания, в процессе которого будут отработаны навыки заполнения тестов и ответов на вопросы. Использование бланковых тестов обеспечивает большую экономичность процедуры и массовость выборок. Квалификация экспериментатора при стандартных процедурах массовых обследований может быть невысока. В соответствии с назначением определяется характер теста.

В комплект бланковых тестов, к которым чаще всего относятся тесты достижений, входят:

- тестовая тетрадь или лист с инструкцией и задачами (вопросами);
- регистрационные бланки (протоколы, ответные листы);
- руководство для проводящего тестирование;
- ключи;
- секундомер (если это необходимо).

Оснащение предметных тестов включает в себя вместо тестовой брошюры необходимый материал для выполнения планируемых действий. Например, палочки, болтики или металлические штифты в тестах на мануальную ловкость и пальцевое проворство, набор материалов и принадлежностей, другие необходимые предметы.

Аппаратные тесты предполагают наличие специальных устройств, приборов для исследований реакций испытуемого.

Поскольку в педагогических целях используются чаще всего тесты достижений, рассмотрим более подробно оснащение бланковых тестов. Бланки обследования фиксируют результаты тестирования. Существуют два вида бланков: рабочий лист для испытуемого (тестовая тетрадь), бланк ответов (регистрационный бланк, протокол), являющийся неотъемлемой частью большинства тестов (особенно бланковых или «тестов карандаша и бумаги»), который представляет собой лист или несколько листов, где отмечаются ответы тестируемого. К этому бланку прилагаются тетради с заданиями. Типовая тестовая тетрадь, как правило, состоит из набора утверждений, вопросов или

задач. На бланке отражена последовательность задач с вариантами ответов или с пропусками в соответствии с типом вопросов.

Приведем форму бланка ответов, разработанную для тестов школьных достижений петербургской лабораторией «ИОС».

Фамилия _____ имя _____ школа _____ класс _____ код

ОТВЕТЫ НА ВОПРОСЫ ПСИХОЛОГО-ПЕДАГОГИЧЕСКОГО ТЕСТИРОВАНИЯ

ТРЕНИРОВКА		Часть I математика		Часть II биология		Часть III физика		Часть IV химия		Часть V география	
1	ABCD	1	1234	1	ABC	1	ABCDE	1	ABCD	1	ABCD
2	AB	2	ABCD	2	ABCD	2	ABCDE	2	ABCDE	2	ABCD
3	ABC	3	ABCD	3	ABCDEF	3	ABCDE	3	ABCDFG	3	ABCD
4	ABC	4	ABCD	4	ABC	4	ABCDE	4	ABCD	4	ABCD

Задачи в тестах расположены, как правило, в порядке возрастания сложности. Перед задачами должна быть инструкция испытуемому по выполнению теста с одним-двумя примерами. На примерах учащийся знакомится с типом и формой задач, с правилами ответа на вопросы, делает первые попытки решения задач подобного типа.

Существуют определенные правила для разработки бланка:

1. Бланк составляется таким образом, чтобы свести к минимуму непродуктивную работу испытуемого по отыскиванию места для проставления ответа.
2. Желательно по возможности свести ответ к подчеркиванию, обведению кружком, обозначению простыми символами (обязательно для задач закрытого типа).
3. Бланк не должен быть перегружен цифрами, делениями, строками и др. обозначениями.
4. Ответы по отдельным сериям (субтестам) должны выделяться на бланке отдельно. Если нет групп заданий, но их количество велико, графы для ответов должны разделяться на отдельные блоки по 10—25 заданий.
5. Из бланка, как правило, исключают название теста.

6. На бланке должны быть предусмотрены места для проставления необходимых паспортных данных (если обследование не анонимно), а также номера регистрации обследования и даты проведения.

7. Иногда целесообразно дважды поместить инструкцию для испытуемого: на бланке и в тетради с заданиями.

8. При переходе к ответам на новую группу заданий необходимо поместить примеры заполнения граф.

Формы для обработки и анализа результатов создаются также с учетом указанных требований и не должны затруднять работу исследователя. Как правило, в методические материалы к тестам включаются образцы бланков. Соблюдение точности бланка необходимо, так как для ряда методик незначительные изменения внешнего вида бланки могут существенно изменить результаты работы (Л. Ф. Бурлачук).

Таким образом, бланк составляется так, чтобы обеспечить наибольшее удобство при его заполнении и обработке.

Руководство к тесту пишется для его пользователей и в нем излагаются основные сведения о тесте:

- назначение и психолого-педагогическое содержание;
- ограничения и показания для применения;
- состав теста;
- описание существующих форм и модификаций;
- ссылка на апробацию теста;
- указания к проведению (условия, инструкции, поведение экспериментатора);
- ключи;
- правила обработки данных;
- устройство шкал;
- данные о надежности и валидности;
- правила интерпретации результатов. Указания к проведению теста, содержащиеся в руководстве для исследователя, состоят из описания:
 1. условий проведения теста;
 2. материалов и приборов, требуемых при проведении (карандаши, резинки, тест-приборы, черновики, бланки для ответов и др.), а также порядка их размещения и предъявления;
 3. временных ограничений;
 4. полного текста сообщения для испытуемых;
 5. поведения экспериментатора во время проведения теста;
 6. ответов на возможные типичные вопросы.

Инструкции для экспериментатора и испытуемого входят в состав всех стандартизированных тестов и определяют их объективный характер.

Инструкция для экспериментатора приводится в руководстве, а инструкция для испытуемого — тестовой брошюре, и, частично, на лицевой странице регистрационного бланка,

Пример несложной инструкции для экспериментатора приведен в приложении 3 «Руководство для исследователя».

Инструкция для испытуемого включает описание назначения теста (если это не противоречит условиям тестирования), правил заполнения бланков для ответов, решения одной-двух простых задач-образцов.

Изложение инструкций должно быть ясным, доступным и подробным. От этого зависит понимание их испытуемыми и отсутствие лишних вопросов. Пример инструкции для заполнения теста приводится в приложении 3 «Руководство для исследователя».

Вне зависимости от происхождения теста (создание нового, модификация старую, адаптация теста) инструкции составляются к каждому из них.

Рекомендуется соблюдать следующий порядок составления инструкций:

1. Написание первого варианта инструкций,
2. Первое опытное проведение теста.
3. Устранение грубых ошибок в инструкциях,
4. Апробация инструкции при проведении проверочного тестирования (желательно, чтобы проводящий тестирование не принимал участие в составлении инструкции).
5. Окончательная редакция инструкций.

Б. К. Гайда и В. П. Захаров указывают, что точное выполнение разработанных таким образом инструкций является обязательным для всех экспериментаторов, желающих получить сравнимые результаты. Изменение инструкций допустимо только в том случае, если они показали свою несостоятельность и не осуществляются с выполнением всех требований к стандартизации теста.

Ключи к тесту представляют собой наборы ответов с оценками к вопросам или к вариантам ответов для определения степени выраженности того или иного свойства личности. В 11 тестах достижений ключом к тесту является упорядоченные наборы правильных (или неправильных) ответов на вопросы или задачи,

По форме ключи могут быть нестандартизированными и стандартизированными.

Пример 1.

Стандартизированный ключ к тесту школьных достижений.

Предметная область — ГЕОГРАФИЯ.

<i>Номера и код заданий</i>	<i>Правильны й</i>	<i>Характер контролируемых умений</i>
<i>У-1</i>	<i>В</i>	<i>работа с планом</i>
<i>У- 2</i>	<i>В</i>	<i>применение знания</i>
<i>У- 3</i>	<i>С</i>	<i>работа с картой</i>
<i>У- 4</i>	<i>Д</i>	<i>работа с картой</i>

Нестандартизированные ключи представляют собой перечень ответов или оценок по каждому варианту ответов.

Пример 2.

Нестандартизированный ключ на основе ШТУР.

Задание. Ком предлагаются два слова. Нужно определить, что между ними общего. В каждом случае необходимо найти наиболее общие признаки этих слов.

дождь — град

Ключ.

2 балла — осадки

1 балл — вода, плохая погода;

0 баллов — остальные ответы.

Это довольно неудобная форма для пользователя, поскольку анализ результатов занимает много времени.

Стандартизированные ключи могут применяться в форме шаблонов с отверстиями, кусков картона с вырезами в виде полосок, спиц к перфокартам, имеющим прорези в определенных местах. Стандартизированные ключи значительно облегчают работу по проверке и анализу данных.

Шаблоны изготавливаются из плотного материала, их размер должен совпадать с размером регистрационного бланка.

Шаблон может иметь специальные отверстия для точного совмещения с бланком (уголки, отверстия). В шаблоне вырезаются отверстия, совпадающие с правильными ответами на регистрационном бланке.

7.2. Подготовка процедуры проведения тестов.

Разработка теста, позволяющего получить объективные результаты, предполагает учет условия, необходимых для успешного его проведения.

Результаты теста во многом зависят от внешних условий — физических, психологических и технологических. При определении таковых учитываются характеристики помещения, его оснащение, наличие отвлекающих факторов, состояние тестовых материалов, субъективно-психологические факторы. В. К. Гайда предлагает учитывать следующие характеристики помещений при выявлении условий проведения теста:

1. Удельный объем помещения на 1 человека (в соответствии с санитарно-гигиеническими нормами, — не менее 2,0 кв.м на 1 чел.);
2. Освещенность (равномерное распределение света по поверхности стола, прибора при интенсивности освещения не ниже 500 люксов);
3. Вентилируемость;
4. Уровень шума (не выше 4 сонов);
5. Акустические характеристики.

Нам такие требования представляются чрезвычайно сложными, поскольку степень их выполнения без специального оборудования и специалистов оценена быть не может.

При проведении группового теста — основного вида тестов достижений — требования к формальной стороне процедуры проведения могут быть следующие:

- обеспечение инвентарием в необходимых количествах в случае, когда используются простые материалы: карандаши, ластик, ручки, фломастеры, необходимо иметь их полуторократный запас;
- наличие столов и стульев в количестве, необходимом для проведения исследования, в соответствии с инструкцией по проведению;
- размещение столов и стульев таким образом, чтобы к каждому испытуемому было удобно подойти;
- обеспечение каждому испытуемому удобного места за столом, за счет подбора оснащенной мебелью помещения необходимых размеров.
- оборудование места с максимальным обзором для экспериментатора и, если это необходимо, для наблюдателя.

Особое место среди факторов, влияющих на индивидуальную и групповую работоспособность испытуемых, занимает время проведения теста и характер деятельности учащихся до тестирования. Наиболее благоприятно время от 9 до 12 или с 16 до 18 часов. Поскольку речь идет о тестах школьных достижений, то наиболее приемлемо время второго или третьего урока первой смены.

Авторами тестов может быть выдвинуто требование не проводить тестирование после занятий физической культурой и спортом. В любом случае требования к процедуре проведения должны быть зафиксированы в инструкциях для исследователя.

При организации обстановки для проведения тестов важно учитывать ситуативные отвлекающие факторы. К ним относятся: шумы (с улицы, из других частей здания, радио и телетрансляции и т. п.), звонки, стук, звук шагов, гудение неисправных ламп дневного света, запах (пищи, краски и пр.), мигание света, неопрятность столов, помещения и т. д.

Перед началом тестирования экспериментатор должен оценить состояние тестируемых. Визуально можно определить состояние возбуждения или депрессии; учащихся, у которых отмечается проявления таких состояний, желательно не допускать к тестированию. Не участвуют в тестировании учащиеся, решительно отказывающиеся от испытаний.

Непосредственная подготовка к проведению теста заключается в проверке состояния помещения, его оснащения, пригодности для размещения определенного числа испытуемых, а также устранении или уменьшении ситуативных отвлекающих факторов и проверке наличия, состояния и размещения тестовых установок и материалов.

Процедура проведения теста должна соответствовать порядку, изложенному в руководстве. Пример руководства для исследователя приведен в приложении 3. В любом случае руководство должно содержать сценарий теста и последовательность его проведения.

7.3. Сценарий проведения исследования.

Сценарий проведения исследования в общем виде должен включать в себя следующие необходимые сведения, которые могут сообщаться испытуемым (в зависимости от условий тестирования что-то может быть сокращено или добавлено, например, в приведенном в приложении сценария достаточно подробно описана процедура раздачи и сбора тестовых тетрадей):

1. Объяснить, зачем нужен тест, сообщить, какие результаты ожидаются.
2. Объяснить, почему испытуемые должны приложить максимум усилий для его выполнения, акцентировать внимание испытуемого на возможности проверки своих сил или подчеркнуть соревновательный мотив, при этом следует подчеркнуть, что слишком сильная мотивировка, равно как и слишком слабая, в одинаковой степени негативно сказывается на результативности выполнения заданий.
3. Медленно, громко, четко, без запинок, естественным ^ голосом прочесть инструкцию к тесту с примерами, если они | имеются. В данном случае возможен вариант, когда испытуемые самостоятельно следят по своим вариантам текста за инструкцией. При таком порядке возможно воспроизведение инструкции по памяти.
4. Дать возможность испытуемым потренироваться, решив самостоятельно одну или более из — задач-образцов, если таковые имеются, проверить, правильно ли понята инструкция.
5. Сообщить о временном ресурсе, о правилах исправления: допущенных ошибок, о том, чего не рекомендуется делать при решении задач, к кому обращаться в случае возникновения вопросов,
6. Вместе с испытуемыми или самому записать, если требуется, паспортные и биографические данные в регистрационных бланках. Проследить за правильностью заполнения.
7. Ответить на имеющиеся вопросы,
8. Дать команду начать решение задач теста, Время начала записать самому или попросить сделать это испытуемых на регистрационном бланке,
9. Во время решения задач или ответов на вопросы следить:
 - за временем решения, если это необходимо;
 - за наличием отточенных карандашей и других материалов;
 - за правильностью заполнения паспортной части регистрационных бланков (если замечена ошибка, своевременно | ее устранить);
 - за тем, чтобы испытуемые не писали на тестовых брошюрах, если иное не предусмотрено, не портили тестовых установок и приборов; — за тем, чтобы соседи не общались между собой, не шептались, не мешали друг другу, не подглядывали друг у друга;
 - за состоянием испытуемых;
 - за тем, чтобы испытуемые своевременно получали ответы на вопросы в ситуациях, связанных с процедурой проведения (ответы не должны служить подсказкой для решения или нарушать указания руководства, возможные варианты ответов должны быть предусмотрены в руководстве),

10. После сигнала к окончанию решения задач теста при групповом проведении дать команду сложить брошюры и бланки для ответов в исходное положение или самому собрать (если участвует не более 30 человек), Если участников тестирования больше 30 человек, то рекомендуется попросить всех оставаться на своих местах, чтобы облегчить сбор материала. Затем попросить передать в начало или конец колонки (ряда) тестовые материалы в следующем порядке: бланки для ответов, брошюры, черновики, После этого пересчитать количество бланков и брошюр, проверить, чтобы их количество совпадало с числом испытуемых.

11. По окончании тестирования просмотреть все брошюры и стереть пометки на них; если это невозможно — брошюры следует уничтожить.

Кроме этого, сценарий должен предусматривать процедуру приветствия и благодарности за выполненную работу, действия экспериментатора с опоздавшими учащимися и просьбами учеников временно покинуть место проведения тестирования, ответы на наиболее часто встречающиеся вопросы и некоторые другие процедурные вопросы по усмотрению авторов.

7.4. Требования к личности экспериментатора.

Проводить тестирование может только специально подготовленный человек. Он должен удовлетворять следующим профессиональным и личностным требованиям;

1. он должен понимать задачи эксперимента, быть компетентным в проведении теста, а если ему предстоит обработка результатов, то и в оценивании результатов;
2. он должен уметь контролировать себя, быть эмоционально уравновешенным, общительным, тактичным.

7.5. Наблюдатели.

При проведении тестов школьных достижений в форме сравнительного исследования достаточно надежным средством повышения объективности и надежности исследования служит введение института наблюдателей.

Наблюдатель — лицо фиксирующее процедуру проведения и соответствие действий исследователя сценарию тестирования. Наблюдателю, как правило, запрещается вмешиваться в процесс тестирования. Присутствие наблюдателя несомненно удорожает проведение исследования, однако при этом значительный выигрыш в качестве может компенсировать все дополнительные затраты. Ведь в случае некачественного проведения все затраты могут оказаться напрасными. Кроме этого, наблюдатели могут присутствовать не на всех процедурах тестирования, однако, сама возможность их присутствия в значительной степени дисциплинирует исследователей.

Присутствие наблюдателей позволяет сравнить качество проведения тестирования у разных экспериментаторов, в тех группах, где присутствовал наблюдатель и там где не присутствовал. Наш опыт использования института наблюдателей говорит о том, что процедурные ошибки могут добавлять до 50%

разброса в результативность выполнения тестов. Особенно важно присутствие наблюдателя на этапе апробации инструментария, поскольку оно дает возможность собрать материал для дальнейшей работы над инструментом. Цели, задачи наблюдателя, процедура его работы будут ясны из приложения 1.

Глава 8. Представление данных и их интерпретация

В данной работе мы не обсуждали тему шкалирования и проблем, связанных с этим. Достаточно подробную информацию по данному вопросу можно найти в литературе. Отметим лишь, что данная проблема хорошо разработана как теоретически, так и практически и, как правило, не вызывает трудностей у разработчиков тестов.

8.1. Формы представления полученных данных.

Помимо рассмотренных процедур, при разработке тестов необходимо предусмотреть форму представления полученных результатов, определяемую как целями тестирования, так и индивидуальными особенностями лиц, которые будут эти результаты использовать, как правило, форма представления результатов содержит в себе текст, таблицы, графики, диаграммы, профили и т. п. Для тестов достижений чаще всего используются вербальная, табличная и графическая форма представления данных.

В приведенных примерах значения округлены до целых значений. В реальной практике пользуются, как правило, вычислениями с точностью до второго знака после запятой. Однако, точность представления данных, строго говоря, определяется точностью вычислений заложенной в исследование при его проектировании, например, при проектировании выборки. Это понятно: нет необходимости указывать в данных цифры, которые лежат ниже порога ошибки.

Табличная форма представления результатов может представлять собой:

1. Значения первичных данных, например:

Результаты выполнения задания для школьников 7—х классов трех школ.

школа	класс	процент выполнения	
		по классам	по школе
209	7А	57	61
	7Б	65	
229	7А	44	66
	7Б	88	
249	7А	89	84

	7Б 7В	85 78	
Среднее по изучаемым школам			70

2. Таблицы перевода первичных оценок в школьные. Пример такого рода таблиц приведен выше, в главе 6.

3. В виде отклонений от среднего для разных по составу групп. Возьмем данные из первого примера:

школа	класс	отклонение от среднего	
		по классам	по школе
209	7А	-13	-09
	7Б	-05	
229	7А	-26	-04
	7Б	+18	
249	7А	+19	+14
	7Б	+15	
	7В	+08	
Среднее по изучаемым школам			0

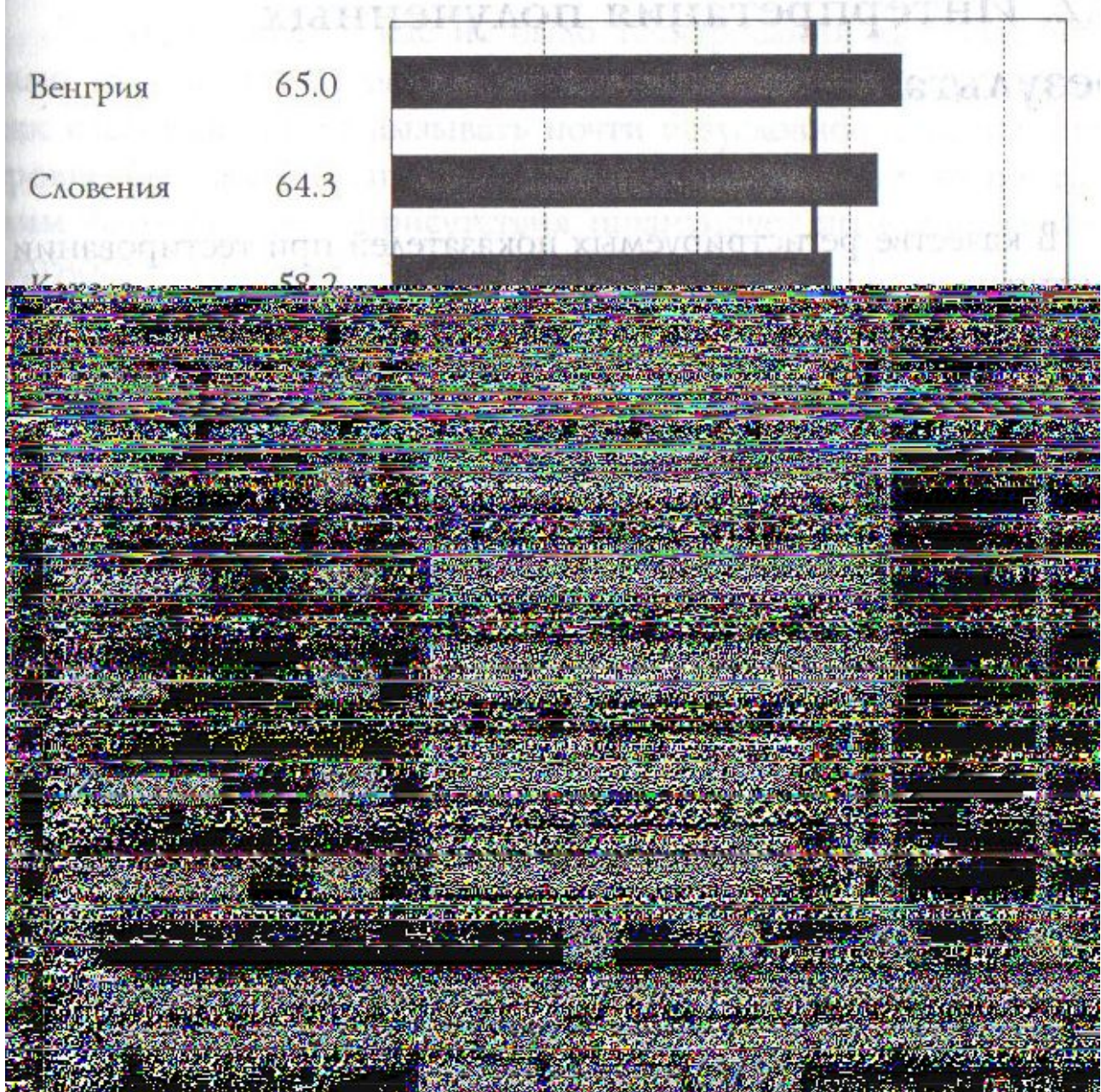
4. В виде таблиц прогноза, на основе которых осуществляется предсказание успехов в обучении, академической успеваемости или продуктивности деятельности. Пример такой таблицы приведен в разделе описания тестов, ориентированных на критерии.

Графическая форма представления данных применяется как для отдельных задач, тестов, так и для тестовых наборов, батарей тестов в виде графиков, различного вида диаграмм, профилей.

1. Графики удобно применять, когда по результатам тестирования необходимо провести предсказание какой-либо измеряемой величины, сделать прогноз. В этом случае выявляется закономерность и применяется метод аппроксимации.

2. Профили используют, как правило, для многокомпонентных тестов. Они позволяют кроме характера распределений наглядно представить вариативность и форму заданий.

3. Разной формы диаграммы используются для представления в наглядной форме оценки по тестам, Разработано значительное количество диаграмм: столбчатые, горизонтальные и вертикальные, круговые, объемные и т. д. Это позволяет облегчить процедуру классификации объектов, обобщения данных для их дальнейшей интерпретации. Диаграммы указывают, и первую очередь, порядок распределения данных. Перед представлением данных на диаграмме их, как правило, ранжируют. Приведем пример представления данных в такой форме.



В психологии применяются и другие формы представления результатов, например, алгебраическая форма линейной множественной регрессии, однако в тестах достижений практически не применяются. Это определяется низким уровнем подготовленности руководителей к восприятию специфической формы информации.

8.2. Интерпретация полученных результатов.

В качестве регистрируемых показателей при тестировании выступают: время реакций и выполнения задач, число правильно выполненных задач, число ошибок при решении теста, динамика числа ошибок во времени, оригинальность ответов, точность реагирования и т. п. На основе этих данных исследователь получает возможность судить о степени выраженности изучаемого признака. Применительно к тестам школьных достижений таковым может быть степень владения определенным кругом знаний, умений и навыков по конкретной дисциплине, для двух и многомерных тестов — это показатели, заложенные в шкалы измерения. На основе количественных результатов тестирования дается их интерпретация.

Вначале остановимся только на одном, весьма узком направлении — интерпретации результатов тестирования школьников при применении ненормированных тестов. Данный случай весьма иллюстративен.

На самом деле может существовать два взгляда на интерпретацию полученных результатов — это интерпретация разработчика, которую можно определить как субинтерпретацию в рамках используемой им терминологии, и интерпретация пользователя результатами, интерпретированных разработчиком, которые могут не совпадать. (Под разработчиком подразумевается осуществляющий тестирование, в любом случае — уполномоченный разработчика, предъявляющий пользователю интерпретацию данных).

В самой выше приведенной формулировке заложены практически три возможности, подразумевающие двойное толкование первого: 1. собственно разработчик, 2. интерпретатор (уполномоченный разработчиком и, соответственно, им подготовленный, «субинтерпретатор») и 3. пользователь, с возможностью собственной интерпретации полученных данных.

За годы, когда у нас не было тестирования как явления, само это понятие получило некую мистическую окраску и, как следствие, стало вызывать почти безусловное доверие. На уровне бытового педагогического мышления тестирование самим фактом своего присутствия гарантирует получение абсолютного и однозначного объективного результата. Одновременно трудно найти работника образования, который бы не был укорен в том, что он-то сам способен действительно объективно оценить результативность работы образовательной системы. Следствием этого стала ситуация, когда результаты тестирования либо однозначно воспринимаются как безусловные, либо немедленно подвергаются сомнению, если не совпадают с личным профессиональным мнением. Но оба случая требуют существенных оговорок. Позиция первая: тест не обязательно объективен, более того, он обязательно субъективен, и об этом будет говорить хотя бы в том случае, если собственно разработчик и уполномоченный разработчика будут разными физическими лицами. Позиция вторая: любой (и вы в том числе) педагог склонен считать объективной свою субъективную интерпретацию любого предъявленного педагогического явления.

Сутью тестирования, в определенном смысле, выступает необходимость получения некоей максимально объективизированной информации, которая выполняет функцию обратной связи в системе, и предназначена для оптимизации работы системы. Интерпретация этой информации тем не менее — набор субъективно выбираемых «оптических инструментов» для определения соответствия данной связи получаемым на выходе системы характеристикам продукта и ожиданиям уровня и объема информации, необходимым при рассмотрении задач предмета. Причем «продукт» однозначно определяется предметом, организующим данную систему. При этом связь должна иметь и психологическую, и сущностную оценку, без которой сама связь не переходит на уровень информации.

Тест, как и любая другая форма контроля учебной деятельности, не есть самоцель. Контроль является обязательной составляющей в управленческом цикле, и его функцией представляется регуляция системы с целью повышения эффективности ее работы и, как следствие, повышение как количества, так и качества продукта на выходе из системы, если такая цель вообще ставится при управлении системой. Продуктом образовательной системы, в частности, выступает приобретение учащимися усредненного и адекватного их возрастному уровню объема информации, реализующегося в учебной деятельности. (Эту информацию можно структурировать, например, как это делается в книге «Материалы к курсу общей психологии для педагогического колледжа»: 1. Смысл — значение понятия; 2. Структура явления, которое описывает понятие; 3. Его свойства; 4. функции; 5. Виды; 6. Цель; 7. Развитие, изменение; 8. Влияние, воздействие, взаимодействие, связь, отношение; 9, Условие, факторы изменения; 10. Критерии, показатели явления. Но в данной главе это опять же не является существенным).

«Усредненность» объема информации изначально детерминирована существующими учебными программами, но, поскольку в последнее время школа получила определенную свободу деятельности, и в учебный процесс хлынули новые подходы, образовательные концепции, предметы, учебники и т. д., необходимо выделить это понятие, если речь не идет о тестировании ограниченной и специфической образовательной системы. Если же мы имеем дело именно с ограниченной и специфической образовательной системой, то адекватность теста системе определяются разработчиком, который, как правило, одновременно является и пользователем результатов теста (упрощенным примером этого может быть тест, разрабатываемый учителем для своего класса). При этом тест может рассматриваться как обязательный набор условностей (конвенций) между разработчиком (тестирующим) и объектом тестирования, в каком бы виде ни был представлен этот объект, а в нашем случае объектом всегда будет выступать какая-либо образовательная система, Продолжая сужать (или умышленно упрощать) круг рассуждений, касающихся интерпретации результатов тестирования учащихся, можно абстрагироваться от всех присущих образовательной системе педагогических понятий, как то: знания, умения, навыки, интеллектуальное развитие и т. п. и остановиться только на двух «рабочих» понятиях, таких как «уровень» и «объем»

информации. Отметим при этом, что и одно и другое выступают, как правило, лишь как индикаторы тестируемых параметров, а не как собственно параметры измеряемого явления. Оговоримся, что они же могут выступать и как заложенные в тестирование непосредственные параметры. Пользователь должен четко разделять одно и другое.

Представим условно само тестирование как систему со входом и выходом, где текст теста выступает в качестве входа, а результат — в качестве выхода из системы. Все остальное — «черный ящик». Итак, система получает на входе 100% закодированной информации, но не может, по определению, получить те же 100% объема на выходе. (Если на выходе получается 100% введенной информации, то тест плох и не обеспечивает объективности результата).

Вход (тест) содержит в себе желаемый разработчиком объем информации, который, по его мнению, соответствует максимально возможному для учащегося данной возрастной группы усвоению учебного материала определенного школьного курса. В главе о целеполагании мы заметили, как множественны подходы ученых в этой сфере, а в реальной педагогике можно найти и небольшую группу учителей, где все сошлись бы в понимании связей или подчиненности понятий «бучение», «образование», «воспитание», «развитие», определении основной функции школы, или реально и объективно (не декларативно) могли бы оценить свое понимание процесса взаимодействия с учащимся. Во избежание упреков будем считать, что мы несколько преувеличиваем. Тем не менее, это имеет непосредственное отношение к интерпретации результатов тестирования,

В тесте реализуются определенные гипотетические ожидания разработчика по осуществлению заданной им информации в виде обратной связи. Сам процесс тестирования не может восприниматься иначе, как связь в системе, а связь (по крайней мере, по мнению психолога) — иначе, как понимание взаимодействующих субъектов. Понимание же при составлении теста не оговаривается сторонами, а закладывается как само собой разумеющееся только одной стороной, то есть разработчиком. Даже в том случае, если разработчиком является коллектив, каждый тест школьных достижений имеет свою специфику по ориентации ожиданий его результатом при однозначности достигнутого понимания. По данным хотя и небольшого опыта нахождения норм при тестировании им является, что гипотетический результат выполнения теста не совпадает с реальным результатом иногда в 80% случаев. Но мы примем во внимание цифру 60%, вероятно, более соответствующую действительности. И не совпадает (всегда) только в большую сторону.

Что это означает на практике, если речь идет о ненормированных тестах и о тестах вообще?

Любой составитель (разработчик) теста проецирует на некий исследуемый объект уровень ожидаемой от него информации, превышающий иногда его собственный.

Позиция третья: заложенная точка отсчета в предлагаемой шкале, как правило, завышена, то есть условная сложность теста превышает педагогические реалии.

Позиция четвертая: у пользователя одновременно завышены ожидания результатов тестирования. Причем, не собствен но ожидания, а, скорее, желаемость высокого результата, если, конечно, пользователь не мотивирован желанием утвердите н в убеждении, что тестируемая система не отвечает предъявляемым ей требованиям.

Это касается количественных показателей объема информации на выходе. Однако вне зависимости от возможных разночтений в интерпретации показателей объема, процесс тести тестирования однозначно вскроет относительное соответствие тестируемой системы заложенным в тест уровням информации.

Если имеется много «простых» заданий, покрывающих тестируемую часть учебного материала, который, повторимся, мы а чаем понятием «информация», то полученный на выходе объем, можно предположить, будет ниже ожидаемого, с одной стороны, и желаемого, с другой. При наличии же «сложных» заданий, или «простых», но относящихся к другому информационному уровню, может получиться, что тест будет выявлять только соответствие тестируемой системы информационному уровню теста, определяемому по всем системным характеристикам.

При интерпретации полученных данных, если таковые отличаются от ожидаемых пользователем результатов, надлежит установить соответствие используемого теста реалиям конкретной образовательной системы — пользователя теста, не обращая внимания на указанную разработчиком специфику данного теста. Но это вариант прямолинейный, когда имеется подсказка в необходимости внесения коррекции. В любом случае необходимо установить соответствие используемого теста реалиям конкретной образовательной системы — пользователя теста, хотя бы на уровне терминологии разработчика. Нельзя предлагаемый тест брать только потому, что разработчик называет свой продукт тестом. Даже если продукт действительно является тестом, пользователь должен убедиться, что данный продукт соответствует измеряемому параметру конкретной образовательной системы. Предложенное утверждение относится к случаям использования одного теста или батареи тестов в одной образовательной системе. Что это означает? В случае получения различных результатов тестирования в блоке (различные результаты тестирования по разным предметам в одном классе), следует делать поправку в интерпретации на возможность различий в ориентации субтестов (отдельных и автономных тестов в блоке) и различия в ожидаемых разработчиком результатах тестирования. Тем не менее, использование отдельного теста на уровне индивидуальных результатов однозначно даст возможность для сравнения вне зависимости от установленной статистической нормы. Сама разница результатов в этом случае является основанием для сравнения. Вторичное сравнение результатов тестирования с показателями текущей успеваемости может дать определенный материал для интерпретации полученных данных.

Другим способом оптимизации может стать параллельное тестирование другой образовательной системы с целью получения результата для сравнения. Точка

отсчета в таком случае может выбираться произвольно, но не может быть определено соответствие результата норме. (Напомним, что тест является стандартизированной процедурой по определению).

В случае параллельного тестирования с использованием отдельного теста, при интерпретации результатов необходимо учитывать целый ряд способных воздействовать на него факторов: санитарно-гигиенического, психологического (эмоционального — выделим внутри психологического), организационного. Не будем на этом останавливаться особо, т. к. это не является предметом предлагаемой книги; для примера, однако, упомянем: место теста в расписании; урок, предшествующий тестированию; урок данного класса в расписании работы учителя в случае, если параллельное тестирование проходит в классах одного учителя. Последнее представляется стоящим более подробного рассмотрения. В ежедневном расписании работы учителя часто получается, что один класс из параллели всегда оказывается первым, а другой — вторым. Однако в силу индивидуальных особенностей учителя (алгоритм вработываемости, личные симпатии к учащимся, возникающие в результате их успешности и, в свою очередь, дополнительно мотивирующие учителя, алгоритм работы с планом урока и способность внесения текущей коррекции и т. д.) один из уроков «сильный», а другой — «слабый». При этом субъективная оценка уроков учителем может не совпадать с реальной. Это может привести к тому, что результаты даже стандартного параллельного тестирования будут нуждаться в коррекции при интерпретации. Но сам, тестируемый на профессиональную эффективность педагог (или он же — разработчик) не может быть ни инициатором, ни тем, кто может осуществить данный вид коррекции. Еще одним случаем необходимости внесения поправок в интерпретацию, а речь идет не только о различиях в интерпретации пользователя и разработчика, но и об интерпретации самого разработчика, может стать несоответствие заложенных в тест и ожидаемых пользователем уровнях информации. В действительности это означает следующее: тест, в силу заложенных в технологию его составления ограничений, оказывается на информационном уровне, отличном от информационного уровня, на котором функционирует образовательная система-пользователь. Мало того, при совпадении информационных уровней, в идеале, в тест закладывается информация, нуждающаяся в подинтерпретации уполномоченным разработчиком в силу абсолютной относительности психологических оценок измеряемой связи. (Если разработчик оригинала и интерпретатор результата не являются одним лицом, то необходимо, чтобы интерпретатор, помимо взаимосогласованного аппарата декодирования информации, имел и соответствующие ключи декодирования. А это практически невозможно на индивидуальном уровне, поскольку даже при согласованности по всему алгоритму интерпретации таковая останется субъективной. На статистическом же уровне объективность как результатов, так и интерпретации результатов вполне достижима.)

В результате интерпретация разработчика отличается качественно, а количественные показатели оказываются только индикатором этого

качественного различия. Сопоставление смысловых нагрузок в заданиях теста, а точнее, уровня среза информации с тем, что имеет место в реалиях системы-пользователя, может в определенных случаях привести к выводу о неадекватности теста системе. Речь здесь не идет о сложности заданий. Речь идет о разночтениях в значимости общепринятой единицы информации — факте: в тест закладывается материал именно данного курса, но важность различных его частей по целеполаганию разработчика и пользователя определяется по-разному.

При параллельном тестировании возможен вариант, когда результаты на выходе в их количественном выражении описывают в одном случае объем информации, а в другом случае являются именно индикатором качественного несовпадения заложенного в тест уровня информации и информационного уровня системы.

В данном случае мы не говорим о качественной интерпретации результатов тестирования, которая предполагает объяснение, толкование, опирающееся на теоретическую базу, положенную в основу теста, и может предусматривать сравнительный анализ. В руководствах к тестам должна даваться их качественная интерпретация, но для тестов достижений она невелика. При интерпретации результатов по тестам достижений, и, в частности, школьных достижений, за основу может браться содержательный практический критерий валидности. Так, если академическая успешность учащегося по химии соответствует успешности выполнения им теста по химии, то оценки по тесту могут служить основанием для вывода об уровне знаний, умений и навыков в данной области, но только в том случае, повторимся, если тест валиден и надежен. При интерпретации нормированных тестов наличие нормы облегчает интерпретацию, поскольку в случае стандартизированного и нормированного теста, норма представляет собой результат, который и является объективным в силу заложенной в него выборки. Другими словами, если результаты, полученные при тестировании, не соответствуют интерпретации пользователя, то нельзя субъективно вводить поправку на субъективность теста, т. к. данный тест по нормированию представляет результаты множественного исследования с уже введенной поправкой на возможные различия в гипотетических и реальных результатах. Здесь выборка и полученные на ней результаты и являются той «объективной» точкой отсчета шкалы, которую следовало бы корректировать при ненормированном исследовании. Сама форма представления говорит о том, существуют ли различия между полученными и нормированными результатами, и, если таковые существуют, каков знак этих различий. В случае, если результаты интерпретатора отличаются от желаемых, следует корректировать, в первую очередь, а свою субъективную оценку, а не пытаться откорректировать возможную интерпретацию разработчика. Если тест по химии дал результаты, которые были менее ожидаемы пользователем, но подтверждены нормированием теста, то именно нормированная оценка по каждому показателю может служить основой для интерпретации, а не субъективная оценка пользователя, тем более, если она не подтверждена (а это

делается особым образом) статистическими данными по (и при) данному явлению тестирования.

Интерпретация результатов требует от исследователя высокой квалификации. Мы выяснили, что на оценку могут влиять субъективные установки, принятые стереотипы и шаблоны оценки как предмета, так: и поведения, состояние, степень адекватности оценки и т. п. С этим связаны основные ошибки при интерпретации результатов тестирования. Часто погрешности связаны с неоправданно широким толкованием полученных тестовых оценок.

Интерпретация результатов для сложных педагогических систем сложна. Отметим некоторые возможные ошибки.

1. Проведение измерений и использование результатов измерений процессов и явлений без учета особенностей их развития. Первый пример — попытки измерения динамических характеристик с последующим формулированием выводов прогнозирующего характера на основе фиксации положения на момент измерения. В качестве примера приведем динамическую характеристику, измерение которой реально дает ошибочные выводы. Это способности, особенно в раннем возрасте, и использование результатов измерения способностей для формирования классов разного уровня преподавания. При этом забывается, что «... способность по самому своему существу есть понятие динамическое. Способность существует только в движении, только в развитии. В психологическом плане нельзя говорить о способности, как она существует до начала своего развития, так же как нельзя говорить о способности, достигшей своего полного развития, закончившей свое развитие.» (Б. М. Теплое).

2. Отсутствие учета вероятностного характера измерений и результатов в социальных системах вообще и в педагогических системах в частности, и многофакторность при описании образовательных систем. Измерениям вообще присущ вероятностный характер, который в области материальных объектов характеризуется точностью измерения, погрешностью измерения, округлением и целым рядом статистических характеристик. Что касается измерений социальных систем, объектов, фактов, то кроме вероятностного характера самих измерений вероятностный характер носят показатели, которые принимаются к измерению, и единицы измерения.

3. Невнимание к проблемам детерминации того или иного явления или факта, к отделению влияния фактора педагогического от влияния иных социальных факторов, обуславливающих то или иное явление. Другими словами, к вопросу: что или кто определяет тот или иной педагогический результат или явление. Вопрос этот достаточно сложный, но без его решения невозможно решение важных для образовательных систем вопросов: что в состоянии выполнить образование, а что нет; за какие результаты образовательной деятельности школа должна нести ответственность; какие из функций, присвоенных школе исторически, ей не свойственны; какие условия необходимо создать и какие преобразования провести, чтобы те или иные задачи стали посильны образовательной системе. Отсутствие четких разграничений приводит к

деформациям в деятельности образовательной системы, фальсификациям, имитации деятельности.

4. Невнимание к проблемам качества проведения измерений и качества существующих измерительных инструментов. К сожалению, используемые в настоящее время измерительные инструменты не отвечают даже минимальным требованиям по качеству. Причем недостатки качества инструмента и методики измерения можно отметить на всех этапах измерения — от построения выборки через способы предъявления до интерпретации данных. Сами понятия надежности и валидности не адаптированы применительно к измерениям в педагогике, а попытки перенесения способов определения показателей о качества из психологии вялы и, как правило, низкопрофессиональны.

5. Проблемы стандартизации в области измерений в педагогике. При всей обширности нерешенных проблем стандартизации измерения необходимо зафиксировать три основных направления:

Первое — создание стандартизированных инструментов для измерения в педагогике, иными словами, объективных инструментов, которые отвечали бы основному требованию — результаты, получаемые при применении этого инструмента, не должны зависеть от личности исследователя, и этот факт должен быть соответствующим образом доказан.

Второе — получение некоторых норм в качестве основания для сравнения. Спор о том, должны быть эти нормы статистическими или критериальными, должны они носить социальный или специфический характер, способы их получения и прочее не представляется существенным. Для практика важен факт существования каких-то статистически и метрологически обоснованных норм для их использования при интерпретации собственных данных.

Третье — это выбор формы проведения исследования в качестве стандартной. Возможны две альтернативы — мониторинг и сравнительные исследования. Повсеместное увлечение мониторинговыми исследованиями проходит. При массе привлекательных сторон этот вид организации исследования обладает одним недостатком — максимальная эффективность применения в закрытых системах. Педагогические системы по своей сути — системы открытые и количество и качество информации, получаемое в сравнительных исследованиях, при одинаковых ресурсных затратах, выше. Время для массового осознания этого факта в педагогике еще не пришло.

6. Несоответствие затрат на измерение полученному результату. В данном случае под затратами подразумеваются два их вида — ресурсные и временные. Если с ресурсными ситуация достаточно ясная и прогнозируемая, то временные затраты зачастую служат преградой для получения ценных данных, имеющих реальную, а не историческую ценность.

Сложности в этой области связаны с двумя причинами. Во-первых, в последнее время значительно возросла динамичность развития образовательных систем, причем эта динамичность достаточно стабильна. В связи с этим проведение громоздких и сложных измерений приводит к тому, что результаты, полученные через год-полтора, что, в общем-то, не такой уж большой срок, уже не представляют ценности. Вторая причина связана с переносом методов

научного исследования, которые разрабатываются, как правило, для проведения диссертационных исследований, в область прикладных исследований — результат зачастую достаточно плачевный: добросовестные измерения неприменимы в силу временных ограничений.

Как указывают В. К. Гайда и В. П. Захаров, при интерпретации результатов необходимо соблюдать сдержанность и осторожность и, по возможности, стремиться к снижению произвольности в толковании, для чего следует опираться только на проверенные, обоснованные факты. При интерпретации можно опираться на следующие общие правила:

— сохранение терминологии, используемой разработчиком теста в руководстве;

— применение описаний и трактовок разработчика только для тех выборок, для которых определены нормы;

— любое отклонение интерпретации пользователя от интерпретации разработчика должно быть обосновано статистически.

Мы согласны с этими правилами.

Приложение 1

Инструкция и анкета наблюдателя

Наблюдателем может быть учитель, привлеченное лицо, не работающее в классе, где проводится тестирование учащихся, прошедший специальную подготовку и инструктаж.

Наблюдатель следит за ходом и качеством проведения тестирования, фиксирует все происходящее во время его проведения. Наблюдателю запрещено вмешиваться в ход работы, он лишь фиксирует и описывает процедуру проведения тестирования.

1. Инструктаж наблюдателей.

Для качественного выполнения работы все наблюдатели должны принять участие в совещании, на котором будет дан подробный инструктаж и описание работы наблюдателей. Вы получите информацию о школах, которые должны посетить и времени проведения в них тестирования.

2. Телефонные звонки.

За неделю до проведения тестирования Вы должны позвонить в школу и уточнить у исследователя его дату и время, а также получить информацию о том, как добраться до школы. Объясните исследователю, что ему следует планировать проведение тестирования независимо от того, будет присутствовать наблюдатель или нет.

3. График посещения школ.

Наблюдатель должен быть в школе более, чем за час до начала тестирования. Это необходимо для того, чтобы отметить, что папка с тетрадями была вскрыта только за 1 час до начал работы.

Если в школе участие в тестировании принимают 2 класса, то оно может проводиться двумя исследователями одновременно в двух разных комнатах или одно за другим в разное время, Если тестирование проводится одновременно в

двух разных комнатах, то наблюдатель посещает только одно из них. В случаях проведения работы в разное время наблюдатель посещает оба занятия.

Наблюдатель остается в школе до тех пор, пока исследователь не соберет и не упакует все материалы. Только после этого он может взять интервью у школьного координатора.

В случае, если в тестировании участвовало менее 85% учащихся, школьный координатор должен назначить дату дополнительного тестирования.

Наблюдатель должен посетить и дополнительное тестирование, заполнив при этом и дополнительную анкету. (Посетить школу во время дополнительного тестирования может и другой наблюдатель.)

Основное внимание следует обратить на:

— исключение возможности ознакомления с содержанием тестов до тестирования;

— точное следование процедуре тестирования;

— соблюдение нормы времени, отведенного на выполнение каждой части теста;

— соблюдение правил ответов учителя на вопросы учащихся.

В процессе наблюдения заполняется анкета наблюдателя.

Вскрытие пачки с тестовыми тетрадями проводится в присутствии наблюдателя,

Анкета наблюдателя. Инструкция по заполнению анкеты.

Наблюдатель получает одну анкету на класс, в котором проводится тестирование. При заполнении анкеты указывается следующая информация:

1. Код класса. В исследовании принята 9—значная система кодирования.

Графа 1. Указывается наличие или отсутствие официального статуса (гимназии, лицея, комплекса, авторской школы и т. д.) образовательного учреждения:

S — статус утвержден в городе;

R — статус утвержден с районе;

N — статус отсутствует.

Графа 2. Фиксируется расположение образовательного учреждения:

A — центр города;

B — промышленно-жилые районы;

C — спальные районы.

Графы 3—5, Указывается номер школы. Если номер одно- или двузначный, то перед номером необходимо добавить соответственно 2 или 1 ноль.

Графы 6 и 7. Указывается номер параллели (класса), если он однозначный, то перед номером добавляется 0.

Графа 8. Заполняется индекс класса в цифровой форме. -, образовательном учреждении принята буквенная индексация, то ее необходимо перевести в цифровую:

A — 1

Б - 2,
В - 3,
Г - 4,
И т.д. .

Графа 9. Указывается код профиля класса:

Ф — физико-математический;
Е — естественнонаучный;
Г — гуманитарный;
Р — профессиональной направленности;
N — без профиля.

Пример:

1	2	3	4	5	6	7	8	9
S	B	0	7	9	1	0	2	F

имеет городской статус

промышленно-жилой район

79я школа

10й класс

Б (2й) класс

физико-математический профиль

2. Дата и время тестирования.

3. Тип тестирования (возраст и предметы) и порядок (прямой или обратный) проведения тестирования.

В анкете наблюдатель отмечает всю процедуру проведения тестирования.

Анкета разделена на три столбца. В первом столбце дается вопрос, во втором — предполагаемые варианты ответа, в третьем оставлено место для необходимых объяснений.

Приведенная анкета и инструкции предназначена для заполнения во время тестирования учащихся.

Сценарий, который исследователь будет использовать во время тестирования, находится в конце руководства для исследователя. Для облегчения наблюдения весь сценарий разбит на части.

Для исследователя является обязательным чтение сценария «слово в слово», так же как и для Вас является обязательным наблюдение за тем, как исследователь следует сценарию. Важность выполнения этой требования для всех школ,

участвующих в исследовании, неоднократно подчеркивалась во всех материалах исследования. При оценке того, как исследователь следует сценарию, вам предлагается выбрать ответы:

А. Дословно.

В. С небольшими добавлениями.

С. Много добавлений (Объясните).

Д. Пропуски или свободная интерпретация (Объясните).

Приведем некоторые разъяснения к этим ответам.

ДОСЛОВНО: Исследователь прочитал Сценарий «слово в слово», не опуская ни одного слова и ничего не добавляя. Непреднамеренные ошибки при чтении не учитываются.

С НЕБОЛЬШИМИ ДОБАВЛЕНИЯМИ: Сценарий был прочитан полностью, без пропусков, но исследователь сделал небольшие отступления или повторил некоторые моменты, чтобы обратить на них внимание учащихся.

МНОГО ДОБАВЛЕНИЙ: Сценарий был полностью прочитан без пропусков, но исследователь сделал довольно-таки существенные добавления к тексту.

Наблюдателю следует отметить в анкете суть этих добавлений.

ПРОПУСКИ/СВОБОДНАЯ ИНТЕРПРЕТАЦИЯ: Исследователь пропустил часть сценария (намеренно или ненамеренно) или пересказал его своими словами. Наблюдатель должен в анкете отметить суть этих изменений.

Запомните, пожалуйста, вам не следует вмешиваться в ход | тестирования, даже если исследователь делает пропуски, добавления или значительно изменяет сценарий.

ХРОНОМЕТРАЖ

При выполнении всех частей теста вам следует осуществлять хронометраж работы. Наиболее существенный элемент процедуры тестирования заключается в том, что учащимся всех школ и классов дано одинаковое время на выполнение отдельных частей теста.

Для каждой части тестирования отмечайте в 3 столбце время, когда исследователь сказал учащимся, что нужно приступить к работе. Подсчитайте время, в течение которого учащиеся работали над данной частью. Эти данные также запишите в столбец 3.

Отмечать время следует с точностью до 1 мин. Для этого лучше всего иметь ручные часы с секундной стрелкой.

Помните, что вам не следует вмешиваться в ход работы даже если хронометраж будет нарушен.

Отвечая на этот вопрос о поведении учащихся, просим Вас обобщить свои наблюдения за поведением учащихся. Хорошему поведению учащихся соответствует минимальный уровень их разговоров и ответственное отношение к работе. Если небольшая группа учащихся иногда разговаривает, но их разговоры не мешают работе класса, то можно выбрать ответ «ДА».

Если Вы наблюдали нарушения дисциплины, то ответьте, предпринимал ли исследователь что-либо для того, чтобы успокоить учащихся (например, делал ли им замечания).

В руководстве для исследователя записано, что он не должен отвечать на вопросы учащихся, касающиеся содержания заданий после того, как они приступили к работе. Исследователь не должен давать учащимся дополнительную информацию или разъяснения на какие-либо вопросы или помогать учащимся в чтении или написании слов. Если ученик задал вопрос, то наилучшим ответом школьного координатора может быть следующий:

«Извините, пожалуйста, я не могу отвечать ни на какие ваши вопросы. Постарайтесь сами справиться с заданием». Можно отвечать на вопросы о том, что делать дальше и как записать ответ. Возможна ситуация, когда учащиеся сами не задают вопрос, но исследователь начинает обсуждать какое-либо задание. Очевидно, что он намеревается оказать некоторую помощь учащимся. В этом случае следует выбрать ответ «Нет».

Ответьте, собирал ли исследователь тетради по одной, проверяя при этом, правильно ли заполнена обложка.

Запишите время, когда Исследователь закончил сбор тетрадей.

Анкета проведения тестирования.

Наблюдения	Ответы	Объяснения
Отметьте время, когда исследователь начал читать Сценарий проведения тестирования		
Следовал ли исследователь сценарию во время «ВВЕДЕНИЯ»?	А. Дословно В. С небольшими добавлениями С. Много добавлений (Объясните) Д. Пропуски/ свободная интерпретация (Объясните)	
Следовал ли исследователь Сценарию при РАЗДАЧЕ ТЕТРАДЕЙ?	А. Дословно В. С небольшими добавлениями С. Много добавлений (Объясните) Д. Пропуски/ свободная интерпретация (Объясните)	

Далее анкета продолжается по всем пунктам сценария, включая, если это необходимо, оценку поведения детей и исследователя во время перерыва. Последняя часть анкеты наблюдателя выглядит так:

Хорошо ли вели себя учащиеся во время тестирования?	А. Да В. Нет (Объясните)	
Если нет, то предпринимал ли школьный координатор попытку контролировать учащихся?	А. Да В. Нет	
Ходил ли школьный координатор по классу для того, чтобы проверить, над какой частью работают учащиеся?	А. Да В. Нет (Объясните) С. Не мог ходить по классу	
Правильно ли школьный координатор выполнял правила, как отвечать на вопросы учащихся?	А. Да В. Нет (Объясните) С. Вопросов не задавали D. Отвечать на вопрос не мог	
Проверял ли школьный координатор, правильно ли заполнена обложка тетради?	А. Да В. Нет (Объясните)	

Запишите время окончания работы.

Приложение 2

Основные трудности внедрения тестовых методик в практику работы образовательных учреждений

Представленные ниже трудности внедрения тестовых методик были выделены в результате выполнения эксперимента по сдаче экзаменов по единым тестам. Их наличие необходимо учитывать тем, кто начинает работы по тестированию учащихся с целью аттестации учащихся.

1. Кадровая. Разработка тестов является сложной междисциплинарной задачей. Для разработки и применения тестов необходимо участие специалистов более 12 профессий, часть из которых для системы образования нетрадиционны. В настоящее время слаба информированность педагогов о том, что такое тесты и тестирование, часто словом «тесты» называют тестовые задания или просто вопросы множественного выбора.

Для составления заданий методистам и учителям необходим набор знаний о возможных формах заданий и правилах их составления. В данной связи встал вопрос о стандартизации форм тестовых заданий и инструкций к ним.

Отсутствие стандарта приводит к значительным потерям времени при занесении заданий на магнитный носитель.

Неожиданно встал вопрос о слабой профессиональной подготовке педагогов. От 10 до 60 процентов заданий (по мнению методистов-экспертов), составленных педагогами, имели содержательные и фактические ошибки. Кроме этого, около 30 процентов заданий было отбраковано из-за тестологических недостатков.

2. Согласования. Вытекает из первой. Для организации совместной, деятельности такого количества специалистов необходим механизм согласования и координации.

3. Проблема заинтересованности в объективности. Основное отличие тестовых методик — объективность, В объективных результатах заинтересованы государство, общественность, однако на уровне конкретных групп может быть против определенная часть родителей, педагогов, учащихся, управленцев, которых современное состояние дел вполне устраивает, Сказать что-либо определенное о размере этой части на сегодня не представляется возможным. Однако предположительно это достаточно большие группы.

4. Психологические особенности оценивания педагогов. Педагоги оценивают учащихся без дифференциации на составляющие оценок, дают обобщенную, зачастую немотивированную оценку. Применение тестов дифференцирует оценивание, Например, ученик может не обладать конкретными знаниями, однако иметь высокий уровень способностей или высокий уровень сформированное-то общих или специальных навыков. Несовпадение оценок педагогов и результатов тестирования не может способствовать его восприятию.

5. Отсутствие внутренних стандартов. Отсутствие единых форм представления заданий, единых стандартов составления и процедур разработки не позволяет сравнивать результаты тестирования разработанные разными авторами, в разных организациях.

6. Отсутствие исторического опыта, опыта обучении тестированию.

7. «Тоталитарность» мышления. Эта причина связана с существованием нескольких достаточно устойчивых стереотипов в сознании как педагогов, так и практически большинства работников образования. Основные из них: — «все или ничего» — для получения какого либо достижения в обучении и в жизни необходимо сделать все, а не максимально возможное. Реальным воплощением этого принципа в педагогике является применение контрольных работ, где необходимо справиться со всеми заданиями, а если справляешься не со всеми, то это уже плохо. Тестирование предполагает выполнение посильной части заданий, и положительная и даже максимальная оценка может быть получена за выполнение части заданий;

— конкретный характер мышления, когда ставится пример, доказательность которого максимальна. Противоположным этому может быть статистический характер мышления — где доказательна статистическая часть, совокупность, а не отдельный пример.

8. Временные. Разработка тестового инструментария чрезвычайно длительный процесс. Разработка одного комплекта тестирования предполагает (по опыту национальных организаций Великобритании и США) от 2 до 5 лет, комплекта на все классы от 5 до 7 лет. Подготовка инструментария международных исследований занимает 4—5 лет, непонимание этого управленцами может привести к подготовке и использованию некачественных материалов.

9. Технические. Для проведения работ по массовому тестированию необходим набор технических средств, часть их которых в нашей стране не производится, например оптический сканер.

10. Финансовые. Разработка одного комплекта тестов по международным оценкам стоит от 20 тысяч долларов США. Стоимость оптического сканера — 30 тыс. ф. ст.

11, Психологические. Достаточно большое количество педагогов относится к тестам неадекватно. По этому отношению можно выделить три группы педагогов.

Первая относительно небольшая не воспринимает тестирование как таковое, считая существующие способы достаточно качественными и их удовлетворяющими.

Вторая считает тестирование единственно верным способом, который при правильном применении решил не только задачу контроля, но и многие другие школьные проблемы.

I *Приложение 2. Основные трудности внедрения тестовых методик...* 266

Третья — те кто считает, что тесты они применяют давно и постоянно, о тестах и тестировании им все известно и ничего нового и полезного им сообщено быть не может.

Все эти три позиции в равной мере вредят внедрению тестирования в работу образовательных учреждений, но, что естественно, по-разному и, как следствие, требуют разных корректирующих воздействий.

Приложение 3

Руководство для исследователя

1. ВВЕДЕНИЕ.

1.1. Основные сведения об исследовании.

1.2. Обязанности школы.

2. ПОДГОТОВКА СПИСКА УЧАЩИХСЯ.

2.1. Списки учащихся.

3. ПОДГОТОВКА К ТЕСТИРОВАНИЮ.

3.1. Составление графика проведения тестирования.

3.2. Проверка материалов тестирования.

3.3. Дополнительное оборудование.

4. ПРОВЕДЕНИЕ ТЕСТИРОВАНИЯ.

4.1. Подготовка.

- 4.2. Сценарий проведения тестирования.
- 4.3. Опоздавшие ученики.
- 4.4. Правила наблюдения за работой.
- 4.5. Сбор тетрадей.
- 4.6. Заполнение протокола проведения тестирования.
- 5. КОНТРОЛЬ ЗА КАЧЕСТВОМ ПРОВЕДЕНИЯ ТЕСТИРОВАНИЯ.
- 5.1. Процедура контроля.
- 6. ВОЗВРАЩЕНИЕ МАТЕРИАЛОВ ТЕСТИРОВАНИЯ.
- 6.1. Проверка и отправка материалов.

1. ВВЕДЕНИЕ

1.1. Основные сведения об исследовании

Целью исследования является оценка подготовленности школьников по биологии, географии, физике, химии, а также сформированность навыков учебной деятельности.

В тестировании примут участие школьники 10—х классов. Им будут предложены тестовые тетради с заданиями по вышеназванным предметам. На выполнение всех заданий учащимся потребуется 1 час. Проведение тестирования, включая раздачу материала, чтение инструкции, перерывы между выполнением отдельных частей теста и сбор тетрадей, занимает 1,5 часа. Исследование проводится исследователем в соответствии с процедурой, одинаковой для всех школ.

Тестирование будет проводиться «...» 199.. г.

1.2. Обязанности школы

Администрация школы выделяет ответственного за подготовку и проведение тестирования исследователя. В его задачи входит:

- составление списков учащихся, принимающих участие в тестировании;
- проверка наличия всех материалов, необходимых для проведения тестирования, и подготовка дополнительных материалов: карандашей, ластиков, часов (таймеров, секундомеров) и пр.;
- проведение тестирования в соответствии с инструкцией, приведенной в данном руководстве;
- заполнение протокола проведения тестирования;
- заполнение анкеты для школы;
- проведение дополнительного тестирования, если это будет необходимо;
- сбор всех материалов тестирования, включая и неиспользованные тетради, возвращение их официальному лицу, отвечающему за проведение исследования в данной школе.

2. ПОДГОТОВКА СПИСКА УЧАЩИХСЯ

2.1. Списки учащихся

Списки учащихся составляются на специальных бланках, которые получает каждая школа, принимающая участие в исследовании. Форма бланка и инструкция по его заполнению приведены ниже.

Исследователь заполняет бланки, внося в них фамилии и имена всех учащихся классов, принимающих участие в тестировании. Для каждого ученика указывается месяц и год рождения, пол, класс и индекс класса (а, б и др.). Списки учащихся составляются по алфавиту.

СПИСОК
учащихся 10 — класса _____ школы
_____ района

Название школы: _____
адрес: _____
телефон: _____

№ п/п	Фамилия и имя ученика	Месяц, год рождения	Примечание
1			
2			

2.2. Инструкция по составлению списка учащихся

1. Укажите в соответствующей строке индекс класса (а, б, если в школе несколько 10—х классов принимают участие в исследовании), № школы, наименование района.
2. Ниже впишите название школы (общеобразовательная, гимназия, лицей, авторская школа, экспериментальная площадка, учебно-территориальный комплекс и др.), адрес и телефон школы.
3. Впишите фамилии и имена учащихся 10—х классов, принимающих участие в исследовании, в алфавитном порядке.
4. Для каждого ученика, внесенного в список, укажите месяц и год рождения.

3. ПОДГОТОВКА К ТЕСТИРОВАНИЮ

3.1. Составление графика проведения тестирования

Исследователь должен определить дату проведения тестирования и согласовать ее с директором школы.

Для выполнения теста потребуется 60 минут, а проведение всей работы, включая раздачу тетрадей, чтение инструкции, перерывы и сбор тетрадей займет 1.5 часа.

Исследователь должен сообщить дату и время проведения тестирования официальному лицу. Если впоследствии произойдут какие-либо изменения в графике, необходимо немедленно сообщить об этом официальному лицу, отвечающему за проведение исследования. Это особенно важно в связи с тем, что в школах во время тестирования будут присутствовать наблюдатели за ходом и качеством проведения тестирования.

Учащиеся школы, которые будут участвовать в тестировании, должны знать о времени и месте его проведения.

3.2. Проверка материалов тестирования

За несколько дней до проведения тестирования школьный координатор получит следующие материалы: — тетради с текстами заданий; — анкета для школы; — ^ протокол проведения тестирования; — руководство для школьного координатора.

Если школа не получила все эти материалы в указанные сроки, то они будут переданы школьному координатору накануне или в день тестирования не менее, чем за 2 часа до его проведения (об этом будет дополнительно сообщено школе официальным лицом, отвечающим за проведение исследования),

После получения материалов исследователь должен проверить их и убедиться в том, что все указанные выше материалы находятся в комплекте. Не разрешается вскрывать пачку с тетрадями

3.3. Дополнительное оборудование о Б дополнение к полученным материалам исследователь должен иметь надежный секундомер для фиксации времени при тестировании, а также дополнительные карандаши и ластик для учащихся.

3.4. В материалы для тестирования включена анкета для школы, с помощью которой будет получена информация об учащихся, а также о программах по предметам. Эта анкета должна быть отдана официальному лицу после получения материалов. Заполненная анкета возвращается с комплектом тетрадей после тестирования.

4. ПРОВЕДЕНИЕ ТЕСТИРОВАНИЯ

4.1. Подготовка

До начала тестирования школьный координатор должен проверить помещение, в котором оно будет проводиться. В помещении должно быть достаточное количество парт для всех учащихся. Желательно, чтобы учащиеся сидели по одному за партой для избежания списывания.

Исследователь должен иметь следующие материалы для проведения тестирования:

- сценарий проведения тестирования (приводится в приложении);
- тетради для тестирования;
- списки учащихся;
- протокол проведения тестирования;
- надежный таймер или секундомер;
- дополнительные карандаши и ластик. Исследователь не должен открывать упаковку с тетрадями раньше, чем за 1 час до начала тестирования. Открыв упаковку, координатор должен пересчитать тетради и записать число полученных тетрадей в протокол проведения тестирования (см. 4.6).

4.2. Сценарий проведения тестирования

Исследователь должен проводить тестирование в соответствии со сценарием, который дается в Приложении 1 данного руководства.

Сценарий включает:

- введение перед тестированием;

- инструкцию по распределению тетрадей;
- инструкцию для учащихся (о том, как выполнять тест), взятую из тетрадей: инструкция должна быть громко прочитана вслух;
- распределение времени на выполнение теста.

Текст сценария должен быть полностью прочитан с тем, чтобы обеспечить одинаковые условия проведения тестирования во всех школах. Исследователь должен потренироваться в чтении сценария до начала проведения тестирования.

Во время тестирования исследователь должен вести протокол проведения тестирования. В протоколе отмечается время начала тестирования, начала чтения инструкции, начала и конца выполнения каждой части теста (см. 4.6.).

4.3. Опоздавшие ученики

Если ученик опоздал, а тестирование еще не началось, то исследователь выдает ученику тетрадь и помогает ему заполнить обложку тетради. Если ученики уже приступили к выполнению заданий или учитель читает инструкцию по выполнению заданий, то опоздавший ученик не допускается к работе.

4.4. Правила наблюдения за работой Исследователь должен соблюдать следующие правила во время тестирования:

— Нельзя отвечать на вопросы учащихся, когда они уже приступили к выполнению теста. Нельзя сообщать им никакой дополнительной специальной информации, давать ответы; на вопросы или инструктировать их, а также помогать в чтении и написании слов. Наилучшим ответом на любую просьбу; ученика может быть следующий: «Извините, пожалуйста. Я не могу отвечать ни на какие ваши вопросы. Постарайтесь сами справиться с заданием».

— Необходимо убедиться в том, что все учащиеся понимают, что они должны делать и знают, как: записать свои ответы.

— Во время работы учащихся следует перемещаться по классу и наблюдать, правильно ли учащиеся заполняют тетради и отвечают ли на вопросы той части, над которой работает весь класс. Номер части напечатан на каждой странице тетради вверху.

— Не следует заставлять учащихся участвовать в тестировании. Их следует заинтересовать в этом. Если ученик отказывается участвовать в тестировании до его проведения или во время его, напомните ученику, что его участие очень важно. Если ученик отказывается работать, уже начав тетрадь, заберите эту тетрадь и напишите на ее обложке «Отказ».

— Если ученик просит разрешения выйти из класса, это разрешается только в исключительных случаях. Если ученик не может закончить работу (например, почувствовал себя плохо), возьмите у него тетрадь и напишите на ее обложке причину прекращения работы. Если ученик вышел из класса, а потом вернулся, запишите на обложке тетради время его ухода и возвращения. Во всех случаях эти тетради следует рассматривать как «заполненные».

— Если при выполнении работы ученик обнаружил дефекты в тетради, замените ее другой тетрадью. Выясните, в чем состоит дефект, запишите номер новой тетради на обложке тетради с дефектом. После окончания работы, вложите новую тетрадь в тетрадь с дефектом.

— О любом необычном происшествии во время тестирования следует сделать запись на обложке тетрадей после их сдачи.

4.5. Сбор тетрадей

Собирайте тетради по одной. Взяв тетрадь, проверьте ее обложку и убедитесь в том, что на ней указана вся необходимая информация.

4.6. Заполнение протокола проведения тестирования.

Во время тестирования исследователь заполняет протокол проведения тестирования. Пример протокола приведен в Приложении данного руководства. Дополнительные экземпляры протоколов будут переданы школе вместе с другими материалами,

Протокол заполняется во время тестирования, в нем необходимо отметить время начала и конца работы над каждой частью. После тестирования надо подсчитать число заполненных тетрадей, занести эти данные в протокол. Кроме того, необходимо отметить проблемы, которые возникли при проведении тестирования,

5. КОНТРОЛЬ ЗА КАЧЕСТВОМ ПРОВЕДЕНИЯ ТЕСТИРОВАНИЯ

5.1. Процедура контроля

Очень важно, чтобы во всех школах во время тестирования выполнялись одни и те же требования, особенно касающиеся:

— исключения возможности ознакомления с содержанием тестов до тестирования;

— точного следования процедуре тестирования;

— соблюдения правила, как отвечать учащимся на их вопросы.

В связи с тем, что в исследовании придается большое значение качеству проведения тестирования, то в исследуемых школах будут присутствовать наблюдатели. Они должны быть в школе за час до проведения тестирования. Наблюдатели не имеют права вмешиваться в процесс работы. Их основные задачи — наблюдение за тем, как проходит тестирование и беседа с исследователем после его проведения.

Для организации посещения школ наблюдателями очень важно, чтобы школьный координатор сообщил точные данные о проведении тестирования в школе официальному лицу, ответственному за проведение тестирования. , , : , ;

6. ВОЗВРАЩЕНИЕ МАТЕРИАЛОВ ТЕСТИРОВАНИЯ

6.1. Проверка и отправка материалов.

Все тетради, как заполненные, так и неиспользованные, должны быть возвращены официальному лицу, отвечающему за проведение тестирования. Не следует смешивать заполненные и неиспользованные тетради. Заполненные тетради должны быть подобраны в соответствии со списками.

Все материалы для отправки должны находиться в следующем порядке: внизу неиспользованные тетради, затем заполненные тетради, контрольные бланки, протокол проведения тестирования и сверху анкета школы. Все материалы

должны быть переданы лицу, отвечающему за проведение исследования сразу же после тестирования.

Текст сценария исследования приведен в приложении 4.

Приложение 4

Сценарий проведения тестирования

ВВЕДЕНИЕ.

Здравствуйте (Представьтесь, если вас не знают учащиеся). Сегодня вы будете участниками исследования по оценке подготовленности школьников. Вы будете отвечать на вопросы по биологии, географии, физике, химии, математике. На работу отводится 1 час. Учащиеся разных школ города будут отвечать на те же самые вопросы, на которые вы отвечаете сегодня. По условиям проведения работы мы не будем показывать ваши тетради в школе. Тем не менее, постарайтесь выполнить работу как можно лучше.

РАЗДАЧА ТЕТРАДЕЙ.

Сейчас я раздам вам тетради. Пусть они лежат на парте лицевой стороной вниз до тех пор, пока я вам не скажу их перевернуть. Вам будет нужен карандаш и ластик. Если у вас их нет, вы можете взять у меня.

Если у вас на парте лежат книги или бумаги, уберите их, пожалуйста, с парты. Какие у вас вопросы? (ответьте на вопросы).

ЗАПОЛНЕНИЕ ЛИЦЕВОЙ СТОРОНЫ ТЕТРАДИ.

Сейчас переверните свои тетради. Внизу тетради, слева, обведите месяц, в котором вы родились (пауза). Затем запишите две последние цифры года вашего рождения (пауза). А теперь девушки обводят слово «девушка», а юноши — слово «юноша» (пауза). Затем обведите класс, в котором вы учитесь (пауза).

ЧТЕНИЕ ИНСТРУКЦИИ.

В вашей тетради дана инструкция о том, как работать с заданиями. Сейчас я прочту ее вам. Для вас очень важно внимательно следить за текстом инструкции во время чтения.

Сейчас откройте свои тетради на первой странице инструкции (пауза). Следите по своей тетради за моим чтением. (Прочитайте медленно всю инструкцию, помещенную в тетради).

ИНСТРУКЦИЯ.

У вас в тетради вопросы по математике, биологии, географии, физике, химии. И еще несколько вопросов о вас самих и ваших занятиях в школе.

Не удивляйтесь, если ваш сосед выполняет задания, не похожие на ваши — у вас разные варианты.

Некоторые вопросы легкие, другие сложные. Внимательно читайте каждый вопрос и постарайтесь ответить на него как можно лучше.

В вашей тетради 6 частей. Начать выполнение каждой части можно только по команде учителя.

Когда увидите знак ,

СТОП

остановитесь!

Если у вас осталось время после выполнения одной части работы, то проверьте, как вы их выполнили. Когда вам скажут остановиться, вы должны закончить работу над этой частью. Работать над другой частью в это время не разрешается.

Прочитайте пример 1.

(разбирают первый пример заполнения задания).

Если вам надо сделать какие-нибудь вычисления, используйте для этого свободное место около задания. Пользоваться калькулятором не разрешается. В некоторых заданиях вам нужно будет записать ответ на специально отведенном для этого месте. В примере 2 показано, как это можно сделать.
(разбирают пример 2)

ПОМНИТЕ!

Читайте каждый вопрос **ВНИМАТЕЛЬНО**.

Отвечая на вопрос, обводите только **ОДНУ БУКВУ** или записывайте свой ответ в специально отведенном месте.

Если вам надо изменить свой ответ, сначала **ПОЛНОСТЬЮ СОТРИТЕ** тот ответ, который вы дали раньше.

Если у вас осталось время после выполнения одной части работы, то **ОБЯЗАТЕЛЬНО ПРОВЕРЬТЕ**, как вы ее выполнили.

ЧАСТЬ 1. Мы начинаем работу с 1 Части.

Пожалуйста, откройте тетради в начале Части 1. (Убедитесь в том, что все учащиеся нашли правильную страницу) Я вам буду говорить, когда можно начать выполнять каждую часть и сколько у вас времени на ее выполнение.

Когда вы дойдете до конца части, вы увидите знак «СТОП». Если вы закончите любую часть раньше, вы можете проверить свою работу. Не разрешается работать над другой частью. Если вам понадобится карандаш во время работы, пожалуйста, поднимите руку. Я не могу отвечать на ваши вопросы во время работы. На выполнение этой части отводится 10 минут, (Установите таймер на 10 минут). Пожалуйста, начинайте.

Остановитесь. Это конец первой части.

Остальные части работы прописаны аналогично.

Остановитесь. Это конец работы. Пожалуйста, закройте тетради и отложите карандаши. Оставайтесь на своих местах, пока я не соберу ваши тетради.

(Соберите тетради по одной. Проверяйте каждую тетрадь, заполнена ли ее лицевая сторона).

БЛАГОДАРНОСТЬ

Я благодарю вас за то, что вы приняли участие в данной работе. Она поможет нам больше узнать о школах и об учениках,
(Отпустите учащихся.)

Приложение 5

Протокол проведения тестирования

При заполнении протокола обводите букву, соответствующую ответу или запишите ответ в специально отведенном месте.

Номер и название школы _____

Код класса (способ кодирования приведен в инструкции для наблюдателя)

Фамилия, имя и отчество школьного координатора _____

Дата проведения тестирования _____

1. Число полученных, заполненных и неиспользованных тетрадей с тестами:

	Тетрадь ?	Тетрадь ?
Число полученных тетрадей (включая дополнительные)	_____	_____
Число заполненных тетрадей	_____	_____
Число неиспользованных тетрадей	_____	_____

2. Укажите время выполнения теста:

	Время		Время
Начало работы	_____	Начало части 3	_____
Начало чтения инструкции	_____	Конец части 3	_____
		Начало части 4	_____
		Конец части 4	_____
Начало части 1	_____	Начало части 5	_____
Конец части 1	_____	Конец части 5	_____
Начало части 2	_____	Начало части 6	_____
Конец части 2	_____	Конец части 6	_____

3. Не было ли каких-либо неожиданных происшествий во время работы?

(А) Нет.

(В) Да. Объясните.

4. Каким был уровень работы и дисциплины учащихся во время выполнения теста?

(А) Очень высокий.

(В) Высокий.

(С) Низкий.

(D) Очень низкий.

5. Возникали ли у учащихся какие-либо проблемы во время работы (например, было слишком трудно, не хватило времени, появились усталость, смущение) ?

(A) Нет.

(B) Да. Объясните

6. Испытывала ли школа какие-либо трудности при подготовке и проведении тестирования (например, при подборе школьного координатора, при выборе помещения, при составлении графика тестирования) ?

(A) Нет.

(B) Да. Объясните.

Исследователь _____

Приложение 6

Глоссарий оснащения тестов для проведения аттестации учащихся

1. Бланки для составления тестовых заданий. Бланки представляют собой место для формализованной и стандартизированной записи тестового задания, кроме текста задания они предлагают автору дать первичные оценки, характеристики задания — сложность, количество времени, которое потребуется на его выполнение, отнесение задания к какой либо теме данного учебного предмета; кроме этого бланки содержат место для экспертных рецензий.

1. 1.1 Бланк для составления тестового задания универсальный

2. 1.3 Бланк для составления тестового задания альтернативных ответов

3. 1.4 Бланк для составления тестового задания множественного выбора

4. 1.5 Бланк для составления тестового задания дополнения

5. 1.6 Бланк для составления тестового задания свободного изложения

6. 1.7 Бланк для составления тестового задания соответствия

7. 1.8 Бланк для составления тестового задания реструктурирования

8. Технологическая матрица. *Технологическая матрица представляет собой инструмент, на основании которого происходит компоновка заданий в тест.*

В данном эксперименте принят сокращенный, одномерный вариант технологической матрицы.

8. 8.1. Форма технологической матрицы 3

9. 8.2. Заполненные технологические матрицы по 11 предметам

10. Рекомендации по составлению тестовых заданий для педагогов.

11. Рекомендации по составлению тестовых заданий для методистов.

12. Дополнения и изменения к рекомендациям по составлению тестовых

заданий для вступительных экзаменов в 10—е классы. Изменения и дополнения

касались в основном тех недостатков в стандартизации форм заданий, которые выявились в процессе работы.

13. Инструкция по экспертизе тестовых заданий.

14. Инструкция по занесению заданий на магнитный носитель.

14. 14.1. Система кодирования заданий, тетрадей, учащихся, образовательных учреждений и пр.

15. 14.2. Приведение заданий к стандартной форме инструкции и предъявления

16. Тестовые тетради. *Тестовые тетради представляют собой набор заданий и место для их заполнения. Предполагается, что, кроме тестовой тетради, для работы ученику не потребуются никаких дополнительных материалов. Всего разработано 5 тестовых тетрадей, в которые сгруппированы 11 школьных предметов, включая экономику.*

17. 16.1. Тестовые тетради по предметам физико-математической группы (математика и физика)

18. 16.2. Тестовые тетради по предметам естественнонаучной группы (химия, биология и география)

19. 16.3. Тестовые тетради по предметам историко-краеведческой группы (истории, обществоведение и история и культура Санкт-Петербурга)

20. 16.4. Тестовые тетради по предметам гуманитарно-филологической группы (русский язык и литература)

21. 16.5. Тестовые тетради по предметам экономического профиля (экономика)

22. Ключи. *Ключи представляют собой правильные ответы на каждое из заданий в тесте, формализованные до вида однозначных,. Ключи составляются на основе технологической матрицы. Кроме этого ключи по математике разнесены еще по некоторым учениям.*

22. 22.1. Ключи по предметам физико-математической группы (математика и физика)

23. 22.2. Ключи по предметам естественнонаучной группы (химия, биология и география)

24. 22.3. Ключи по предметам историко-краеведческой группы (истории, обществоведение и история и культура Санкт-Петербурга)

25. 22.4. Ключи по предметам гуманитарно-филологической группы (русский язык и литература)

26. 22.5. Ключи по предметам экономического профиля (экономика)

27. Инструкции для ведущих, *Инструкция для ведущих содержит требования к ведущему его обязанности, набор действий в нестандартных ситуациях, сценарий проведения тестирования. ведущий обязан строго придерживаться сценария.*

27. 27.1. Инструкция для ведущего предметы физико-математической группы

28. 27.2. Инструкция для ведущего предметы естественнонаучной группы

30. 27.3, Инструкция для ведущего предметы историко-краеведческой группы

31. 27.4. Инструкция для ведущего предметы гуманитарно-филологической группы
32. 27.5. Инструкция для ведущего предметы экономического профиля
33. Инструкция для учащихся по заполнению тестовых тетрадей. *Эта инструкция содержит образцы заполнения заданий, которые встречаются в тестировании по данному предмету, сведения о том как исправлять неверно отмеченные задания и как исправлять исправления, кроме этого инструкция содержит временную разбивку теста и общие рекомендации по заполнению.*
33. 33.1. Инструкция для учащегося предметы физико-математической группы
34. 33.2. Инструкция для учащегося предметы естественнонаучной группы
35. 33.3. Инструкция для учащегося предметы историко-краеведческой группы
36. 33.4. Инструкция для учащегося предметы гуманитарно-филологической группы
37. 33.5. Инструкция для учащегося предметы экономического профиля
38. Инструкции для наблюдателя. *Наблюдатель — 'это специально подготовленный педагог в задачу, которого входит фиксация процедуры проведения экзамена. Наблюдателю запрещено вмешиваться в действия ведущего.*
38. 38.1, Инструкция для наблюдателя предметы физико-математической группы
39. 38-2. Инструкция для наблюдателя предметы естественнонаучной группы
40. 38.3. Инструкция для наблюдателя предметы историко-краеведческой группы
41. 38.4. Инструкция для наблюдателя предметы гуманитарно-филологической группы
42. 38.5. Инструкция для наблюдателя предметы экономического профиля
43. Анкета для наблюдателя.
44. Инструкция по использованию анкеты наблюдателя.
45. Инструкция по обработке результатов тестирования. *Эта инструкция объединена с двумя, другими единым блоком.*
45. 45.1, Инструкция о проверке работ
46. 45.2. Инструкция по занесению результатов на магнитный носитель
47. 45.3. Инструкция по использованию компьютерной программы обработки результатов
48. Компьютерная программа для занесения хранения и анализа данных результатов тестирования.
49. Рекомендации для детей и родителей. *«Как успешно пройти тестирование». Брошюра. Содержит рекомендации по стратегии и тактике заполнения, информации о видах и формах тестовых, заданий. Тесты с ключами для самостоятельно подготовки. Объем брошюры 2 п. л.*
50. Заявки для участия в эксперименте.
51. Заявки на тестирование.

52. Заявки на тестовые тетради,

53. Примерная инструкция о проведении экзаменов в образовательных учреждениях и учреждениях начального профессионального образования в форме тестирования.

Приложение 7 „ „ „ „ „ Таксономии учебных целей в когнитивной и аффективной областях

(По М. В. Кларину.)

Категории учебных целей в когнитивной области.

Основные категории учебных целей	Примеры обобщенных типов учебных целей
1. Знание	Ученик
Эта категория обозначает запоминание и воспроизведение изученного материала. Речь может идти о различных видах содержания — от конкретных фактов до целостных теорий. Общая черта этой категории — припоминание соответствующих сведений.	Знает употребляемые термины знает конкретные факты, знает методы и процедуры, знает правила И принципы.
2. Понимание	Ученик
Показателем способности жег служить преобразование (трансляция) материала из одной формы выражения в другую, * перевод» его с одного «языка» на другой (например, из словесной формы — в математическую). В качестве показателя понимания может также выступать интерпретация материала учеником (объяснение, краткое изложение) или же предположение о дальнейшем ходе явлений, событий (предсказание последствий, результатов). Такие учебные результаты превосходят простое запоминание материала.	Понимает факты, правила и принципы, интерпретирует словесный материал, интерпретирует схемы, графики, диаграммы, преобразует словесный материал в математические выражения, предположительно описывает будущие последствия, вытекающие из имеющихся данных.
3. Применение	Ученик

<p>Эта категория обозначает умение использовать изученный материал в конкретных условиях и новых ситуациях. Сюда входит применение правил, методов, понятий, законов, принципов, теорий. Соответствующие результаты обучения требуют более высокого уровня владения материалом, чем понимание.</p>	<p>Использует понятия и принципы в новых ситуациях, применяет законы, теории в конкретных практических ситуациях, демонстрирует правильное применение метода или процедуры.</p>
<p>4. Анализ</p>	<p>Ученик</p>
<p>Эта категория обозначает умение разбить материал на составляющие так, чтобы ясно выступала его структура Сюда относятся вычленение частей целого, выявление взаимосвязей между ними, осознание принципов организации целого, Учебные результаты характеризуются при этом более высоким интеллектуальным уровнем, чем ку требуют осознания как содержания учебного материала , так и его внутреннего строения.</p>	<p>Выделяет скрытые (неявные) предположения, видит ошибки и упущения в логике рассуждения, проводит различия между фактами и следствиями, оценивает значимость данных.</p>
<p>5. Синтез</p>	<p>Ученик</p>
<p>Эта категория обозначает умение комбинировать элементы, чтобы получить целое, обладающее новизной. Таким новым продуктом может быть сообщение (выступление, доклад), план действий или совокупность обобщенных связей (схемы для упорядочения имеющихся сведений). Соответствующие учебные результаты предполагают деятельность творческого характера с акцентом на создание новых схем и структур.</p>	<p>Пишет небольшое творческое сочинение, предлагает план проведения эксперимента, использует знания из разных областей, чтобы составить план решения той или иной проблемы.</p>

6. Оценка	Ученик
<p>Эта категория обозначает умение оценивать значение того или иного материала (утверждения, художественного произведения, исследовательских данных) для конкретной цели. Суждения ученика должны основываться на четких критериях. Критерии могут быть как внутренними (структурными, логическими), так и внешними (соответствие намеченной цели). Критерии могут определяться самим учащимся или же задаваться ему извне (например, учителем). Данная категория предполагает достижение учебных результатов по всем предшествующим категориям плюс оценочные суждения, основанные на ясно очерченных критериях.</p>	<p>Оценивает логику построения материала в виде письменного текста, оценивает соответствие выводов имеющимся данным, оценивает значимость того или иного продукта деятельности, исходя из внутренних критериев, оценивает значимость того или иного продукта деятельности, исходя из внешних критериев.</p>

Категории учебных целей в аффективной области.

Основные категории учебных целей	Примеры обобщенных типов учебных целей
1. Восприятие	Ученик
<p>Эта категория обозначает готовность и способность ученика воспринимать те или иные явления, поступающие из окружающего мира стимулы. С позиции учителя путь к достижению таких целей состоит в том, чтобы привлечь, удержать и направить внимание ученика. Входящие сюда субкатегории (1.1 осознание, 1.2 готовность или желание воспринимать и 1.3 избирательное — произвольное — внимание) образуют диапазон восхождения от пассивной позиции ученика до более активного (хотя на этом уровне еще не полностью осознанного, целенаправленного) отношения к содержанию обучения</p>	<p>Проявляет осознание важности учения, внимательно слушает высказывания окружающих в классе, в беседе и т. д., проявляет осознание эстетических факторов в одежде, интерьере, архитектуре, живописи, проявляет восприимчивость к проблемам и потребностям других людей, к проблемам общественной жизни.</p>

2. Реагирование (отклик)	Ученик
<p>Эта категория обозначает активные проявления, исходящие от самого ученика. На данном уровне он не просто воспринимает, но и откликается на то или иное явление или внешний стимул, проявляет интерес к предмету, явлению или деятельности. Субкатегории: 2.1 подчиненный отклик, 2.2 добровольный отклик, 2.3 удовлетворение от реагирования.</p>	<p>Выполняет заданную учителем домашнюю работу, подчиняется внутришкольному распорядку и правилам поведения, участвует в обсуждении вопросов в классе, самостоятельно знакомится с освещением общественно-политических и международных проблем, добровольно вызывается выполнять задание, проявляет интерес к учебному предмету.</p>

3. Усвоение ценностной ориентации	Ученик
<p>В эту категорию входят различные уровни усвоения ценностных ориентаций (отношения к тем или иным объектам, явлениям или видам деятельности): 3.1 принятие ценностной ориентации (в обиходе это соответствует понятию «мнение»); 3.2 предпочтение ценностной ориентации и 3.3 приверженность, убежденность.</p>	<p>Проявляет устойчивое желание, например, овладеть навыками грамотной устной и письменной речи, целенаправленно изучает различные точки зрения с тем, чтобы вынести собственное суждение, проявляет убежденность, отстаивая тот или иной идеал.</p>

4. Организация ценностных ориентаций	Ученик
<p>Эта категория охватывает осмысление и соединение различных ценностных ориентаций, разрешение возможных противоречий между ними и формирование системы ценностей на основе наиболее значимых и устойчивых. Сюда входят две субкатегории: 4.1 концептуализация ценностной ориентации, т. е. осмысление своего отношения; 4.2 организация систе-</p>	<p>Пытается, например, определить основные черты своего любимого произведения искусства, принимает на себя ответственность за свое поведение, понимает свои возможности и ограничения, строит жизненные планы в соответствии с осознаваемыми им самим собственными</p>

5. Распространение ценностной деятельность	Ученик
Эта категория обозначает такой уровень усвоения ценностей, на котором они устойчиво определяют поведение индивида, входят в привычный образ действий, или жизненный стиль. Обобщенный характер ценностных ориентации и их соединение в целостное мировоззрение отражены в субкатегориях: 5.1 обобщенная установка и 5.2 полная интернализация (усвоение), или распространение ценностных ориентации на деятельность.	УСТОЙЧИВО проявляет самостоятельность в учебной работе, проявляет стремление к сотрудничеству в групповой деятельности, проявляет готовность к пересмотру своих суждений и изменению образа действий в свете убедительных аргументов, постоянной гигиены и здорового образа жизни, формулирует устойчивое и последовательное жизненное кредо.

Литература

1. Аванесов В. С. Тесты в социологическом исследовании. М., 1982.
2. Аверин В. А. Психология детей и подростков. СПб., 1994.
3. Агафонова И. Н., Колеченко А. К. и др. Методика изучения интеллекта. Часть 1. СПб., 1991.
4. Айзенк Ю. Г. Проверьте свои способности. СПб., 1993.
5. Алексеев А. А., Громова Л. А. Поймите меня правильно, или книга о том, как найти свой стиль мышления, эффективно использовать интеллектуальные ресурсы и обрести взаимопонимание с людьми. СПб., 1993.
6. Аprobация учебной литературы. Информационный сборник. Вып. 1. М., 1994г.
7. Ашмарин Б. А. О тестах и тестировании // Физ. культура в школе. 1985, № 3, с. 60-62.
8. Беспалько В, П. Слагаемые педагогической технологии. М., '1989.
9. Битинас Б. Методические проблемы выявления научной информации в педагогическом исследовании, Автореферат докторской диссертации. Вильнюс, 1972.
10. Блинов В. М. Эффективность обучения. (Методологический анализ определения этой категории в дидактике). М., 1976.

11. Бурлачук Л. Ф., Морозов С. Н. Словарь-справочник по психологической диагностике. Киев, 1989,
12. Бюллетень состояния образовательной системы Санкт-Петербурга. Под ред. Майорова А. Н. Вып. 2(6), СПб. 1995.
13. Видинеев Н. В. Природа интеллектуальных способностей человека. М., 1989.
14. Всероссийская конференция «Измерения в педагогике». Материалы. /Под ред. Майорова А. Н. Часть 1—4. СПб, 1994/.
15. Вопросы психологии обучения. /Под ред, Н. А. Менчинской. М., 1960/.
16. Выготский Л, С. Избранные психологические исследования, М., 1956.
17. Гайда Б. К., Захаров В. П. Психологическое тестирование. Л., 1982.
18. Глассер у. Школы без неудачников. М., 1991.
19. Гильбух Ю. З. Метод психологических тестов: сущность и значение. // Вопросы психологии, 1986, № 2, с. 30—42.
20. Граник Г. Г., Соболева О. В. Понимание теста: проблемы земные и космические (психологический анализ) // Вопросы психологии, 1993, № 5, с. 50-55.
21. Закс Л. Статистическое оценивание. М., 1976
22. Занков Л. В. Развитие учащихся в процессе обучения. М., 1963.
23. Занков Л. В. Дидактика и жизнь. М., 1970.
24. Знание, мышление, умственное развитие. /Под ред. П. Я. Гальперина. М., Педагогика, 1972/.
25. Калмыкова З. И. К вопросу о методах диагностики обучаемости школьников. // Вопросы психологии, 1968, № 6.
26. Кабардин О, Ф., Землякова А. Н. Тестирование знаний и умений учащихся (школ) // Советская педагогика, 1991, № 12, с. 27-33.
27. Как провести социологическое исследование. М, 1990.
28. Кларин М. В. Инновационные модели обучения в зарубежных педагогических поисках М., 1994 г.
29. Клименко А. Практика тестирования. М., 1991.
30. Краевский В. 5-1 Высоцкая С. И., Шубинский В. С. Умения и навыки как компоненты содержания общего и среднего е образования. //Советская педагогика, 1981, № 10.
31. Крутецкий В. А. Развитие умственных способностей § школьника в процессе обучения. //Советская педагогика, 1971, № 8.
32. Кривошапова Р. Ф., Красновский Э. А., Резникова В. З. Требования к знаниям, умениям и навыкам учащихся. //Советская педагогика, 1978, № 2.
33. Кулагин Б. В. Основы профессиональной психодиагностики. Л., 1984.
34. Левитов Н. Д. О психических состояниях человека. М., 1964.
35. Лернер И. Я. Базовое содержание общего образования,, 1991.
36. Лучшие психологические тесты для профотбора и профориентации. Описание и руководство по использованию. /Под ред. Кудряшова А. Ф., Петрозаводск, 1992.
37. Мельников В. М., Ямпольский Л, Г. Введение в экспериментальную психологию личности. М., 1985,

38. Методика и техника статистической обработки первичной социологической информации. М., 1968.
39. Методики диагностики апробируемых учебных материалов. Информационный бюллетень. Вып. 1—3. /Ред. Алексашина И. Ю. СПб., 1994-95.
40. Методические разработки для лабораторных занятий по психодиагностике интеллекта. Л., 1990.
41. Мешалкина К. Н. Эффективность обучения и развитие аналитических способностей. //Педагогика, 1994, № 10.
42. Мышко С. А. Тестирование как психолого-педагогическое средство оценки академических способностей в системе образования США, //Высшие и средние школы за рубежом, 1980, Вып. 6.
43. Навокина З. Снова «интеллектуальные тесты (в системе образования США). //Народное образование, 1985, № 1, с. 86-87.
44. Обиря психодиагностика. /Под ред. Бодалева А. А., Столина В. Б. М. 1987.
45. Огорелков В. И. Надежность измерений и оценки знаний учащихся на основе элективных заданий. //Советская педагогика, 1970, № 7.
46. Основные математико-статистические понятия и формулы в экономическом анализе. М., 1979.
47. Петербургская школа. Теория и практика формирования многовариантной образовательной системы. /Под ред. О. Е. Лебедева. СПб, Центр педагогической информации, 1994.
48. Психодиагностика — учителю Научно-методический сборник. Харьков, 1992.
49. Психологическая диагностика: Проблемы и исследования. /Под ред. К. М, Гуревича, М., 1981/
50. Психология (словарь). М., 1991.
51. Психологический словарь. /Под ред. А. В. Петровского и М. Г. Ярошевского. М., 1990, с. 396—399/
52. Рабочая книга метролога. М, 1991.
53. Рабочая книга социолога. М., 1983.
54. Решетников М. М. Профессиональный отбор в системе образования, промышленности и армии США: краткий исторический обзор. //Психологический журнал, 1987, № , 1, с. 113-126.
55. Романова Е. С., Усапова О. Н. и др. Психологическая диагностика развития школьников в норме и патологии. М., 1990.
56. Санкт-Петербургская школа накануне 1993—94 учебного года. СПб., 1993.
57. Симонов В. П. Директору школы об управлении учебно-воспитательным процессом, М., 1987.
58. Собчик Л. Н. Методы психологической диагностики. Вып. 1-3. М., 1990».
59. Социологический словарь. Минск, 1991.
60. Тарасун В. В. Тесты учебных способностей для 1 класса. Киев, 1992.
61. Тесты успешности. Сост. В. Муравьев, М., 1979-1981.
62. Урунтаева Г. А., Афонькина Ю. А. Практикум по детской психологии, М., 1995.

63. Цатурова И. А. Из истории развития тестов в СССР и за рубежом. Таганрог, 1969.
64. В. Черны, Т. Колларик, Компендиум психодиагностических методов. В 2-х т. Братислава, 1988.
65. Ярошевский М. Г. История психологии. /3-е изд., М., 1985.
66. Клайн П. Введение в психометрическое программирование. Справочное руководство по конструированию тестов. Киев, 1994.
67. Агапов В. Ю., Мишакова Л. В. Алгоритмы целеполагания в современных педагогических технологиях. Рязань, 1994.
68. Гимназические основы гуманитарного образования в школах Петербурга. /Ред. Лебедев О. Е. СПб, 1995.

Ответы на задания

Ответ на задание 1.

Правильными ответами в данном задании будут:

1. стандартизированные да
2. нестандарттизированные нет
3. специальных способностей нет
4. достижений да
5. аппаратурные нет
6. бланковые да
7. индивидуальные нет
8. тесты скорости нет
9. смешанные тесты да
10. гетерогенные нет
11. нормированный нет

Стандартизированные тесты имеют стандартную процедуру проведения, поэтому они нам подходят, нестандарттизированные тесты то же могли бы нам подойти, если бы в условии задания не стояло условие проведения аналогичной работы другими педагогами. В данном случае подходят ненормированные тесты, поскольку по условию задания будет проведено межгрупповое сравнение. Все остальные классификационные признаки достаточно ясны.

Ответ на задание 2.

Ответ Е.

Ответ на задание 3.

Основной недостаток этого задания — несоответствие основному требованию к тестовым заданиям. Смысл этого задания заключается в том, что бы сделать правильный ответ однозначным. Решение этой задачи возможно несколькими путями, при чем условие требует оставаться в рамках заданий свободного изложения и дополнения:

Задание свободного изложения. Вариант 1;

Инструкция: Закончите предложение, записав вместо многоточия необходимые слова

Вопрос: Комплексное использование сырья и создание производств с замкнутым циклом без сброса сточных вод и без выброса в атмосферу вредных веществ называется

Ответ: Безотходной технологией.

Задание свободного изложения. Вариант 2:

Инструкция: Закончите предложение, записав вместо многоточия необходимые слова

Вопрос: Безотходная технология — это ...

Ответ: Правильным считается ответ, если в нем использованы совместно три элемента: слово «комплексное» и словосочетания: «замкнутый цикл» и «без выброса (сброса)».

Задание дополнения может выглядеть примерно так; (точный набор пропущенных слов определяется требованием: «Дополнять наиболее важное, то, что нужно проверить»):

Инструкция: впишите на место многоточий три пропущенных слова.

Вопрос: Безотходная технология — это ... использование сырья и создание производств с ... без сброса сточных и без выброса в атмосферу вредных веществ,

Ответ: комплексное, замкнутым, циклом,

Ответ на задание 4.

Первое, что бросается в глаза при анализе этого задания, — большое количество пропусков — 4. С формальной точки зрения это так, однако два последних пропуска однотипны и их вполне можно считать за один.

Кажется, что большое количество пропусков лишает данное задание смысла, поскольку возможен еще один вариант правильного ответа: "прилагательное, признак предмета, какой, чей. Действительно при данной к этому заданию инструкции («вместо многоточий впишите пропущенные слова») не ясно сколько слов можно вписать вместо каждого из многоточий. При такой инструкции вариант ответа для прилагательного представляется вполне правомерен, Таким образом, основной недостаток этого задания в противоречии между предполагаемым ответом и инструкцией. Вариант инструкции: «вместо каждого многоточия впишите только одно слово*» делает ответ на это задание однозначным.

Ответ на задание 5.

1. Инструкция к этому заданию не содержит указание, каким образом необходимо выбирать ответ, какие конкретные действия должен совершить испытуемый — дописать, отметить знаком в тестовой тетради, вписать ответ в бланк и т. д,

2. Формулировка вопроса содержит лишнюю деталь — слова «как известно», без которых вполне можно обойтись.

3. Формулировка вопроса содержит неопределенность — не ясно утверждение да или нет относится к словам «одобряли» или «не одобряли». Лучше были бы варианты:

- а) Бояре жившие в то время это событие — одобряли/не одобряли.
б) Одобряли бояре жившие в то время это событие — да/нет.
4. Сама форма задания представляется не совсем удачной, поскольку важнее знать не столько отношение одобрения или не ободрения, а причины этого явления, интересы сторон и пр.

Ответ на задание 6.

1. Варианты ответов на данное задание содержат только утвердительные варианты ответа.
2. Четвертый вариант ответа не согласуется грамматически с вопросом.

Ответ на задание 7.

Это задание содержит в условии излишнюю информацию (до 1990 г. это была школа № 99, в 1994 году на экзаменах в данном образовательном учреждении присутствовали работники РОНО), проверка которой, представляется некорректной, потому вопросы типа: «Связано снижение успеваемости с посещением экзаменов работниками РОНО?», «Увеличилась ли успеваемость с присвоением школе статуса гимназии?», «На какое количество баллов педагоги завышают оценки?» и т. д. нельзя признать правильными, кроме того, данные формулировки противоречат самому условию задания — составить вопросы на понимания диаграммы, то есть предметом данного задания служит диаграмма.

Правильными можно считать вопросы типа:

- Б 1993 году в гимназии была
максимальная успеваемость да/нет
- В 1989 году успеваемость была
минимальной да/нет
- Успеваемость в 1990 и 1984 годах
была одинаковой да/нет
- С 1990 по 1994 год успеваемость
непрерывно увеличивалась да/нет
- и т. д.

Ответ на задание 8.

Недостатками этого задания являются:

1. Наличие лишних деталей — это первое предложение задания-
2. Неоднозначна формулировка вопроса — основной элемент неоднозначности слова «самые крупные», сколько крупных надо отметить? Представляется целесообразным данный вопрос либо переформулировать — выберите из приведенного перечня стран два наиболее крупных экспортёра сахара, что представляется тоже не очень удачным, лучше бы было перевести это задание в задание с переструктурированием данных, предложить ранжировать страны в зависимости от экспорта сахара, либо в задания альтернативных ответов, где противопоставить экспортёров и импортёров.
3. Лучше, если задание множественного выбора содержит один вариант правильного ответа.
4. Лучше названия стран давать единообразно, придерживаясь либо аббревиатур, либо общеупотребительных названий.

Ответ на задание 9.

Представляется крайне неудачным объединение двух вопросов в один. Что порождает путаницу.

Порядок дат вопроса и ответа не соответствует друг другу—в вопросе Санкт-Петербург стоит первым, а в правильном ответе — вторым.

Даты необходимо упорядочить хотя бы по одному из рядов — первому или второму.

Из приведенных цифр недостоверным представляется вариант ответа В, который существенно отличается от остальных — варианты ответов должны быть более правдоподобными.

Уважаемый читатель!

Мы — издательство— будем рады узнать ваше мнение об этой книге. Наше издательство пока еще очень молодо, и отклик заинтересованных читателей для нас жизненно необходим. Бы можете сделать наши книги лучше — а, они того стоят, право же!

Выберите время написать нам, как вы оцениваете качество оформления, подготовки текста, общий уровень издания. Кроме того, мы будем благодарны за любые замечания и предложения, касающиеся цены и доступности этой книги, эффективности ее 'рекламы. Не забудьте указать свой адрес — мы с удовольствием будем держать вас в курсе наших издательских планов. И конечно же, ваши пожелания относительно издания книг по определенной тематике или книг определенного автора станут ценным подспорьем при формировании нашего издательского портфеля. Наши координаты: факс. (812) 311 3403 e-mail: postmaser@omsy.spb.su 191025, Санкт-Петербург, а/я 83



Лаборатория изучения
образовательных систем
Комитета по образованию
Санкт-Петербурга

ПРЕДЛАГАЕТ

комплект научно-Методического

обеспечения программы «Педагогический Мониторинг»,
*предназначенный для использования информационно-диагностическими
службами сферы образования.*

Комплект— результат шестилетней работы с руководителями образовательных учреждений всех Типов и уровней от детского сада до гимназии и от заместителя директора школы до Министерства образования России.

Ваши управленческие решения будут иметь информационную основу соответствующую реальному положению дел. Вы будете иметь объективную информацию о развитии, функционировании и эффективности деятельности образовательной системы для;

- принятия решений по текущим проблемам;
- оценки эффективности и коррекции программ развития;
- выявление наиболее успешных направлений деятельности; «• прогноза развития Вашей образовательной системы;
- сканирования неблагополучных сфер деятельности.

Система педагогического мониторинга прошла двукратную апробацию с 14 регионах России. На основе этой апробации получены средние показатели - нормы, в сравнении с которыми Вы. можете точно и объективно определить положение своей образовательной системы, отдельных ее элементов.

Уникальными особенностями нашего комплекта являются:

- высокий уровень стандартизации, обеспечивающий воспроизводимость результатов;
- наличие норм, обеспечивающих основания для сравнения.

В комплект поставки входят:

1. оригинал-макеты анкетно-тестового материала для родителей, учеников, учителей и администрации образовательного учреждения;

2. инструкции по организации;

3. инструкции по проведению;

4. сценарий проведения;

5. нормированные показатели;

6. пакет компьютерных программ для обработки результатов исследования; А инструкция по анализу результатов.

По Вашей просьбе мы осуществим авторское сопровождение нашей продукции:

- обучим персонал;
- поделимся особенностями анализа;
- совместно определим сферы применения;
- определим оптимальные формы распространения полученной информации с учетом региональных особенностей.

На все Ваши вопросы мы будем рады ответить по тел. (812) 219-41-37, факс (812) 311-34-03.

Заявки отправлять по адресу: 192007, Санкт-Петербург, а-/я 102.

Консультации и материалы можно получить по адресу: 198025, СПб., ул. Марата, в. 25.

Приобретать готовые тесты для нужд образования представляется нерациональным поскольку:

- тесты необходимо достаточно часто обновлять;
- готовые тесты не учитывают школьной и региональной специфики;
- готовые тесты нормируются на выборке разработчиков, а не пользователей.

Поэтому мы не продаем готовые тесты — мы продаем готовые технологии. Ежегодно в январе и апреле мы организуем и проводим курсы.

Для метода стон, руководителей образовательных учреждений специалистов управления, педагогов курсы «Введение в тестологию» (36 часов, 3-5 дней).

Окончив эти курсы Вы сможете:

- делать осознанный выбор на рынке тестов;
- составлять тестовые задания и ненормированные тесты;
- использовать тесты школьных достижений в своей

работе,

По окончании курсов Вы получите

- Сертификат о прохождении курсов;
- образцы тестов и тестовых заданий;
- образец стандартизированного инструментария;
- комплект нормативных документов по использованию тестов для аттестации учащихся;
- комплект компьютерных программ;
- право на использование тестов школьных достижений, разработанных для промежуточной аттестации учащихся.

Для методистов, специалистов служб аттестации, разработчиков тестов курсы **«Конструирование и практическое применение тестов школьных достижений»** (52 часа, 57 дней). Окончив эти курсы Вы сможете:

- делать осознанный выбор на рынке тестов;
- составлять тестовые задания и нормированные тесты;
- организовать взаимодействие разработчиков тестов для создания тестов школьных достижений;
- организовать проведение аттестации школьников с использованием тестов;
- использовать тесты в различных сферах педагогической деятельности.

По окончании курсов Вы получите:

- Сертификат о прохождении курсов;
- образцы комплектов тестового инструментария для сравнительных исследований и аттестации учащихся;
- комплект компьютерных программ;
- комплект нормативных документов по использованию тестов для аттестации учащихся;

- право на использование тестов школьных достижений, разработанных для промежуточной и итоговой аттестации учащихся.

На все Ваши вопросы мы будем рады ответить по тел. (812) 219-41-37, факс (812) 311-34-03.

Заявки на участие отправлять по адресу:

192007, Санкт-Петербург, а/я 102.

Консультации и материалы можно получить по адресу:

198025, СПб., ул. Марата, д. 25.

Майоров Алексей Николаевич

ТЕСТЫ ШКОЛЬНЫХ ДОСТИЖЕНИЙ

конструирование, проведение, использование

издание второе

научно-методическое издание

ВВЫ 8-88857-021-4

«Образование и культура» 190000, Санкт-Петербург, ул. Декабристов, д. И

Лицензия ЛР № 071128

Сдано в набор 10.05.96 г. Подписано в печать 06.09.96 г. Формат 60x88/16.

Печать офсетная. Усл. печ. л. 19. Тираж 3000 з

Издательство ОМ5 191025, Санкт-Петербург, а/я 83, пр. Ветеранов, 55-2

Отпечатано в типографии Университета Технологии и Дизайна Санкт-

Петербург, ул. Моховая, 26