

РОЗДІЛ «ТЕОРІЯ І МЕТОДИКА НАВЧАННЯ МАТЕМАТИКИ, МЕХАНІКИ І ІНФОРМАТИКИ»

КРЫЛОВА Т.В., д-р пед. наук, професор
СТЕБЛЯНКО П.А., доктор физ.-мат. наук, професор

Днепродзержинский государственный технический университет

ТЕСТИРОВАНИЕ, ЕГО ВОЗНИКНОВЕНИЕ И РАЗВИТИЕ

*Тест в дословном переводе
с английского test – проверка,
испытание, проверочная
контрольная работа.*

Приведен обзор истории развития тестирования, в частности, педагогического. Выделены исторические периоды его возникновения и применения, а также этапы развития педагогического тестирования. Рассмотрены виды, типы, требования к тестам и их составлению, преимущества и недостатки тестирования.

An overview of the history of testing, in particular, teachers. Highlighted the historical periods of its origin and application, as well as the stages of teacher testing. The types, requirements and test preparation, advantages and disadvantages of testing.

1. История возникновения и развития. Тестирование имеет давнюю историю. Его возникновение и применение имеет три исторических периода.

Первый период (до конца XIX века) – предыстория с донаучными формами контроля приобретенных знаний и способностей.

Второй период (1920-1960г.г.) – классический, в течение которого создается классическая теория тестов.

Третий период (с 1970г.) – технологический, в течение которого разрабатываются методы адаптивного тестирования и обучения на основе Item Response Theory, представляющей, по сути, методологию эффективной разработки тестов и тестовых заданий для параметрической оценки испытуемых по изменяющемуся латентному (скрытому) качеству.

Тесты начали использовать с 1864 года в Великобритании.

Английский ученый Фрэнсис Гальтон (1822-1911) был одним из первых, кто пытался измерить различия между людьми в области элементарных технических процессов. Он определил три основных принципа контроля с помощью тестирования, которые используются и по сей день, а именно:

1. Применение серии одинаковых испытаний к большому количеству испытуемых.
2. Статистическая обработка результатов.
3. Выделение эталонов оценки.

Для измерения умственных способностей при обработке результатов испытаний Ф. Гальтон использовал работы бельгийского статистика А. Кеттелла.

На основе серии испытаний определяются различного вида нормы для оценки результатов тестирования. Все современные тесты построены на основе статистической теории измерений, они являются стандартизированным инструментом.

В лаборатории Ф. Гальтона некоторое время работал американский психолог, первый в мире профессор психологии (1888 г.) Джеймс Маккин Кеттелл (1860-1944). В 1890 году он первым употребил английское слово «тест» для обозначения психодиагностической методики. Дж. Кеттелл считал тест средством для проведения научного эксперимента и разработал требования для проведения тестирования:

1. Одинаковость условий для всех испытуемых.
2. Ограничение времени тестирования приблизительно одним часом.
3. Отсутствие зрителей в лаборатории, где проводятся эксперименты.
4. Оборудование должно быть хорошим и располагать людей к тестированию.
5. Одинаковые инструменты и четкое понимание испытуемых, что нужно делать.
6. Результаты тестирования подвергаются статистическому анализу, находится минимальный, максимальный и средний результаты, рассчитывается среднее арифметическое и среднее отклонение.

Эти идеи Дж. Кеттелла составляют основу современной тестологии.

В 1895-1896 годах в США были созданы два национальных комитета, призванных объединить усилия тестологов и придать общее направление появившимся тестологическим работам.

Родоначальником современных тестов для определения уровня развития интеллекта считается французский психолог Альфред Бинэ (1857-1912).

Тестологи продолжали развивать методики обработки результатов тестирования. К началу 1896 года английский математик-статистик Карл Пирсон (1857-1936) заложил основы теории корреляции.

Ф. Гальтон первым разработал метод сравнения двух рядов переменных и ввел коэффициент корреляции. Ему также принадлежит заслуга изучения взаимосвязи двух переменных и построения линий регрессии.

В 1904 году известный английский психолог Чарльз Эдуард Спирмен (1863-1945) сделал следующий шаг в использовании метода корреляции для психологии. Он явился основателем двухфакторной теории («G-фактор» («General ability») – «общая способность», «S-фактор» («Special abilities» – «частичные способности»)) и показал, что корреляция между двумя переменными свидетельствует о наличии общего фактора – причины, от которой зависят эти переменные.

Дальнейшее развитие тестологии во многом определялось нуждами практики.

Массовый характер тестирования вызвал необходимость перехода от индивидуальных тестов к групповым. В 1914-1918 годах в США появились первые групповые тесты для нужд армии. Наибольшее применение нашли тесты А. Отиса, а именно: Альфа-тесты для владеющих английским языком и Бета-тесты («немые» тесты) для невладеющих языком.

Разрабатывались тесты для профессионального отбора в различных сферах производственной деятельности, а также активно развивалось направление, связанное с созданием тестов для определения уровня интеллектуального развития.

Наиболее известным тестом такого рода в нашей стране является индивидуальный тест интеллекта американского психолога Д. Векслера. Наряду с вербальными заданиями тест содержит также задания, требующие невербальных действий. Оценка дается в форме коэффициента интеллектуальности.

С начала XX века выделилось и педагогическое направление в развитии тестологии. Американский ученый В.А. Макколл разделил тесты на психологические (Intelligence Test) и педагогические (Educational Test). Задачей психологических тестов являлось определение уровня умственного развития человека. Основной задачей педагогических тестов являлось измерение успешности учащихся по тем или иным

дисциплінам за определенный период обучения, а также успешности применения методов преподавания и организации учебного процесса.

Известный американский психолог Эдуард Ли Торндайк (1874-1949) выделил три этапа внедрения тестирования в практику американских школ.

Первый этап (1900-1915 г.г.) – период поисков. На этом этапе происходило осознание и начальное внедрение тестов памяти, внимания, восприятия и других тестов, предложенных А. Бинэ. Разрабатывались и перепроверялись тесты интеллекта, позволяющие определить коэффициент умственного развития.

Второй этап (1916-1930 г.г.) – годы широкого распространения тестов в Америке. Первый стандартизованный педагогический тест был создан под руководством основоположника педагогических измерений Э.Л. Торндайка. Это был тест на решение арифметических задач, впервые снабженный «нормами». Были разработаны и внедрены тесты О. Стоуна по арифметике, Б. Зекингема для проверки правописания, Э. Торндайка по диагностике большинства школьных предметов. Т. Келли разработал способ измерения интересов и склонности учеников при изучении алгебры.

В 20-х годах XX столетия под руководством Льюиса Медисона Термена (1877-1956), автора основного американского варианта системы тестов Бинэ-Симона, предназначенного для изучения умственных способностей детей, были проведены длительные исследования одаренности. Теория «умственной одаренности» существенно повлияла на организацию американской школы (создание специальных школ для одаренных детей).

Ч. Спирмен предложил общие основы корреляционного анализа для стандартизации тестов.

За эти 15 лет были разработаны тесты успешности по всем учебным дисциплинам, созданы «интеллектуальные» тесты для всех школ США. Эти тесты назывались «национальными», перед каждым разделом тестов американскими тестологами впервые были даны примерные задания.

В Европе распространение тестов происходило далеко не так благополучно.

Россия в начале XX столетия безоговорочно приняла тестовую основу объективного школьного контроля.

В СССР практическое значение тесты приобрели после 1925 года, когда была создана особая тестовая комиссия при педагогическом отделе Института методов школьной работы, в задачи которой входила разработка стандартизованных тестов для советской школы. На основе американских были разработаны тесты по природоведению, обществоведению, счету и решению задач, знанию географической карты, на понимание чтения и на правописание. Разработкой тестов вплотную занимались видные советские психологи и педагоги, в том числе П.П. Блонский, М.С. Бернштейн, С.М. Василейский и другие.

Основные принципы, использованные при составлении тестов, были систематизированы М.С. Бернштейном и в последствии легли в основу методологии групповых тестов:

1. Ограничение во времени – только 5 % испытуемых могут окончить за отведенное время проработку всего теста, то есть показатель уровня развития интеллекта прямо зависит от скорости выполнения заданий испытуемых.

2. Наличие детализированной инструкции как в отношении проведения, так и в отношении оценки.

3. Введение тестов с заданием выбрать правильный ответ с указанием в случае незнания или сомнения подчеркивать наугад.

4. Тщательная статистическая обработка и экспериментальная проверка.

С 1931 года начинается третий, современный этап развития школьного тестирования. Поиски тестологов направлены на повышение качества школьного

тестирования, создания непрерывной системы школьной тестовой диагностики и новых способов обработки тестов, накопление диагностических сведений.

Однако, несмотря на полученные советскими тестологами теоретические результаты, отношение к тестированию в СССР кардинально изменилось. В 1936 году вышло постановление ЦК ВКП(б) «О педагогических извращениях в системе Наркомпроса», которое негативно сказалось на развитие педагогики в целом и тестологии в частности. Метод тестов был признан буржуазным орудием дискриминации учащихся и «изгнан» из советской школы.

В школах развитых стран внедрение и усовершенствование тестов шло достаточно быстро. Распространялись диагностические тесты успешности, накапливался статистический материал, публиковались сборники тестов.

В зарубежных школах контроль и учет успеваемости, диагностирование обученности имеют характер объективной констатации результатов. Эти мероприятия не предусматривают заботу об образовании каждого ученика, выполняется один из подходов принципа индивидуализации обучения, а именно: каждый идет своим путем и темпом, обучается по своим возможностям, способностям, потребностям.

Только в шестидесятых годах в СССР вновь стали изучаться возможности оценивания достижений с помощью тестов. Однако тесты достижений тогда не нашли своего должного применения.

В 90-х годах отмечается повышение интереса к этой проблеме. Этот интерес к проблеме тестирования был связан с определенными недостатками традиционных методов диагностирования успеваемости учащихся:

- лишь на основе ответа на вопросы одного билета, представляющего весьма незначительную часть программы, преподаватель вынужден судить об уровне усвоения всей программы,

- субъективизм и ошибка в оценке знаний преподавателем снижают мотивацию учебной деятельности студента, вызывают стрессы и способствуют появлению эмоциональных барьеров в отношениях между студентами и преподавателями,

- экзаменационная оценка зависит не только от уровня усвоенных студентом знаний, но и от индивидуальной толерантности к тревоге, умения общаться, приспособляться к экзаменационному стрессу, а также от типовых ошибок экзаменаторов (чрезмерная снисходительность или строгость, искусственное занижение экзаменатором оценки, ошибки, связанные с уровнем подготовленности группы (в хорошо подготовленной группе преподаватель склонен занижать оценки, так как стремится воспроизвести свойственную ему схему распределения оценок), ошибки, получившие название «гало-эффекта», в которых обнаруживается влияние на оценку контакта учащегося с преподавателем на экзамене и экзаменационные установки преподавателя по отношению к учащемуся),

- трудности, связанные с особенностями преподавательской работы (несовпадение требований разных преподавателей, различие в профессиональной квалификации, при организации текущих контрольных мероприятий при оценивании усвоенных знаний, приобретенных навыков и умений большого числа студентов) наблюдается загруженность преподавателя рутинной нетворческой работой, связанной с большим объемом сведений, которые требуется подготовить, обработать и проанализировать за относительно короткий промежуток времени,

- трудности, связанные со спецификой традиционной формы проверки усвоенных знаний (например, отсутствие четко сформулированных стандартов знаний и конкретно очерченных объемов умений, достаточных для каждой положительной оценки),

- трудности, связанные с поведением студентов на экзамене (использование шпаргалок, списывание, «взаимопомощь» сокурсников), что искажает достоверность оценки и мешает преподавателю объективно взглянуть на качество своей педагогической работы,

– отсутствие объективных критериев оценки и эффективных механизмов сравнения результатов обучения по данной дисциплине в различных вузах, что особенно актуально для выработки верной стратегии подготовки кадров.

Вопросы создания и применения тестов были отражены в работах В.А. Аванесова, В.П. Беспалько, Н.Н. Лосевой, И.П. Подласого, З.И. Слепкань, И.А. Цатуровой и других.

2. Тесты, их определение, виды, формы. Существует несколько современных определений тестов для педагогической области. Приведем некоторые из них.

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

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.

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

Как правило, стандартизация подразумевает единообразие процедуры проведения и оценки выполнения теста.

Тест достижений – это система заданий специфической формы, определенного содержания возрастающей трудности, позволяющая качественно оценить структуру и эффективно измерить уровень усвоенных знаний, приобретенных умений и навыков и представлений по изучаемой дисциплине.

Тест – совокупность заданий по определенной области знаний или учебного предмета, позволяющая количественно оценить знания, умения, учебные достижения, компетентность учеников и студентов.

Существует немало истолкований педагогического теста, отображающие широкий спектр подходов и мнений относительно корректного его определения – от широкого его понимания как метода исследования до узкого восприятия как средства измерения.

Тест может быть гомогенным (содержательно охватывать один предмет), гетерогенным (содержательно направленным на несколько дисциплин, которые можно с его помощью оценить комплексно) и интегрированным (построенным на междисциплинарных, интегрированных знаниях и умениях, с помощью которых можно оценить степень обобщения (синтеза) усвоенных знаний и умений, профессиональную компетентность, сформированность мировоззрения).

В сфере педагогического тестирования существует два основных подхода: нормативно-ориентировочный и критериально-ориентированный.

Их главное отличие друг от друга заключается в способе интерпретации результатов измерения тестовых баллов.

Для нормативно-ориентировочного подхода характерно сопоставление индивидуального результата с результатами других учащихся. Эти тесты лучше всего подходят для решения задач, связанных со сравнением уровня учебных достижений, ранжированием и отбором:

- сравнение индивидуальных результатов учащихся со среднегрупповыми,
- сравнение учебных достижений отдельных испытуемых, проведение на этой основе ранжирования учащихся,
- сравнение между собой учебных классов (групп) и учебных заведений,
- отбор фиксированного количества испытуемых, например, при приеме в учебное заведение или переходе на следующую ступень (уровень) образования.

В основе критериально-ориентированного подхода лежит сопоставление индивидуальных результатов тестирования с общим объемом знаний, который должен быть усвоен учащимися на данном этапе обучения. Эти тесты используются для решения

задач, зв'язаних з оцінкою доли навчального матеріалу, усвоеного випробуємым. Основні з цих задач наступні:

- оцінка доли навчального матеріалу, усвоеної учасниками,
- порівняння з вимогами стандартом освіти досягнення окремих учасників навчальних груп,
- атестація навчальних закладів на основі порівняння усвоених знань з вимогами стандартом,
- вибір стратегії подальшого навчання,
- діагностика труднощів в навчанні,
- вибір випробуємих, досягли необхідного рівня знань, наприклад, при завершенні навчального закладу.

Від чіткого усвідомлення завдань педагогічного контролю залежить вибір способу інтерпретації тестових балів, методи конструювання і особливості застосування педагогічних тестів. Важливо відзначити, що можливо поєднання обох підходів в одному тесті. Такі тести найбільш інформативні, але їх розробка відрізняється підвищеною складністю.

Існує багато різновидностей тестів, які розділяються на групи за наступними підставами:

- за предметом тестування (тому якості, яку оцінюється з допомогою тесту),
- за особливостями використовуваних в тесті завдань,
- за матеріалом, пред'являемому випробуємому,
- за об'єктом оцінювання.

За предметом тести діляться на інтелектуальні (оцінюють ступінь розвитку людини різних пізнавальних процесів), особистісні і міжособистісні.

За особливостями використовуваних завдань виділяють практичні, образні і вербальні (словесні) тести.

За характером тестових матеріалів, пред'являемим випробуємому, тести діляться на бланкові і комп'ютерні (апаратні).

За обсягом оцінювання – на процесуальні тести, тести досягнень, тести станів і властивостей.

Інтелектуальні тести призначені для оцінювання рівня розвитку мислення (інтелекту) людини.

Особистісні тести пов'язані з психодіагностикою стійких індивідуальних здібностей людини, визначають його поведінку (тести темперамента, характеру, мотивацій, емоцій, здібностей). Тести, дають різносторонню, комплексну оцінку стану особистості або ступеня розвитку стану особистості, називають комплексними, наприклад, тест Кеттелла, ММПІ і інші. До числа окремих можна віднести тести, з допомогою яких оцінюються окремі риси особистості.

Міжособистісні тести дозволяють оцінювати людські стосунки в різних соціальних групах, наприклад, соціометричний тест, тест соціально-психологічної самооцінки групи як колективу.

Практичні тестові завдання включають в себе задачі і вправи, які випробуємый повинен виконати в наочно-дійовому плані, тобто практично маніпулюючи реальними матеріальними предметами або їх заміниками.

Образні завдання містять в собі вправи з образами, з картинками, малюнками, схемами, представленнями. Вони передбачають активне використання уявлення, мислених перетворень образів.

Вербальні тести включають в себе знання на орієнтування словами. Вони передбачають, наприклад, визначення понять, умозаключення, порівняння обсягу і змісту різних слів, виконання з поняттями різних операцій і т.п.

Многие задания, которые используются в тестах, имеют комплексный характер, т.е. включают в себя и практические, и теоретические, и вербальные действия и упражнения. Это связано с тем, что большинство задач, с которыми сталкивается человек в реальной жизни, являются комплексными по характеру, а тестирование проводится для того, чтобы предсказать поведение человека и его возможные достижения в реальной жизни.

Бланковыми называются тесты, при использовании которых испытуемый получает тестовый материал в форме различных бланков (рисунков, схем, таблиц, опросников и т.п.).

Аппаратурные – это такие тесты, в которых используется различного рода аппаратура для предъявления и обработки результатов тестирования, например, аудио- и видеотехника, электронно-вычислительные машины.

Процессуальными называют тесты, с помощью которых используется какой-либо психологический или поведенческий процесс и ему в результате дается точная качественная или количественная характеристика, например, процесс запоминания человеком материала, процесс межличностного взаимодействия индивидов в группе.

К группе тестов достижений относятся такие тесты, в итоге применения которых оцениваются успехи человека в том или ином виде деятельности, в той или иной сфере познания, например, продуктивности памяти, логичности мышления, устойчивости внимания, уровня развития вербального мышления и др.

Тесты состояний и свойств касаются диагностики более или менее стабильных психологических качеств человека (черты личности, свойства темперамента, способности и т.п.).

В особую группу выделяются проективные тесты. Они основаны не на прямой, а на косвенной оценке тех или иных психологических качеств человека. Такая оценка получается в результате анализа того, как данный человек воспринимает и интерпретирует многозначные объекты: сюжетно-неопределенные картинки, бесформенные пятна, незавершенные предложения и др.

Предполагается, что в оценке к инструкции подобных объектов человек бессознательно «вкладывает» или «проецирует» самого себя. К этой группе тестов относятся тест Рармаха, тематический апперцептивный тест и многие другие.

Были сформулированы шесть основных критериев качества тестов: объективность, надежность, валидность, дифференцированность, сложность, точность.

Объективность теста достигается минимизацией влияния субъективных факторов на результат тестирования, в частности, путем унификации условий выполнения теста, регламентацией процедур тестирования, формированием коллегиальной оценки, выставляемой специально созданной комиссией, или использованием стандартных тестовых программ, технических средств контроля, в частности, с помощью компьютера.

Надежность теста (повторяемость (стойкость), согласованность его результатов, полученных теми же самыми учащимися при повторном тестировании по тому же самому тесту или его эквивалентной формой) обеспечивается правильным выбором инструментария, его апробацией и логическим совершенством содержания задания.

Валидность (valid – пригодный, годный, имеющий значение, ценный) или соответствие (адекватность) требованиям и цели тестирования, т.е. соответствие тестового задания тому уровню усвоения понятий, признаков и т.п., который запланирован как цель. Валидность является статистической характеристикой, определяющей содержательные и функциональные параметры метода измерения – отклонения результатов тестирования от закона нормального распределения, что характеризуется асимметрией и эксцессом получения данных.

Сложность тестового задания определяется эмпирически по качеству правильно / неправильно данных ответов; чем больше количество неправильных ответов, тем выше его сложность.

Точность означает минимизацию ошибки измерения.

Различают тесты в закрытой и открытой формах.

Тестовые задания закрытой формы могут быть с множественным выбором ответа, с простым множественным выбором, альтернативным, по установлению последовательности.

Тестовые задания открытой формы предусматривают свободное конструирование ответа.

По форме проведения тесты могут быть индивидуальными и групповыми, устными и письменными, бланковыми, предметными, аппаратурными, в частности, компьютерными, вербальными и невербальными. При этом каждый тест имеет несколько составных частей: руководство по работе с тестом, тестовую тетрадь с заданиями и, если необходимо, стимульный материал или аппаратуру, лист ответов (для бланковых методик), шаблоны для обработки данных.

В руководстве приводятся данные о целях тестирования, выборке, для которой тест предназначен, результатах проверки на надежность, способах обработки и оценки результатов. Задания теста, сгруппированные в субтесты (группы заданий, объединенные одной инструкцией), помещены в специальной тестовой тетради (тестовые тетради могут быть использованы многократно, поскольку правильные ответы отмечаются на отдельных бланках). Если тестирование проводится с одним испытуемым, то такие тесты носят название индивидуальных, а если с несколькими – групповых. Каждый тип теста имеет свои достоинства и недостатки. Преимуществом групповых тестов является возможность охвата больших групп испытуемых одновременно (до нескольких сот человек), упрощение функций экспериментатора (чтение инструкций, точное соблюдение времени), более единообразные условия проведения, возможность обработки данных на ЭВМ и др.

Основным недостатком групповых тестов является снижение возможностей у экспериментатора добиться взаимопонимания с испытуемыми, заинтересовать их. Кроме того, при групповом тестировании затруднен контроль за функциональным состоянием испытуемых, таким, как тревожность, утомление и др. Иногда для того, чтобы понять причины низких результатов по тесту какого-либо студента, следует провести дополнительное индивидуальное собеседование. Индивидуальные тесты лишены этих недостатков.

Тестирование широко используется в учебных заведениях для тренировочного, промежуточного и итогового контроля знаний, а также для обучения и самоподготовки учащихся.

Как уже указывалось, результаты тестирования могут выступать как и оценка качества преподавания, а также как оценка самих испытательных материалов.

Не меньший интерес представляет изучение результатов тестирования для определения качества лекции или семинара. Например, пусть в потоке у лектора несколько групп, и во всех проведено тестирование по заданному разделу курса. В тесте имеется определенное количество теоретических вопросов и практических задач. Каждый вопрос соответствует какой-либо теме. По этой же теме в тесте прилагается практическая задача. Если студенты во всех группах плохо справились с каким-либо теоретическим заданием и практической задачей к этому вопросу, следовательно, на лекции и на семинарах не уделено достаточно внимания этой теме (хотя необходимо учитывать, что группы неравномерны по контингенту).

В настоящее время более часто используются следующие варианты тестовых контрольных мероприятий:

- «автоматический», когда обучаемый выполняет задание в непосредственном диалоге с ЭВМ, результаты сразу переносятся в блок обработки,
- «полуавтоматический», когда задания выполняются письменно, а ответы со специальных бланков вводятся в ЭВМ (решения не проверяются),

– «автоматизированный», когда задания выполняются письменно, решения проверяются преподавателем, а в ЭВМ вводятся результаты проверки.

Особенностью первых двух является отстраненность преподавателя от проверки результатов испытаний. В этом случае, казалось бы, их объективность повышается. Однако, при этом утрачивается значительная часть информации, которую можно было бы получить при анализе результатов тестирования с использованием человеческого фактора.

В «автоматическом» режиме такой потери можно избежать. Но при использовании такого метода на сегодняшний день возможно появление некоторых специфических проблем:

– отсутствие достаточного парка ЭВМ. Не все учебные заведения могут позволить себе оснастить классы дорогостоящим компьютерным оборудованием в достаточном количестве,

– отсутствие навыков пользователя ЭВМ у обучаемых. Иногда приходится работать со студентами, у которых по каким-то причинам нет достаточной компьютерной подготовки, или же они вообще никогда не общались с компьютером,

– сложность и дороговизна разработки программного обеспечения,

– существование проблемы распознавания ответов произвольной формы в открытых текстовых заданиях.

И.П. Подласый классифицировал тесты по различным аспектам (компонентам) развития и формирования человеческих качеств:

1. Тесты общих умственных способностей, умственного развития.

2. Тесты социальных способностей в различных областях деятельности.

3. Тесты обученности, успеваемости, академических достижений.

4. Тесты для определения отдельных качеств (черт) личности (памяти, мышления, характера и др.).

5. Тесты для определения уровня воспитанности (сформированности общечеловеческих, нравственных, социальных и других качеств).

3. **Разработка и формирование тестов.** При разработке тестов важно, насколько они соответствуют запроектованным целям обучения, образования, развития обучаемых.

Тест обученности – это совокупность заданий, ориентированных на определения уровня усвоения определенных частей содержания образования.

Для создания теста необходимо провести его спецификацию, предусматривающую следующее:

– определение цели тестирования, выбор способа проведения (бланковое или аппаратное, компьютерное),

– анализ предметной области знаний, подлежащих контролю, конкретизация элементов знаний, подлежащих оцениванию,

– определение структуры теста и принципы размещения заданий (от простейших до более сложных, смешанная подача, формирование субтестов),

– построение матрицы теста, с помощью которой определяется объем материала, репрезентирующую данную область знаний, когнитивные умения или структура компетентности учащихся, уровень усвоения материала,

– выбор формата заданий и ответов,

– определения процедур проведения тестирования и их регламентация, подготовка инструкций,

– конструирование предварительных вариантов теста, которые апробируются по выбранной методике,

– пилотное тестирование, сбор эмпирических данных и их психометрический анализ,

– подбор тестовых заданий и конструирование вариантов теста, определение его длительности,

- инструктаж и контрольное тестирование,
- обработка результатов контрольного тестирования, их шкалирование,
- анализ и интерпретация полученных данных.

Важнейшими критериями диагностических тестов обученности являются действенность (валидность, показательность), надежность (вероятность, правильность), дифференцированность (различимость).

Действенность теста по своему содержанию близка к требованию полноты, всесторонности проверки, пропорционального представления всех элементов изучаемых знаний, умений. Термин действенность имеет, как минимум, два синонима – валидность и показательность, представительность, трактуемые также, как репрезентативность. Составитель теста обязан тщательно изучить все разделы учебной программы, учебники, учебные и учебно-методические пособия, хорошо знать цель и конкретные задачи обучения. Тогда он сможет составить тесты, которые будут действенными для определенной категории обучаемых. Неотъемлемым условием действенности теста является четкая и ясная постановка вопроса в пределах усвоенных знаний. Если тест выходит за пределы усвоенного содержания или же не достигает этих пределов, повышает запроецированный уровень обучения, то он не будет действенным для тех обучаемых, которым он адресован.

Действенность теста определяется статистическими методами. Величина 0,7-0,9 коэффициента корреляции свидетельствует о высокой действенности тестов обученности. Если же коэффициент корреляции достигает значений 0,45-0,55, то действенность теста считается удовлетворительной, при более низких значениях она считается неудовлетворительной.

Степень надежности характеризуется стабильностью, устойчивостью показателей при повторных измерениях с помощью такого же теста или его равноценного заменителя. Количественно этот показатель характеризуется вероятностью достижения запроецированных результатов (правильностью значений). Грамотно составленные и апробированные тесты обученности позволяют достичь коэффициента надежности 0,9. Установлено, что надежность теста повышается при увеличении количества тестовых заданий.

Установлено также, что чем выше тематическое, содержательное разнообразие тестовых заданий, тем ниже надежность теста. Это следует понимать так: тест, нацеленный на проверку усвоение конкретной темы, всегда будет более надежным, чем тест, направленный на проверку всего раздела (курса), охватывающий значительное количество материала – закономерностей, концепции, фактов. Происходит это именно потому, что содержательное разнообразие последнего выше.

Надежность тестов обученности значительно зависит от трудности их выполнения. Трудность определяется по соотношению правильных и неправильных ответов на тестовые вопросы. Включение в состав тестов таких заданий, на которые все обучаемые отвечают правильно или же, наоборот, резко снижает надежность теста в целом. Наибольшую практическую ценность имеют задания, на которые правильно отвечают 45-80 % обучаемых.

Характеристика дифференцированности (различимости) связана с использованием таких тестов, где нужно выбирать правильный ответ из нескольких возможных альтернатив. Если, скажем, все учащиеся безошибочно находят правильный ответ на один вопрос и также дружно не могут ответить на другой, то это сигнал для совершенствования теста в целом. Его необходимо дифференцировать, сделать различимым. Иначе подобные задания не помогут отделить тех, кто усвоил материал на необходимом уровне, от тех, кто заданного уровня не достиг. Практически дифференцируют тесты по результатам статистического анализа, сравнивая результаты выполнения теста в целом с результатами выполнения отдельных заданий. Если коэффициент корреляции между ответами на

конкретные задания и на тест в целом больше 0,5, то это свидетельствует о достаточной дифференцированности теста.

При подготовке материалов для тестового контроля необходимо придерживаться таких основных правил:

1. Нельзя включать ответы, неправильность которых на момент тестирования не может быть обоснована учащимися.
2. Неправильные ответы должны конструироваться на основе типичных ошибок и должны быть правдоподобными.
3. Правильные ответы среди всех предлагаемых ответов должны размещаться в случайном порядке.
4. Вопросы не должны повторять формулировок учебника.
5. Ответы на одни вопросы не должны быть подсказками для ответов на другие.
6. Вопросы не должны содержать «ловушек».

При составлении заданий теста важным является решение проблемы правдоподобности неправильных вариантов ответов. При этом возможны два аспекта проблемы – внешняя правдоподобность и правдоподобность содержания, то есть неправильными вариантами ответов могут быть:

- правдоподобные внешне и правдоподобные по содержанию,
- правдоподобные внешне, но неправдоподобные по содержанию,
- неправдоподобные внешне, но правдоподобные по содержанию.

Нельзя использовать ответы неправдоподобные и внешне, и по содержанию.

Наиболее целесообразной структурой при построении выборочных тестов нужно считать:

- один вариант правильный, остальные неправильные (неполные), но обязательно правдоподобные,
- все варианты неправильные (неполные), но обязательно правдоподобные, при наличии варианта «правильного ответа нет»,
- один вариант неправильный (неполный), остальные варианты правильные.

Тесты обученности должны удовлетворять следующим требованиям:

- не требовать больших затрат времени (относительная краткосрочность),
- не допускать произвольного толкования тестового задания (однозначность),
- исключить возможность формулирования многозначных ответов (правильность): задания должны быть 1) относительно краткими, требующими сжатых ответов, 2) информационными, т.е. обеспечивающими возможность соотнесения количественной оценки за выполнение теста с порядковой или даже интервальной шкалой измерений, 3) удобными, т.е. пригодными для быстрой математической обработки результатов, 4) стандартными, т.е. пригодными для широкого практического использования – измерения уровня обученности возможно более широких контингентов обучаемых, овладевающих одинаковым объемом знаний на одном и том же уровне обучения.

Отработана методика формирования теста:

1. Вводятся понятийные индикаторы первой и последующих ступеней: «Знание учебного предмета (курса, раздела, темы)». Понятийные индикаторы выбираются с учетом требований квалификационной характеристики будущего специалиста, а также межпредметных связей, используемых при изучении контролируемой дозы учебного материала. При этом очерчивается поле действия преподавателя и учащегося на этапе контрольных процедур.

2. Выбирается (задается) количество заданий теста. Методисты рекомендуют 30-60 тестовых заданий на раздел рабочей программы, однако в отдельных случаях могут быть значительные отклонения от этой рекомендации в зависимости от структуры программного материала.

3. Составляется план теста в зависимости от важности контролируемых вопросов в структуре учебного материала.

Нормируются задачи тестирования: умение давать определения, знание формул (теории), умение применять теорию, умение работать с графическим материалом, умение составлять программы.

4. Составляются предварительные тестовые задания, количество которых может в 2-3 раза превышать требуемое.

Количество предварительных заданий по контролируемым разделам и умениям должно быть пропорционально плану теста.

5. Осуществляется опытная проверка тестовых заданий. Контролируется группа с помощью одного теста, за правильный ответ обучаемый получает 1, за неправильный – 0.

Особые требования к тесту: он должен дифференцировать обучаемых по уровням их знаний, поэтому задания, на которые ответили все или не ответил никто, должны быть удалены из теста.

Рекомендуется иногда оставить по одному трудному заданию, на которое может ответить хотя бы один из тысячи обучаемых. В результате этой процедуры количество тестовых заданий доводится до плановой цифры.

6. На основе результатов проверки оставленных тестовых заданий строится матрица теста.

Общественные методы статистической обработки результатов испытания позволяют определить качество теста в целом и каждого в отдельности. Наиболее характерными показателями являются:

1. Средний балл теста

$$M = \frac{\sum_{i=1}^N X_i}{N}$$

где X_i – результат контроля по i -ому заданию;

N – количество заданий.

2. Дисперсия заданий теста

$$\sigma^2 = \sum_{i=1}^N \rho_i q_i,$$

где ρ_i – доля правильных ответов;

q_i – доля неправильных ответов.

3. Средняя квадратическая ошибка (стандарт):

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - M)^2}{N - 1}},$$

Наиболее распространенной формой оценки ответов на тесты является трех-канальная: правильный ответ – плюс один балл, «не знаю» – 0 баллов, неправильный ответ – минус один балл.

4. Преимущества и недостатки тестирования. Преимуществами тестирования являются:

– большая объективность и большее позитивное стимулирующее воздействие на познавательную деятельность учащихся,

– исключение воздействия негативного влияния на результаты тестирования таких факторов, как настроение, уровень квалификации и другие характеристики конкретного преподавателя,

- ориентированность на современные технические средства в среде компьютерных обучающих систем,
- универсальность, охват всех стадий процесса обучения,
- многофункциональность, что позволяет быстрее понять, как дальше работать с данным учащимся, а также помогает преподавателю скорректировать курс.

Недостатками тестирования являются:

- создание тестов, их унификация и анализ – это большая кропотливая долговременная работа,
- доведение теста до полной готовности к применению – это сбор статистических данных за несколько лет, хотя бы с потоком студентов 100-120 человек,
- значительный субъективизм в формировании содержания тестов, в отборе и формулировке тестовых вопросов, в их структуре,
- отсутствие достаточного парка ЭВМ, недостаточность компьютерных классов,
- отсутствие навыков пользователя ЭВМ у студентов,
- сложность и дороговизна программного обеспечения,
- проблема распознавания ответов произвольной формы в открытых тестовых заданиях.

Замена устного экзамена тестированием – не панацея, поскольку тестовые задания и их выборочная система ответов не дают возможности в полной мере проверить умственное развитие студента, его логическое и абстрактное мышление, устную речь, выяснить причины ошибок. Кроме того, разные дисциплины неодинаково поддаются формализации и поэтому плохо вкладываются в систему тестовых заданий. Несколько ответов, среди которых нужно выбрать один, есть не что иное, как «наводящие вопросы», что облегчает и делает однообразными интеллектуальные усилия тестируемых. Это может привести к нарушению равновесия между индуктивным и дедуктивным мышлением.

Регулярное тестирование так или иначе формирует стандарты в образовательной системе региона, страны. В этом есть, на наш взгляд, определенная опасность. Реальная школа, хотим мы этого или не хотим, как правило, подменяет учебную цель: ею становится суммирующий контроль. При переходе на всеобщее тестирование очень быстро для учителя основной целью стала подготовка учеников к сдаче теста. Открываются курсы по подготовке к сдаче тестов. Возникли соответствующие методики, позволяющие, не изучая предмет по-настоящему, готовить школьников к сдаче теста.

Тесты пригодны и нужны для проверки механизма узнавания и исполнения, но не на умение владеть знаниями и умениями.

Следует помнить слова известного математика П.Л. Чебышева: «Новое в преподавании полезно только тогда, когда на опыте проверено, что оно лучше старого».

Все это означает, что нужно разумно объединять традиционные формы контроля с тестированием.

ЛИТЕРАТУРА

1. Аванесов В.С. Композиция тестовых заданий. – М., 2002.
2. Аванесов В.С. Теоретические основы разработки заданий в тестовой форме. – М., 1995.
3. Анастаси А., Урбина С. Психологическое тестирование. – М. ПИТЕР, 2003.
4. Буравлев А.И., Переверзев В.Ю. Методы подсчета индивидуальных баллов при математическом моделировании процесса педагогического тестирования // СПО. – 2001. – № 4.
5. Гулидов И.Н. Педагогический контроль и его обеспечение: Учебное пособие. – М.: Форум, 2005. – 240 с.
6. Гулидов И.Н., Шатун А.Н. Методика конструирования тестов. – М., 2003.

7. Дзундза А. Досвід застосування тестів об'єктивного контролю рівня освітньо-професійної підготовки бакалаврів у Донецькому національному університеті // Освіта і управління. – 2002. – Т.5. – № 4. – С. 157-165.
8. Енциклопедія освіти / Акад. пед. наук України; головний ред. В.Г. Кремень. – К.: Юрінком Інтер, 2008. – 1040 с.
9. Ингенкамп К. Педагогическая диагностика: Перев. с нем. – М.: Педагогика, 1991. – 240 с.
10. Корсак К.О. О качестве систем педагогических измерений // Народное образование. – 2002. – № 4.
11. Крилова Т.В., Орлова О.Ю. Історія розвитку тестування та його застосування // Дидактика математики: проблеми і дослідження: Міжнар. збірник наукових робіт. – Вип. 32. – Донецьк: Вид-во ДонНУ, 2009. – С. 138-142.
12. Лосева Н. Тестування в умовах багатоступеневої підготовки фахівців у вищій школі // Освіта і управління. – 2002. – Т.5. – № 4. – С. 150-156.
13. Майоров А.Н. Тесты школьных достижений: конструирование, проведение, использование. – СПб: Образование и культура. – 1996.
14. Михайлычев Е.А. Дидактическая тестология. – М., 2001.
15. Подласый И.П. Тестирование в учебном процессе: его история и возможности // Андрогогика. Повышение квалификации. Управление знаниями. – www. elitarium. ru. – 2006.
16. Подласый И.П. Новый курс: Учебник для студентов педагогических вузов: В 2 кн. – М.: Гуманит. изд. центр ВЛАДОС, 1999. – 576 с.
17. Педагогика. Учебное пособие для студентов педагогических вузов и педагогических колледжей / Под ред. П.И. Пидкасистого. – М.: Педагогическое общество России, 1998. – 640 с.
18. Симонов В.Л. Новая философия оценки степени обученности личности // Специалист. – 2000. – № 4.
19. Система управління якістю медичної освіти в Україні. – Д., 2003.
20. Слепкань З.І. Наукові засади педагогічного процесу у вищій школі: Навч. посібник. – К.: Вища школа, 2005. – 239 с.
21. Стоунс Э. Психопедагогика. – М., 1984. – 238 с.
22. Хлебников В.А., Михалева Т.Г. Централизованное тестирование в России: необходимость, возможность, проблемы. // Школьные технологии. – 1999. – № 1-2. – С. 213-219.
23. Хубаев Г. О построении шкалы оценок в системах тестирования // Высшее образование в России. – 1996. – № 1. – С. 122-125.
24. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. – М.: ЛОГОС, 2002.
25. Keeves J. P. (Ed) Educational Research, Methodology and Measurement: An International Handbook. – Oxford, 1998.
26. www.refine.ord.ua.
27. www.bks.ru/russian.htm.