

Раздел 2. Развитие системы тестирования в России и за рубежом. Психолого-педагогические аспекты тестирования

Текст лекций

Цели:

- изучить исторические предпосылки современного тестирования в отечественном образовании;
- изучить развитие тестирования в зарубежных странах;
- рассмотреть особенности организации тестирования; условия, необходимые для реализации системы тестирования;
- рассмотреть психолого-педагогические аспекты тестирования;
- изучить основные определения.

2.1. Развитие тестирования в России и за рубежом

Исторические предпосылки современного тестирования в отечественном образовании

Тесты в России в конце XIX первой половины XX в. История тестов, пришедших на смену донаучным формам и обыденным представлениям педагогов о проверочных средствах, началась в России, как и в других странах, в конце XIX — начала XX в. В этот период тесты широко применялись сторонниками экспериментальной педагогики, которые всячески пропагандировали их наравне с другими психометрическими методиками в педологии. Видные специалисты того времени (А. П. Болтунов, А.Ф.Лазурский, А.П.Нечаев, Ф.Е.Рыбаков и др.) в своих исследованиях личности обучаемого широко использовали тесты.

На волне стремления к прогрессу в России в первые годы советской власти был востребован потенциал, накопленный российской наукой в образовании. У истоков советской педологии, в рамках которой развивались тестовые методики, стояли известные ученые того времени (М.Я. Басо, Л.С. Выготский, А.П.Болтунов, П.П. Блонский, К.Н. Корнилов, А.П. Нечаев, А.Р. Лурия и др.), посвятившие ряд своих работ оценке результатов обучения и развития детей. Однако период активной деятельности педологов и тестологов в советской России оказался коротким. В 1936 г. вышло известное постановление ВКП(б) «О педологических извращениях в системе Наркомпросов», в котором тесты были объявлены вредным методом. Многие годы, вплоть до начала 90-х гг. XX в., отношение к тестам было негативным. Педагоги-практики вернулись к тестам в период прекращения массовых политических репрессий. Уже к началу 60-х гг. XX в. многие учителя, заботящиеся не о формальных показателях, а о реальном качестве учебного процесса, стали использовать при контроле наборы заданий в тестовой форме, которые чаще всего назывались по-разному, но только не словом «тест».

Промежуток времени с 60-х гг. до начала 90-х гг. XX в. был для отечественной школы периодом постепенной либерализации, когда учителя изыскивали различные резервы в борьбе за повышение качества обучения на фоне снижения авторитарного характера учебного процесса. Тесты в это время существовали полулегально, официально запрет на них не был отменен, но уже предпринимались попытки диссертационных исследований по проблемам тестирования, появлялись работы педагогов-новаторов, в которых тесты и анкетные опросы использовались как инструментарий для подтверждения эффективности нововведений. В это время тесты разрабатывались без должного знания теории педагогических измерений, на основе опыта и здравого смысла, поэтому чаще всего тестами их было можно назвать лишь условно.

Столь же необоснованными по надежности и валидности были наборы заданий с выбором ответов, создаваемые практически повсеместно на рубеже 60-70-х гг. XX в. в связи с интенсивным развитием программированного обучения. Только к середине 80-х гг. XX в. в нашей стране в научных и методических работах преподавателей стали утверждаться основные положения теории педагогических измерений, декларирующие необходимость эмпирической верификации качества измерений и статистического анализа характеристик тестов.

В целом рассматриваемый временной промежуток в развитии тестов был периодом прогресса, когда на смену представлениям о тесте как о простом наборе заданий пришло научное понимание этого термина. В этот период появляются научные издания по тестовой и смежной с ней проблематике, учебные пособия и многочисленные статьи отечественных авторов, как поддерживающие, так и осуждающие тесты в образовании.

Развитие тестов в конце XX в. и в наши дни. Новая история тестов в России началась в 90-е гг. XX в., когда стали больше говорить не о руководстве, а о научно обоснованном управлении учебным процессом, в информационном обеспечении которого важная роль по праву принадлежит тестам. Новое понимание возможностей тестов в образовании способствовало росту научных исследований. В 90-е гг. XX в. появляются работы по проблемам измерений и тестирования в образовании, защищаются многочисленные кандидатские и докторские диссертации, издаются монографии, учебные пособия, журналы, проводятся конференции и симпозиумы.

Последнее десятилетие XX в. в нашей стране совпало с периодом бурного развития структур, занимающихся практической работой по созданию и применению тестов. К основным событиям этого периода можно отнести открытие в 1990 г. первой в стране кафедры педа-

гогических измерений в Исследовательском центре проблем качества подготовки специалистов (директор Н. А. Селезнева) Московского института стали и сплавов; создание при Московском государственном университете в этом же году Центра тестирования «Гуманитарные технологии» (проект по компьютерному аттестационному тестированию для старшеклассников «Теле-тестинг» - научный руководитель А.Г.Шмелев); привлечение в 1991 г. Центра качества образования Института общего среднего образования РАО к участию в сравнительных международных исследованиях по оценке учебных достижений (директор Центра - Г. С. Ковалева); открытие в 1995 г. Центра тестирования выпускников общеобразовательных учреждений, преобразованного впоследствии в Федеральный центр тестирования со статусом государственного учреждения Министерства образования и науки Российской Федерации; основание в 1998 г. государственной системы тестирования иностранных граждан по русскому языку как иностранному при поддержке ведущих вузов Москвы, Санкт-Петербурга и других городов России. Перечень этот можно было бы смело продолжить. Начиная с 90-х гг. XX в. практически во всех регионах России создаются центры тестирования и аттестации учащихся, методические лаборатории по диагностике и центры качества образования.

Значимым событием в области подготовки кадров по педагогическим измерениям стало открытие в 2001 г. на факультете повышения квалификации Российского университета дружбы народов (декан Т. М.Балыхина) кафедры тестологии (заведующая кафедрой М.Б. Чельшкова). На этой кафедре впервые в нашей стране началась реализация профессиональной образовательной программы, рассчитанной на 1480 часов, для получения дополнительной квалификации «Тестолог (специалист в области педагогических измерений)».

Для повышения качества тестовых материалов, разрабатываемых в России, в 2000 г. при Исследовательском центре проблем качества подготовки специалистов Московского института стали и сплавов (технологического университета) был открыт Центр сертификации педагогических тестовых материалов (ПТМ) (директор В.И. Звонников) и создан Координационный совет Минобрнауки России (в настоящий момент - Минобрнауки) по вопросам сертификации качества педагогических тестовых материалов.

Коренные изменения в отношении учителей к тестам произошли в 2001 г. в связи с началом эксперимента по введению единого государственного экзамена, благодаря которому тесты получили официальное признание в России. За годы эксперимента значительно повысилась степень доверия к результатам тестирования со стороны органов управления образованием, образовательных учреждений, самих учителей, которые убедились в высокой объек-

тивности и обоснованности баллов ЕГЭ, в их высокой прогностичности при отборе абитуриентов вузов. В 2002 г. был открыт Федеральный институт педагогических измерений (директор А.Г.Ершов), приоритетным направлением деятельности которого является научное, методическое и организационное сопровождение процесса создания контрольных измерительных материалов (КИМ) для ЕГЭ. В целом, подводя итоги развития тестирования в России, можно сказать, что на сегодняшний день среди ученых-педагогов в нашей стране, наконец, появилось понимание того, что теория педагогических измерений - наука, обладающая своей методологией, методами и аппаратом, необходимым для разработки качественных педагогических тестов. Сегодня большими тиражами издаются инновационные работы по педагогическим тестам отечественных ученых и зарубежных авторов, а также сборники материалов КИМ ЕГЭ, специальные журналы по тестовой проблематике. Таким образом, можно считать, что на данный момент в России сформировалось сообщество профессионалов - специалистов по разработке и применению тестов.

Развитие тестирования в зарубежных странах

Становление тестов в психологии, образовании и армии. Появление педагогических тестов за рубежом нередко связывают с именем французского врача и психолога А. Бине. Его работа по диагностике интеллектуальных способностей, вышедшая в 1905 г., считается точкой отсчета в становлении основных научных подходов к измерениям в психологии и образовании [69]. Конечно, были и более ранние попытки создания подобных тестов. Британские исследователи Ф. Гальтон и Дж. Кэттелл в 1890 г. использовали термин «тест на интеллект». Им же принадлежит заслуга введения в научный оборот специальной характеристики качества теста, указывающей на его способность дифференцировать испытуемых по измеряемой переменной.

Немного позднее идеи А. Бине были использованы немецким психологом и философом В.Штерном, который предложил специальный коэффициент для оценивания интеллекта - коэффициент 10. Этот термин используется до сих пор, правда, несколько в ином контексте. В США работы А. Бине по измерениям продолжил Л.Терман. Он создал новые тесты (Стэнфорда-Бине), представляющие собой оригинальную модификацию работы его предшественника.

Достижения второго десятилетия XX в. в сфере измерений были связаны с проблемой распределения большого количества призывников по различным родам войск в США во время Первой мировой войны. Правительство Соединенных Штатов организовало «мозговой центр»

из специалистов — составителей тестов, статистиков и экспертов по измерениям — и направил его работу на решение задач, продиктованных военной необходимостью. В этом центре свои усилия по отбору и распределению новобранцев объединили такие известные специалисты, как Э.Торндайк, Р. Годдард, П.Йеркс и Л.Терман. Результатом их работы стали знаменитые батареи Альфа и Бета тестов для армии, которые после стали использовать (вплоть до настоящего времени) в качестве примеров при обучении разработчиков тестов в образовании.

Первые стандартизированные измерители в образовании США были созданы для оценки качества почерка, правописания и выполнения арифметических действий в конце 20-х гг. XX в. Тогда же в Америке появилась батарея тестов, разработанных по инициативе Совета колледжей и предназначенных для отбора абитуриентов. Несколько позже, в конце 40-х гг. XX в., были созданы батареи тестов достижений, реализующие идеи многомерных измерений и обеспечивающие сопоставимость результатов по разным школьным предметам.

Развитие классической (традиционной) теории педагогических измерений и тестирования в XX в. В 1904 г. англичанин Ч. Спирмен опубликовал фундаментальный теоретический труд по исследованию общих интеллектуальных способностей. Используя школьные оценки по различным предметам, Ч. Спирмен применил к данным тестирования аппарат новой для того времени теории корреляции и выявил примеры ковариации оценок школьников. Тем самым он заложил основы научных подходов к обоснованию качества тестов, соединив в своем исследовании теорию физических измерений, корреляционные методы и накопленный его предшественниками-психологами опыт оценивания способностей детей.

Прообразом научных положений теории педагогических измерений послужила далекая от образования работа Н.Р.Кэмпбелла «Основы физики» (1920), благодаря которой был разработан теоретический аппарат для анализа качества измерений, а обыденное представление о тесте и его научное определение стали заметно различаться. Для обоснования качества педагогических измерений, их надежности и валидности была создана классическая теория тестов, получившая впоследствии название традиционной, и базирующаяся на концепции параллельных измерений и теории корреляции. На основе этой теории в 30 - 40-е гг. XX в. интенсивно разрабатывались количественные методы для анализа качества тестовых заданий, строились стандартизированные тесты учебных достижений и осваивались методы шкалирования результатов выполнения тестов.

В конце 40-х гг. XX в. увлечение чистой теорией сменилось пониманием важности правильного применения измерений в сфере психологии и образования. Исследователи осознали

тот факт, что к тестируемым нужно относиться бережно и осторожно и что любая ошибка в оценках может привести к необратимым последствиям. В этой связи в теории педагогических измерений активизировался поиск эффективных методов повышения надежности результатов тестирования, стали разрабатываться методы факторного, дисперсионного и корреляционного анализа данных, получили широкое развитие методы формирования репрезентативных выборок, необходимых для стандартизации тестов.

Значимым событием конца 40-х гг. XX в. для развития теории и практики измерений в образовании стало создание в 1947 г. в США Службы образовательного тестирования. Сегодня эта организация имеет представительства практически во всех странах мира. Из числа созданных тестов, наибольшую известность в связи с расширением программ международного обмена в обучении приобрел, выявляющий уровни владения различными видами речевой деятельности для обучения в англоязычных странах.

Создание современной теории тестов. В истории тестов были этапы подъемов и спадов. В частности период ожесточенной критики педагогических тестов наблюдался в США в конце 60-х гг. XX в. Широкая публичная дискуссия по проблемам использования тестирования совпала по времени с научной критикой, вызванной недостатками классической теории тестов. Высказывались опасения в том, что тестирование служит инструментом подавления инициативы и творчества учащихся, слишком упрощенно трактует такие сложные конструкты, какими являются учебные умения, не объективно оценивает учебные достижения. Приводились и другие аргументы против широкого использования тестов. Ученые-тестологи занимались в основном конструктивной научной критикой традиционных методов разработки тестов, что привело к построению новой теории, которую нередко в наши дни называют современной теорией тестов.

Неоспоримые преимущества связанные с возможностью прогноза надежности измерений, увеличения эффективности тестирования и получением оценки параметров подготовленности учащихся, не зависящих от трудности заданий теста, стали широко применяться на практике с конца 80-х гг. XX в. В это время были разработаны алгоритмы оценивания параметров испытуемых и заданий на основе математических моделей, создано программное обеспечение и стали широко использоваться ПК. Сочетание аппарата классической теории и ИКТ при разработке тестов открыло новые возможности повышения качества педагогических измерений и применения тестирования в образовании.

Современный период в истории развития тестов. Современный период в истории развития тестов характеризуется интенсивным развитием теории, созданием новых моделей и методик ее применения, внедрением в тестирование компьютерных технологий, адаптивного тестирования, различных инноваций в области разработки и применения тестов. В настоящий момент наметился ряд направлений исследований, нацеленных на расширение возможностей педагогических измерений, разработку инновационных измерителей и повышение качества тестов. К их числу относится создание новых моделей педагогического измерения, инновационных форм тестовых заданий для проверки творческих и практико-ориентированных аспектов подготовленности учащихся, методов калибровки тестовых заданий и методик компьютерного моделирования тестов, обеспечивающих планируемую точность измерений.

Большое внимание уделяется психолого-педагогическим проблемам тестирования (проблемам тревожности, мотивации), развитию специальных процедур для выявления размерности пространства измерений, созданию специальных методик, позволяющих обоснованно использовать результаты педагогических измерений в управлении качеством образования.

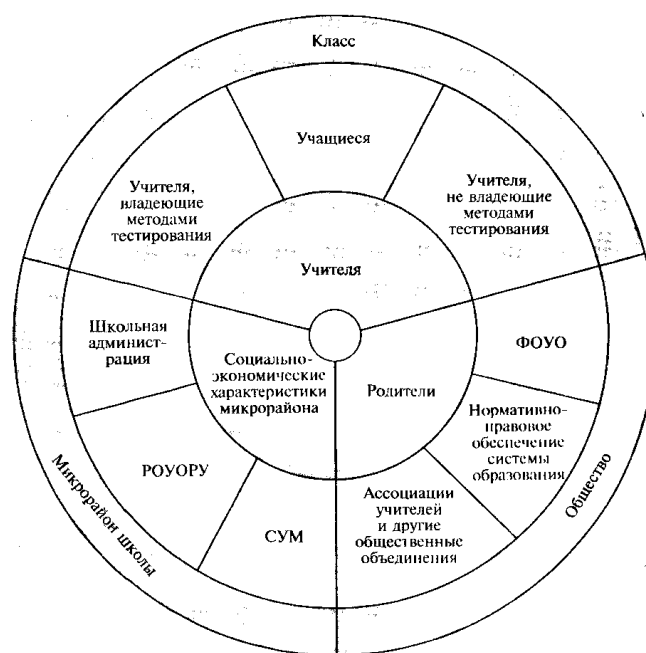
Особенности организации тестирования. Условия, необходимые для реализации системы тестирования

Использование тестовых методик предполагает наличие определенных *условий*, которые имеются далеко не во всякой школе:

1. Прежде всего *требуются сами тестовые задания*, разработка которых не может быть инициативной работой одного учителя в ущерб своему свободному времени. Желательно, чтобы в создании тестов *принимали участие почти все преподаватели школы*, чтобы они обменивались своими достижениями, совместно преодолевали трудности и получали благодаря поддержке администрации *денежное вознаграждение* за свою дополнительную работу.
2. Необходимо также иметь *программно-инструментальное обеспечение* для ведения баз данных тестирования, хранения банков заданий, выдачи заданий учащимся в компьютерной форме или распечатки заданий на бланках и обработки данных тестирования. Как уже отмечалось, перечисленные условия есть далеко не во всех учебных заведениях, поэтому нередки случаи, когда по возвращении в школу после очередного повышения квалификации по тестовым методикам учитель сразу забывает все, чему научился, и включается в повседневный учебный процесс.

Какие тесты необходимо разрабатывать и применять всем учителям? Математико-статистический аппарат, используемый профессионалами в полном объеме при разработке тестов для административно-управленческих решений, довольно сложен, но он учителям и не нужен. В основном в учебном процессе используются тесты для входного и текущего контроля, не нуждающиеся в серьезном статистическом обосновании. От разработчиков таких тестов требуется лишь владение методикой отбора содержания, знание требований к тестовым формам и простейших показателей дескриптивной статистики, необходимых для выполнения основных требований к качеству теста. Поскольку на тесты для текущего контроля приходится основная доля по времени и объему использования, то будет верным считать, что учитель - ключевая фигура в разработке и применении педагогических тестов.

Факторы, влияющие на интерпретацию результатов педагогического тестирования



РОУОРУ – руководители органов управления образованием районного уровня

СУМ – сообщество учителей микрорайона

ФОУО – федеральные органы управления образованием

Факторы, влияющие на интерпретацию результатов тестирования. Помимо разработки тестов для текущего контроля учителю также нужно обязательно овладеть методикой применения педагогических тестов, интерпретацией их результатов, в первую очередь, для разъяснения родителям учеников проблем и трудностей, с которыми сталкиваются их дети. Та-

кая интерпретация не может иметь изолированный, полностью беспристрастный характер. На нее оказывают влияние многочисленные факторы и сопутствующие элементы социально-экономической среды, окружающей учащихся, учителя, класс и в целом школу (см. рис.).

Результаты тестирования влияют на отношения педагога с учащимися и их родителями и на его статус в профессиональном сообществе. Поэтому умения применять педагогические тесты, шкалировать и интерпретировать результаты их выполнения необходимы каждому учителю для самооценки эффективности своей работы, выявления в ней слабых мест и поиска факторов профессионального роста. Игнорирование или отрицание потребности в тестировании не отменяют эту потребность, поэтому, учитывая, что педагог работает в условиях ограниченного времени, правильнее было бы начать осваивать разработку и применение тестов еще в студенческие годы, до начала педагогической деятельности.

2.1. Психолого-педагогические аспекты тестирования

Взаимосвязь психологических и педагогических измерений. Взаимодействие психологов и педагогов в области разработки тестов имеет свои исторические причины и научные предпосылки. В научном плане оно предопределено близостью методологий измерений в эмпирических науках и единством методик разработки тестов, общностью объекта исследования, тесной взаимосвязью и частичным пересечением субъектов измерения. Исторические аспекты этого взаимодействия обусловлены особенностями становления теории педагогических измерений, сложившимися за рубежом в 20 —40-е гг. XX в. В то время потребность в стандартизированных педагогических тестах постоянно увеличивалась, а сообщество тестологов в сфере образования практически еще не сложилось, поэтому психологи зачастую привлекались к работе над педагогическими тестами.

В наши дни область совместных работ педагогов и психологов расширяется. Во многих учебных заведениях в России и за рубежом работают психологи, создаются специальные службы консультирования школьников и студентов. В их функции входит не только психологическая поддержка учащихся в проблемных ситуациях, но и участие в проведении мониторинга, анкетных опросов и в подготовке учебного заведения к государственной аттестации. Педагоги, в свою очередь, проявляют интерес к оцениванию способностей обучающихся, когнитивных аспектов подготовленности, психологических особенностей учащихся в усвоении знаний и нередко выполняют свои исследования на стыке психодиагностики и измерений в образовании.

Тенденция вхождения психологов в педагогические измерения в нашей стране наметилась в середине 90-х гг. XX в. В это время возрос интерес к педагогическим тестам, и отсутствие тестологов в образовании ощутилось особенно остро. Образовавшийся кадровый вакуум в значительной степени заполнили психологи. Попутно, сами того не желая, они привнесли в образовательную среду устойчивые представления о тестах как о наборе довольно тривиальных по содержанию заданий, носящих косвенный характер и имеющих форму с выбором ответа. Преодолению искаженного представления о тестах способствует эксперимент по введению ЕГЭ. Однако, несмотря на то что он начался еще в 2001 г., многим пользователям до сих пор неизвестны широкие возможности педагогических тестов.

Различия между психологическими и педагогическими тестами. Несмотря на имеющиеся точки соприкосновения, между педагогическими и психологическими тестами есть существенные различия. Подробно они представлены в книге А. Анастаси [1]. В частности автор обращает внимание на то, что педагогические тесты оценивают результаты усвоения общего для всех обучаемых программного курса, а психологические — отражают индивидуальное восприятие жизненного опыта, различающегося по социальным и экономическим условиям его приобретения.

Результаты тестирования в психологии используются в основном для прогнозирования успешности какой-либо деятельности в зависимости от оценок личностных характеристик и склонностей или носят рекомендательный характер при выявлении личностных проблем. Педагогические тесты в итоговом контроле дают количественные оценки уровня подготовленности обучаемых, которые нередко служат основой для принятия административно-управленческих решений в образовании, поэтому к качеству результатов педагогических измерений и их сопоставимости должны предъявляться повышенные требования, обеспечивающие корректность управленческих выводов и административных решений.

Внешние различия между тестами в психологии и в образовании проявляются в формулировках заданий, которые в педагогических тестах всегда носят прямой, а в психологических тестах, чаще всего, косвенный характер. Например, в психодиагностике встречаются ситуации, когда неправильные ответы испытуемых расцениваются как положительные характеристики их личности, а правильные, наоборот, свидетельствуют о необходимости принятия коррекционных мер. Иногда границы между тестами стираются, особенно в тех случаях, когда педагогические тесты выступают в качестве предикторов успешности профессиональной деятельности или нацелены на проверку логического мышления и творческих способностей уча-

щихся. На существующее сходство между измерителями в психологии и в образовании обычно указывает высокая корреляция, которую можно оценить по эмпирическим результатам выполнения тестов.

Основные определения

Измерение в образовании, латентные переменные. Согласно наиболее распространенному определению, введенному в 1946 г. американским психологом С.Стивенсом,

Измерение - это процедура приписывания чисел некоторым характеристикам объектов в соответствии с определенными правилами. Данное определение - результат формального обобщения опыта количественных измерений, широко применяемых в физике и других естественных науках, однако на протяжении многих лет его брали за основу и в эмпирических науках.

По мере развития педагогики, психологии и социологии возникла потребность во введении не только количественных, но и качественных оценок для величин, отличающихся по степени проявления того или иного свойства. Качественные оценки являются менее точными по сравнению с количественными в силу применяемых способов и инструментов измерения. Например, классифицирующие понятия в образовании («знающий», «подготовленный» и др.), которые дифференцируют обучающихся по уровню знаний и играют важную роль в учебном процессе, определяются субъективно учителем или группой учителей. Нередко качественные оценки выражают с помощью чисел, которые выбирают на основе экспертных суждений и соглашений. Приписываемые числа могут трактоваться по-разному. Так, в традиционном педагогическом контроле у каждого учителя есть свои представления о том, за что нужно ставить «5», «4», «3» и т.д.

Неоднозначность оценивания в образовании усугубляется латентным (скрытым, исключаящим возможность непосредственного измерения) характером измеряемых переменных. В силу латентности оцениванию подвергаются не сами характеристики обученности и обучаемости, а их эмпирические референты — наблюдаемые признаки измеряемых характеристик. Выбор последних происходит интуитивно, поэтому их соответствие латентным характеристикам нуждается в доказательстве на основе экспертного и статистического анализа эмпирических результатов измерения.

Современная трактовка понятия «педагогическое измерение». Современная теория измерений появилась в 80-х гг. XX в. Она строится на более строгой аксиоматической основе.

В соответствии с новыми представлениями, измерение трактуется как конструирование числовой функции, осуществляющей изоморфное отображение некоторой эмпирической структуры в соответствующим образом подобранную числовую структуру.

Изоморфизм - важное понятие математики, которое определяет ряд условий взаимно однозначного отображения двух множеств с сохранением их свойств в процессе такого отображения. Хотя это понятие впервые появилось в высшей алгебре, в наше время оно используется довольно широко, хотя и не вполне строго, например в педагогических измерениях. Поскольку эмпирическая структура и строящаяся по результатам оценивания числовая структура (шкала) изоморфны, имея шкалу, можно, не обращаясь непосредственно к измеряемым объектам, восстановить все их свойства, характерные для эмпирической структуры.

Компоненты процесса педагогических измерений. *Процесс педагогических измерений включает:*

1. *Выбор предмета измерения (латентных характеристик объектов) и их числа;*
2. *Выбор эмпирических референтов (наблюдаемых характеристик объектов);*
3. *Выбор измерительных процедур;*
4. *Конструирование и использование измерительных инструментов;*
5. *Выбор шкалы (если измеряемая переменная одна) или шкал (если измеряют более одной переменной при многомерных измерениях);*
6. *Построение отображения результатов измерения на шкалу (шкалы в случае многомерных измерений) по определенным процедурам и правилам;*
7. *Обработку, анализ и интерпретацию результатов измерения.*

В силу неизбежности ошибок измерения оцениваемые характеристики объектов могут принимать более или менее точные значения, поэтому эти характеристики принято называть переменными измерения. Любые отклонения от стандартизированных условий измерения, обработки, анализа и интерпретации полученных результатов увеличивают ошибки измерения, которые представляют наибольшую опасность в эмпирических науках в силу латентного характера переменных. Поэтому так важен анализ устойчивости и точности (надежности) результатов тестирования, что выгодно отличает тесты от традиционных оценочных средств.

Валидность отражает адекватность эмпирических результатов поставленным целям измерения. В силу многогранности целей анализ валидности должен быть многоаспектным, но в любом случае важное место занимает доказательство адекватности эмпирических референтов концептуально выделенной переменной (переменных) измерения (конструктивная ва-

лидность).

Валидность – характеристика теста, отражающая адекватность системы тестовых заданий содержанию учебного материала и результату тестовых испытаний.

Высокая надежность теста это необходимое, но недостаточное условие получения высококачественного теста. Тест еще должен быть валидным. Валидность – это важнейшая характеристика теста, без указания которой, его нельзя считать измерительным инструментом.

Анализируя сложную ситуацию с валидностью педагогических тестов, Е.Михайлычев¹⁶ отмечает, что педагогу, заинтересовавшемуся валидностью, трудно будет разобраться в том, что же это такое.

Ниже мы приведем несколько определений валидности теста.

ВАЛИДНОСТЬ означает пригодность тестовых результатов для той цели, ради чего проводилось тестирование (В.Аванесов)¹⁷.

ВАЛИДНОСТЬ - это характеристика способности теста служить поставленной цели измерения (М.Челышкова)⁷.

ВАЛИДНОСТЬ - определяет, насколько тест отражает то, что он должен оценивать (А.Майоров)¹⁸.

Приведенные определения в целом перекликаются и являются практически равноценными. Мы несколько уточним определение, сделав акцент на цель тестирования. Тестирование как измерительная процедура, дает информацию, на основе которой в дальнейшем должно быть принято то или иное управленческое решение. Обоснованность этих решений, зачастую сильно влияющих на судьбу испытуемых, определяется надежностью и валидностью теста.

ВАЛИДНОСТЬ – это характеристика теста, отражающая его способность получать результаты, соответствующие поставленной цели и обосновывающая адекватность принимаемых решений.

После создания теста начинается процесс его валидизации. Приведем определение:

ВАЛИДИЗАЦИЯ – процесс накопления подтверждений для доказательства валидности теста¹⁹.

По нашему мнению **ВАЛИДИЗАЦИЯ** – это не столько сбор доказательств валидности теста, сколько процесс выполнения действий, повышающих его валидность. Вследствие этого будет расти и доказательная база валидности теста.

Выделяют три вида валидности – содержательную, критериальную и конструктивную²⁰.

А.Майоров приводит следующую диаграмму видов валидности¹⁸:



Рис.3.9.1. Виды валидности.

КОНСТРУКТИВНАЯ ВАЛИДНОСТЬ (концептуальная валидность) определяется в случаях, когда представление об измеряемом свойстве существует в форме абстрактного образа, модели. Для объяснения определенных качеств личности создается концептуальная модель, которая с помощью тестов подтверждается или опровергается.

КРИТЕРИАЛЬНАЯ ВАЛИДНОСТЬ (эмпирическая валидность) предполагает наличие внешнего критерия, корреляция с которым определяет валидность теста.

Имеется два вида критериальной валидности – текущая и прогностическая.

Текущая критериальная валидность (*concurrent validity*) характеризует способность теста измерять некоторые качества личности. Валидность теста подтверждается корреляцией с некоторым внешним критерием, существующим в данное время. Допустим, тест показал для некоторого испытуемого отличные знания по предмету, а школьные отметки, выставленные учителем – неудовлетворительные. Если мы в качестве внешнего, независимого и достоверного критерия выберем школьные отметки, то критериальная валидность теста – низкая, даже если он имеет высокую надежность.

Прогностическая критериальная валидность (*predictive validity*) характеризует способность теста предсказывать будущие качества, формирующихся в результате воздействия внешних обстоятельств или целенаправленной собственной деятельности. Этот тип валидности характеризует корреляцию результатов тестирования с внешним критерием, который появится в будущем.

СОДЕРЖАТЕЛЬНАЯ ВАЛИДНОСТЬ (*content validity*) характеризует тест по степени его соответствия предметной области.

Согласно А.Анастаси, содержательная валидность означает систематическую проверку содержания теста, с тем чтобы установить, соответствует ли оно репрезентативной выборке измеряемой области поведения. Такая процедура валидации обычно применяется для тестов достижений²⁰.

Содержательная валидность необязательно означает полноту отображения изучаемой

дисциплины. Например, для нормативно-ориентированного теста, полнота охвата всех тем может быть меньше, чем для критериально-ориентированного. Здесь важнее глубина проработки отдельных подтем, вопросов. Это позволит с большей эффективностью дифференцировать обучаемых. Под содержанием понимается не только совокупность фактов, понятий, терминов, но и умение применять имеющиеся знания, оценивать информацию, выполнять действия, соответствующие верхним уровням таксономии Блума.

Для обеспечения содержательной валидности необходим детальный анализ учебных программ, на основании чего составляется спецификация теста. Спецификация содержит перечень учебных тем, их важность, количество и тип тестовых заданий. Оценка содержательной валидности выполняется экспертом в данной предметной области.

Согласно П.Клайну содержательная валидность определяется следующим образом:

- 1) указать категорию лиц, для которой предназначен тест;
- 2) составить список знаний, умений, навыков, подлежащих тестированию;
- 3) выполнить внешнюю экспертизу полученного списка на предмет его полноты и обоснованности;
- 4) на основе списка составить перечень заданий;
- 5) выполнить внешнюю экспертизу полученных заданий;
- 6) после проверки преобразовать их в задания в тестовой форме. В дальнейшем, на этой основе создать тестовые задания, образующие тест, который будет содержательно валидным.

Проблема валидации педагогического теста является, видимо, самой сложной в процедуре создания высококачественного измерительного инструмента.

Банк тестовых заданий – логически упорядоченная структура конечного числа программно-дидактических тестовых заданий, позволяющая генерировать множество культурных форм тестов.

Задание – минимальная составляющая единица теста.

Педагогический тест — это инструмент, предназначенный для измерения обученности учащегося, состоящий из системы тестовых заданий, стандартизированной процедуры проведения, обработки и анализа результатов.

Тесты можно разделить на две категории — адаптивные и традиционные тесты.

Измерительный инструмент. Измерительный инструмент включает два компонента. Первый компонент - *само измеряющее устройство*, роль которого в педагогических измерениях чаще

всего, но не всегда выполняет тест. В самом обобщенном виде под тестом можно понимать совокупность контрольных заданий в стандартизированной форме, обладающих необходимыми системообразующими статистическими характеристиками и обеспечивающих обоснованные оценки концептуально выделенной переменной (переменных) измерения с высокой объективностью. Таким образом, в самом определении теста заложены требования к его качеству, отсутствующие в традиционных оценочных средствах.

Второй компонент измерительного инструмента - *заранее подготовленная шкала*, которая служит для фиксации результатов измерения и на которой откладываются оценки (количественные или качественные) измеряемой переменной. В процессе упорядочения оценок каждому элементу совокупности наблюдаемых эмпирических данных ставится в соответствие определенный балл, устанавливающий положение наблюдаемого элемента на шкале, где можно размещать сырые (первичные) баллы (результаты суммирования оценок по отдельным заданиям теста) или производные баллы, получающиеся в результате преобразования первичных оценок для повышения сопоставимости и удобства интерпретации результатов учащихся.

Шкала с отложенными оценками переменной является целью измерения. При измерениях с высокой надежностью и валидностью она адекватно отображает оцениваемые характеристики и представляет их без существенных искажений. В зависимости от количества оцениваемых характеристик объекта можно говорить об *одномерных* (одна переменная) или *многомерных* (более одной переменной) *измерениях*. Соответственно по результатам измерения строится одна шкала или несколько шкал, число которых в последнем случае обычно бывает равно числу переменных измерения.

Обработка и анализ данных измерения. Последний компонент процесса педагогических измерений, включающий обработку, анализ и интерпретацию данных, служит для выявления обеспечиваемого качества результатов измерения, коррекции тестов и представления полученных данных в форме, удобной для интерпретации и сравнения. Благодаря сопоставимости тестовых баллов, достигаемой в процессе обработки, по результатам педагогических измерений можно выстраивать качественный анализ результатов учащихся, проводить мониторинг и принимать обоснованные управленческие решения в образовании.

Объективность педагогических измерений. Может ли быть абсолютная объективность? Появление первых стандартизованных тестов в образовании вызвало массовую позитивную реакцию, поскольку первоначально они рассматривались как средство получения объективных оценок подготовленности обучаемых, преодолевающее субъективизм традиционных

оценочных средств. По мере развития теории педагогических измерений и накопления опыта применения тестов пришло понимание того, что абсолютная объективность — это недостижимая характеристика результатов любых, в том числе и педагогических, измерений в силу существования ошибочных компонентов, неизбежно смещающих оценки. Поэтому при использовании тестов можно говорить лишь о высокой или низкой объективности, степень проявления которой связана с величиной надежности теста.

Наиболее полно трактовка термина «объективность измерений» представлена в исследованиях Е.Вебстера, предложившего восемь толкований этого понятия. Три из них - процедурная объективность, классическая (традиционная) объективность и инвариантная (специфическая) объективность - непосредственно относятся к педагогическим измерениям.

Процедурная объективность. Под процедурной объективностью (первая трактовка термина «объективность измерений») понимается независимость результатов тестирования от субъективных суждений педагога, использующего тест. Эта независимость обеспечивается благодаря равенству условий тестирования, использованию для одной группы тестируемых параллельных (совпадающих по трудности и другим характеристикам) вариантов теста, стандартизации процедуры проверки результатов и максимальной ее автоматизации, исключающей влияние педагога на оценки.

Сведение всех видов объективности только к процедурной недопустимо, поскольку при таком подходе не выдвигается никаких требований к качеству теста. В этом случае может создаться впечатление, что для получения объективных данных о подготовленности испытуемых достаточно перейти от традиционных экзаменов к любым, в том числе некачественным, тестам, устранив влияние педагога на оценку тестирования путем автоматизации процедуры подсчета баллов испытуемых.

Классическая, или традиционная, объективность. Второе, углубленное, понимание объективности измерений рассматривается в классической теории тестов и основывается на понятиях «сырой балл» и «истинный балл», отличающихся друг от друга на величину ошибки измерения.

Сырой (первичный, наблюдаемый, индивидуальный) *балл* получается простым суммированием результатов испытуемого по отдельным заданиям теста. При дихотомической оценке результатов по заданиям (1 или 0) индивидуальный балл равен количеству правильно выполненных заданий теста. *Истинный балл* в классической теории отождествляется с абсолютно объективной оценкой свойств испытуемого, свободной от влияния любых ошибок измерения.

В отличие от сырого балла, который меняется в зависимости от теста и способа подсчета результата испытуемого, истинный балл трактуется как не зависящая от средств измерения константа, характеризующая оцениваемое свойство испытуемого в момент измерения, но меняющаяся в процессе обучения.

Согласно основной аксиоме классической теории тестов, любой наблюдаемый балл равен сумме истинного балла и ошибки измерения. В тех случаях, когда ошибка измерения не превышает выбранных пределов точности измерений, говорят о высокой объективности результатов тестирования, а оценки испытуемых принимают за их истинные баллы. Таким образом, углубленное понимание объективности измерений требует оценивания величины ошибки измерения, на размер которой влияют не только условия проведения тестирования, но и качество теста.

Инвариантная, или специфическая, объективность. Третья трактовка объективности основана на современной теории конструирования тестов. Преимущества ИКТ, позволяющие оценить подготовленность обучаемых независимо от трудности заданий теста, приводят к достижению так называемой инвариантной объективности измерений, которая предпочтительнее объективности, обеспечиваемой классической теорией тестов.

Для достижения специфической объективности необходима подгонка данных тестирования к требованиям моделей теории ИКТ и длительная серьезная работа над тестом. Поэтому на практике тестологи часто сталкиваются с тем, что эффект инвариантной объективности либо реализуется со слишком большими затратами, либо не реализуется вообще в силу недостаточности высокого качества теста.

Размерность пространства измерений, одномерные и многомерные конструкты, латентные переменные. Концептуальные и реальные переменные измерения, конструкты. Измерение начинается с постановки цели, в соответствии с которой выбирают одну (одномерный случай) или несколько (многомерный случай) переменных. В последнем случае для обозначения измеряемых характеристик часто используют обобщающий термин — «конструкт». Каждый разработчик теста уверен в том, что он ясно представляет себе измеряемые характеристики и способен на основании своего педагогического опыта точно подобрать задания, обеспечивающие оценивание конструкта. Многие тестологи так и остаются в полном заблуждении относительно того, что на самом деле измеряет тест, поскольку не проверяют соответствие задуманного конструкта и реальных результатов измерения.

Анализ такого соответствия является необходимым этапом оценивания валидности из-

мерения. Нередко его пытаются провести априорно, до начала тестирования, экспертным путем, и в результате, как правило, получают недостоверную информацию. Для корректного оценивания валидности необходим статистический анализ эмпирических данных тестирования, поэтому понять, что же мы на самом деле измеряем, можно лишь после применения теста.

Сложность процедуры установления размерности пространства измерений увеличивает проблемы, связанные с неоднозначной трактовкой многих конструктов в образовании и в других социальных науках (рис. 2). Каждый педагог вкладывает в оценивание учебных достижений свое видение оптимального набора переменных измерения. Неоднозначность трактовки конструкта усугубляется по мере продвижения от начальных ступеней образования к более высоким ступеням, когда содержание большинства учебных курсов приобретает междисциплинарный характер.

Операционализации. Операционализация заключается в придании оцениваемым латентным характеристикам подготовленности учащихся формы, удобной для фиксации определенными правилами измерения. При педагогическом измерении в качестве таких характеристик подготовленности обучаемых обычно выступают знания, умения, навыки, компетентности и т.д.

В процессе операционализации происходит выделение набора эмпирических индикаторов, в роли которых выступают задания теста. Количество правильно выполненных заданий, подсчитанное и преобразованное по определенным правилам, дает основание для присвоения испытуемому определенного места на шкале переменной измерения.

Визуализация результатов педагогического измерения. Визуализация - геометрическая интерпретация связи между латентной переменной A (одномерной или многомерной) и наблюдаемой переменной B - показана на рис. 3.

Стрелки на рисунке указывают характер связи между переменными. Латентная переменная A является первопричиной, порождающей множество наблюдаемых результатов выполнения теста. Однако при измерениях всегда ставят обратную задачу - по наблюдаемым результатам тестирования найти достаточно точные оценки латентных переменных.

Взаимосвязь результатов измерения и положения испытуемого на шкале переменной для одномерного случая представлена на рис. 4. Каждая оценка переменной измерения для учащихся из тестируемой группы соответствует одной из точек оси. В свою очередь каждая точка определяет положение испытуемого или группы испытуемых с одинаковым тестовым баллом, полученным по результатам выполнения теста.

На изображенной оси более высокие баллы располагаются правее, а более низкие — левее. Крайний слева результат отражает случай, когда испытуемый выполнил правильно лишь несколько заданий теста. Противоположной ситуации, когда ученик выполнил все или почти все, соответствует крайняя правая точка на оси переменной измерения. Остальные точки занимают некоторое промежуточное положение на отрезке, где лежат тестовые баллы учащихся.

Если правильно выполненные задания теста соотнести с результатами учащихся и расположить их вдоль оси переменной измерения, то можно предположить, что более трудные задания сместятся вдоль оси вправо, так как их, скорее всего, будут выполнять правильно наиболее сильные учащиеся в классе. И наоборот, более легкие задания будут смещены влево — они по силам ученикам с низким уровнем подготовки (см. рис. 5).

Из дидактических соображений на рисунке показано выполнение четырех заданий, однако все выводы, получаемые с помощью этого примера, применимы к любому числу заданий в тесте. Расположение тестового балла первого учащегося говорит о том, что он выполнил верно два самых легких задания, но не справился с третьим и четвертым заданиями. Второй учащийся имеет более высокий тестовый балл и подготовлен лучше. Он не выполнил только самое трудное - четвертое задание теста.

Ошибки измерения. Локализация места расположения результата ученика на оси переменной зависит в основном от соотношения между величиной его истинного балла и трудностью заданий теста. Если балл довольно высок, а задание довольно легкое, то у ученика все основания для успешного выполнения этого задания теста. В противном случае ученика скорее всего ждет неудача.

Конечно, наверняка предугадать ничего нельзя в силу действия различных смещающих факторов (эффект забывания, подсказки и т.д.), поэтому обычно говорят лишь о некоторой вероятности успеха или неуспеха.

Вероятностный характер наблюдаемых результатов выполнения теста обусловлен влиянием случайных и неслучайных ошибок измерения. В число последних входят те, которые появляются из-за просчетов разработчиков в процессе создания теста. К ошибкам систематического характера могут также привести нарушение требований к сбору статистических данных, некачественная интерпретация результатов выполнения теста и ряд других причин. К случайным факторам можно отнести настроение испытуемого, поведение экзаменатора, обстановку при тестировании в классе и многое другое - словом, все, что учесть и предвидеть при тестировании попросту невозможно.

Одномерные измерения. Чаще всего при планировании измерений в образовании выбирают одномерные конструкты. Это упрощает процесс построения шкалы, но не всегда бывает адекватно содержанию тестов. Рис. 6 иллюстрирует случай одномерных измерений, который может быть интерпретирован следующим образом: одна латентная переменная T — истинный уровень подготовленности каждого обучаемого — приводит к возникновению одной оценки наблюдаемой переменной X - уровня подготовленности обучаемого. Помимо переменной T на оценку X оказывает влияние фактор E - ошибка измерения.

Чтобы принять гипотезу об одномерности теста, необходимо выявить связь между теоретическим конструктом и эмпирическими индикаторами, роль которых выполняют задания теста. Оценка связи требует ответа на вопрос - есть ли разница между доказательством одномерности конструкта и доказательством одномерности заданий теста?

На рис. 7 представлена измерительная модель для одномерного случая, иллюстрирующая связь между конструктом, обозначенным символом Γ , и четырьмя заданиями (X_1, X_2, X_3, X_4). Числа, стоящие у каждого луча, показывают меру предполагаемой корреляционной связи между конструктом и эмпирическими индикаторами — видеть при тестировании попросту невозможно.

При анализе модели важно понимать, что конструкт является латентным (скрытым от возможностей непосредственного измерения) фактором, взаимодействие которого с заданиями порождает наблюдаемые результаты выполнения теста. Влияние конструкта на наблюдаемые переменные показано на рис. 7 с помощью направленных лучей.

Поскольку каждое задание в рассмотренном гипотетическом примере измеряет только один конструкт, то справедлив вывод об одномерности заданий теста. Обратный вывод, в общем случае, неверен: из одномерности заданий не следует одномерности теста.

Многомерные измерения. Если конструкт включает не одну, а несколько переменных, то измерения называются многомерными. Совокупность переменных образует пространство переменных измерения, размерность которого равна их числу. Иногда при проведении многомерных измерений создают несколько субтестов, каждый из которых является одномерным и измеряет свою переменную с помощью одномерных заданий.

Примером такого подхода является полидисциплинарный тест, состоящий из набора одномерных субтестов. В другом случае в многомерных измерениях используют междисциплинарный тест, задания которого не являются одномерными. Каждое из заданий измеряет свою совокупность переменных, которые могут отличаться как по количеству, так и по содержа-

тельной трактовке конструкта.

В практике педагогических измерений существуют специальные методы анализа размерности пространства измерений. Такую группу методов предоставляет исследовательский и конфирматорный факторный анализ, применение аппарата которого основано на использовании соответствующего программного обеспечения, например статистического пакета SPSS.