

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ДАЛЬНЕВОСТОЧНЫЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ОТКРЫТЫЙ УНИВЕРСИТЕТ ДВГУ

И. А. МОРЕВ

**ОБРАЗОВАТЕЛЬНЫЕ
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
Часть 2. Педагогические измерения**

Учебное пособие
для учащихся педагогических специальностей вузов
и слушателей курсов повышения квалификации
педагогических и управленческих кадров

Владивосток
Издательство Дальневосточного университета
2004

УДК 37.012.8

ББК 74.202.2

М 79

Рецензенты:

Михина Галина Борисовна –

канд. пед. наук, доцент, декан педагогического факультета ДВГУ;
зав. кафедрой общей педагогики и психологии ДВГУ

Грудин Борис Николаевич –

докт. техн. наук, директор Дальневосточного
Центра Федерации Интернет-образования

Морев И. А.

М 79

Образовательные информационные технологии. Часть 2.

Педагогические измерения: Учебное пособие. – Владивосток: Изд-во Дальневост. ун-та, 2004. – 174 с.

Пособие содержит учебные и методические материалы необходимые для изучающих и применяющих информационные технологии в системе образования. Приложения, расположенные на компакт-диске, содержат электронный тренажер для проверки усвоения знаний, списки дополнительной литературы, документы, описания электронных педагогических средств, концепции, обзоры, тематические перечни Интернет-ссылок.

Для студентов педагогических специальностей вузов, аспирантов и слушателей курсов повышения квалификации педагогических и управленческих кадров. Может использоваться для самостоятельного изучения.

М $\frac{4309000000}{180(03)-04}$

ББК 74.202.2
М 79

© Морев И. А., 2004
© ТИДОТ ДВГУ, 2004
© Издательство Дальневосточного
университета, 2004

АННОТАЦИЯ.....	6
СТУДЕНТУ	7
ВВЕДЕНИЕ. Измерения качества и количества в образовании	11
Глава 01. Компьютерные технологии аттестации	13
01.01. Педагогическое и психологическое тестирования	16
01.02. Бланковое тестирование. Десять несовершенств	17
01.03. Двадцать преимуществ компьютерного тестирования	20
01.04. Общие достоинства и недостатки тестовых технологий	22
01.05. Культура и философия адаптивного тестирования	24
Глава 02. Технологии массовой компьютерной аттестации ЗУН. Цели, задачи, характеристики, этапы, пути, результаты	27
Цель и три десятка задач педагогического тестового контроля с шести точек зрения	27
Десять функций ОИТ контроля учебной деятельности	29
02.01. Виды, типы, этапы, требования, характеристики компьютерного тестирования ...	30
Полдюжины видов и пятнадцать направлений ОИТ контроля учебной деятельности	30
Десять характеристик ОИТ аттестации ЗУН	31
Десять общих этапов компьютерной аттестации	32
Десять путей достижения целей компьютерной аттестации	32
Дюжина типов сценариев компьютерного тестирования	33
Два десятка вопросов теории и практику	36
Десять общих и тридцать частных требований к компьютерному тестовому комплексу	38
03.01. Четыре группы недостатков интерфейса комплексов тестового контроля	40
02.02. Полдюжины замечаний пользователю	41
02.03. Результаты компьютерной аттестации ЗУН и факторы влияния	42
Три интерпретации результатов тестирования	43
Длительность сеанса тестирования	44
Количество заданий в тесте	47
Спецификация теста	47
02.04. Артефакты	48
02.05. Динамика результатов педагогических измерений	52
02.06. ОИТ и менеджмент качества. Принцип непрерывного улучшения	61
02.07. Каким быть компьютерному комплексу для обучения и контроля ЗУН?	62
Глава 03. Тестовые испытания и управление образовательными результатами	64
03.01. Мотивация учащихся к повышению качества ЗУН	65
03.02. Технология мягкого тестирования. Десять задач и десять решений	66
03.03. Мотивация студента к повышению качества ЗУН	70
03.04. Полторы дюжины привлекательных черт процедуры компьютерного тестирования	71
Глава 04. Информационная безопасность процедур массового компьютерного тестирования	73
Десять основных требований информационной безопасности тестирования и обеспечение их выполнения	73
04.01. Два удивительных явления из опыта массовых компьютерных тестирований	74
04.02. Дюжина путей обеспечения информационной безопасности массовых компьютерных тестирований	75

04.03.	Десять каналов утечки конфиденциальной информации. Обеспечение секретности БТЗ	77
04.04.	Полдюжины стимулов сохранения конфиденциальности тестовых материалов в США	79
04.05.	Стимулы сохранения конфиденциальности тестовых материалов в России	80
04.06.	Обеспечение конфиденциальности теста в масштабе региона и страны	81
	Десять принципов безопасной технологии создания БТЗ	81
	Десять шагов безопасной технологии тестирования	83
	Десять шагов безопасной технологии обработки результатов и публикации	84
	Десять направлений снижения общей себестоимости и себестоимости мер безопасности	84
	Десять известных в практике тестирования способов подлога	85
Глава 05.	Количество и качество ЗУН	86
05.01.	Является ли традиционная оценка мерой качества и количества? Сепарабельность оценки	87
05.02.	Что измеряют в процессе тестирования?	88
	Как измерить количество ЗУН?	91
	Как измерить качество ЗУН?	92
	Как оценить качество оценивания качества?	93
05.03.	Тестирование – средство объективизации оценки качества и количества. Пять предпосылок	95
05.04.	Разрешающая способность и объективность измерительных процедур	97
05.05.	Систематические ошибки в тестировании	99
05.06.	Рейтинг и оценка. Их определение и назначение	101
	Системы кумулятивного индексирования	103
	Определение скалярного рейтинга	105
Глава 06.	Основные сведения из области тестологии	108
06.01.	Тридцать три принципа конструирования теста, создания тестовых заданий и выбора технологии тестирования	109
06.02.	Педагогический тест	115
06.03.	Семь оснований для классификации и двадцать видов компьютерных педагогических тестов	117
06.04.	Десять недостатков педагогических тестов	120
06.05.	Валидность и надежность теста	121
	Надёжность теста. Контролируемые и неконтролируемые факторы влияния	121
	Три типа надежности теста и три способа ее определения	123
	Развитие понятия валидности и надежности теста	124
06.06.	Информационный граф	127
06.07.	Два типа ошибок измерений в тестологии и их источники	128
06.08.	Педагогическое тестовое задание	129
06.09.	Четыре формы тестовых заданий	131
	Первая форма – закрытая форма задания	132
	Вторая форма – открытая форма задания	134
	Третья форма – задания на установление соответствия	134
	Четвертая форма – задания на установление последовательности	135
06.10.	Вес тестового задания. Полдюжины факторов снижающих ценность понятия веса ТЗ	136
06.11.	Трудоемкость и сложность тестового задания	137
06.12.	Две дюжины оснований типологии тестовых заданий	138
06.13.	Четыре группы из восемнадцати характеристик ЗУН, выявляемых путем тестирования	142

06.14.	Два десятка требований к тестовым заданиям	143
06.15.	Две дюжины недостатков тестовых заданий и полдюжины их причин.....	144
Глава 07.	ОИТ для защиты качества и управления качеством образования	147
	ОИТ и управление качеством. Антураж компьютерных тестирований.....	147
	ОИТ и организация управления качеством. Пять важнейших характеристик антуража массовых компьютерных тестирований	149
	ОИТ и кадровое обеспечение	150
	ОИТ и защита качества образования	150
	ОИТ и проблемы финансирования.....	152
Глава 08.	Применение компьютерных технологий массового тестирования	154
	Массовые тестирования знаний в России.....	155
	Полдюжины средств профилактики негативных ситуаций при компьютерном тестировании	156
	Дидактический принцип индуктивной мотивации в организации массовых компьютерных тестирований.....	158
08.01.	Мониторинг качества образования в России. Международное сотрудничество ..	159
08.02.	Единый государственный экзамен в России. Чертова дюжина препятствий	164
08.03.	Мониторинг, аттестация, сертификация. Опыт США.....	166
08.04.	К развитию технологий тестирования	168
	Адаптивное тестирование. Три варианта и семь возможностей.....	170
	Необычные формы тестирования	171
СПИСОК ЛИТЕРАТУРЫ.....		173

АННОТАЦИЯ

Понять – значит простить.

Анна Луиза Жермена де СТАЛЬ, aforizm.kaminplus.ru

Пособие предназначено для работников системы образования, желающих быстро и без «ужасающих» технических подробностей получить ответы на вопросы, касающиеся образовательных информационных технологий (ОИТ):

- Что такое ОИТ?
- Зачем мне ОИТ?
- Зачем моим ученикам ОИТ?
- Зачем моим подчиненным ОИТ?
- Как изменится образовательный процесс при внедрении ОИТ?
- Какой нужен инструментарий для работы с ОИТ?

Пособие может быть использовано для самостоятельного изучения, а также для обучения слушателей курсов или студентов педагогических специальностей.

Всем читателям будет полезно проверить свои знания с помощью электронного тренажера, включенного в электронную версию пособия.

Материал пособия и приложений содержит много практических советов для педагогов и руководителей образовательных учреждений. Отдельно собрана информация для родителей и учащихся, которая может быть использована учителем при ведении внеурочной работы. Большое количество Интернет-ссылок на российские и зарубежные образовательные сайты позволило сократить объем описательной информации и ограничиться при разборе специальных вопросов обсуждениями и рекомендациями. Приложения, насыщенные такими ссылками, можно рассматривать как своеобразное приглашение в Интернет.

Приложения помещены на прилагаемом компакт-диске.

Разделы пособия содержат обсуждения действенности тех или иных методов обучения и контроля ЗУН (знаний, умений, навыков), применимости их в условиях наличия компьютерной поддержки.

Все разделы пособия снабжены вступлениями, формулирующими тему. Иногда это – цитата, заостряющая (одобряющая, высмеивающая) обсуждаемую проблему. Цитаты мы взяли из открытых источников Интернет и из своих старых записей. Они наверняка пригодятся Вам при подготовке к лекциям. Авторы цитат везде указаны, источники цитат – иногда. Это означает, что восстановить источник, где цитата появилась впервые, мы не смогли. За это мы заранее приносим авторам свои извинения и уверения в почтении.

Пособие не претендует на полноту. Мы будем пополнять его, в том числе и материалами, присланными по адресу morev@vido.dvgu.ru. За это мы Вам заранее благодарны.

Здесь не рассматриваются специальные вопросы, касающиеся технического обслуживания и правил пользования программными средствами.

Здесь не содержится рекламы технических и инструментальных средств. Эта информация, несомненно, необходимая читателю, представлена в приложениях Интернет-ссылками.

СТУДЕНТУ

Народу нужны не отвлеченные идеи, а прописные истины
Антуан РИВАРОЛЬ, французский писатель, aforizm.kaminplus.ru

Это пособие предназначено для учащихся по Программе «Менеджмент в образовании». Курс «Образовательные информационные технологии» является одним из основных разделов этой программы, поскольку нацелен на подготовку специалистов, которые непосредственно реализуют новые проекты в области образовательных информационных технологий (ОИТ) и технологий управления качеством образования.

Задачи пособия:

- способствовать формированию потребности в применении новых технологий на практике в процессе педагогической или управленческой деятельности;
- познакомить учащихся с ОИТ;
- познакомить учащихся с мировыми контрастами в развитии ОИТ;
- расширить круг известных учащимся направлений развития ОИТ;
- дать обзор наиболее значимых публикаций и ресурсов в области ОИТ.

Для удобства учащихся содержание курса организовано в виде последовательности равнозначных по насыщенности материалом учебных модулей. Приложения в равной степени относятся ко всем модулям.

В Приложениях приведен материал для практического применения изложенных теоретических разработок, положений, сведений. Представлены нормативные документы, опубликованные в Интернет и открытой печати. Материалы Приложений не являются обязательными для Вас именно в момент обучения, поэтому тестирование знаний их не предусмотрено. Материал Приложений пригодятся Вам потом, в реальной деятельности.

Дальневосточный государственный университет (ДВГУ) и Министерство образования РФ (МОРФ) создали в 2001 году Приморский краевой центр новых информационных технологий. Некоторые изложенные здесь материалы являются результатом деятельности ПК ЦНИТ.

В тексте пособия и Приложениях можно найти информацию о деятельности и принципах функционирования Открытого университета ДВГУ, о новых возможностях, возникающих с появлением открытых форм образования.

*Кто хочет достигнуть того, чего желает;
пусть желает достижимого*
Хуан МАНУЭЛЬ, aforizm.kaminplus.ru

В соответствии с технологией дистанционного обучения, принятой в Открытом университете ДВГУ, пособие содержит компьютерную тестирующую программу, с помощью которой учащиеся смогут подготовиться к сдаче зачета и экзамена. Тестовые задания разбиты на группы, соответствующие Главам пособия. Проверить свои знания учащиеся смогут как в рамках каждого модуля, так и в целом, после завершения обучения.

Тестирование знаний производится на компьютере по следующему плану:

- После изучения материала каждой главы учащийся подвергает свои знания тестированию с помощью компьютерного тренажера и после этого проходит промежуточную аттестацию. В ходе тренировки учащийся может увидеть правильные отве-

ты на текущие вопросы теста. По окончании тренировки учащийся получает перечень вопросов, ответы на которые он не знает;

- При прохождении промежуточной аттестации обучающийся получает оценку «зачет / нечет», которая определяется долей (процентом) правильно выполненных заданий;
- После прохождения всего курса учащийся либо представляет (через руководителя или методиста своего территориального представительства) в учебный отдел Открытого университета ДВГУ файл-отчет итоговой аттестации, либо все файлы-отчеты промежуточной аттестации;
- Оценку по пятибалльной шкале учащемуся выставляет преподаватель на основе анализа файлов-отчетов и практической работы.

Учащиеся программы обязаны выполнить и защитить одну практическую работу. Оценка работы осуществляется комиссией, в соответствии с приказом ректора.

*Желание высказаться почти всегда бывает сильнее;
чем желание чему-нибудь научиться.
ПИСАРЕВ Дмитрий Иванович ,aforizm.kaminplus.ru*

Защита работы может быть заочной, очной либо публичной.

Результаты и текст работы представляются преподавателю в электронном виде (передаются на дискете, либо пересылаются по электронной почте через свое территориальное представительство).

Учащиеся выбирают тему практической работы самостоятельно, придерживаясь приведенного здесь списка и учитывая актуальность её для образовательного учреждения, территориального управления образованием, Открытого университета ДВГУ.

Практическая работа оформляется в соответствии с принятым в Открытом университете ДВГУ стандартом, аккуратно, в электронном виде, в формате Word. Графики, таблицы, схемы выполняются с использованием инструментария MS Office.

ТРЕБОВАНИЯ к оформлению практической работы

*Легче сделать более, нежели то же
Марк Фабий КВИНТИЛИАН, aforizm.kaminplus.ru*

Работа содержит титульный лист, Оглавление, Введение, три Главы, Заключение, список литературных источников, Приложения.

На титульном листе должны быть указаны следующие сведения:

- **Первая строка:** Министерство образования и науки РФ
- **Вторая строка:** Дальневосточный государственный университет
- **Третья строка:** Открытый университет ДВГУ
- **Четвертая строка:** Фамилия, имя, отчество автора (полностью)
- Название работы (не более 10 слов)
- Данные о руководителе и соруководителе работы (ФИО полностью, место работы, звание, должность) с помеченными местами для их подписи
- Адрес и наименование образовательного учреждения, на базе которого исполнялась работа, адрес электронной почты (номер телефона) автора
- Последняя строка: год исполнения

Во Введении описываются предпосылки работы, формулируется тема, ставятся задачи. Здесь же делаются предварительные пояснения к тексту работы, если это необходимо.

Первая глава, как правило, посвящается обзору литературы, обсуждению научной и практической ситуации в избранной области.

Вторая глава содержит используемые и выдвигаемые автором теоретические идеи, формулировку технологии работы.

Третья глава – результаты работы. Они оформляются в виде таблиц и графиков со словесным описанием. Часть результатов, для сохранения целостности восприятия текста Главы, можно вынести в Приложение.

Выводы (или Заключение) должны включать 3-4 положения с краткими формулировками того, что автором сделано самостоятельно в процессе работы. Именно эти положения автор будет защищать публично перед комиссией.

Всюду, где в тексте включен материал источников, помещаются соответствующие ссылки – номер источника (согласно списку литературных источников) в квадратных скобках.

Текст работы создается на компьютере с использованием MS Office (или аналога).

Объем работы: около 20-40 страниц печатного текста (Times New Roman, 12). Оформление полей – верхнее и нижнее поля: 2,5 см, левое поле: 3 см, правое поле: 1 см. Межстрочный интервал – 1.

К публичной защите учащийся обязан подготовить компьютерную презентацию работы в формате MS Power Point для демонстрации комиссии.

ПЕРЕЧЕНЬ рекомендуемых направлений практических работ

*Если кажется, что работу сделать легко, это непременно будет трудно.
Если на вид она трудна, значит, выполнить ее абсолютно невозможно.
Теорема Стокмайера. Артур БЛОХ, Мерфология*

Мы предполагаем, что в приведенном ниже списке Вам удастся выбрать направление по душе. Если этого не произошло, предложите свой вариант. Ваше предложение должно включать предполагаемые название, план и объем работы.

Обращаем внимание, что право названия будущей работы принадлежит Вам.

1. Постройте и обоснуйте подробный план внедрения компьютерной технологии (обучения, аттестации, управления качеством) для применения в конкретном образовательном учреждении. План включает финансовую, материальную, кадровую, временную составляющие. Примените знания, полученные при изучении других дисциплин Программы.
2. Постройте модель внеурочной работы школьного компьютерного класса. Обоснуйте модель, имея в виду главную цель – удовлетворенность родителей качеством образования, получаемого детьми в Вашем образовательном учреждении. Модель включает финансовую, материальную, кадровую, временную составляющие.
3. Создайте тест по материалам избранных разделов какой-либо учебной дисциплины на основе знаний, полученных при прохождении данного курса. Тест должен содержать не менее 100 тестовых заданий, задания должны быть разнесены по темам, уровням сложности, направленности, значимости. Разнесение должно быть обосновано письменно в тексте работы.
4. Создайте обзор известных Вам тестовых систем, содержащий обоснованную критику и перечень положительных характеристик. Среди этих систем должны быть и не отраженные в Приложениях настоящего пособия.
5. Создайте действующий макет программной оболочки для самостоятельного (группового) обучения и тестирования знаний, обладающей чертами, выгодно отличающими ее от известных комплексов.
6. Напишите сценарий программной оболочки для тестирования знаний и обучения, обладающую чертами, выгодно отличающими ее от известных. Сценарий должен содержать текстовое описание и формализованную рабочую блок-схему. Допустимы нестандартные обозначения элементов блок-схемы.

Справки и консультации можно получить через местное представительство Открытого университета ДВГУ, либо по электронной почте prog@vido.dvgu.ru

Авторы будут признательны читателям, которые найдут возможность поделиться своим собственным опытом в области проведения образовательного мониторинга.

Предложения и информацию присылайте по адресу: morev@vido.dvgu.ru

ВВЕДЕНИЕ. Измерения качества и количества в образовании

*Нет ничего более безнадежного,
чем развлечение по плану*
Самюэль ДЖОНСОН, aforizm.kaminplus.ru

В отличие от многих сфер деятельности человека, категории качества и количества в образовании не вполне уживаются с понятием измерения. Может быть, именно это и не позволяет педагогике получить всеобщее признание как полноправной науки.

Вопрос о формализации определения качества и количества в образовании стал насущным в связи с произошедшим во второй половине XX века кардинальным пересмотром отношения к образованию, «превращением» образования в товар. Вслед за этим превращением, в образовательную сферу проникли понятия, давно ставшие традиционными для сферы товарного производства – сертификация, менеджмент качества, мониторинг качества, рентабельность и пр. Происходит организационное оформление международного рынка образовательных услуг.

Сфера образования отличается от традиционных сфер производства и потребления, поэтому школы никогда не превратятся в подобие гастрономических магазинов, где люди вместо колбасы покупают информацию, а издательства не превратятся в фермы, которые вместо молока производят и доставляют в магазины для продажи тексты и картинки. Тем не менее, многими признано, что происходящий в сфере образования пересмотр подходов и понятий, придание им нового «экономического» смысла, может принести сюда много полезного для дальнейшего развития.

Формальные определения качества и количества образовательной продукции (т. е. образованности, обученности и др.), используемые в педагогике, не позволяют однозначно сопоставить им объективные процедуры (технологии) измерения. Мы не приводим их здесь по причине сложности выбора лучшего.

Обычная экзаменационная процедура и пятибалльная шкала оценки качества специалистов («продукции»), часто не устраивают заказчиков («покупателей продукции») по причинам необъективности и низкой «разрешающей способности».

Необъективность традиционной процедуры оценки ЗУН связана, в частности, с неизбежным привнесением в процесс измерения особенностей, связанных с человеческими отношениями:

- между учениками и педагогами;
- между педагогами и родителями;
- между педагогами и руководством образовательного учреждения;
- между педагогами;
- между экзаменаторами.

На экзаменационные оценки может непредсказуемо повлиять настроение экзаменуемых и экзаменаторов, внешний вид учащегося, погода и пр. Эта необъективность оценки может сыграть как негативную, так и позитивную роль. Необъективность оценки может привести, например, к следующим последствиям:

- никудашный специалист получит высокую оценку своим умениям, получит ответственный пост в управляющей или проектной организации, что приведет к непредсказуемым техническим и гуманитарным катастрофам;
- «слабый» учащийся получит положительный эмоциональный заряд, самостоятельно освоит нелюбимую им ранее область деятельности, что приведет к появлению в его лице уважаемого ученого, писателя, инженера;

и пр.

*Мозг, хорошо устроенный, стоит больше,
чем мозг, хорошо наполненный.
(Из фольклора бывалых физиков)*

Скорее всего, автоматизированная процедура массовой оценки, не несущая эмоциональной окраски и применяемая единообразно ко всем испытуемым, не вызовет подобных последствий. Хотя и здесь есть о чем подумать методистам и сценаристам будущих компьютерных обучающих систем.

Верный психологический ход, грамотно построенный автоматизированный анализ результатов теста сопровождаемый полуинтерактивным общением с учащимся, может привести к положительному результату – «зажечь» будущего специалиста.

*Народ, презревший историю, не имеет будущего.
(прописная истина)*

Одним из первых ученых, попытавшихся измерить различия между людьми с помощью тестов, был Ф. Гальтон. Он определил три принципа тестирования:

- применение серии одинаковых испытаний к большому количеству испытуемых;
- статистическая обработка результатов;
- выделение эталонов оценки.

Дж. Кеттелл считал тест хорошим средством для проведения научного эксперимента по выявлению характеристик людей. Он выделил такие требования к сеансу тестирования:

- одинаковость условий для всех испытуемых;
- ограничение времени тестирования приблизительно одним часом;
- отсутствие зрителей в лаборатории, где проводится эксперимент;
- оборудование должно даже своим видом располагать людей к тестированию;
- одинаковые инструкции и четкое понимание испытуемыми плана действий;
- результаты тестирования должны подвергаться статистическому анализу.

Наиболее полно эти требования могут быть соблюдены при компьютерном тестировании.

Компьютерное тестирование – путь к адаптивному обучению, к эффективным и не применяемым в России формам учебного процесса. Эти возможности могут широко проявиться лишь при условиях изменения общей организации обучения.

Использование компьютерного тестирования позволяет учащимся самостоятельно обнаруживать пробелы в своих знаниях и помогает принимать меры для их ликвидации. Это говорит о значительном обучающем потенциале тестовых заданий.

Глава 01. Компьютерные технологии аттестации

*Ничто так не путает понятий об искусстве,
как признание авторитетов.
Лев Николаевич ТОЛСТОЙ*

Среди продукции системы образования – знания, умения и навыки, переданные учащимся и воспринятые ими. Качество этой образовательной продукции играет в развитии общества первостепенную роль.

Без измерения и сравнения качества продукции невозможно развитие производства. В образовании для измерения качества применяют поурочный, рубежный, итоговый и прочие виды контроля. Каждый вид контроля характерен своей технологией.

Наиболее объективная технология массового контроля, по мнению многих, – автоматическая, однообразная, не зависящая от каких-либо субъективных влияний. Хотя, у этой точки зрения есть и противники, считающие, что контроль должен проводиться «глаза в глаза».

Слабые места первой точки зрения: невозможно автоматическими средствами проверить и гарантированно оценить умения

- излагать мысли;
- находить нестандартные решения;
- учить и учиться.

Вторая точка зрения также имеет слабые места:

- высокая себестоимость;
- отсутствие гарантии исключения субъективизма;
- отсутствие в регионах страны достаточного количества кадров гарантированно высокой квалификации, способных гарантированно беспристрастно и единообразно оценить свойства большого количества претендентов.

В споре между точками зрения следует занимать «философскую» позицию, находить золотую середину. Например, такую:

- контролировать автоматическими средствами знания рутинных сведений, умений производить стандартные действия – т. е. то, что в вузе относят к «зачету»;
- контролировать силами специалистов умение мыслить и излагать мысли, а также другие свойства, не поддающиеся контролю со стороны автомата.

Аттестация ЗУН – многосторонний контроль, проводящийся по наиболее объективной из возможных технологий, сопровождающийся оформлением признаваемого в обществе документа. Аттестация ЗУН – важнейший элемент учебного процесса. Объективность аттестации – одна из гарантий развития и поддержания жизнедеятельности общества.

В мире разработано и внедрено множество подходов к аттестации ЗУН, основанных на применении компьютеров и телекоммуникаций технологий. Многие из них признаны достаточно объективными.

Однако среди известных компьютерных обучающих и тестирующих комплексов трудно найти абсолютно выдерживающий критику с точки зрения

- объективности;

- удобства;
- привлекательности;
- полноты набора необходимых функций контроля и представления результата;
- объемности содержания.

Счастливым исключением – профессиональные военные тренажеры (к слову сказать, часто созданные за рубежом руками именно российских умельцев).

Это объективно сложившееся положение связано со многими причинами:

- во-первых, в России педагогическое тестирование долгое время было запрещено нормативными документами и, в результате, России приходится догонять другие страны в этой технологической области. Также, как в кибернетике и генетике, которые более «на слуху»;
- во-вторых, период первоначального развития технологий компьютерного тестирования пришелся в России на период экономического кризиса;
- в-третьих, развитие качеств российских компьютерных комплексов тормозит прагматичное отношение к результату их применения, имеющее простую экономическую подоплеку. Наделение продукта перечисленными выше качествами – дорогостоящее удовольствие;
- в-четвертых, культура производства в России значительно ниже, чем в других странах, где предпринимателю не все равно, каких работников он принимает на работу, и где предприниматели давно применяют автоматические методы для контроля ЗУН. Редкие российские работодатели прониклись важностью этих мероприятий.

Перечислим недостатки традиционных систем оценки и аттестации ЗУН:

- субъективизм;
- нерегулярность;
- несогласованность требований и режимов контроля;
- сложность ведения статистической обработки (мониторинга);
- сложность обеспечения режима секретности хранения экзаменационных материалов и защиты от подлога;
- наличие случайных, несистематических ошибок измерений;
- отсутствие чётких математических критериев оценки.

Многое из перечисленного вполне может быть преодолено при внедрении компьютерных технологий. Однако, пока это – только пожелание, подтвержденное теоретически. Безупречной системы автоматического контроля ЗУН в настоящее время нет.

*Самую большую и самую искреннюю часть
наших молитв составляют жалобы.
Джонатан СВИФТ, aforizm.kaminplus.ru*

Связано это со многими причинами:

- разнородность, частая смена и относительная незрелость (молодость) как ИТ, так и соответствующего технического обеспечения;
- отсутствие должной государственной поддержки разработки и внедрения;

- отсутствие специалистов;
- отсутствие адекватной теоретической базы.

Путь преодоления этих проблем связан с новыми управленческими решениями и мероприятиями, в частности, такими:

- государственная поддержка развития тестологии как отрасли знаний и науки;
- создание индустрии систем тестового контроля.

В современном научном мире сложилась противоречивая ситуация с признанием тестологии как полноправной науки. С одной стороны, общее признание тестового контроля работоспособным направлением исследования широкого спектра педагогических и психологических проблем; с другой – затянувшаяся сдержанность, недостаток информации, непонимание сущности и возможностей.

Может быть, следует более конкретизировать понятие образования, более наглядно представить его в качестве товара? Там, где товар, там и сертификация, там и тестирование.

В России считается нормальным тестировать водку, шпроты, огурцы, квас и многое другое. Нормально, когда весь товар имеет сертификат и производитель товара пользуется защитой качества своего товара со стороны государства. Соответствующие гарантии, предоставляемые независимыми центрами тестирования, позволяют производителям, покупателям и продавцам цивилизованно защищать свои права. Те же гарантии позволяют непротиворечиво оценивать продукцию, вести стратегическое планирование производства.

Почему такого нет на российском рынке образовательных услуг? Причин тому много:

- Образование не имеет четкого математического определения.
- Управленцы в системе образования не имеют достаточного образования.
- Проникновение математики в педагогику – далекая перспектива.
- Существуют в обществе силы, препятствующие этому (с компьютером труднее договориться о подлоге, чем с экзаменатором).
- Понятие и разъяснение о защите качества отсутствует в Законе «Об образовании». Система образования вооружена только понятиями аккредитации, сертификации, аттестации – т. е. тем, что относится к формальному засвидетельствованию качества.

01.01. Педагогическое и психологическое тестирования

«А такой анекдот вы слышали?»

- «От вас – еще нет».

Юрий НИКУЛИН

Во многих областях практической психологии широко используются измерительные психодиагностические методики, к которым относятся тесты на измерение способностей, достижений, аппаратные методики, основанные на стандартизированном самоотчете – опросники и техники субъективного шкалирования. Корректность в применении этих методик обеспечивается не только содержательными представлениями, но и выполнением особых требований психометрики ... Тестовые методики призваны решить определенный ограниченный круг задач. Это задачи массовой экспресс-диагностики. Здесь не исключены ошибки в индивидуальных случаях, диагноз и прогноз даются лишь с вероятностной точностью.... Прежде чем применить какой-то тест-опросник на особом контингенте лиц, психолог должен постараться взглянуть на каждый вопрос глазами испытуемого. Это требует немалой профессиональной интуиции ... (Шмелев А. Г., Похилько В. И. Анализ пунктов при конструировании и применении тест-опросников: ручные и компьютерные алгоритмы. – Вопросы психологии, № 4, 1985. – С. 126-134.)

Этот текст позволяет окунуться в круг проблем, которые решают психологи с помощью тестирования. Психологу, в отличие от педагога, приходится работать не с группами учащихся, а с совершенно разными людьми. Тесты психологов отличаются от педагогических удивительностью, неожиданностью и простотой заданий. По результатам выполнения заданий, психолог решает сложнейшую задачу – определяет тип человека и его характеристики, принимая во внимание множество шкал и определений, в том числе противоречивых, изобретенных как им самим, так и его предшественниками и коллегами.

Перед психологом, составляющим тест, всегда стоит трудноразрешимая задача: создать минимальное количество тестовых заданий, позволяющих получить максимально достоверную и обширную информацию о человеке. Тестовые задания психологов, поэтому, многофункциональны – результат выполнения их «говорит», частично и сразу, о нескольких характеристиках человека. Их составление – это путь проб и ошибок. Каждое задание проверяется на больших группах испытуемых и только потом делается заключение о включении или не включении его в тест. Поскольку возможности эксперимента всегда ограничены, нет гарантии, что задание одинаково надежно будет играть свою роль всегда. Кроме определения типа человека, психолог определяет, с долей достоверности:

- профессиональную пригодность;
- деловые качества;
- образ мышления;
- склонности;
- предрасположенности;
- психические болезни и расстройства;
- совместимости,

и другие характеристики. К образованию эти характеристики имеют отдаленное отношение.

01.02.Бланковое тестирование. Десять несовершенств

*В строю мы привыкли к порядку;
Нам по сердцу творческий труд!*

В. МАТВЕЕВ

*Левицкая Н. В., Луганская Л. И., Лаврова К. И. Русский язык.
Учебное пособие для солдат, не владеющих или
слабо владеющих русским языком. Воениздат, 1982.*

Ранние попытки автоматизировать контроль ЗУН во многих странах в начале XX века привели к созданию бланковых технологий тестирования и «индустрии тестов», а затем – к появлению международно-признанных организаций, проводящих ежегодно десятки миллионов сеансов бланковых тестирований. В России одной из таких организаций является Центр тестирования Министерства образования Российской Федерации. (ЦТ МОРФ), многие годы проводящий ежегодное Централизованное бланковое тестирование (ЦТ) среди школьников и абитуриентов.

Несовершенства бланковых технологий тестирования стали очевидны после появления достаточно качественных компьютерных дисплеев (мониторов) в 80-х годах прошлого века.

К техническим несовершенствам бланковых технологий можно отнести:

- длительность тиражирования;
- высокая вероятность подлога;
- сложность сохранения конфиденциальности;
- высокая вероятность угадывания верных вариантов путем простого логического сравнения;
- трудоемкость тиражирования, хранения и транспортировки бланков и соответствующих вопросников;
- сложность обеспечения должного уровня секретности экзаменационных материалов и относительная простота подлога;
- сложность обеспечения верного восприятия правил претендентами и их выполнения;
- трудоемкость извлечения и обработки результатов;
- дороговизна сеанса.

и др. К методическим несовершенствам бланковых технологий мы здесь отнесем еще одно, на наш взгляд, самое существенное. Суть этого несовершенства легко понять, если внимательно сравнить две популярные телевизионные игры – «Кто хочет стать миллионером?» и «Своя игра». В первом случае один играющий указывает верный ответ на вопрос в предъявленном списке. Во втором – игроки соревнуются друг с другом в скорости ответа на вопрос, зная лишь его тему и используя лишь свою память.

С методической точки зрения, эти два сценария кардинально различны. В первом случае указать правильный ответ несравненно легче:

- во-первых, видя четыре варианта ответа, среди которых всего один верный, можно осуществить выбор просто логическим путем, не имея специального знания;

- во-вторых, даже если нет возможности вспомнить верный ответ или догадаться путем логического перебора вариантов, можно выбрать любой ответ и с вероятностью 25% добиться успеха.

Во втором же случае, т. е. в «Своей игре», ни логика, ни удача не помогут победить – помогут только память и знание.

Применяемые тестовые задания называют:

- в первом случае – заданиями закрытой формы;
- во втором случае – заданиями открытой формы.

Надо сказать, что, согласно математической теории вероятностей, приведенную выше величину – 25% – можно уменьшить, т. е. уменьшить вероятность простой удачи. Например, четырьмя способами:

- увеличением количества вариантов выполнения заданий;
- помещением среди демонстрируемых вариантов не одного, а нескольких верных ответов, которые требуется указать все;
- созданием дистракторов;
- применением технологий психофизиологического контроля.

Однако, даже при применении этих способов, при тестировании:

- все равно сохраняется, хотя и уменьшается, степень возможности достижения успеха с помощью простой логики, поскольку испытуемый видит одновременно все варианты и имеет возможность их сравнить;
- в связи с наличием в результатах тестирований определенного высокого процента успеха, списываемого «на удачу», много труда составителей тестовых заданий оказывается напрасным (иногда даже свыше трети);
- значительно «раздувается» объем бумажных бланков, и, соответственно, увеличивается вероятность сбоев при проверке;
- увеличивается непроизводительное время знакомства испытуемых с заданиями, соответственно должно быть уменьшено количество заданий в связи с объективно накапливающейся усталостью и ограниченным временем сеанса.

В технологиях автоматизированного бланкового тестирования обычно заложен именно вариант выбора из списка (задания закрытой формы) с его **главным недостатком** – возможностью увидеть одновременно все правильные и неправильные варианты ответа и воспользоваться этим, проявив не знания, а умение логически мыслить.

На взгляд организаторов таких тестирований, все отмеченные выше недостатки тестов с заданиями закрытой формы окупаются простотой обработки результатов. Бог им судья.

Для «обмана» человеческой логики создатели бланковых тестовых материалов придумали множество приемов. Эти приемы позволяют конструировать варианты ответов на вопросы так, что неверный ответ кажется правдоподобным, а верный – неправильным. Совокупность этих приемов иногда называют **теорией дистракторов**.

Создатели этих приемов, подобно мастерам рукопашного боя, сохраняют их в тайне. Мы не укажем доступное изложение всех этих приемов, изданное в открытой печати.

Самое удивительное, что не часто удастся встретить среди педагогов желающих познать тайны «теории дистракторов». Не смотря на то, труд по созданию тестовых материалов и становится в России все популярнее. Российский преподаватель – сам себе голова. Гово-

рят, что российская система образования потому так хороша, потому она так магически устойчива, потому ее не смогли разрушить реформы, что устроена она на удивительных традициях:

- не верить авторитетам;
- учить так, как придет в голову в данный момент, не оглядываясь на поурочные планы;
- подчиняться приказам начальников «по-итальянски»;
- ожидать от учеников только лучшего, подозревая худшее;
- надеяться на «авось».

Процедура массового бланкового тестирования слабо защищена от подлога. Одно дело, когда педагог проводит тестирование для облегчения собственного труда и сам наблюдает за выполнением процедуры, и совсем другое – когда тестирование ведется большим коллективом людей, которые не заинтересованы в качестве результатов. Или, еще хуже, которые заинтересованы лишь в успехе «выделенных» испытуемых. Такие явления можно увидеть, например, при проведении Централизованного тестирования или Единого государственного экзамена, организуемых МОРФ по всей территории России уже несколько лет. Особо яркие случаи коллективного подлога известны в педагогической печати.

Возможных путей подлога в бланковом тестировании довольно много:

- «выделенного» учащегося можно вместе с заданием поместить в отдельный кабинет, где он все правильно выполнит под присмотром, например, репетитора;
- задание может выполнить репетитор без участия самого испытуемого и потом подложить бланк в общую пачку;
- в процессе своего изготовления бланки проходят через сотни рук и глаз, поэтому тестовые ключи «уплывают» на рынок. Их можно купить предварительно, и особенно легко это сделать в Москве, поближе к Министерству образования. Тестовые ключи «уплывают» одновременно с отправкой тонн бланков в регионы России.

При желании и опыте можно придумать еще больше вариантов теневых путей к успеху в бланковом тестировании.

*В России две беды – дураки и дороги.
Н. В. ГОГОЛЬ*

Перевод тестирований на компьютеризованную основу способен резко поднять объективность результатов. Над решением этой проблемы работает много ученых и организаций. Например – Центр тестирования МОРФ, где объединяются лучшие тестологические силы России. Тем не менее, несмотря на очевидность, магическим образом, связанные с человеческим фактором несовершенства бланковой технологии вот уже несколько лет сохраняются и здесь. И охраняются нормативными документами.

Здесь далее будут подробно рассмотрены известные технологические особенности, позволяющие обеспечить уровень информационной безопасности массовых компьютерных тестирований на уровне учреждений, территорий, регионов, государств.

01.03. Двадцать преимуществ компьютерного тестирования

De minimis non curat computer –

Компьютер не занимается малозначительными делами (лат.)

Появление персонального компьютера, а также сопутствующих ему принтера и сканера, привело к значительной модернизации бланковой технологии и резкому подъему массовости бланковых тестирований. Компьютер стал использоваться для подготовки (распечатки) бланков и для статистической обработки информации, которую испытуемые на бланки заносили в виде крестиков, галочек, и даже букв. По сравнению с ранними бланковыми технологиями, где результат обрабатывался вручную, снизились:

- трудоемкость подготовки материалов и обработки результатов;
- вероятность внесения случайных ошибок в результат.

Вместе с тем повысились требования к аккуратности испытуемых, поскольку:

- используемое программное обеспечение не всегда верно позволяло идентифицировать нанесенные карандашом на бланк знаки;
- помятость бланка могла привести к его деформации в считывающих механизмах и появлению дополнительных погрешностей.

Трудно восстановить, кому первому пришла в голову мысль о тестировании ЗУН, минуя бланки, только с помощью компьютера. Скорее всего, это произошло одновременно во многих странах. Так же, как и мысль о компьютеризации игр. Гениальная идея о компьютерном тестировании ЗУН для педагогики столь же значительна, как и изобретение паровой машины для техники.

За последние два десятилетия в мире обозначилась тенденция к переходу от бланковых технологий к полностью электронным технологиям тестирования. В конце XX – начале XXI вв. исторически первой в России получила известность Всероссийская компьютеризованная олимпиада «Телетестинг», проводившаяся Московским государственным университетом (Центр «Гуманитарные технологии»).

Компьютерные технологии тестирования позволили «забыть» о многих недостатках бланковых технологий. С помощью компьютерной техники оказалось возможным:

1. Быстро сравнивать результаты измерений для большого количества испытуемых на больших территориях;
2. Внедрять и разрабатывать новые технологии адаптивного тестирования;
3. Повысить дифференцирующую способность тестов;
4. Оперативно использовать методы статистической обработки для повышения уровня объективности результатов;
5. Объективизировать результат измерения путем применения новых способов демонстрации тестовых заданий, например:
 - исключением возможности для испытуемого видеть все варианты ответов на вопрос задания закрытой формы и сравнивать их логическим перебором;
 - заменой фиксированных вариантов наборов заданий случайными выборками;
6. Оптимизировать продолжительность тестовых измерений;

7. Устранить некоторые, существовавшие в бланковых технологиях, возможности для возникновения случайных и технических ошибок;
8. Устранить необходимость перевозок бланков и связанные с этим потери, опоздания, внешние влияния;
9. Устранить некоторые возможности подлога;
10. Снизить количество персонала, необходимого в момент проведения мероприятий;
11. Оперативно следить за объемом и качеством ЗУН больших контингентов учащихся, сохраняя при этом все индивидуальные характеристики, выявленные для каждого;
12. Вести мониторинг качества образования в масштабе страны, использовать его результаты для целей аттестации учреждений и работников системы образования;
13. Увеличить скорость поиска и предоставления заданий в соответствии с алгоритмами адаптивного тестирования;
14. Упростить хранение, сортировку и формирование больших тестовых массивов в соответствии с алгоритмами;
15. Увеличить скорость обработки результатов;
16. Реализовать возможность автоматической массовой коррекции параметров заданий;
17. Реализовать возможность интерактивного адаптивного взаимодействия системы с претендентом;
18. Упростить введение алгоритмических ограничений на действия претендента;
19. Повысить технологичность проведения тестирований и снизить уровень требований к персоналу центра тестирований;
20. Упростить ведение архива тестирований.

Применение компьютерных технологий вполне может положительно повлиять на уровень качества российского образования в самое ближайшее время.

01.04. Общие достоинства и недостатки тестовых технологий

*Собственные недостатки – это котомка за плечами: не увидишь.
Эразм РОТТЕРДАМСКИЙ, aforizm.kaminplus.ru*

С внедрением тестовых технологий в педагогике появилась возможность использования точных статистических методов анализа качества, позволяющих повысить объективность суждений о том, в какой степени усилия преподавателей и учащихся достигают цели.

Внедрение тестового контроля:

- ведет к совершенствованию приемов обучения, позволяя оперативно оценивать их результативность на основе объективных критериев;
- превращает педагогику в (допускающую измерение) точную науку;
- влечет появление новых воспитательных и обучающих методик;
- превращает преподавателя из транслятора знаний в разработчика новых программно-педагогических средств, в организатора самостоятельного обучения.

Как бланковое, так и компьютерное тестирование ЗУН нередко подвергаются критике со стороны педагогов. Изготовленные в России тестирующие программы мало отличаются от тех, что предлагают зарубежные производители. К основным недостаткам автоматизированных систем тестирования часто относят

- двужначность (не многозначность) логики вариантов ответа в наиболее популярных тестовых заданиях закрытого типа;
- невозможность контроля навыков устной речи;
- жесткость контроля, ведущая к психическому напряжению и сверхнормативной усталости претендентов;
- невозможность диагностики навыков общения (коммуникации);
- невозможность диагностики изобретательских качеств;
- невозможность диагностики оригинальности мышления в решении учебных проблем и задач;
- невозможность учета всех случайностей и факторов, которые всегда учтет преподаватель-профессионал на экзамене – настроение, усталость, темперамент, возраст, пол, национальность экзаменуемого;
- невозможность диагностики ассоциативного и образного мышления, способности к обучению, желания обучаться.

Поэтому совместно с использованием тестовой формы контроля необходимо практиковать и традиционные формы – семинары, конференции, диспуты, обсуждения, деловые игры. Общепринято, что оптимального сочетания форм контроля можно достичь так:

1. С помощью компьютерных тестов проводится предварительный контроль (зачет) знаний определений и сведений, навыков решения задач, навыков восприятия речи на слух, грамотного написания слов и формул и т. п.;
2. После зачетных испытаний учащимся предлагается встреча с преподавателем, где они продемонстрируют уже не формальные знания, а навыки коммуникации и устной речи, остроту и оригинальность мышления, способность к учению.

Более подробно проблемы двузначности логики компьютерных тестеров и тренажеров мы разберем ниже. Пока же приведем следующие соображения. Поскольку создателями сценариями компьютерных тренажеров, преимущественно, были и остаются мужчины, они привнесли в практику создания свойственный себе логический стиль общения, где ответы на вопросы могут быть только верными либо неверными.

Что есть истина?

В религии – это чудом сохранившаяся точка зрения.

В науке – это сенсация.

В искусстве – чье-то вчерашнее настроение.

Оскар УАЙЛЬД, aforizm.kaminplus.ru

В реальном учебном процессе, как правило, такой ситуации не бывает. Даже односложный ответ ученика грамотный учитель способен ранжировать по пятибалльной шкале, опираясь на свой опыт общения с ним и в зависимости от контекста урока.

Не бывает однозначной научной истины. Не бывает абсолютно точных ответов на все случаи жизни. Так устроена жизнь. Грамотный специалист, изобретатель, ученый может появиться лишь в такой учебной среде, где логика общения многозначна (с ответами, например: «да», «нет», «может быть», «не совсем точно», «не вполне»), где есть противоречия в знаниях, где есть место для фантазии. Это одна из причин, из-за которых современная, созданная по традиционной логической схеме, компьютерная обучающая программа не сможет его научить и воспитать.

Парадоксальность ситуации с применением многозначной логики в компьютерном обучении и контроле связана еще и со следующим: проводя урок, учитель без труда задает классу подразумевающий многозначные ответы вопрос, распознает и оценивает эти ответы, а выразить их письменно – может только с большим трудом.

Создавать тест, содержащий задания с многозначительными (многозначными) вопросами и ответами, оказывается, очень сложно!

Конечно, известные технологические разработки, посвященные применению техники для контроля знаний, не исчерпывают всех возможностей компьютера. Возможности компьютера в сфере образования нам еще предстоит открывать и изучать.

01.05. Культура и философия адаптивного тестирования

Развитие каждой науки сопровождается философским осмыслением. Тестология – наука молодая. Особое внимание в ней уделяется адаптивным технологиям, которые не только оптимизируют процесс тестирования, но и сближают процессы создания и измерения образовательной продукции. Ниже приведена совокупность формулировок, позволяющая судить о современном уровне развития философского осмысления развития тестологии.

Торопить женщину – то же самое, что пытаться ускорить загрузку компьютера.

Программа все равно должна выполнить все очевидно необходимые действия и еще многое такое, что всегда остается сокрытым от вашего понимания.

Смас ЯНКОВСКИЙ, <http://syi.narod.ru/wordsd.htm/>

Сущность, способы, формы и содержание компьютерного дидактического оценивания уровня ЗУН, интеллектуальных умений и практических навыков культурология изучает в связи с ответами на вопросы:

- Каков смысл компьютерного тестирования?
- Почему необходим переход к адаптивному оцениванию?

и др. Дефицит культурных оснований в традиционных системах контроля учебных достижений связан со скудным учетом рефлексии личности. Этому рациональному представлению отвечает культурологический подход, связывающий меру и критерий прогресса в сфере образования с массовой и объективной оценкой личности.

Философия культуры адаптивного тестирования (КАТ) – это система методологических и теоретических знаний, основным содержанием которой являются:

- закономерности установления действительного состояния обученности испытуемых;
- правила и стандарты продуцирования культурных дидактических систем;
- установления механизмов функционирования и развития систем тестирования.

На первый план философия КАТ выдвигает проблемы духовного и социального смыслов деятельности, гуманизма и нравственности.

Философия КАТ формируется под влиянием двух подходов: методологического и технологического. Объектами методологического подхода являются концептуальная модель тестирования, система правил конструирования феноменов культуры, стратегия оценивания и представления результатов тестирований, конкретные формы представления тестовых заданий, допустимые границы классификации и аттестации тестируемых и т. п.

Эмпирический и теоретический подходы к проектированию интеллектуальных систем тестирования различают по методам. В первом случае разработчик оперирует содержательными понятиями в образе идеальных объектов, обладающих набором признаков. Например, в процессе тестирования в среде АСТ -тест подвергаются установлению только учебные достижения испытуемых, а другие признаки человека остаются вне поля зрения.

Концепция культурологии тестирования понимается как сжатое описание социальной информации, которая создается, преобразуется, накапливается, передается и модифицируется с помощью создаваемых участниками образовательного процесса знаковых и инструментальных средств, образующих культурное пространство.

Элементарным феноменом тестовой культуры (ЭФТК) считается тестовое задание, которое создано с применением правил и стандартов, ранжировано по степени трудности,

обладает заданным значением валидности и надежности, характеризуется универсальными, общими и специфическими чертами.

Универсалии происходят оттого, что в культуре проектирования тестовых заданий существует устойчивый порядок, образованный основными положениями тестологии. Общие черты, характерные для тестовых заданий, состоят в том, что тестовые утверждения из определенной области знаний, созданные различными тестологами, связаны с культурой мышления разработчиков и возможностью их общения в процессе конструирования.

Имеет место культурное многообразие в различном представлении одного и того же содержания учебного материала с помощью уникальных строений тестовых утверждений. Это несет в себе творческие потенции, характерные для мышления каждого разработчика проблемной тестовой ситуации. Но главным здесь является применение стандартных схем конструирования, что придает заданиям специфические черты и делает их узнаваемыми. Узнаваемость формы тестового утверждения имеет принципиальное значение, поскольку она способствует концентрации внимания именно на содержании проблемной ситуации.

Понятие **стандартизированной формы задания** включает такие связи и отношения между элементами, которые характеризуются признаками объективности и повторяемости. В отличие от методологических правил, стандарт не имеет исключений и его действию подвержены проблемные тестовые ситуации из любой области знаний.

Культурное пространство тестирования представляет собой открытую систему, развивающуюся на основе объективных социальных законов, форму интроспекции тестируемых, способ их рациональной организации.

Наличие банков тестовых заданий (БТЗ), охватывающих отдельные направления или специализации, – один из источников многообразия культурных форм тестов, придающий им окраску различных научных школ. Неповторимость каждого БТЗ означает, что в отношении учебного содержания, БТЗ, созданные в разных учебных заведениях, эквивалентны.

Есть более развитые, более мощные и менее развитые культурные формы и культурные системы. В рамках культурного пространства создается возможность непрерывного взаимообогащения территориально разобщенных культурных систем, синтеза различных культурных форм и модификации элементарных феноменов тестовой культуры.

В этом пространстве можно выделять координаты знания, ценности и регулятивы, то есть требования и правила, по которым участники строят свое поведение. Эти координаты образуют три пересекающиеся плоскости – духовной, технической и социальной культуры, на которое «проецируется» содержание культурного пространства.

Тестология характеризуется комплексом разнообразных смыслов –

- духовных (например, в личностных планах);
- технических (например, в использовании тех или иных инструментальных сред);
- социальных (например, в отношениях с информационными службами).

Предмет культурологии адаптивного тестирования: изучение объективных закономерностей развития культурного пространства стандартизированного компьютерного адаптивного оценивания уровня учебных достижений в различных аспектах свободной творческой деятельности, создание и реализация символически обозначаемых и общественно значимых для всех участников процесса тестовых испытаний ценностей и смыслов.

Объективность оценки достигается:

- одинаковым содержанием норм и правил для всех претендентов;
- применением единой автоматической шкалы оценивания;

- возможностью представления содержания ТЗ в различных формах;
- стандартизацией тестовых материалов, технологии и условий;
- устранением влияния мешающих факторов и артефактов;
- научной обоснованностью структуры культурных форм тестов;
- валидностью БТЗ;
- защитой от несанкционированного доступа;
- достаточной величиной генеральной выборки ТЗ для формирования множества культурных форм тестов;
- применением механизмов адаптации;
- жесткой связью между результатом и содержанием.

С философской точки зрения процесс тестирования не позволяет достичь абсолютно точного измерения уровня достижений. Это связано со следующим:

- не все понятия областей знаний, можно представить в виде ТЗ;
- любое наблюдение за поведением объекта не может служить доказательством установления уровня его абсолютной обученности;
- конечное количество ТЗ накладывает ограничения на точность оценки.

Объектом культурологии являются участники создания моделей, методологических правил и стандартов, культурных систем и артефактов, объединяющие различные феномены культурного пространства в целостную систему.

В качестве собственных приложений культурология выделяет следующие подразделы:

- историю становления культуры адаптивного тестирования;
- философию культуры адаптивного тестирования;
- методологическую и прикладную культуру дидактического оценивания достижений.

Культурное пространство КАТ представляет собой упорядоченные элементарные феномены и формы культуры, выполненные с применением методологических правил и стандартов, хранящиеся в культурных системах и образующие целостную структуру благодаря наличию интеллектуальных инструментальных средств и систем коммуникации.

Сущностью поведения участников в этом пространстве является деятельность по правилам.

*Истина хорошо усваивается только в разбавленном виде.
Ее содержание в информации должно быть в диапазоне от 4 до 40%.
Смас ЯНКОВСКИЙ, <http://syy.narod.ru/wordsd.htm/>*

При формировании текста параграфа использованы материалы: Васильев В. И., Тягунова Т. Н. *Теория и практика формирования программно-дидактических тестов.* – М.: Издательство МЭСИ, 2001 . – 130 с.

Глава 02. Технологии массовой компьютерной аттестации ЗУН. Цели, задачи, характеристики, этапы, пути, результаты

*Не занимайся наукой со звериной серьёзностью,
науку надо делать весело и красиво, иначе нечего в неё и соваться.
Николай Владимирович ТИМОФЕЕВ-РЕСОВСКИЙ*

Этот раздел обычно является в учебных пособиях формальным. Мы постарались сделать его полезным, не скучным и информативным.

Тестовый контроль – надежный и объективный способ оценивания. Он ориентирован на возможно более точное измерение степени (уровня) усвоения понятий и разделов учебной программы, умений, навыков.

В России педагогическое и психологическое тестирование долгое время было запрещено. История возникновения идей тестирования в России уходит корнями в педологию, научное течение, возникшее среди педагогов в начале XX века и официально закреплённое в России как «лженаука» подобно кибернетике и генетике. Сейчас России приходится догонять мир.

Цель и три десятка задач педагогического тестового контроля с шести точек зрения

*Математику ошибочно считают наукой трудной,
а иногда даже подозрительной только потому,
что она имела несчастье быть неизвестной отцам церкви.
Роджер БЭКОН*

Процедура аттестации является важным элементом образовательного процесса. Переходя к автоматической аттестации, имеют в виду много целей, которые, в конечном счете, сводятся к одной главной:

Цель педагогического тестового контроля:

- опосредованное поддержание и повышение уровня качества образования.

Каждый участник учебного процесса, а также потребитель образовательной продукции, используя тестовый контроль, или подвергаясь тестовому контролю, решает достаточно разные задачи. Перечислим некоторые из них.

Задачи тестового контроля, как эффективного инструмента управления качеством, с точки зрения руководителя учебного заведения:

- Повышение уровня качества учебного процесса;
- Освобождение преподавателей от рутинного труда (ни один преподаватель не сможет так, как это делает компьютерный комплекс, за 15-20 минут задать всем учащимся группы по 40-50 вопросов и объективно оценить ответы);

- Объективная дифференциация абитуриентов, как минимум, на три категории: на тех, кто умеет, и кто будет учиться, на тех, кого не следует учить, и на «полупроходников»;
- Объективная дифференциация учащихся на успешно завершивших обучение, второгодников, и тех, кого следует исключить из учебного процесса;
- В случае функционирования в учебном заведении нескольких учебных программ, а также других нестандартных ситуациях (посылка лучших учащихся на олимпиаду, подготовка к сложной аттестации учреждения) решение задачи тонкой дифференциации учащихся;
- Объективное выявление изъянов учебного процесса (нерадивость преподавателей, несогласованность учебных планов и др.);
- Повышение аудиторного потенциала, снижение неэффективной аудиторной нагрузки преподавателей;
- Снижение потока апелляций;
- Измерение уровня образовательной подготовки учащегося по дисциплине;
- Диагностика достижения целей обучения, установленных в соответствии с ГОС и нормативными документами;
- Обеспечение привлекательности учебного процесса за счет мотивации, побуждения и поддержания желания учащихся повышать качество личной образованности, снижения аудиторной нагрузки;
- Привлечение контингента учащихся из числа занятых на производствах;
- Генерация и представление рекомендаций по управлению учебным процессом;
- Снижение себестоимости учебного процесса;
- Повышение уровня престижности учебного заведения и привлекательности его для инвестиций.

Задачи тестового контроля с точки зрения преподавателя:

- Облегчение рутинного труда по проверке стандартизованных знаний. С помощью компьютера можно усилить контроль не только знаний фактов и определений, но и навыков решения типовых задач;
- Прогноз успешности прохождения учащимися выходного тестового контроля, получение оперативной информации об изъянах в качестве ЗУН учащихся;
- Повышение уровня качества обучения (за счет, например, включения элементов игры и соревнования) путем применения автоматизированного контроля.

Задачи тестового контроля с точки зрения учащегося:

- Выявление изъянов в собственных знаниях и навыках, дать рекомендации;
- Решение психолого–педагогические проблем (неприятие личности преподавателя, боязнь экзаменационной лотереи и необъективности преподавателя и др.).

Задачи тестового контроля с точки зрения руководителя управления образованием:

- Решение проблемы контроля качества образования;
- Объективизация аттестации учреждений образования и их работников;

- Объективизация сравнительной картины качества образования во всех учреждениях, моментальный снимок, «срез».

Задачи тестового контроля с точки зрения родителей учащегося:

- Устранение экзаменационных стрессовых ситуаций из жизни ребенка;
- Повышение комфортности обучения;
- Упрощение профориентации (оценки проф. пригодности) и прогноза направлений будущей учебы путем объективного соотнесения уровня достижений учащегося по дисциплинам и отдельным разделам с уровнем достижений коллег;
- Снижение стоимости обучения.

Задачи тестового контроля с точки зрения работодателя будущих выпускников:

- Объективное ранжирование будущих сотрудников по уровню ЗУН с целью облегчения выбора предоставляемого им рабочего места, а также мотивированного отказа в предоставлении рабочего места;
- Открытие сведений о результатах тестового контроля учащихся вузов – будущих выпускников для упрощения поиска новых молодых перспективных сотрудников;
- Объективная ориентация в выборе достойного вуза в качестве базового (например, для массовой подготовки специалистов в региональном проекте) путем знакомства с результатами сравнительного мониторинга качества образования.

Десять функций ОИТ контроля учебной деятельности

Функции автоматизированного контроля (в алфавитном порядке):

- Интегративная – способствование выработке представлений о науке как едином комплексе знаний и методов;
- Исследовательская – автоматическая обработка информации об изучаемой системе с целью выработки рекомендаций по управлению качеством;
- Информативная – представление информации о педагогической системе;
- Контролирующая – осуществление контроля качества учебного процесса;
- Оценочная – сопоставление результатов контроля с эталоном;
- Организующая – представление автоматизированных средств качественного и своевременного выполнения учебного плана;
- Реорганизующая – реорганизация труда педагогов, устранение рутинного труда, реорганизация учебного процесса, увеличение доли занятий с применением ОИТ;
- Прогностическая – прогноз успешности деятельности учащихся;
- Регулятивная – представление рекомендаций по регулированию учебного процесса
- Стимулирующая – поощрение творческой деятельности, побуждение обучаемых и педагогов к самообразованию и поиску новых форм и путей обучения.

Время добавит к этому списку новые функции, которые пока находятся за горизонтом нашего восприятия.

02.01.Виды, типы, этапы, требования, характеристики компьютерного тестирования

*Ohne tierisch Ernst
(Без звериной серьезности).
Нильс Бор*

Полдюжины видов и пятнадцать направлений ОИТ контроля учебной деятельности

Контроль усвоения учебного материала и оценка ЗУН является составной частью системы оценки качества. Автоматизация контроля тесно связана с автоматизацией всего образовательного процесса и обеспечивает обратную связь в автоматизированных системах управления образовательным процессом.

Виды автоматизированного контроля:

1. Входной контроль предназначен для:

- определения начального уровня ЗУН обучаемых с целью отбора кандидатов для последующего обучения;
- поддержания высокого уровня познавательного интереса во время обучения на основе определения стиля обучения, оптимального для обучающегося;
- составления индивидуального маршрута обучения.

2. Текущий контроль предназначен для:

- проверки степени усвоения обучаемыми содержания обучения.
- степени готовности обучаемых к изучению последующего материала;
- коррекции индивидуальных маршрутов обучения;
- обеспечения регулятивной функции.

3. Периодический контроль предназначен для:

- проверки степени усвоения обучаемыми содержания обучения. Проводится реже, чем текущий контроль, охватывает целые разделы учебной программы. Основная форма – зачетные тесты. В дополнение используются курсовые работы, контрольно-зачетные задания и т. д. Для периодического контроля, как правило, используются те же тесты, что и для итогового контроля.

4. Итоговый контроль предназначен для:

- проверки качества выполнения обучаемыми учебного плана и проводится в форме экзамена по всей дисциплине;
- решения вопроса о присвоении выпускникам соответствующей квалификации.

5. Самоконтроль предназначен для:

- определения претендентами успешности своего обучения;

- установления претендентами пробелов в изучении дисциплины;
- корректировки претендентами индивидуального стиля обучения.

6. Взаимоконтроль предназначен для:

- формирования у претендентов оценочных алгоритмов и качеств;
- повышения объективности оценивания за счет участия в этом процессе самих претендентов.

Автоматизация позволяет значительно повысить роль всех видов контроля, особенно самоконтроля и взаимоконтроля, открыть новые аспекты их применения. Не все методы контроля могут быть в равной степени подвергнуты автоматизации, но с развитием ИТ появляются новые возможности для этого.

Десять характеристик ОИТ аттестации ЗУН

Ad disputandum
- для обсуждения (лат.)

Важными характеристиками процедуры компьютерной аттестации являются (в алфавитном порядке):

1. Дифференцирующая способность;
2. Валидность;
3. Надежность;
4. Объем результата (количество полезных сведений, которые можно из него извлечь);
5. Оптимальность соотношения времени и точности измерения;
6. Оптимальность соотношения себестоимости процедуры (материалы, техника, оплата труда) и степени достижения цели (объема и объективности результатов);
7. Привлекательность;
8. Простота (уровень необходимой предварительной подготовки персонала);
9. Технологичность и простота (понятность) представления результата;
10. Удобство (простота) применения технологии.

Характеристики эти, естественно, взаимосвязаны. Опыт показывает, что при прагматическом подходе организаторы тестирований менее всего обращают внимание на характеристики №№ 3,4,7.

Десять общих этапов компьютерной аттестации

Argumentum ad ignorantiam
- рассчитанный на неосведомленного собеседника (лат.)

Процедура компьютерной аттестации – довольно продолжительное многоэтапное мероприятие, состоящее из этапов:

1. Регистрация учащихся в специализированной БД;
2. Тренировочное тестирование с элементами обучения, включающее автоматическую генерацию промежуточных результатов и рекомендаций учащимся;
3. Основное тестирование;
4. Статистическая обработка результатов групп учащихся;
5. Публикация рейтингов;
6. Подготовка аттестационной документации (ведомостей) на утверждение преподавателю;
7. Подготовка и представление преподавателю информации о качестве усвоения учащимися разделов дисциплины;
8. Пополнение банка результатов новыми данными;
9. Статистическая обработка результатов, в том числе – сравнение их с полученными ранее и параллельно, т. е. с результатами тестирования иных коллективов испытуемых;
10. Представление результатов в виде, способствующем росту уровня качества образования (в виде статей, докладов, выводов, приказов, отчетов и пр.).

Десять путей достижения целей компьютерной аттестации

*Количество электронов во Вселенной конечно
в смысле конечности нашего знания,
и бесконечно в смысле бесконечности пути познания
Владимир СВИДЕРСКИЙ*

Для достижения целей процедуры аттестации, необходимо проведение мероприятий:

1. Обеспечение валидности теста;
2. Обеспечение вариативности теста;
3. Обеспечение обновляемости базы тестовых заданий;
4. Обеспечение максимальной разрешающей способности теста;
5. Обеспечение достоверности результатов (исключение внешних воздействий);
6. Обеспечение полного и доступного представления результатов;
7. Включение элементов деловой игры в процедуру аттестации и интерфейс компьютерного тестового комплекса;

8. Обеспечение эффективности сеанса тестирования (максимально полный и достоверный результат при минимальных затратах времени и ресурсов);
9. Включение в процедуру аттестации элементов состязательности (перманентная публикация списков претендентов, получивших высшие рейтинги в печати и на Интернет-сайтах);
10. Обеспечение востребованности рекомендаций, получаемых учащимися и преподавателями по результатам тестирований.

Обращаем внимание на пункты №9 и №10. В случае массовых профориентационных и образовательных мероприятий, это – публикация результатов, награждение участников, предоставление победителям престижных рабочих мест и пр.

Дюжина типов сценариев компьютерного тестирования

*Я в этом разбираюсь.
Я знаю, что нужно делать – но каждый раз, когда я
пытаюсь заняться технической проблемой,
какой-нибудь идиот требует, чтобы
я принял решение насчет грузовика,
или телефонов, или другой такой же чертовщины.
Роберт ХАЙНЛАЙН, “The Man who sold the Moon”*

Компьютерные комплексы для тестирования ЗУН могут быть построены с применением разных сценариев. Нам не приходилось встречать в литературе общепринятой их классификации. Приведем классификацию, представляющуюся нам приемлемой сегодня:

1. **«Стандарт».** Самый распространенный, ставший стандартным сценарий, где задания из фиксированного списка предъявляются на экране последовательно, все варианты выполнения задания видны одновременно, и претенденту в каждом случае предлагается выбрать один из предъявленных вариантов выполнения. Задания часто «взвешены», рейтинг претендента вычисляется как приведенная к 100-балльной шкале сумма весов верно выполненных заданий. Здесь велика вероятность

- простого списывания;
- случайного угадывания;
- угадывания путем сравнения вариантов.

Применяются ТЗ I формы.

«Стандарт. Случайный выбор». Сценарий, усложненный и отличающийся от «Стандарта» тем, что задания из фиксированного списка предъявляются претендентам в случайном порядке, а их варианты выполнения также перемешиваются при каждом предъявлении. Такой подход значительно снижает вероятность

- простого «списывания».

Применяются ТЗ I формы.

2. **«Стандарт. Множественный выбор».** Усложненный, по сравнению со «Стандартом», сценарий, где требуется указать в списке не один, а все верные варианты выполнения задания. Такой подход значительно снижает вероятность

- случайного угадывания.

Применяются ТЗ I формы.

«Стандарт. Множественный случайный выбор». Усложненный, по сравнению со «Стандарт. Множественный выбор» сценарий, где задания предъявляются претендентам в случайном порядке, а их варианты выполнения также перемешиваются при каждом предъявлении. Такой подход значительно снижает вероятность

- случайного угадывания;
- простого «списывания».

Применяются ТЗ I формы.

3. **«Единственный выбор».** Усложненный, по сравнению со «Стандартом», сценарий, согласно которому вместе с текстом задания демонстрируется только один из возможных вариантов выполнения. От претендента требуется оценить этот единственный вариант по двузначной шкале (правильно – не правильно). В этом варианте очень низка вероятность угадывания путем сравнения, однако велика вероятность простого угадывания. Для исправления последнего недостатка применяется система «штрафов». Такой подход позволяет

- не только констатировать знания, но и контролировать их глубину; но и значительно снижает вероятность
- случайного угадывания;
- угадывания путем сравнения вариантов;
- простого «списывания».

Применяются ТЗ I формы.

«Поиск однозначного соответствия». Сценарий, где от учащегося требуется однозначно сопоставить термины (изображения, звукозапись) из двух списков (подобно тому, как в математике определяется изоморфное соответствие множеств). В таком подходе очень низка вероятность

- угадывания путем сравнения вариантов;
- случайного угадывания.

Применяются ТЗ III и I форм.

4. **«Поиск неоднозначного соответствия».** Это усложненный вариант, где требуется построить такое соответствие, когда одному термину (изображению, звукозаписи) из первого столбца могут соответствовать 2-3 термина из второго, и наоборот. Такой подход позволяет

- не только констатировать знания, но и контролировать их глубину; но и делает ничтожной вероятность
- угадывания путем сравнения вариантов;
- случайного угадывания.

Применяются ТЗ III и I форм.

5. **«Создание сообщения».** Сценарий, в котором от учащегося требуется ввести в указанном месте слово, словосочетание или совокупность цифр и букв – т. е. закодировать сообщение, являющееся решением поставленной в задании задачи. Этот сценарий сложен как для претендента, так и для автора, поскольку часто такая вводимая претендентом совокупность может быть представлена в нескольких разных, но абсолютно верных вариантах. Закодированное сообщение может указывать закон взаимного соответствия элементов двух или трех множеств, а также последовательность элементов множества.

В таком подходе очень низка вероятность

- угадывания путем сравнения вариантов;
- случайного угадывания.

Применяются ТЗ II, III, IV форм.

6. **«Поиск и исправление ошибки».** Сценарий, усложненный по сравнению с седьмым тем, что место ввода последовательностей знаков не определено, и учащемуся предла-

гается его найти (нет явного указания на место в таблице или тексте на место расположения ошибки). Такой подход позволяет

- не только констатировать знания, но и контролировать их глубину;
- но и делает ничтожной вероятность
- угадывания путем сравнения вариантов;
- случайного угадывания.

Применяются ТЗ II формы.

7. **«Диктант».** Расширенный и усложненный, по сравнению с восьмым, мультимедийный сценарий, в котором требуется написать диктант, слушая звук в наушниках, либо перевести прослушанный текст на другой язык. Такой подход лишен большинства недостатков, связанных со случайностью и позволяет

- контролировать знание правил написания математических выражений, словосочетаний, химических формул и пр.

Применяются ТЗ II формы.

8. **«Путешествие».** Игровой мультимедийный сценарий, где для выполнения задания требуется проделывать сложные манипуляции с клавиатурой и мышью. Например, «двигаться» по анимированному помещению, вводить с клавиатуры словосочетания в специальных «окнах» и «разыскивать» объекты на основании инструкций. Такой подход позволяет

- не только констатировать знания, но и контролировать их глубину;
- но и делает ничтожной вероятность
- угадывания путем сравнения вариантов;
- случайного угадывания.

Применяются мультимедийные ТЗ II формы.

9. **«Определение оценки».** Сценарий близкий сценарию **«Единственный выбор»** и усложненный тем, что претендент должен оценить степень верности единственного предъявленного ответа по какой-либо из многозначных шкал. Например:

- «да» – не знаю – нет»;
- «верно – не точно – не верно»;
- «верно – не полно – не точно – не верно».

Такой подход

- позволяет не только констатировать знания, но и контролировать их глубину;
- делает ничтожной вероятность угадывания путем сравнения вариантов;

Сценарий может быть игровым, имитировать учебные действия учителя и учеников.

Применяются градуированные ТЗ I формы.

10. **«Учитель»** или **«Дидактор».** Игровой мультимедийный полуинтерактивный сценарий, усложненный по сравнению с **«Поставь оценку»**, где претенденту предлагается провести устный опрос в виртуальном классе и поставить оценки виртуальным ученикам. Усложнение состоит в следующем:

- варианты выполнения ТЗ являются двухуровневыми, после выбора варианта и его оценки, демонстрируются варианты, уточняющие уже оцененный вариант, которые так же предлагается оценить;
- среди выполняемых заданий присутствуют двухуровневые задания II формы (открытая форма).

Сценарий отличается крайне малой вероятностью «угадывания» верного варианта. Такой подход позволяет

- не только констатировать знания, но и контролировать их глубину;
- но и делает ничтожной вероятность
- угадывания путем сравнения вариантов;

- случайного угадывания.

Применяются мультимедийные двухуровневые градуированные ТЗ I и II формы.

11. «Диалог». Игровой мультимедийный сценарий, в котором претенденту, в определенном порядке, предъявляются высказывания персонажей в утвердительной форме, которые требуется оценить по двузначной либо многозначной шкале (ответить персонажу). Можно использовать и прямой ввод краткого ответа с клавиатуры.

Такой подход позволяет

- контролировать глубину знаний;
- однако велика вероятность
- случайного угадывания.

Применяются ТЗ I формы, либо градуированные ТЗ I и II форм.

12. «Адаптивное тестирование». Тестирование ведется формально по одному из приведенных выше сценариев, но последовательность предъявления заданий зависит от того, как претендент справился с предыдущими заданиями. Психологическое состояние претендента оценивается автоматически, в зависимости от оценки состояния меняется сложность и трудоемкость заданий, включаются блоки психологической разгрузки. Информация поступает преимущественно по слуховому каналу. Этот сценарий считается наиболее перспективным для применения в практике индивидуальных и массовых тестирований.

Применяются ТЗ I, II, III, IV форм.

По нашему мнению, эту классификацию можно сегодня считать исчерпывающей. Варианты №10 и №11 в настоящее время существуют только в Дальневосточном государственном университете (Открытый университет, Центр новых образовательных технологий), где с ними активно экспериментируют в настоящее время. Вариант №12 описан в тестологической литературе, но в полном реализованном виде пока не предстал «на суд зрителей».

*Каждый раз, когда я заполняю вакантную должность,
я порожаю сотню недовольных и одного неблагодарного.
Людовик IV, aforizm.kaminplus.ru*

Многие известные и распространенные БТЗ, практически без корректировки, могут быть применены с любым из описанных сценариев.

Два десятка вопросов теоретику и практику

*Об уме человека легче судить по его вопросам, чем по его ответам.
Гастон де ЛЕВИС*

Сейчас вокруг проблем технологий тестирования сконцентрировались мысли многих российских ученых и практиков – приверженцев разных, почти антагонистических воззрений. В частности, в современных спорах поднимаются вопросы:

1. Может ли (должен ли) процесс тестирования приносить учащимся удовольствие, быть увлекательным?
2. Может ли (обязательно ли) результатом тестирования стать моральное удовлетворение?
3. Можно ли (обязательно ли) публиковать результаты тестирования?

4. Может ли (должен ли) процесс тестирования носить соревновательный оттенок?
5. Следует ли ограничивать время тестирований? А если ограничивать, то, какое время – общее время работы или время выполнения конкретных заданий?
6. Какова оптимальная длительность сеанса тестирования в разных случаях?
7. Может ли (должен ли) процесс тестирования быть поучающим?
8. Является ли тестирование обязательным участником образовательного процесса? Какое время следует отвести тестированию в учебных планах? Как тарифицировать время тестирования и время подготовки тестирования для персонала и преподавателя?
9. Нужны ли (обязательны ли) психологические разгрузки в процессе тестирования? Следует ли организовывать эти разгрузки с помощью компьютера или компьютер на это время следует исключать (выключать)?
10. Может ли (должен ли) процесс тестирования опираться на личностно-ориентированную сюжетную линию?
11. Следует ли разрешать претенденту вносить исправления в уже выполненные задания?
12. Следует ли выделять претенденту лимит подсказок?
13. Следует ли указывать в процессе тестирования учащемуся на верно и неверно выполненные задания?
14. Может ли (должен ли) процесс тестирования быть мягким, позволяющим учащемуся самостоятельно выбирать план действий – последовательность выполнения заданий, возвращаться и исправлять результаты выполнения заданий?
15. Следует ли позволять претенденту видеть весь спектр возможных вариантов выполнения тестового задания одновременно?
16. Могут ли (должны ли) тестовые задания содержать элементы сюрпризности, юмора?
17. Какой сценарий тестирования эффективнее для образовательного процесса в целом – «серьезный» или «игровой»?
18. Какой тестирующий комплекс лучше – тот, где результаты претендентов растут с каждым новым сеансом, или тот, где результаты воспроизводятся неизменными от сеанса к сеансу?
19. Должен ли тест, будучи измерительным инструментом, быть одинаковым для всех или его следует адаптировать? И как потом учитывать то, что для разных претендентов тест был адаптирован по-разному?
20. Должны ли претенденты тестироваться «все сразу» или можно использовать режим «живой очереди»? Как это повлияет на реализацию дидактических целей и задач?

Ответы на эти вопросы весьма важны. Роль тестирований стремительно возрастает. От этих ответов, без преувеличения, зависит уровень будущих специалистов и, следовательно, будущее нашего государства.

Цель введения педагогического тестового контроля в образовательном учреждении можно сформулировать и так: повысить эффективность управления качеством образования. Эта сжатая формулировка объединяет многие аспекты влияния на кадровую политику, на вложения средств, на партнерские взаимоотношения и пр. Ответы на перечисленные вопросы могут зависеть от понимания того, что такое эффективность управления и в чем она выражается:

- в малом штате;

- в малых затратах;
- в высокой доходности;
- в скорости обучения;
- в соответствии учебного процесса утвержденным стандартам;
- в высокой вариативности;
- в доле учащихся перешедших на высшие степени обучения;
- в доле учащихся работающих по получаемой специальности.

Десять общих и тридцать частных требований к компьютерному тестовому комплексу

*... в отметках мало педагогики, в них больше власти,
отметки – это костыли хромой педагогики.*

Ш. А. АМОНАШВИЛИ

Компьютерный тестовый комплекс – это совокупность компонент:

- База тестовых заданий, из которых по какому-либо алгоритму собирается тест;
- компьютерная программа, реализующая сценарий тестирования и обеспечивающая процесс тестирования испытуемых в сети или на одном персональном компьютере;
- пакет компьютерных программ обеспечивающих сбор, статистическую обработку и представление результатов тестирования;
- компьютерная база данных, аккумулирующая результаты тестирования;
- необходимые для процедуры тестирования и не включаемые в традиционную офисную конфигурацию ПК устройства (аксессуары).

К тестовым комплексам предъявляют, следующие требования:

1. Эффективность (оптимальность):

- точность измерения;
- скорость измерения.

2. Дифференцирующая способность:

- объективность;
- адаптируемость к условиям.

3. Мобильность:

- скорость доставки и развертывания до готовности);
- количество, доступность, стоимость, скорость предоставления необходимых для сеансов и их обработки материалов и услуг (доступ к сети, размножение инструкций и справочных материалов, дополнительные аксессуары).

4. Дружелюбность интерфейсов:

- претендента;

- руководителя;
- обслуживающего персонала;
- автора;
- редактора.

5. Качество БТЗ:

- наполненность (количество заданий);
- полнота (соответствие имеющихся ключевых заданий всем разделам дисциплины);
- вариативность, фасетность (количество возможных вариантов);
- сепарабельность (отделимость заданий в соответствии с темами и условиями).

6. Диапазон измеряемых характеристик:

- широта диапазона;
- вариативность характеристик в соответствии с условиями.

7. Себестоимость:

- создания;
- наполнения;
- редактирования;
- валидации;
- обслуживания.

8. Привлекательность:

- процедуры измерения;
- интерфейса претендента.

9. Скорость и простота

- наполнения;
- редактирования;
- корректировки шкал и задания условий.

10. База результатов:

- представительность (наполненность, понятность) результатов;
- конвертируемость результатов в связи с изменениями системы оценки.

Таким образом, качество результата тестирования, т. е. степень соответствия его дидактическим целям и задачам, зависит от многих факторов.

03.01. Четыре группы недостатков интерфейса комплексов тестового контроля

*Русский человек всегда найдет способ
сократить любую дорогу на несколько минут с тем,
чтобы потом часами отмывать внедорожную грязь.
Стас ЯНКОВСКИЙ*

Перечислим четыре группы основных недостатков интерфейса известных компьютерных тестирующих комплексов.

1. Недостатки интерфейса учащегося:

- отсутствие свободы выбора режимов самообучения (недружественность);
- отсутствие соревновательного компонента (снижающее эффективность учебного процесса);
- отсутствие элементов сюрпризности и привлекательности (нудность);
- однообразие представления информации (вызывающее эффект «ровной дороги»).

2. Недостатки интерфейса составителя тестовых заданий

- сложность освоения;
- большие затраты времени на ввод информации и конструирование заданий;
- сложность ввода графической информации;
- необходимость технического сопровождения, т. е. присутствия дополнительного специалиста;
- отсутствие анимационных средств;
- сложность перевода тестовых заданий в электронный вид;
- сложность редактирования тестовых заданий.

3. Недостатки интерфейса экзаменатора

- сложность освоения;
- отсутствие средств оперативного контроля успехов учащихся;
- отсутствие средств интегрального наблюдения за успехами группы во времени;
- сложность сравнения результатов тестирования групп учащихся по разным дисциплинам.

4. Недостатки интерфейса руководителя учебного заведения

- сложность освоения;
- отсутствие средств систематизации и хранения результатов тестирований;
- отсутствие средств сопоставления успехов учебных групп по разным дисциплинам и во времени;
- сложность оперативного контроля успеваемости учащихся заведения в целом.

02.02.Полдюжины замечаний пользователю

*Все пчелы прилетали с медом,
а одна – такая маленькая и вредная – с дегтем.
Андрей КНЫШЕВ*

Практика показывает, что наименее всего люди, проводящие тестирования, обращают внимание на выполнение следующих требований:

- привлекательность процедуры тестирования, психологически правильное выстраивание инструктажа и пр.;
- дружелюбность интерфейса;
- представительность и конвертируемость результатов;
- подбор формы заданий (неправильная форма приводит к неправильному выражению содержания и пониманию смысла задания претендентами);
- соответствие инструкций форме и содержанию заданий;
- конвертируемость результатов.

Это, более всего, связано с бытующей среди «занятых» людей прагматичностью взглядов на необходимость следования дидактическим целям.

Если тестирования в образовательных учреждениях проводятся не эпизодически, и люди, которые их проводят, рассчитывают на дидактическую полезность мероприятий, то, кроме приведенных замечаний, абсолютно необходимо учесть следующие. Процедура тестирования станет привлекательнее и полезнее для системы образования, если придать ей элементы

- сюрпризности;
- соревновательности;
- публичности.

Проводите тестирования как длительную олимпиаду (четвертьфинал, полуфинал, финал и пр.). Маленькие и большие успехи претендентов должны быть сразу доступными всем учащимся, родителям, учителям. Ответственные работники администраций в обязательном порядке должны упоминать фамилии победителей тестирований в своих докладах и отчетах. Портреты лучших учащихся и их родителей должны висеть в актовом зале. Информация о прошедших тестированиях должна отображаться в газетах с обязательной публикацией таблиц рейтингов. Для того чтобы победителей было больше, поступайте, как в спорте – размножайте номинации: длительное, скоростное, тематическое тестирование и пр.

ПОМНИТЕ: образовательная олимпиада важнее для развития территории и государства, чем спортивная.

В спортивной олимпиаде побеждает сильный и ловкий, а в образовательной – умный и знающий.

02.03. Результаты компьютерной аттестации ЗУН и факторы влияния

*Весело входили в пасть змея Агхасуры
мальчики – пастушки, уверенные, что Кришна защитит их...
Источник вечного наслаждения, Шри Шримад*

Результатом процедуры компьютерной аттестации знаний (но не практических умений!) являются численные значения следующих четырех параметров обученности учащегося:

1. Знание общих сведений по дисциплине и умение среди них ориентироваться;
2. Знание определений, фактов, дат, названий, формулировок;
3. Знание способов решения и оформления решений типовых учебных задач, навыки рассуждений и доказательств;
4. Умение решать сложные учебные проблемы.

На основании значений этих параметров, для каждого учащегося строится рекомендуемая оценка по пятибалльной шкале и рейтинг. После обработки результатов строится таблица рейтингов учащихся в пределах группы. Рейтинг, обычно, – одно число, не дающее представления о частностях в структуре знаний претендента. Мы называем такой рейтинг скалярным. Более точно отразить ситуацию может векторный рейтинг, состоящий, например, из 4-х чисел, каждое из которых соответствует значению одного из означенных выше параметров обученности. Соответственно, предъявляемые студенту тестовые задания делятся на 4 группы.

Проведя тестирование тех же испытуемых несколько раз, перемежая его определенными и соответствующим образом обустроенными перерывами на отдых, с использованием того же (такого же) теста, можно получить дополнительную информацию, динамическую:

1. Способность испытуемого к самостоятельному обучению;
2. Способность испытуемого к адаптации в режиме тестирования;
3. Развитость коллективизма среди испытуемых.

Отправляя испытуемых на перерыв можно разрешать и не разрешать им читать учебники, общаться. Можно варьировать продолжительность перерыва. Можно перерыв устраивать в библиотеке, полной учебной литературы. Можно варьировать стимул. Вариантов перерыва, как видим, много и, соответственно, варьируются результаты повторных измерений. Это – не менее ценные результаты, чем те, которые получены в однократном измерении.

Иногда способность испытуемого к самостоятельному обучению ценится значительно выше, чем тот багаж ЗУН, который испытуемый принес на сеанс тестирования.

*Полиция в жизни каждого государства есть.
Козьма ПРУТКОВ*

После окончания сеанса тестирования проводятся мероприятия аттестационного блока. Среди них:

1. Претендент получает оценки своей подготовки по каждому из параметров, рекомендуемую интегральную оценку и рейтинг по 100-балльной (или 1000-балльной) шкале, а также перечень своих ошибок;
2. Преподаватель знакомится с характерными ошибками претендентов, сохраняющимися в соответствующей БД, корректирует параметры итоговой интегральной оценки;

3. Руководитель учреждения получает статистические данные, характеризующие результат в сравнении с аналогичными (предыдущими и параллельными);
4. Руководство территориального управления образованием получают интегральную сводку результатов, позволяющую оценить характеристики качества системы образования и пригодные для построения прогнозов;
5. В случае если такое же тестирование параллельно проводилось в ином регионе или за рубежом, например, оформленное в виде межрегионального или международного турнира, соответствующая комиссия сравнивает результаты, подводит итоги и награждает победителей в межрегиональном или международном масштабе.

Последний пункт здесь пока выглядит фантастично. Наш опыт вызова на турнир, т. е. параллельное тестирование, американскую сторону (образовательные учреждения штата Вашингтон, США) не увенчались успехом. Причин несколько:

- Мероприятие кто-то должен оплатить, а образовательные фонды не ведут деятельности в этом направлении;
- Американцы не любят соревнований ради соревнований, так же, как россияне;
- Отсутствуют прецеденты;

и др.

Результаты компьютерной аттестации должны освещаться в СМИ и стенной печати с обязательным указанием победителей и:

- набранных ими баллов;
- полученных ими наград и привилегий (от школы и вузов);
- их учителей;
- директоров их школ.

Общеизвестным должен быть и адрес центра, где в течение всего учебного года можно попробовать свои силы в порядке живой очереди.

Это значительно повысит популярность не только тестирований, но и образования в целом, как среди учащихся, так и среди их родителей. Не лишним будет отметить и лучших работников образования.

Согласитесь, образовательная олимпиада, с точки зрения развития региона и государства, не менее важна, чем спортивная. В спортивной олимпиаде побеждает сильный и ловкий, в образовательной – умный и знающий. В отсталых государствах культ силы всегда выше культа ума. Собственно, потому они и отсталые.

Три интерпретации результатов тестирований

Самым известным и научно обоснованным методом педагогического измерения является тест. К настоящему времени только в отношении тестов утвердилась рефлексивная норма обязательной проверки их качества. Это, пожалуй, самое существенное требование, выгодно отличающее тесты от экзамена и от остальных методов педагогического контроля.

В. С. АВАНЕСОВ

Выделяют 3 типа интерпретации результатов тестирования на языке тестологии:

1. Интерпретация результата тестирования предметно-педагогическая;
2. Интерпретация результата тестирования критериально-ориентированная;
3. Интерпретация результата тестирования нормативно-ориентированная.

Охарактеризуем их кратко, следуя В. С. Аванесову.

Интерпретация результата тестирования предметно-педагогическая – интерпретация с главной задачей: выяснить, какие элементы учебной дисциплины усвоены испытуемым. При этом анализируется выполнение большого количества заданий, сравнивается содержание аттестационных материалов с результатами тестирования, с принятыми правилами и делается вывод: можно ли аттестовать испытуемого на основании полученных данных.

Интерпретация результата тестирования критериально-ориентированная – возникает при решении задачи аттестации, приема в вуз, профотбора и т. п., где важно определить соответствие испытуемых заранее критерию. Для аттестации выпускников, им дают только задания из области минимально допустимой компетентности. Особое внимание обращается на определение зачетного уровня трудности. Содержание заданий принципиально облегчено. Задания должны выполнять все выпускники, допущенные к аттестации. Такие тесты нередко применяют органы управления образованием, стоящие перед необходимостью: в короткое время проверить состояние образования в большом количестве учебных заведений, и не позволить им опуститься ниже допустимого уровня. От испытуемых требуют выполнить небольшое количество заданий, чтобы определить – что испытуемый знает, и что не знает из заданного стандарта. Интерпретация результатов ведется работниками органов управления образованием и теми педагогами, на мнения которых управленцы опираются при аттестации.

Интерпретация результата тестирования нормативно-ориентированная – интерпретация с опорой на такие принятые в математике статистические показатели, как среднее арифметическое, процентильная норма и др. Тестирование без интерпретации результатов не имеет научного смысла. Главные вопросы такой интерпретации – не «кто что знает?», а «кто выше нормы и кто на каком месте?» В этом суть нормативно-ориентированной интерпретации тестовых результатов. Для ответа на эти вопросы не требуется тест большим количеством заданий. Нужно с использованием минимума заданий получить максимум дисперсии надежных тестовых баллов, рассчитать процентильные нормы и рейтинг.

Длительность сеанса тестирования

Примечательно, что время, уделяемое самкой воспитанию и обучению сыновей и дочерей, как и степень их социального взаимодействия, примерно одинаковы. То есть дело тут не в том, как учат, а в том, кто и как учится. Самки учатся быстрее, они более ориентированы на конечный результат.

А самцы – творчески и ... и разгильдяйски, что ли?

Elizabeth V. Lonsdorf, Linn E. Eberly, Anne E. Pusey. Sex differences in learning in chimpanzees. Nature. Vol. 428. 15 Apr. 2004. p. 715

Одна из основ тестологии – идея создания инструмента для быстрого и относительно точного оценивания характеристик больших контингентов испытуемых. Требование экономии времени естественно в массовых процессах. Направление современной организации тестового педагогического контроля – адаптация приводит к значительному сокращению времени сеансов. От времени сеанса существенно зависит качество результатов. Каждый тест характерен оптимальным временем сеанса, изменение которого снижает уровень ка-

чества измерений. Оптимальное время сеанса определяется эмпирически. Со временем сеанса связано количество заданий теста. Длительность тестирования ограничивается исходя из соображений:

- удобства процедуры тестирования;
- темпа накопления усталости испытуемых;
- достижения оптимального уровня дифференциации претендентов.

Большинство тестов, применяемых в педагогической практике, являются скоростными. Временной режим задается

- на весь сеанс тестирования;
- на субтесты;
- на отдельные тестовые задания.

Активно обсуждается проблема методов согласования заданий на скорость, необходимых для повышения надёжности тестов.

Известнейший тест «на время» – американский TOEFL. Ограничения во времени выполнения там вполне обоснованы. Известны скоростные тесты для водителей. Однако то, что хорошо для отбора пилотов и водителей, не всегда хорошо для аттестации ЗУН ученика. Современные концепции построения педагогического процесса требует индивидуализации всех звеньев обучения, в том числе и контроля.

Там, где тест напрямую не связан с функциональными ограничениями для видов работ, нет необходимости вводить фактор времени как лимитирующую характеристику оценивания деятельности испытуемого. При контроле трудовых навыков определение лимита времени целесообразно, так как выполнение трудовых операций имеет установленные нормативы. Время выполнения заданий важно при контроле практических умений в электротехнике и т. п. Но в таких дисциплинах, как история или литература, введение временных ограничений во многих случаях проблематично и для подобных тестов следует экспериментально установить такие нормативы, в которые могли бы укладываться и сангвиники, и флегматики – ведь у каждого из них свой индивидуальный стиль деятельности.

Важен фактор времени при проведении массовых тестирований – это вопрос экономии ресурсов и ограничения рамок исследования. Не менее он важен и при проведении зрелищных мероприятий – викторин, конкурсов – где используются тестовые подходы. Например – телевизионные шоу типа «Кто хочет стать миллионером?» и пр.

Наши многолетние эксперименты на тысячах школьников, студентов, учителей показали:

- **Школьники:** результаты 15-ти и 30-ти минутных компьютерных сеансов, где количества заданий относились как 1:2, относятся, в среднем, как 1:2. При увеличении же времени сеанса до 45 минут (с соответствующим увеличением количества заданий) приводит к ощутимому «завалу» зависимости на 15-20%. В экспериментах использовались гомогенные тесты с однотипными заданиями на общие знания и знания определений;
- **Школьники:** если тесты включают задания на решение численных либо графических задач, время сеанса должно быть увеличено до 2 часов. Однако при этом компьютер используется всего несколько минут – для регистрации и ввода результата. Компьютер только мешает испытуемому – шумит, ограничивает площадь стола, распространяет излучения. Поэтому такие задания лучше использовать в ином режиме – получить вариант, выполнить работу в спокойной обстановке в аудитории, а затем перейти в компьютерный класс и ввести результат;

- **Студенты:** после 40 минут сеанса тестирования общих знаний, навыков решения простых задач, знаний определений – включается эффект «ровной дороги» и большинство тестируемых начинают работать автоматически, не задумываясь, стремясь к скорейшему завершению сеанса. Стремление к лучшей оценке для них сменяется стремлением к тройке;
- **Учителя:** не выдерживают длительности сеанса больше 20 минут. Происходит срыв на обсуждение заданий, автоматическое «дотыкивание» теста и пр.;
- **Школьники, студенты, учителя:** Описанная выше ситуация коренным образом меняется при включении в сеанс тестирования элементов деловой игры, неожиданности, сюрпризности. Например, путем применения необычных для практики тестирований интерфейсных элементов и формулировки заданий в разговорном тоне с использованием «необычных для серьезной работы» словарных оборотов. Это позволяет значительно продлить период активности внимания, как у студентов, так и у школьников, и, следовательно, повысить точность измерений за счет включения большего количества заданий.

Таким образом:

1. При использовании комплексов с «прямоугольными интерфейсами» без элементов психологической разгрузки, при тестировании общих знаний и знаний определений и навыков решений простых задач ограничивать длительность сеанса
 - для учащихся 2-8 классов 10-20 минутами,
 - для учащихся 9-11 классов 20-30 минутами,
 - для студентов – 30-40 минутами;
2. При тестировании умений и навыков решения сложных численных и графических задач целесообразно основную работу проводить в учебной аудитории и использовать компьютер лишь для ввода результатов и регистрационных данных. В этом случае время работы можно ограничить несколькими часами, т. е. периодом наступления усталости контролирующего персонала. Можно даже устраивать большие перерывы, но с применением средств информационной безопасности – вариативность теста, контроль шпаргалок, ограничение общения и пр.;
3. Возможно, «золотая середина» здесь: тренировочные компьютерные сеансы, с использованием заданий, подобных контрольным, следует проводить без жесткого ограничения времени, а контрольные – ограничивать;
4. Следует шире внедрять в практику компьютерных тестирований элементы деловых игр и психологической разгрузки. Это позволяет значительно продлить время сеанса, предложить испытуемым больше заданий без опасения эффектов типа «ровной дороги» и, тем самым, поднять уровень надежности измерений и качества образования в целом.

Вопрос о связи интеллекта со скоростью выполнения операций при тестировании не прост и в настоящее время обсуждается психологами. Ежемесячно появляются новые научные и методические публикации на эту тему.

Количество заданий в тесте

Если ты уж вышел на сцену, постарайся, чтоб в зале кто-то был.
Михаил ЖВАНЕЦКИЙ

От количества заданий в тесте прямо зависит качество измерения, поскольку каждое задание позволяет оценить знание отдельного информационного блока учебной дисциплины.

В тестологии нет четких ограничений на количество заданий дидактических тестов: оно доходит в стандартизированных американских тестах до 200 и более, не ограничиваясь каким-то удобным для обработки числом (особенно с появлением и применением для тестирования компьютерной техники). Заданий нужно столько, чтобы по возможности полно отразить основное содержание диагностируемого объема знаний. (Е. А. Михайлычев, 2001).

Наш опыт показывает, что для осознанного выполнения более 40 (даже простых) заданий учащемуся необходимо иметь очень высокую мотивацию. Если сеанс измерения не приводит к возможности получения учащимся уважения, желанного рабочего места либо ощутимого приза – можно ожидать, что через 10-15 минут сеанса учащийся переключится на выполнение тестовых заданий методом «случайного тыка». Сохранить внимание и выполнить 100 заданий многим не под силу даже при очень высокой мотивации.

То, что сказано выше, относится к распространенным формальным процедурам тестирования, где применяются комплексы с «не отвлекающими» интерфейсами, демонстрирующими на экране однообразные прямоугольники с однообразными текстами. Иное дело – когда сеанс тестирования представлен в виде деловой игры с сюрпризами, действующими персонажами и пр. В этом случае удастся достичь значительного продления внимания и увеличить количество заданий в сеансе. Игровые технологии позволяют повысить уровень объективности измерений.

Спецификация теста

Спецификация теста является обязательной операцией эмпирического анализа теста и необходимым условием создания нового теста. Спецификация заключается в том, что строится таблица, в которой указываются:

- номера заданий и их принадлежность к субтестам;
- направленность заданий;
- сложность, трудоемкость, вес заданий;
- какую из характеристик каждое из заданий конкретно диагностирует.

В спецификации должны быть сформулированы диагностические цели теста. Спектр диагностических задач отражён в номенклатуре и типологии тестов и тестовых заданий.

При создании сложных дидактических тестов необходимо составить более подробную таблицу, в которой будут, наряду с указанной выше информацией, перечислены дидактические цели (усвоение формул и алгоритмов, понимание таблиц и графиков и т. п.), а также представленные в БТЗ разделы, параграфы, темы. Это прояснит, на какие диагностические цели тест ориентирован, и что он измеряет.

02.04. Артефакты

*Норма – это то, что встречается лишь изредка
Сомерсет МОЭМ, aforizm.kaminplus.ru*

Большое влияние на результаты тестового контроля оказывает так называемый эффект «привыкания». Он выражается в начале сеанса тестирования, где применяется тестовый комплекс с непривычным для претендента интерфейсом: замедление реакции, повышение суетливости, повышение вероятности арифметических и орфографических ошибок, неверные манипуляции. Это может быть связано и с тем, что претендент

- не знаком с инструкций и порядком проведения процедуры;
- участвует в тестированиях часто, но с использованием иных комплексов, и ему требуется некоторое время для освоения нового интерфейса;
- уже тестировался с использованием данного комплекса, но база тестовых заданий изменилась – создана другими авторами, оформлена в другом литературном стиле или содержит непривычные жаргонные выражения.

Исключить влияние «привыкания» на результаты тестирования несложно – нужно провести предварительное тестирование, дать претендентам освоиться. Согласно нашим измерениям, для этого достаточно от 5 до 10 минут.

Кроме этого, результаты могут быть искажены по приведенным ниже причинам.

- Группа претендентов предварительно ознакомилась с заданиями теста. Это может произойти случайно, когда им просто «попалась» книга с подобной информацией или репетитор провел тренировку, опираясь на сыгравшую нужную роль интуицию, и намеренно, когда произошла утечка информации. В обоих случаях на кривой распределения результатов тестирования может «вырасти» удивительный «горб». Методы дезавуирования этого артефакта известны и приведены ниже в главе посвященной информационной безопасности.
- Тест не вариативен и правила допускают повторные сеансы. В этом случае, претендент, прошедший сеанс два-три раза, да с основательными перерывами на общение с книгой и преподавателем, находится в выигрышной ситуации. Кривая распределения «оживет», с каждым сеансом в конце ее будет увеличиваться «горб». Для устранения перекоса, следует поставить всех претендентов в равные условия: все тестируются вместе, варианты теста меняются, желательно с применением процедур случайного выбора. В росте «горба» должны иметь возможность принять участие все претенденты. Такой многосеансный «марафон» может принести и положительные плоды, естественным образом дифференцировав претендентов на «упорных», «безразличных» и «среднячков».
- Среди претендентов оказалась достаточно большая группа, для которой этот тест не предназначен. Это могут быть ученики специальной школы «с уклоном», или, наоборот, ученики школы, где данная дисциплина не изучалась. В первом случае «горб» на кривой распределения вырастет в ее конце, в области высших рейтингов, во втором – в начале, в области «случайного тыка».
- Среди претендентов есть неспособные выдержать достаточно длительное напряжение в течение сеанса. Увеличение длительности сеанса действует угнетающе на всех претендентов и по-разному на степень объективности их результатов. А для обеспечения требуемого уровня качества измерений необходимо включение в тест достаточно большого количества заданий, для выполнения которых необходимо

достаточное время. Для устранения этого противоречия есть два пути: применение технологий адаптивного тестирования, позволяющих оптимально распорядиться временем сеанса и применение игровых технологий, позволяющих за счет видоизменения интерфейса повысить привлекательность процедуры.

- В формулировках заданий допущено использование жаргонных, иноязычных либо специфических для какой-то группы испытуемых выражений, которые повышают вероятность неоднозначного восприятия условий. Скажем, автор этих строк помнит экзаменационный случай, произошедший во Владивостоке в 80-х годах, когда заминка в решении задач случилась потому, что учащийся не знал отличий сосиски от сардельки. Он их никогда не видел.
- В кадре заданий встретились неоднозначно воспринимаемые претендентами активные элементы, неадекватно описанные в инструкции, а претендентам поставлено жесткое условие – ничего ни у кого не спрашивать под угрозой прекращения сеанса. В такой ситуации происходит столкновение безграмотности персонала с безалаберностью составителей теста, реально влияющее на распределение результатов.
- Специалист, проводивший тренировочные занятия, не вполне адекватно воспринимает дисциплину и, в процессе тренировки, допустил презрительные замечания в адрес составителей, свои комментарии о качестве заданий, свои фантазии относительно верных и неверных вариантов. Если этот специалист пользуется достаточным авторитетом среди учащихся, он нанесет такими действиями урон знаниям своих подопечных и объективности результатов тестирования. Бороться с этим трудно, такая ситуация случается, например, когда преподаватели всю ответственность за тренировку перекладывают на инженера компьютерного класса, говоря: «я все равно там ничего не понимаю». Для деавуирования влияния этого артефакта нужно поступать по правилу: в процессе тренировок в классе обязательно должен находиться опытный преподаватель дисциплины.

К факторам, ставящим под угрозу чистоту тестовых измерений, относят (Михайлычев Е. А. *Дидактическая тестология*. М.: Народное образование, 2001. 432 с.):

- реактивный эффект – уменьшение или увеличение восприимчивости испытуемых к экспериментальному воздействию под влиянием предварительного тестирования;
- взаимодействие выборки и экспериментального воздействия;
- условия организации эксперимента, вызывающие реакцию испытуемых, которая не позволяет распространить полученные данные о влиянии на лиц, подвергающихся такому же воздействию в не экспериментальных условиях;
- интерференция экспериментальных воздействий, возникающая, когда одни и те же испытуемые подвергаются нескольким воздействиям, поскольку влияние более ранних, как правило, не исчезает.

Рассмотрим эти факторы подробнее, следуя Е. А. Михайлычеву:

Реактивный эффект. Проявляется в двух вариантах.

Первый вариант. Учащиеся тренировались на прохождение тестов по-разному. На тренировки ходили не все, а для тех, кто тренировки посещал, различные преподаватели акцентировали внимание на тестах разного вида. При валидации абитуриентских тестов без учета этого фактора эффект гарантирован. Можно контролировать эффект, если:

- предварительно опросить учащихся о тренировочных занятиях;

- параллельно проводить визуальное наблюдение за поведением претендентов и сопоставлять данные, собираемые по программе такого формализованного наблюдения, с данными анкетирования. Не исключено использование видеокамеры при условии, что она будет хорошо замаскирована (иначе получим дополнительный «визит-эффект» с непредсказуемыми последствиями – от жалоб до истерик).

Это позволит сделать и повысить точность поправки.

Второй вариант. Имеется хорошее психодиагностическое сопровождение, настраивающее обследуемого на успешную работу с дидактическим тестом. Вариант сделать такое сопровождение для учащихся привычным – заманчив. Но в контрольных группах возможен отказ от применения сопровождения, что влияет на результат. После стандартизации теста – его валидизации, строгой проверки надёжности, когда будут просчитаны все основные поправочные коэффициенты (в том числе на реактивность к психодиагностическому сопровождению по данным контрольных групп или ретеста, параллельной формы теста), психодиагностику можно будет спокойно подключать к дидактическому тесту и использовать как стимулирующие средства обучения. Ведь с позиции гуманистической педагогики главная диагностическая задача – не поймать учащегося на незнании, а помочь ему усвоить материал, развить мышление и способности.

Взаимодействие выборки и экспериментального воздействия. Здесь опасность искажения результатов заключена в том, что экспериментальная и контрольная группы, на которых валидируется тест, существенно отличаются от выборочной совокупности, для которой он предназначен.

Первый вариант. Суть проблемы – в репрезентативности выборки. Особенно если выборка состоит из добровольцев. Их демографические характеристики, успеваемость, психологические особенности можно просчитать и учесть. А как быть с теми, кто отказался быть добровольцем при адаптации теста? Насколько их личностные характеристики адекватны участникам эксперимента?

Второй вариант. Рассмотрим реальную ситуацию. Тест в вузе проходит адаптацию на контингенте четырёх групп с двух факультетов, в то время как на этих факультетах имеются лишь по 4 группы на том же курсе, а всего факультетов в вузе 5; вузов же аналогичного профиля 10. Вопрос о результативности и обоснованности мотивации отбора групп для участия в эксперименте неоднозначен. Смещения выборки повлияют на результатах валидизации, и исследователь рискует получить только данные о локальной валидности теста именно для тех групп, которые участвовали в эксперименте. При валидизации теста в вузе необходимо либо стремиться не к подбору удобных для экспериментаторов послушных кураторских групп, в которых работает сам разработчик теста, а к возможно более репрезентативной выборке, либо проводить массовый эксперимент на всём контингенте курса, изучающего в это время данную дисциплину по данной программе. Если по этой же программе работают в других вузах, то на репрезентативных выборках желательно организовать выверку теста, строго соблюдая внешние условия процедуры, а затем учесть эти данные при сопоставлении результатов и подсчёте валидности.

Третий вариант. Возможен отсев обследуемых по ходу эксперимента из экспериментальной и контрольной групп. Если эксперимент тщательно подготовлен и сбор данных проходит быстро, то такая опасность для валидизации практически ничтожна.

Негативная реакция испытуемых на эксперимент. Реакция испытуемых на эксперимент может быть вызвана любым аспектом процедуры. Например, самым фактом приглашения к участию в эксперименте, которого многие сокурсники не получили. Таинственное появление незнакомых экспериментаторов возбуждает учащихся и повышает вероятность стрессовых ситуаций. Рекомендуется замаскировать эксперимент различными средствами: например, не оповещать о нём, если он внешне похож на привычные мероприя-

тия. Наиболее сложная проблема – обеспечение случайного отбора для получения репрезентативной (по отношению к контингенту, для которого предназначен тест) выборки. Если эксперимент предусматривает индивидуальное обучение (парную работу диагноста с испытуемым), то в разведении класса, учебных групп по аудиториям (экспериментальная – налево, остальные – направо) необходимости нет. При бланковом варианте дидактического тестирования этого избежать можно, но затраты времени возрастут во столько раз, сколько будет участников эксперимента. Возможно, постепенно сократить необходимое количество самих экспериментаторов – они станут быстро инструктировать и оперативнее реагировать на обратную связь. Но это может стать побочным фактором: так, первого испытуемого инструктировали 10 мин, пятидесятого – 2 мин, последнего – скороговоркой. Иногда для смягчения воздействия фактора предоставляют учащимся записываться в вузе на различные факультативы, читаемые в разных помещениях в одни и те же часы. Контроль записи позволяет составлять по этим данным эквивалентные экспериментальные группы и нетрадиционно использовать процесс естественной рандомизации. Наилучшим решением считается использование в качестве единицы рандомизации целых классов – чтобы и экспериментальная, и контрольная группы состояли только из случайно образом отобранных. Следует также обращать внимание, как претенденты принимают саму ситуацию тестирования и задачи разного типа. Игнорирование отношения испытуемых к ситуации эксперимента и степени принятия ими задачи, обычно формулируемой в инструкции, ведет к ошибочной интерпретации результатов и к низкой толерантности разрабатываемых методик. Вполне вероятно, что учащиеся, привыкшие к закрытым тестовым заданиям, будут хуже выполнять их, чем те, кто не имел дела с тестами. Надо знать аудиторию.

Интерференция экспериментальных воздействий. Под интерференцией в психологии понимается ухудшение сохранения в памяти материала. В контексте экспериментальной терминологии под интерференцией имеются в виду незафиксированные в эксперименте наложения факторов. Интерференция воздействий следует предвидеть и учитывать при анализе результатов выборки (рандомизации) учебных групп, участвующих в эксперименте. Например, проведение эксперимента в одном из выбранных трех вузов может совпасть с внедрением именно в этом вузе новой системы стимулирования успеваемости. Это существенно повысит их интерес к тесту как средству диагностики и корректировки собственных знаний. Или же, среди групп одного из факультетов, отобранных для апробации теста, будет активно распространяться информация, что созданное в городе совместное предприятие с участием крупной иностранной фирмы собирается оплатить учебу за границей победителям химической олимпиады, или же что ректорат решил резко сократить контингент студентов в связи с отсутствием заказов на специалистов, и неудачники в тестировании могут быть отчислены. Естественно, в первых случаях неожиданный позитивный интерес к тесту экспериментаторов вызовет недоумение, а последний будет способствовать росту стрессовых ожиданий, что отразится и на экспериментаторах. В любом случае предварительный опрос кураторов, доброжелательная ориентировочная беседа для выяснения возможности влияния таких факторов абсолютно необходимы, равно как и гарантия того, что никаких негативных селекции для неудачников не будет и т.д. И, конечно же, обязательна фиксация всех «подозрительных» внешних факторов в отчете с выдвижением и последующей проверкой гипотез об этом воздействии, его причинах и возможных последствиях.

02.05. Динамика результатов педагогических измерений

Наука и религия обязаны принимать догматику.

*Этой ценой покупается возможность
получать ответы на поставленные вопросы.*

Ю. ШРЕЙДЕР "Заметки о философии"

Этот параграф посвящен дискуссии о целесообразности некоторых понятий и действий тестологии. В качестве предмета обсуждения мы выбрали понятия «надежность» и «вес задания», а в качестве действий – их измерение.

Прочитав методическое пособие по тестологии, каких немало, начинаешь верить, что одна из главных характеристик педагогического теста – надежность. Тестологи считают надежным (то есть, в просторечии, хорошим) тестом тот, результаты прохождения которого не меняются (или мало меняются), сколько бы раз претендентов не тестировали. Как в технике: хороша та линейка, которая не меняет толщину кирпича в процессе измерения. Ниже мы покажем, что высокая надежность – это как раз тот идеал, которого нельзя достичь в педагогических измерениях в условиях массовых тестирований.

Основной метод измерения надежности – ретестинг, дублирующее тестирование, которое проводят в течение одного или двух сеансов.

В течение одного сеанса повторное тестирование устраивают путем:

- либо дублирования заданий в составе теста, представление их в разной форме и разными формулировками;
- либо деления теста пополам и принятия утверждения, что половинки эквивалентны.

Если повторное тестирование устраивают в два сеанса, то применяют в обоих случаях один тест либо тесты с подобными, но измененными заданиями (варианты теста). При этом группы претендентов могут быть разными, но должны быть «статистически равными». «Статистическую равность» устанавливают путем предварительного опроса либо предварительного тестирования.

При повторном тестировании в течение одного сеанса, учащиеся часто проявляют смекалку и, наперекор изобретательным составителям теста, находят дубли – одинаковые задания. Дальше они поступают так: если есть уверенность в точности решения, оба раза выбирают один (верный на их взгляд) вариант. Если нет – выбирают разные варианты. Этот простой алгоритм, применяемый мотивированными претендентами, влияет на чистоту эксперимента однозначно: значение «надежности» завышается.

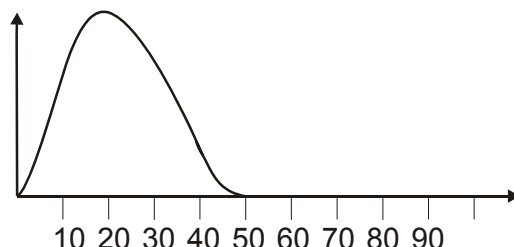
Как бы не старались экспериментаторы соблюсти чистоту эксперимента при тестировании в два сеанса:

- учащиеся почему-то вдруг начинают консультироваться с преподавателем, читать учебники и повышать свои знания в перерывах между сеансами;
- учащиеся выходят после сеанса и рассказывают стоящим в очереди «новичкам» смысл заданий и вероятные решения;
- учащиеся почему-то учатся в разных школах и невозможно предсказать заранее, чему их там учили лучше или хуже;
- учащимся может все надоесть, и они начнут тестироваться методом «тыка» (кстати, самый лучший метод получения надежных результатов).

Технология тестирования, тот антураж, который создается вокруг сеансов, обязательно приводит к усилению и ослаблению мотивации претендентов к успеху. По-другому не бывает, ведь тестирование – образовательный акт. Мотивация играет важнейшую роль в измерениях надежности. Мотивация непосредственно влияет на величину коэффициента надежности.

Если претенденты не мотивированы – результаты измерения сконцентрируются у точки случайного «тыка» области определения результатов тестирования:

Рис. 1.



Сколько бы сеансов не проводил экспериментатор – результаты там и останутся. Это – тривиальный вариант; здесь, после измерения, надежность теста получится очень высокой.

Если претенденты мотивированы, во втором сеансе средний результат будет выше. Таковы люди. Изменить их трудно. Они – не машины. Одинакового результата не получается. Остановить изменения результатов не удастся, каким бы прекрасным не был используемый тест. Проведите третье тестирование, и Вы увидите – результаты «плывут», причем, для разных претендентов – в разные стороны. Чем выше мотивация претендентов – тем ненадежнее тест. Надежных, в смысле прямого следования определению, дидактических тестов не бывает.

Парадокс: чем выше дидактические качества технологии тестирования, чем более она усиливает мотивацию претендентов к дополнительному обучению, тем она ненадежнее. Тестирование, как часть учебного процесса, должно нести дидактическую нагрузку. Какую технологию тестирования выберет педагог: усиливающую или снижающую мотивацию учащихся?

*Если какая-нибудь неприятность
может произойти, она случается.
Артур БЛОХ. Закон Мэрфи*

И что же – не бывает хороших, в смысле традиционного понимания надежности, тестов? И тестология – не наука? Или мы не правильно воспринимаем понятие «чистоты эксперимента»? А может, тестологическое понятие «надежности» просто не адекватно действительности?

Давайте рассмотрим определение надежности, данное В. С. Аванесовым [Композиция тестовых заданий. Учебная книга. 3 изд., доп. М.. Центр тестирования, 2002г. -240с.]:

Надежность теста – показатель точности и устойчивости результатов измерения при его многократном применении. Надежность теста тем выше, чем «одинаковее» результаты его применения при тестировании однородных групп претендентов.

Теперь рассмотрим обсуждение надежности, данное тем же автором (текст приводится в сокращении):

Надежность характеризует степень адекватности отражения тестом соответствующей генеральной совокупности заданий. Раньше предполагалось, что мера надежности является устойчивой характеристикой теста. На надежность сильно влияет степень гомогенности

групп испытуемых, уровень их подготовленности, а также другие факторы, связанные не столько с тестом, сколько с условиями его проведения. Поэтому в последние годы стали чаще писать о надежности измерения, имея в виду тест как результат тестирования в конкретно определенных условиях, а не общую характеристику теста как метода.

Подробно ситуацию с определением надежности теста в тестологии описал Е. А. Михайлычев [*Дидактическая тестология. М.: Народное образование, 2001. – 432 с.*].

Оказывается, недоверие к понятию «надежность» давно существует в среде тестологов. Так нужна ли она – «надежность»? Может, она должна уступить место другой характеристике, более адекватно отображающей свойства теста?

При внимательном рассмотрении тестологических работ можно заметить, что сами значения «коэффициента надежности», тщательному измерению которых посвящается так много сил, после измерения и вычисления нигде реально не применяются. Используется лишь мнение, порожденное в процессе измерения – «этот тест лучше, а тот – хуже». Но разве нельзя для упрочения мнения придумать другие, менее спорные критерии?

Понятие «надежность», оказывается, не одиноко, в смысле нашего критического рассмотрения. При повторении сеансов, так же, как значения коэффициента «надежности», «плывут» и значения весов заданий. При этом даже не важно – те же претенденты пришли на повторный сеанс, или другие. Важно лишь, чтобы сеансы были последовательными, а претенденты имели возможность общаться.

Весы заданий занимают одно из центральных мест в парадигме IRT. С помощью значений весов темперируют тест, упорядочивают задания по критерию сложности.

Рассмотрим реальную ситуацию. Возьмем тест, составленный в полном согласии с канонами учебной дисциплины и тестологии. Возьмем две группы претендентов, которые прекрасно подготовлены по двум разным учебникам этой дисциплины соответственно. Пусть группы будут разной численности. Разные учебники потому и разные, что разные вопросы дисциплины там освещены по-разному, в том числе и по-разному несовершенно. Следовательно, часть заданий гарантированно правильно выполняют претенденты из первой группы, другую часть – из второй. Тогда, вычисленный после сеанса вес каждого задания окажется зависимым от соотношения численности групп. Сколько бы мы не рассуждали о «репрезентативности», от этой определенности никуда не денешься. Чем меньше претендентов выполнивших конкретное задание, тем выше его вес. Следовательно, при условии равновероятности соответствия заданий тому и/или иному учебнику, вероятность победы выше у тех претендентов, которые принадлежат меньшей группе.

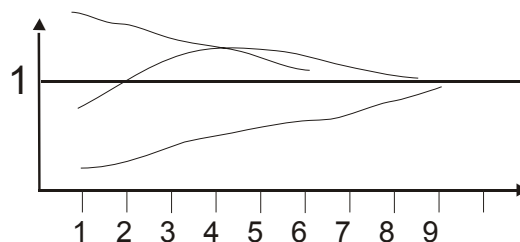
Выбор учебника для учения – дело субъективное. Получается, что значения весов также субъективны. Но тогда зачем тратить столько сил на расчет этих значений?

Период вычислений весов долог. Пока пройдут тестирования сотни и тысячи учащихся проходят недели и месяцы. Но это не вся трудность. За это время в учебных заведениях многое меняется, появляются новые учебники, меняются учебные программы, сами учебные дисциплины становятся другими. «Плывет» основательность знаний учащимися тех или иных разделов дисциплины. При всем этом, естественно, должны меняться и базы заданий. Вместе с базами, должны измениться и значения весов заданий. И это – еще не все аргументы.

В тестологии разъясняют такую ситуацию, т. е. «плывущие» результаты измерений, «нерепрезентативностью» выборки претендентов. Это можно было бы так и оставить, если бы не видны были явные закономерности «заплывов» этих чисел. Можно утверждать: если претенденты мотивированы, а временные интервалы между сеансами достаточно велики и претенденты имеют возможность общения, то после нескольких сеансов мы увидим -

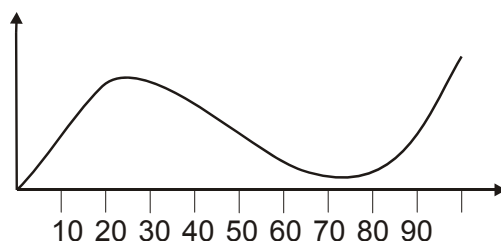
А) веса заданий стремятся стать равными:

Рис. 2.



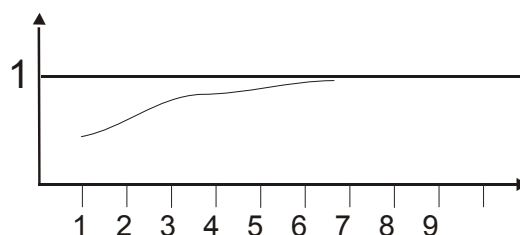
Б) результаты претендентов имеют тенденцию к группировке в двух районах области определения кривой распределения – рядом с точкой случайного «тыка» и рядом с пределом 100%:

Рис. 3.



В) надежность теста растет и постепенно становится самой высокой:

Рис. 4.



Веса заданий, по изначальному замыслу, вычисляются для мифического «усредненного» претендента. Ситуация с этим вычислением напоминает притчу о больнице, где в качестве показателя успешности работы вычисляли среднюю температуру больных, и средний больной был здоров ($T_{\text{ср}} = 36,6^{\circ}$). Вместо того чтобы сравнить результаты учащихся двух школ между собой и прямо определить реального лидера, тестологи, в рамках действующей парадигмы, сравнивают их результаты с результатом «среднего ученика» (который не только не известен, но еще и «тестировался» в прошлом году). В итоге может случиться, например, ситуация, когда некое, легкое для «среднего ученика» задание, реальные, учившиеся у конкретного учителя, школьники не выполнили, но это мало повлияло на их итоговый результат. И наоборот.

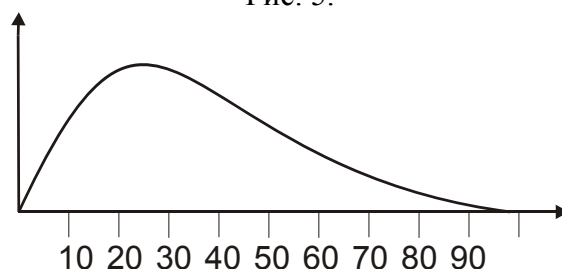
Некоторые тестологи, видя, как «плавают» значения коэффициентов, придумали выход для «укрепления» теории: взять и запретить повторные тестирования. Так и случилось в первых экспериментальных всероссийских тестированиях «Телетестинг», проводившихся Центром «Гуманитарные технологии» (г. Москва). Но школьники оказались хитрее ученых мужей – они запоминали задания и передавали их, вместе с решениями, своим коллегам, ожидавшим сеанс в очереди в коридоре. «Впереди танков по минному полю шли тральщики». Каждый последующий сеанс проводился с более подготовленными претендентами. И веса поплыли. Если бы их значения вычислялись каждые два часа и по всей стране, мы бы увидели чудесные метаморфозы со школьниками, умнеющими на глазах, и с одними и теми же тестовыми заданиями, которые, в течение дня, становятся все проще и проще. Более того, мы бы увидели, что качество образования в России чудесным образом

зависит от часового пояса (отгадайте, почему). Был ли смысл в скрупулезном расчете весов, длившемся несколько недель?

Давайте рассмотрим динамику результатов многократного прохождения одного и того же теста одной и той же группой претендентов. Для того чтобы картина не стала тривиальной, пусть тест будет вариативным, т. е. его задания каждый раз будут меняться, но спектр их типов будет сохраняться. Пусть веса заданий вычислены заранее так, как это требует теория, и не меняются.

После первого сеанса, если тест «достаточно» хорош, а группа претендентов «достаточно» представительна и велика, согласно канонам тестологии, должно получиться следующее распределение результатов:

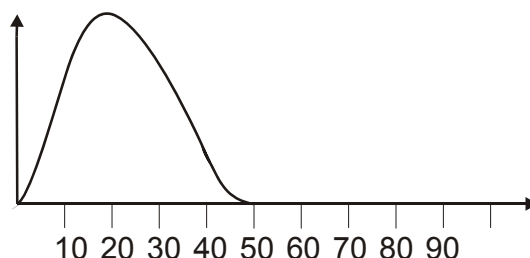
Рис. 5.



Каноны тестологии (теории IRT) требуют, чтобы вершина у кривой распределения была одна. Начинаться она должна от нуля, затем переваливать через вершину и плавно снижаться к нулю на границе 100%. Так оно иногда и происходит. Если в группе претендентов 300-400 человек и более, претенденты мотивированы, задания заранее не известны, свойства претендентов более-менее однородно распределены, то график получится более-менее плавным, как на этом рисунке (Рис. 5.).

Бывает и иначе. Скажем при тестировании по какой-либо коллективно «нелюбимой» дисциплине или при отсутствии у претендентов мотивации:

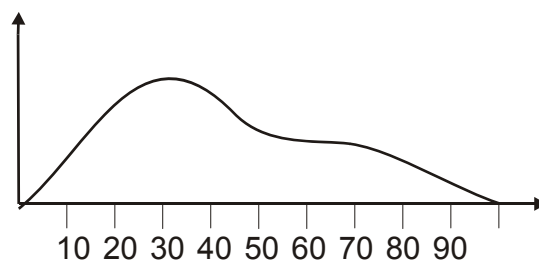
Рис. 6.



Здесь вершина распределения расположена над неким средним значением, которое получится, если проходить тестирование методом случайного «тыка». Ширина кривой в этом случае зависит от свойств самого теста и может быть вычислена по стандартным формулам теории ошибок.

А бывает и так, что в группу претендентов попали учащиеся с существенно разным уровнем и спектром ЗУН. Скажем, когда школьники гуманитарного и физико-математического класса одновременно тестируются по тригонометрии. Кривая может стать двугорбой:

Рис. 7.



Горб, что поближе к нулю, соответствует гуманитариям, а другой, подальше – физикам и математикам. Горбы могут быть разделены резко, а могут и слиться в широкую вершину. Степень различения горбов зависит от свойств теста и учебной программы, которую выполнили претенденты.

Можно предположить и другие ситуации, когда кривая распределения результатов будет многогорбой. Тестологи, как правило, «отмечают» такие случаи, говорят, что выборка не представительная, претендентов мало и вообще – претенденты попались не те, что надо.

Причина «любви» тестологов к одногорбым кривым лежит у истоков IRT – одной из распространенных теорий, описывающих результаты тестирования. Именно с IRT они «впитывают» «технарские» идеи о надежности, репрезентативности и пр., именно подружившись с IRT, тестологи начинают жить в фантастическом мире «надежностей» и «репрезентативностей». В этом – истоки сюрпризов, которые преподносит тестологам обычная «серая» повседневность.

*Пользователь не знает, чего он хочет,
пока не увидит то, что он получил.*

Э. ЙОДАН

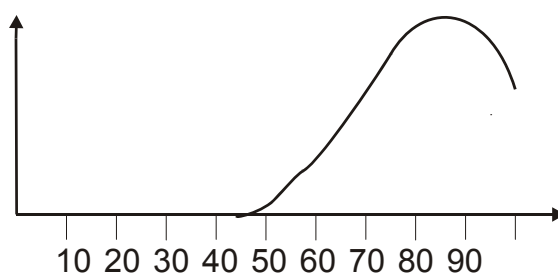
Реальность богаче теоретических представлений. Редко кому придет в голову проследить, как меняется кривая распределения результатов с течением времени, с каждым новым сеансом. Это – интересно.

Происходящие сдвиги и их скорость зависят от следующих причин:

- степени внешней мотивации (если претенденты – абитуриенты престижного вуза);
- степени внутренней мотивации (если претенденты следят за своими личными успехами, стараются развивать свой интеллект, стремятся к знаниям);
- под влиянием мотивации учащиеся оперативно (или не оперативно) консультируются с учителем, читают учебники в перерывах между сеансами либо вечером дома;
- при отсутствии мотивации учащимся с течением времени может все надоесть.

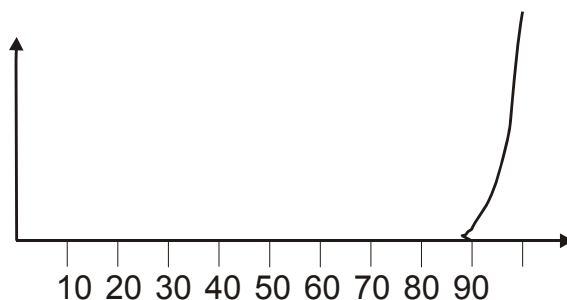
В первом и втором случае, через 3-5 сеансов кривая может стать такой:

Рис. 8.



А потом, еще после нескольких сеансов, результаты большинства претендентов окажутся сконцентрированными у границы 100%:

Рис. 9.

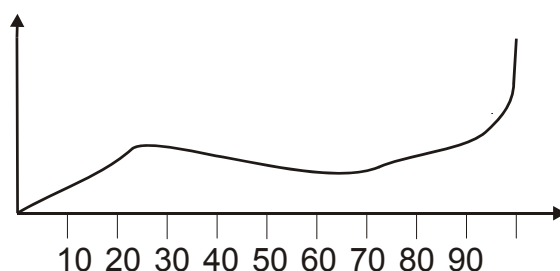


Претенденты – разные. Неоднородность группы претендентов может проявиться в динамике кривой распределения результатов так:

- Претенденты, потерявшие интерес к тестированию, не имеющие достаточной мотивации и просто уставшие покажут результаты вблизи уровня случайного «тыка»;
- Результаты сильно мотивированных претендентов, активно выяснявших методы выполнения заданий и правильные решения, окажутся через 2-3 сеанса вблизи отметки 100%;
- Результаты мотивированных, но слабо подготовленных к стрессовому дообучению претендентов будут медленно расти, двигаться к отметке 100%.

Картину распределения тогда можно представить таким рисунком:

Рис. 10.

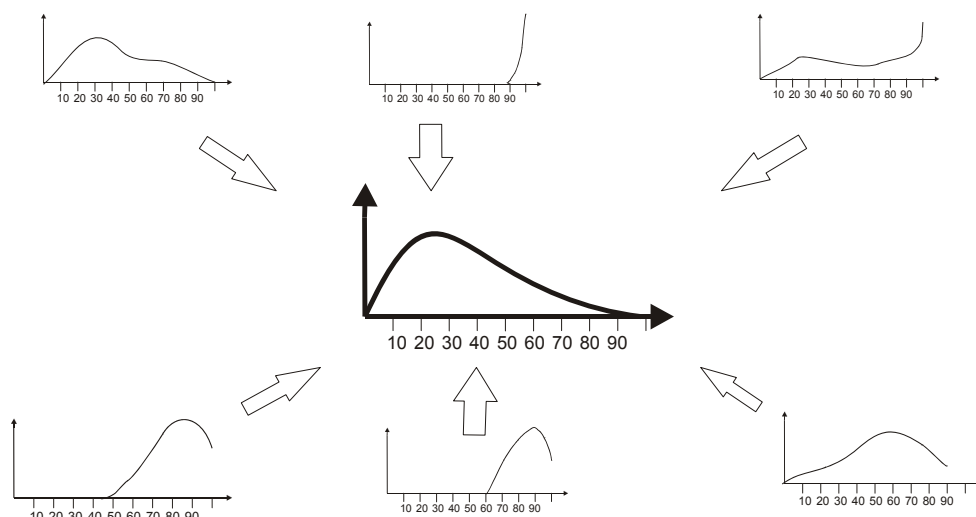


Конечно, эта кривая совершенно не похожа на те, что изучают поклонники IRT.

Рассмотрение, которое мы провели, будет таким при условии оговоренного нами выше постоянства весов заданий.

В тестологии существуют хорошо разработанные методы пересчета весов на основании полученных распределений результатов. Представленные выше кривые (Рис. 6-10.) могут быть трансформированы к каноническому виду (Рис. 5.) путем «перевзвешивания», пересчета значений весов:

Рис. 11.



Если после каждого сеанса тестирования приводить кривые распределения результатов к каноническому виду путем «перевзвешивания», то, вместе с «плывущими» кривыми распределения результатов (теперь – медленно плывущих), мы получим «плывущие» веса.

Из всех неприятностей произойдет именно та, ущерб от которой больше.

Артур БЛОХ. Третье следствие из Закона Мэрфи

Несложно предположить, куда «плывут» веса заданий:

- «разочаровавшиеся» претенденты нажимают клавиши случайно и, в соответствии с законами случайности, все задания будут выполнять ими, правильно и неправильно, равномерно;
- «стремящиеся к успеху» претенденты, рано или поздно, узнают все верные варианты выполнения заданий и выполняют все верно;
- «середняки» вносят разноречивую в описанную равновероятность, однако их количество постепенно уменьшается за счет продолжающегося самообучения.

Следовательно, веса «плывут» друг к другу, постепенно становясь равными.

На процесс «плавания» могут влиять артефакты, скажем такой: одна часть участников поверили своему учителю, научившему их ответам на вопросы теста, другая часть – соседу студенту, который решил задания по-своему. В итоге, совершенно случайно, часть заданий частью претендентов будет в течение нескольких сеансов выполняться неверно, то есть окажется очень сложной для совершенно конкретного и неизменного количества претендентов.

Может быть, мы просто не замечаем «объективной реальности»? Может, человек не машина, и к нему нельзя подходить с «технарскими» мерками и определениями? Может, лучшая тестовая технология как раз та, которая прекрасно выполняет дидактическую функцию, где результаты претендентов растут, и, чем быстрее – тем лучше? Или та, где веса заданий, многократно пересчитываемые при многократных тестированиях, становятся равными быстрее? Ведь главная цель дидактики как раз в этом и состоит – научить учащихся путем тренировок. Чем быстрее растут результаты – тем лучше применяемая технология (тест + программный комплекс + сценарий сеанса) приспособлена для тренировок, тем выше ее дидактическая ценность.

Свойство, отражающие скорость роста результатов претендентов, вполне можно назвать **релевантностью**, т. е. степенью соответствия теста дидактическим целям. А соответствующий коэффициент – **коэффициентом релевантности**.

Для повышения надежности тестологи рекомендуют провести предварительный опрос претендентов. Можно ли поверить, что большинство «претендентов» правдиво ответят перед сеансом на вопросы:

- Не знает ли он из каких-либо источников верных вариантов выполнения заданий?
- Не проходил ли он это тестирование ранее и сколько раз?
- Будет ли он проходить тест как следует, или просто потыкает в клавиши?

Нет. Они не заинтересованы в правдивом ответе по разным причинам. Такой опрос имеет лишь умозрительный интерес, его результату нельзя доверять.

Человек – не машина. Результат измерения свойств человека зависит от применяемого инструмента и самого процесса измерения. А свойства человека меняются в зависимости от количества и последовательности измерений. Следовательно, теряется смысл скрупулезного взвешивания сложности и трудоемкости заданий применительно к группам претендентов с неизвестной предысторией.

Есть ли смысл в изменении свойств инструмента измерения применительно к разным группам претендентов? Да. Во всяком случае, физики часто так поступают, измеряя свойства разнородных веществ. Но при этом должны быть адекватными процедуры пересчета результатов, чего, как мы видим, добиться довольно сложно в массовых тестированиях.

*Интуитивно ясно, что наличие ошибки в программе
не скажется на результатах тестирования;
если содержащая ошибку программная компонента
при тестировании не выполнялась.*

Дж. ХУАНГ

Измерения надежности в тестологии напоминают квантовомеханические эксперименты, где прибор неизбежно взаимодействует с системой в процессе измерения и меняет систему. И чем точнее прибор – тем сильнее он систему меняет. Физики наработали неплохой опыт, и им нужно пользоваться. Например, понятия дуальности характеристик и соотношения неопределенностей, вероятно, могли бы сослужить хорошую службу тестологии.

Подобно квантовой механике, тестологическая теория должна объединить в рамках изучаемой системы и тест, и технологию тестирования, и проходящих сеанс тестирования претендентов, и тех претендентов, которые ожидают своей очереди, и окружение претендентов (учителей и иных советчиков, продавцов шпаргалок и пр.). Нельзя адекватно описать поведение урезанной системы, рассматривая только совокупность невзаимодействующих претендентов и теста.

Поскольку сложность и трудоемкость тестовых заданий необходимо учитывать при подведении итогов тестирований, а существующие понятия «надежности» и «веса задания» не выдерживают критики, надо от парадигмы, связанной со странной «надежностью» и частым «перевзвешиванием» заданий, перейти к иной, более объективной и более адекватно соответствующей реальности.

02.06. ОИТ и менеджмент качества. Принцип непрерывного улучшения

Прогресс человечества пришел к необходимости систем менеджмента качества. Таких систем создано много. Описание такой системы представляет собой много томов, содержащих схемы предприятий, схемы руководства, перечни принципов, примерные инструкции и перечни обязанностей и пр. Это такой образец, что если, следуя ему устроить производство продукции, можно с большой вероятностью добиться успеха. Среди общепризнанных систем – ISO-9000. Один из принципов ISO-9000 – принцип непрерывного улучшения, согласно которому система менеджмента качества считается хорошей, если качество продукции растет.

В большинстве сфер человеческой деятельности с понятием качества и измерением качества проблем нет. Иное дело в образовании. Как показано в предыдущем разделе (Динамика результатов педагогических измерений), в педагогических измерениях нет места «надежности» в традиционном техническом понимании. Результаты тестирования не повторяются так, как это хотелось бы, например, инженеру, работающему с механизмами.

Результаты тестирования ЗУН, с той или иной степенью определенности, отражают состояние качества образования претендентов. Рост результатов тестирования ЗУН, если процесс тестирования не осложнен негативными артефактами, свидетельствуют о росте уровня качества.

Изменения результатов тестирований, проведенных с одним и тем же (возможно, вариативным) тестом, можно подразделить как общие и локальные. Под локальными изменениями можно понимать те, которые происходят при многократных тестированиях одной группы претендентов в рамках одной темы учебной дисциплины. Под общими изменениями можно понимать те, которые происходят при тестированиях разных групп претендентов, либо одной группы, но по разным темам. Локальные изменения характеризуют, скорее, личностные характеристики претендентов, а общие – учебный процесс. Положительные общие изменения свидетельствуют о росте качества учебного процесса.

Как показано в предыдущем разделе, стремление к «надежности» теста, редко приводит к результату, а «надежность», как численная характеристика теста, реально не используется в тестологических расчетах. Логично вместо «надежности» ввести в научный обиход характеристику, отражающую степень локальных изменений результатов тестирования. Поскольку рост образовательных результатов – это дидактическая цель, эту характеристику можно назвать «релевантность», понимая под этим названием **степень соответствия дидактической цели**.

В отличие от «надежности», измерить «релевантность» несложно. Например, как разность усредненных результатов последовательных сеансов тестирования одной и той же группы претендентов по двум вариантам одного теста в рамках одной технологии.

Положительная релевантность означает рост результатов, отсюда – связь с принципом непрерывного улучшения системы ISO-9000: тест, технология тестирования хороши в случае их положительной релевантности. Принцип «положительной релевантности» лучше характеризует тест и технологию тестирования, чем принцип «надежности».

Целесообразно применить термин «релевантность» не только к тесту, но и к технологии тестирования, и к технологии компьютерного обучения. Подробное рассмотрение возможности применения термина «релевантность» к технологии обучения (и тестирования) приведено в **томе I** настоящего пособия.

02.07. Каким быть компьютерному комплексу для обучения и контроля ЗУН?

Хотелось бы, чтобы не только хотелось...

Борис КРУТИЕР

Какими должны быть образовательные компьютерные комплексы, чтобы быть востребованными и удобными в пользовании для учителя, учащегося, управленца? Как сделать так, чтобы прохождение сеанса компьютерного тестирования вызывало у учащихся рост интереса к обучению и повышению качества своих знаний? Как «скрыть» от учащегося саму процедуру тестирования, снять атмосферу напряжения и волнения, сконцентрировав его на раскрытии творческих качеств и реализации знаний? Как устранить или свести до минимума возможности подлога, «списывания»?

Деловая компьютерная игра многими воспринимается как своеобразная панацея. Вероятно, что именно в этой области нужно искать пути построения сценариев компьютерного обучения и контроля ЗУН – привлекательных и вызывающих желание исследовать, учиться. Удачные попытки построения таких сценариев существуют. Можно привести в качестве примера известный сценарий и игру Сиды Мейера «Цивилизация». Однако следует заметить, что такие игры обычно содержат довольно мало по-настоящему учебной информации и могут быть рассматриваемы лишь как основа для дальнейшей методической и сценарной работы.

К негативным последствиям таких игр можно отнести вероятную неправильную ориентацию игроков в реальных сочетаниях величин в реальной жизни: сколько должно быть тех или иных ресурсов, чтобы производство других развивалось таким-то образом.

Создание профессиональных компьютерных игр-тренажеров – вполне разработанная и освоенная специалистами область. Существуют прекрасные тренажеры и средства для их создания – конструкторские программные комплексы, банки реализаций. Вызывает сожаление, что это относится исключительно к специальным тренажерам – для летчиков, танкистов, судоводителей, автогонщиков. Рынок же тренажеров для учащихся общеобразовательных учебных заведений – компьютерных лабораторных работ и учебных практикумов – пока заполнен изделиями несравненно низшего качества. Они малоинформативны, однотипны, не вызывают интереса к обучению.

*Если отладка – процесс удаления ошибок,
то программирование должно быть процессом их внесения.*

Э. ДЕЙКСТРА (Артур Блох. Законы Мэрфи)

Разработкой и изготовлением компьютерных тестирующих комплексов в России занимаются, в основном, коллективы, базирующиеся при вузах. Среди таких вузов следует назвать МГУ, СПбГУ, НГУ, МФТИ, МИФИ, МЭСИ, ТГУ, ДВГУ и др. Как правило, коллективы состоят из физиков, математиков, инженеров-компьютерщиков. Педагоги либо психологи к сценарным работам практически не привлекаются.

Одним из счастливых исключений здесь является МГУ, где в таких разработках принимают активное участие преподаватели и студенты психологического факультета. Созданная ими система «Телетестинг» завоевала в 1997-2002 гг. в России огромную популярность. Скажем, в Приморском крае в 2000-2002 гг. любое компьютерное тестирование (какая бы организация его не проводила) родителями и учащимися однообразно нарекалось «телетестингом».

В качестве другого примера, приведем опыт разработчиков ДВГУ. Здесь с 1994 года ведутся научные и экспериментальные работы в области разработки компьютерных обу-

чающих и тестирующих средств. В научных семинарах и работах здесь принимают участие сотрудники, студенты и преподаватели – физики, химики, математики, историки, юристы, педагоги, географы, экологи, филологи, экономисты. Создан ряд разновидностей компьютерных тестирующих комплексов, реализующих известные педагогические идеи, восходящие к Платону и Сократу.

Созданные в ДВГУ игровые тестирующие комплексы популярны в школах Приморского края. На их основе ежегодно проводится около пятидесяти Приморских краевых образовательных олимпиад, выпускные и переводные экзамены в школах, вступительные экзамены для абитуриентов ДВГУ. Созданный в ДВГУ в 1998 году компьютерный комплекс – деловая игра «Дидактор» является удачной попыткой реализации игровой концепции обучения и контроля. Работа с ним не утомительна, интерфейсы учащегося и преподавателя просты и понятны. Идея «Дидактора», заключающаяся в предложении не отвечать на вопросы, а оценивать ответы виртуальных учеников, привлекательна для учащихся. «Дидактор» предлагает учащемуся шпаргалки (развернутые ответы и пояснения) – в этом его обучающая функция. «Дидактор» умеет работать и на компакт-диске, и в Интернет.

Составлять учебные материалы для наполнения «Дидактора» проще, чем для многих известных комплексов. Для этого не нужно пользоваться дополнительными средствами. Файлы с наборами тестовых заданий полностью изготавливаются в формате RTF. Тестовые задания «Дидактора» – мультимедийные – это позволяет значительно расширить возможности обучения и тестирования. Все компоненты заданий могут содержать изображения, формулы, звуковые и видеофайлы. Комплект «Дидактор» содержит блок конвертации составленных так тестовых заданий в необходимую для использования форму.

В серии изготовленных в ДВГУ тренажеров – «Кроссворд», «Фракон» (конструктор фраз), «Файндер» (найди ошибку), «Диктант», «Вордик» (узнай слово), «Ковер-самолет», «Толмач» – использованы оригинальные, известные и переработанные сценарии. Тренажер-игра «Диалог» основан на идее вовлечения ученика в разговор между анимированными виртуальными персонажами. Контроль ЗУН и обучение происходят завуалировано, в процессе диалога. Интерес представляет процесс обучения на основе виртуальных диалогов представителей разных культур и времен (диалоги культур).

Широкий круг дисциплин, материал которых послужил наполнением для вышеописанных изготовленных тренажеров – десятки школьных и более двухсот вузовских дисциплин.

Студенты ДВГУ с 1999 освоили компьютерное обучение и часто предпочитают проходить процесс обучения и контроля на компьютере, контактируя с преподавателем только на консультациях либо по электронной почте – это удобно, ведь многие учащиеся работают, многие одновременно получают два и три образования.

Некоторые студенты, рассматривающие качество обучения как особую ценность, высказали мнение, что избранная Открытым университетом форма контроля знаний (тренажер Дидактор) – лучшая из виденных ими, поскольку

- игра позволяет продлить период внимательной работы;
- демонстрация только одного варианта выполнения задания на экране делает невозможным пользование простым логическим сравнением для отгадывания;
- необходимость оценивания вариантов выполнения не по двузначной, а по четырехзначной шкале, заставляет глубже задуматься над учебным материалом.

Такого же мнения придерживаются и многие школьники и учителя.

Глава 03. Тестовые испытания и управление образовательными результатами

*«В современных условиях – говорил на XXVI съезде КПСС товарищ Л. И. Брежнев, – многократно возрастает значение дисциплины, значение личной ответственности»
Материалы XXVI съезда КПСС, М.: 1981, с. 50.*

Одна и та же группа учащихся, проходящая одну и ту же процедуру тестирования, может в разное время показать разные результаты. Это может быть связано и с постепенным привыканием учащихся к технологии, и с экологическими условиями, и с другими факторами, такими, например, как усиление или снижение мотивации под влиянием общения с педагогами или неконтролируемого общения со сверстниками.

Именно эта разность результатов часто не позволяет считать тот или иной комплекс надежным (согласно определению), не смотря на все его видимые достоинства. Это серьезное противоречие заставляет исследователей искать иные, нежели традиционно понимаемая надежность, критерии качества тестового комплекса.

Экологические и погодные условия вполне поддаются контролю. Поэтому их влияние на результаты тестирования может быть экспериментально измерено. Без особых усилий можно исключить влияние эффекта привыкания, проведя предварительно тренировки достаточной длительности.

Совсем другое дело с мотивацией. Мотивированные люди совершают невероятные действия, совершают подвиги, сдвигают горы. Люди, умеющие мотивировать себе подобных и знающие, как это делается, становятся известными и знаменитыми – организаторами производства, педагогами, миллионерами, тренерами команд, преступниками.

Результаты тестирования учащихся, не имеющих выраженной мотивации к повышению своего рейтинга, не выйдут за пределы области «наиболее вероятного» результата. Результаты тех же учащихся, но стремящихся получить приз, очевидно, улучшатся.

Мотивация может быть случайной. Мотивация может быть не только позитивной, но и негативной, например, она может быть связана с потребностью «выделиться» любым путем – скажем, получив по итогам тестирования наименьший в группе рейтинг.

Исключить влияние негативной и случайной мотивации можно только одним путем: дать учащимся новую, мощную, общую для всех мотивацию, которая способна «перебить» своим влиянием все остальные. Это не только исключит досадные случайности, но и приведет к достоверному получению самых «высших» результатов, на которые способны учащиеся.

03.01. Мотивация учащихся к повышению качества ЗУН

Можете всегда положиться на убийцу в отношении затейливости прозы
Владимир НАБОКОВ, aforizm.kaminplus.ru

Мотивация учащихся разнообразна. Она должна быть известна организаторам тестирования до начала тестирований и начала анализа результатов. Мотивация может быть открытой (известной) и скрытой (неизвестной организатору тестирования). Собственная мотивация может быть скрытой и от самих претендентов, в силу каких-либо психологических обстоятельств – недостаточности навыков самоанализа и пр. Влияние скрытой мотивации может быть принято организаторами тестирований, например, за систематическую ошибку процедуры измерений.

Неожиданно появившаяся и неожиданно исчезнувшая мотивация может «спутать карты» тестологов, измеряющих надежность и валидность теста. Мотивацию может создать неформальный лидер коллектива испытуемых, впечатление от увиденного кинофильма и пр. Прекрасный тест будет признан негодным по причине, которая никогда не будет известной.

Особенно заметно на результаты тестирования влияет мотивация, связанная с:

- соревновательным азартом;
- игрой;
- стремлением познания интересных сведений;
- безразличием к результату.

Причем, игра и соревнование здесь могут происходить не только с коллегами, но и с виртуальными персонажами, с собой.

Для проявления мотивационных эффектов необходимо, чтобы претенденты имели возможность тестироваться несколько раз, а тест должен обладать вариативностью.

Начиная с 1994 года, Дальневосточный государственный университет проводит массовые компьютерные тестирования знаний школьников Приморского края. Это делается в рамках договора с департаментом образования и науки администрации Приморского края (ранее – краевым управлением народного образования). Зимой, в январе, компьютерные тестирования проходят на базе Всероссийского детского центра «Океан», находящегося на побережье полуострова Муравьева-Амурского в пригороде Владивостока. В январе там собираются от 400 до 900 школьников – победителей районных олимпиад. Тестирование проводится бесплатно, без вручения внушительных призов – победителям выдаются просто памятные дипломы и сертификаты с указанием набранных результатов. Некоторые учебные заведения выставляют на основании этих сертификатов оценки за вступительные экзамены.

В 1994-1996 году силами коллектива Центра новых образовательных технологий ДВГУ были проведены эксперименты по выявлению и повышению мотивации школьников к участию и победе в тестированиях по общеобразовательным дисциплинам. На основе этих экспериментов была создана относительно простая технология мотивации, которая впоследствии показала свою высокую эффективность не только среди школьников, но и среди университетских студентов и школьных учителей. Она названа создателями технологией «мягкого тестирования» и описана ниже.

03.02. Технология мягкого тестирования. Десять задач и десять решений

*Per ludos ad scientias –
Через игру к знаниям (лат)*

К недостаткам существующих технологий тестирования ЗУН относят, например:

- двузначность логики оценки тестов, не позволяющую распознать, насколько точно претендент чувствует нюансы определений и описаний явлений;
- отсутствие моментов психологической разгрузки, постоянно и неоправданно поддерживаемое у претендента в течение сеанса состояние крайней ответственности за каждое неправильное действие, приводящее к случайным ошибкам;
- односторонность рейтинговой оценки, отсутствие в результатах тестирования подробной информации о склонностях претендента;
- сложность определения весов заданий, требующая предварительных массовых экспериментов и временных затрат, оказывающихся неоправданными после очередной смены преподавателя или учебника по дисциплине.

Технология мягкого тестирования (ТМТ) создается в «противовес» традиционным жестким:

- когда претендентам запрещают повторное тестирование;
- когда задание можно выполнить только правильно или неправильно, без дифференциации;
- когда в процессе тестирования нет места обучению;
- когда сеанс тестирования скучен и зануден;
- когда до успехов претендентов нет никому дела;
- когда претендента ведут тестироваться не тогда, когда он готов, а когда назначено время сеанса.

Перевести термин «мягкое тестирование» на английский язык можно как «Soft testing» или «Soft grading», в соответствии с известной аналогией (В. С. Аванесов).

Создатели ТМТ решали десять приведенных ниже задач.

1. Ввести в ТЗ многозначные логические отношения, создать критериально-ориентированную технику оценки выполнения заданий, включающую не только полюсные («верно» и «не верно») варианты оценки, но и более широкий спектр, в том числе двумерную, матричную шкалу.
2. Снизить вероятность логического угадывания претендентом правильных действий.
3. Ввести в сценарий тестирования игровые моменты, позволяющие безболезненно увеличить время сеанса и количество предъявляемых заданий, а также осуществить психологическую разгрузку в течение сеанса.
4. Найти и ввести в сценарий моменты, побуждающие претендента к самообразованию.
5. Создать и внедрить технику подсчета многомерного рейтинга, исключая необходимость вычисления относительных весов заданий и позволяющую непротиворечиво судить о разных склонностях претендентов.
6. Реализовать в едином программном комплексе возможности единообразного создания и применения ТЗ всех известных форм, дав возможность автору реализовать их достоинства без привлечения дополнительных программных средств.

7. Максимально упростить труд авторов, переложив труд по конвертации и сортировке ТЗ на плечи компьютера.
8. Реализовать наиболее прогрессивные методики обеспечения конфиденциальности и информационной безопасности.
9. Реализовать в практике тестирований достоинства мультимедийных технологий.
10. Сделать программный комплекс мобильным, позволяющим немедленно приступить к работе в сети и на персональном компьютере, с компакт-диска и в сети Интернет без прохождения процедуры инсталляции.

Вот основные положения технологии мягкого тестирования (их можно считать решениями поставленных выше задач). Их десять.

1. Применяемая в тестированиях БТЗ должна быть достаточно велика для того, чтобы сидящим за соседними компьютерами (или повторяющим сеанс) участникам не попадались одинаковые задания.
2. Участникам следует разрешить тестироваться не только официально, но и в свободное время, посещая компьютерный класс «живой» очередью. Господствует принцип: «Тестируйся по любой дисциплине, сколько хочешь раз, не ущемляя этого права своих коллег». Заключительная часть этой фразы оказалась абсолютно необходимой в связи с ограниченностью технических и временных ресурсов. Все результаты, полученные в присутствии ответственного лица, учитываются. Как итоговый, засчитывается лучший результат. Разрешение тестироваться в произвольное время приводит к повышению результатов, поскольку даже на обычном очном экзамене, как известно, большую роль играет настроение учащегося.
3. Отказ от сеанса или получение малой суммы очков не приводит к негативным последствиям для учащегося.
4. Интерфейс компьютерного комплекса должен быть игровым, не занудным. Задания и интерфейс должны содержать элементы сюрпризности. Сначала это привлекает, а потом – не отвлекает. Это абсолютно необходимое требование, которое, вместе с обеспечением привлекательности процесса контроля, позволяет
 - «отодвинуть» период начала влияния действия эффекта «ровной дороги»;
 - увеличить количество предъявляемых заданий и, соответственно, время сеанса.
5. Время активного компьютерного тестирования школьников не должно превышать 15-30 минут. Количество заданий не должно превышать 30-60. Это выявленный нами предел, после которого даже отличники и взрослые учителя начинают тестироваться «методом тыка». Если задания короткие, то удвоение времени тестирования от 20 до 40 минут не дает качественных изменений или методических выгод: чаще всего при этом просто удваивается количество набранных участниками очков. Длительность сеанса может играть решающую роль в привлекательности технологии, например, сидение за компьютером более часа значительно уменьшает аудиторию желающих проверить силы.
6. В том случае, если среди тестовых заданий есть вычислительные задачи, т. е. в случае неактивного тестирования, время можно увеличить. Важно, чтобы в процессе решения задач учащийся смотрел не на экран (усталость зрения), а на листок бумаги. Лучше, если задачи учащиеся получают предварительно, решают их в аудитории, а потом ответят на вопросы теста о ходе решения и результатах.
7. По истечении времени тестирования, необходимо оперативно проанализировать результаты и громко, так, чтобы это слышали стоящие в коридоре учителя и участники следующей смены, назвать фамилии школьников, которые набрали лучшие результаты

в течение сеанса. Если при этом попросить всех присутствующих (входящих и выходящих, сидящих за компьютерами участников) устроить аплодисменты победителю смены, они это с удовольствием сделают. Эти аплодисменты – абсолютно обязательный атрибут. Это действие, как показали наши последующие опросы участников, «подстегивает» и победителей, и побежденных. Это характерно в среде и студентов, и учителей. Такие эксперименты мы проводим регулярно в разных городах и разных аудиториях с одинаково положительным результатом.

8. Тестирование следует вести в течение нескольких дней. Утром каждого нового дня на доске объявлений должны появляться списки вчерашних участников, разбитые по классам и дисциплинам, отсортированные по величине результатов. Это – важнейший фактор. Именно эти списки станут наиболее значимым объектом сегодняшних обсуждений в столовой, коридорах. Кстати, созданная таким образом база данных может пригодиться приемным комиссиям вузов.
9. На следующий день, после появления списков результатов, у компьютерного класса всегда вырастает живая очередь. Именно эта очередь всегда становится для нас своеобразным камнем преткновения. С одной стороны – это показатель успеха технологии, с другой – присутствие в очереди и ожидание своего «звездного часа» для школьников становится значительно более важным, чем посещение лекций, экскурсий, столовой. Для «уничтожения» очереди приходится привлекать сторонние силы – дежурных учителей и пр.
10. ТМТ так проста, что трудно поверить в ее действенность. Но факты упрямая вещь. Эта экспериментально подтверждаемая информация говорит «сама за себя»:
 - результаты претендентов растут (в среднем) в течение периода испытаний при каждом повторном сеансах;
 - ожидающие сеанс претенденты внимательно изучают учебники и дополнительную литературу, заинтересованно обсуждают и задают связанные с учебными дисциплинами вопросы друг другу и учителям;
 - появляющиеся рядом с очередью претендентов учителя (сначала случайно, потом – постоянно) спонтанно устраивают тут же консультации по всем дисциплинам, эти консультации востребованы и желанны.

Такую (п.10) спонтанно создающуюся продуктивную учебную обстановку трудно встретить в учебном заведении в стандартных условиях.

ТМТ предложена и реализована в ДВГУ, варианты ее многократно представлялись и обсуждались на научных конференциях всероссийского уровня.

Результаты экспериментов с ТМТ и тестовым комплексом позволили выработать относительно новый для тестологии критерий: **технология организована тем лучше, и тестовый комплекс тем лучше, чем выше скорость роста результатов участников при повторных тестированиях.** Этот критерий «перекликается» с известным в статистической физике принципом «производства энтропии». Схожесть подчеркивают и ежедневно получаемые кривые распределений результатов тестирований – динамика дифференциации выборки претендентов.

*Кто хвастает перед тем, кто его знает,
- заслуженно поднимает себя на смех.
ЭЗОП, aforizm.kaminplus.ru*

В 1999 году, когда отмечалось 200-летие со дня рождения А. С. Пушкина, на базе ВДЦ «Океан» проводилось экспериментальное внедрение компьютерного игрового тестера-тренажера «Пушкинистика». В один из дней нам довелось случайно услышать разговор библиотекаря и двух учительниц. Они, не зная, что мы проводим эксперимент, живо восторгались тем, что в библиотеке «нарасхват» книги Пушкина и о Пушкине, что книги да-

же приходится «придерживать» и разрешать пользоваться ими только в читальном зале! Более того, они заметили, что творчеством и жизнеописанием Пушкина совершенно неожиданно и в массовом порядке интересуются «физики», «математики», «информатики».

Подчеркнем, что описанное характерно не только для среды победителей районных олимпиад, но и для любой сельской и городской школы, где проводятся длительные компьютерные тестирования знаний по этой технологии.

Вместе с основным результатом эксперимента, успешной проверкой работоспособности технологии, нами были проведены измерения и получены следующие сведения:

1. Период «привыкания» школьника, не знакомого ранее с компьютером, к процедуре компьютерного тестирования составляет 30-40 минут. Иными словами, если у школьника нет учебника, и он пользуется только собственными знаниями в течение одного дня, его результаты от первого к третьему сеансу растут, а результаты третьего и четвертого сеансов одинаковы. Если школьник знаком с компьютером, «привыкание» наступает после первого сеанса, а результаты второго и третьего сеансов одинаковы.
2. В случае если школьник имеет возможность воспользоваться консультацией учителя или учебником для самостоятельного повторения материала, его результаты тестирования в течение 3-4 дней могут достичь определенного технологией максимума. Иными словами, в течение вполне обозримого периода применения описанной технологии растет и обучаемость, и объем приобретаемых школьниками знаний. Причем, это происходит самопроизвольно, без применения каких-либо дополнительных стимулов.

Основным результатом описанных здесь и много раз воспроизведенных экспериментов является следующий: **ТМТ – адекватный инструмент управления ростом уровня качества образования.**

Элементы ТМТ применялись нами в течение двух лет при организации Всероссийских компьютеризованных олимпиад «Телетестинг». В Приморском крае ДВГУ проводил «Телетестинг» в рамках договорных отношений с Центром тестирования «Гуманитарные технологии» МГУ им. М.В.Ломоносова. Договором и инструкциями не оговаривалась возможность массового знакомства участников с результатами – просто каждый участник должен был получить в указанное время из Москвы свой собственный сертификат. Первые дни тестирования шло довольно вяло, не смотря на многомесячную рекламную кампанию. Ситуация кардинально изменилась, когда после каждого из дня тестирований в вестибюле ДВГУ стали появляться списки участников с результатами и было получено разрешение на повторные тестирования. Применение этой «маленькой хитрости» вывело ДВГУ в российские лидеры по массовости участия в централизованных тестированиях.

Участники смен в ВДЦ «Океан» и участники «Телетестинга», ставшие студентами ДВГУ, вспоминают дни тестирований и, особенно, свои эмоции и эмоции своих конкурентов, возникавшие при знакомстве с ежедневными «боевыми листками» – списками результатов тестирований.

Более подробно идеология ТМТ описана в Приложениях на примере сценариев эксплуатирующихся тренажеров. Естественным развитием ТМТ является описанный в Приложениях и реализуемый в настоящее время в ДВГУ сценарий двухуровневого тестирования.

03.03. Мотивация студента к повышению качества ЗУН

*О рефлексах и замещении...
Многие девочки, мечтающие о собаках, выходят замуж за военных.
(украинский фольклор, прилетело из Интернет)*

Описанная выше технология мягкого тестирования «работает» и в студенческой аудитории. Но здесь ее можно и нужно видоизменить.

Студент – не школьник. Он не только поглощает знания, он еще и вынужден искать будущее место работы. Он также желает, чтобы будущий работодатель также его искал и выбирал. Давайте посмотрим, как можно помочь в этом студенту, а заодно и работодателю. А, может быть, удастся попутно и качество образования улучшить?

Сейчас все большую популярность приобретают компьютеризованные экзамены по многим дисциплинам. В ДВГУ в 2000-2003 гг. компьютеризованны экзамены по двум сотням дисциплин, включая специальные. В год сервера ДВГУ фиксируют более полумиллиона сеансов тестирований.

Давайте представим себе, что результаты тестирований студентов помещаются на специализированный сайт в Интернет, который доступен и студентам, и их друзьям, и родителям, и работодателям.

Создание сайта – это первый этап внедрения технологии.

Наверное, не каждый студент сразу согласится с широкой публикацией рекламы своих знаний. Но это не проблема, можно организовать демонстрацию результатов «по желанию». Студентов, заранее ищущих контактов с работодателями, наверняка такой способ саморекламы заинтересует. Им не нужны аплодисменты в аудитории, их интересует личное будущее. Студенты же, рядом с фамилиями которых на сайте нет результатов тестирований, быстро пересмотрят свои взгляды на необходимость публикации. Доводы здесь просты: если результатов нет, значит – нечем гордиться.

Второй этап внедрения – следует разрешить студентам тестироваться в порядке живой очереди произвольное количество раз, публикуя в Интернет их лучшие результаты.

Третий этап внедрения – привлечь к проводимым Вами тестированиям студентов других вузов. Вряд ли они откажутся прорекламировать себя и свои знания в Интернет, пусть даже за оплату.

Когда в базе данных результатов тестирований появятся результаты тестирований студентов 3-4 вузов, возникнет хорошая база для работ по мониторингу качества высшего образования в регионе (городе). Публикация результатов мониторинговых исследований качества подхлестнет конкуренцию вузов и, в целом, неизбежно приведет к росту уровня качества.

Это и есть тот самый главный результат, к которому мы все стремимся – повысить эффективность управления качеством образования с помощью информационных технологий.

03.04.Полторы дюжины привлекательных черт процедуры компьютерного тестирования

*Искусство – в умении скрыть искусство.
ОВИДИЙ, aforizm.kaminplus.ru*

Вряд ли параграф с таким названием можно признать традиционным для тестологической литературы. У тестологов не принято обращать внимания на привлекательность процедуры измерения. Эта особенность перекочевала в современность из опыта бланкового тестирования, где результат мало зависит от того, эстетичности бланков.

На самом деле это достаточно серьезный вопрос. Давайте представим себе два профтехучилища. В мастерских одного – традиционная затхлость, исцарапанные инструменты и станки, выцветшие плакаты, разбитый кафель на полу. Это легко представить. В мастерских другого училища – все наоборот, все как в рекламной телепередаче. Где будет выше количество желающих учиться качество выпускников?

Интересно влияние внешнего антуража не столько на результат тестирования, сколько на побуждение учащихся проходить тестирование самостоятельно и неоднократно, стремясь показать все более высокий результат. В способности генерировать стремление людей к самообразованию – главное свойство и достоинство нововведений в системе образования.

Привлекательность трудно и не нужно описывать числом, разные процедуры тестирования можно сравнить экспериментально, с достаточно представительной выборкой участников. При этом следует использовать один и тот же тест. Эксперимент следует сопровождать грамотным опросом участников, в котором будет учтено мнение всех участников процесса:

- Руководителей учебного заведения;
- Исполнителей наполнивших базу данных тестовых заданий;
- Преподавателей;
- Персонала компьютерного класса;
- Претендентов.

Мнение претендентов будет во многом зависеть от того, насколько удачно пройден тест. Второй тест они почти всегда пройдут более удачно, поскольку получили опыт при первом тестировании. Чтобы исключить этот артефакт, следует варьировать последовательность двух процедур тестирования для разных претендентов, или разделить тестирование на этапы в соответствии с разными процедурами.

1. Руководителям учебного заведения понравится та процедура, после которой им представят более представительные, понятные и «засерьезненные» таблички отчета.
2. Создателям базы данных тестовых заданий понравится та процедура, где им потребовалось меньше времени для проведения работ (при условии одинаковости вознаграждения).
3. Преподавателям понравится та процедура, где они меньше «портили свои нервы», и где они получили более четкую информацию о претендентах.
4. Персоналу компьютерного класса понравится та процедура, где группы претендентов были более дисциплинированными и на подготовку которой (инсталляцию, инструктирование) ушло меньше времени.
5. Претендентам понравится та процедура тестирования, где

- проще достичь более высокого личного результата;
- меньше приходится отвлекаться на диалоги с персоналом;
- меньше длительность сеанса;
- результаты объективны;
- представление результатов понятно;
- понятна и кратка инструкция;
- присутствует психологическая разгрузка, соответствующая возрасту учащегося и ситуации;
- личные успехи сразу становятся известны, а неудачи – нет;
- можно быть успешным лично и в составе команды, по одной дисциплине и по комплексу;
- реакция окружающих на личные успехи эмоциональна и незамедлительна;
- есть возможность посоревноваться в знаниях с виртуальным противником и коллегами – учащимися;
- в интерактиве присутствует сюрпризность;
- меньше психологической усталости накапливается к концу сеанса;
- меньше непроизводительные потери личного времени на стояние в очереди, повторное выслушивание без того понятных инструкций и пр.;
- меньше компьютерных сбоев;
- выше качество монитора, мыши, мебели, воздуха и пр.;
- можно вернуться к выполненным заданиям и исправить свои ошибки;
- участники находятся в равных условиях (у всех одинаковая возможность пользования шпаргалками, никто никому не подсказывает, никто не может извне повлиять на результат тестирования).

Как видим, на мнение претендентов о привлекательности процедуры тестирования влияет огромное количество факторов, которые могут быть неожиданными и о которых редко задумываются организаторы тестирований.

Дайте архитектору, который всю жизнь прожил в сарае, много хрусталя и попросите построить дворец. И он построит хрустальный сарай.

ПРИТЧА

Вы, наверное, видели телеигры «О, счастливчик!» или «Кто хочет стать миллионером?». Что привлекает сюда телезрителей? Не радость же за чужой успех или энциклопедические сведения, которые можно прочесть и без того? Привлекают два момента: чувство юмора ведущего и изредка встречающие неожиданные варианты ответов – сюрпризы. А что привлекает зрителя к игре «Поле чудес»? То же самое – юмор ведущего и выходки игроков. Кто придумал эти сценарии? Есть подозрение, что это – хорошие педагоги и психологи. Эти игры будят желание соревноваться в знаниях, копаться в энциклопедиях.

... помнить о том, что есть и ирония, и эзопов язык, и метафора, и метонимия, и синекдоха, и оксюморон, и еще много других выразительных средств в языке.

М. ГОРБАНЕВСКИЙ, Журналист, 2004, № 2. С. 83 – 85

У этих игр есть аналоги – прекрасные компьютерные тренажеры. Дети и взрослые с удовольствием работают с ними. И в одиночку, и группами. Прямая дорога им – в сферу образования. Нужно только включить туда необходимые для учебного процесса тесты. А потом проводить тренаж, скажем в форме соревнований.

Повысить привлекательность компьютерного учебного пособия или процедуры тестирования не так просто, это всегда – область профессиональной деятельности. Но игра стоит свеч, поскольку высока вероятность положительного влияния привлекательности тестирования на качество образования в учебном заведении.

Глава 04. Информационная безопасность процедур массового компьютерного тестирования

*Как идут две параллели, да не сходятся...
Как стоят два перпендикуляра, да не наклонятся... (мат.)*

Десять основных требований информационной безопасности тестирований и обеспечение их выполнения

- Секретности баз тестовых заданий;
- Секретности перечней лиц имеющих доступ к тестовым заданиям;
- Недоступности лиц, имеющих доступ к информации о тестах, в период конструкторских работ;
- Запрета несанкционированного доступа к базам тестовых заданий и тестовым программным оболочкам во время их перемещений, установок, хранения, использования;
- Запрета несанкционированного доступа к результатам тестирований во время их создания, пересылки, обработки и хранения;
- Недопущения использования несанкционированных справочных материалов;
- Применение специальных методик размещения претендентов в аудиториях, исключающих возможности взаимного подсказывания, подсматривания и пр.
- Недопущения несанкционированных подсказок и пояснений со стороны персонала;
- Недоступности для посторонних лиц вовлеченных в аудитории, коридоры, туалеты и пр., в т.ч. заблаговременная проверка помещений на наличие шпаргалок;
- Предотвращение и ликвидация последствий нарушения названных выше пунктов.

Само же обеспечение предполагает:

- создание документационного сопровождения (инструкции, положения, приказы);
- отбор персонала;
- обучение и инструктаж персонала;
- повышенное вознаграждение персонала за обеспечение особого режима;
- проведение инспекций, в том числе и путем привлечения лиц – «инкогнито»;
- оснащение техникой видеонаблюдения и сейфами;
- привлечение сил милиции либо вневедомственной охраны;
- меры по выявлению нарушений и административному наказанию виновных лиц.

Далее мы расскажем о некоторых неожиданных удивительных явлениях, выявленных коллективом Открытого университета ДВГУ в процессе многолетних тестирований студентов, школьников, учителей.

04.01. Два удивительных явления из опыта массовых компьютерных тестирований

Ultra possi nemo tenetur –

Никто не может быть обязан выше своих возможностей (лат.)

Внедрение технологий тестирования в России и, в частности, в ДВГУ, оказалось связанным с рядом обстоятельств, имеющих отношение к правовым аспектам. Появились, например, хакеры-«доброжелатели», пытающиеся влиять на результаты тестирования – зачеты, экзамены, оценки вступительных испытаний.

Хакеры смогли повлиять на ритмичность учебного процесса вузов, участвующих в подобных экспериментах:

- производились попытки вскрытия засекреченных файлов;
- распространялись компьютерные вирусы;
- публиковались ответы на тестовые вопросы в сети Интернет.

Хакеры пытались вскрывать защищенную информацию не только «из интереса», но и с целью наживы. В этом – элемент преступления.

Бороться с информационным мошенничеством правовым путем затруднительно, особенно в России, где хакерство давно превратилось в своеобразный национальный вид интеллектуального спорта. Однако следует обратить внимание на специфику системы образования.

Мы столкнулись с новым, интересным противоречивым явлением. Вместе с активизацией «преступной» хакерской деятельности, параллельно, необычным путем иницируется массовый процесс повышения уровня качества образования:

- хакеры самостоятельно постигают новые, порой не изучаемые в вузе, знания в области компьютерных наук;
- «потребители» хакерской продукции вполне сознательно и старательно изучают «купленные» тексты с ответами на вопросы тестов, причем, изучают так, как не изучали ни один из своих или чужих конспектов.

В упомянутых же текстах просто содержится полный текст курса, изложенный в несколько иной форме, нежели в учебнике – в виде вопросов и ответов.

Этим положительным свойством нового явления, безусловно, следует воспользоваться в благих целях, попутно устранив возможность возникновения негативных явлений.

04.02.Дюжина путей обеспечения информационной безопасности массовых компьютерных тестирований

*Мужчины на удивление нелогичны:
твердят, что все женщины одинаковы, и постоянно меняют одну на другую.
Сидония КАЛЕТТ*

Здесь мы перечислим нетрадиционные пути устранения негативных последствий информационных «преступлений» и обеспечения информационной безопасности массовых компьютерных тестирований.

1. Не следует «засекречивать» ответы на вопросы компьютерного теста. Большую часть текстов базы тестовых заданий следует издавать типографским способом. Это приведет к исчезновению объекта несанкционированной купли-продажи.
2. Вопросов и ответов в базе тестовых заданий должно быть так много, чтобы невозможно было быстро и полноценно воспользоваться бумажной или электронной шпаргалкой. Для отображения в базе материала объемом 36 учебных часов, как показывает опыт, достаточно 300-400 заданий (10 заданий на час). Зазубрить их невозможно.
3. Рядом сидящие претенденты должны видеть на своих мониторах разные тестовые задания. При повторном тестировании каждый претендент получает новый вариант. Экзамен – это лотерея; претендент не должен заранее знать варианты заданий, которые «выпадут» именно ему. Выполнение этих условий достигается путем случайной генерации вариантов из базы тестовых заданий.
4. Варианты тестов не должны повторяться, однако программа – генератор вариантов должна составлять списки заданий для каждого претендента не абсолютно случайным образом, а исходя не из принципа равномерности покрытия учебного материала – по-немногу из каждой темы курса.
5. Следует предпринимать меры защиты не только на программном уровне. Например, текущие результаты тестирования необходимо контролировать в интерактивном режиме, а программные модули должны храниться на физически независимых носителях и обновляться с периодом меньшим, чем время сеанса. Этим повышается уровень защиты от несанкционированного проникновения.
6. Следует периодически обновлять базы тестовых заданий. Можно не только дополнять базу количественно, но и, например, просто видоизменять словарные обороты в заданиях. Этим устраняются многие из возможностей применения шпаргалок.
7. «Сговор» претендентов с представителями персонала, которые «закроют глаза на шпаргалки», будет исключен, если разрешить использование любых бумажных учебных пособий во время тестирования. При этом следует ограничить время сеанса тестирования так, чтобы было невозможно успеть воспользоваться этой информацией в полной мере. Этим не только устраняется возможность «преступного» сговора претендентов с персоналом компьютерного класса, но и создается стимул к дополнительному изучению материала именно тех пособий, которые рекомендованы преподавателем курса.
8. Контрольное тестирование должно происходить для всей учебной группы одновременно. Защиту от внешних воздействий можно обеспечить присутствием заинтересованного преподавателя, например того, который будет вести последующие учебные курсы для этих студентов в будущем семестре.

9. Следует исключать на время сеанса тестирования связь класса с внешними сетями.
10. Следует использовать процедуру случайного выбора номера компьютера и псевдонима для каждого претендента. Претенденты в классе должны сидеть в случайном порядке и иметь достаточно длинные псевдонимы, которыми они не смогут быстро обмениваться.
11. Тестирование должно вестись на рейтинговой основе. Это влияет как на качество образования, так и на установление внутренней системы контроля. Если студенты будут знать, что количество пятерок и четверок, которые они могут получить ограничено, постепенно устранятся подсказки.
12. Результаты рейтинга (лучшие) должны публиковаться в Интернет и быть доступными для вероятных работодателей выпускников, что станет дополнительным стимулом к учебе. При этом в соответствии с законодательством, каждый тестирующий комплекс должен предоставить студенту право выбора – разрешить или не разрешить подобную публикацию.

Последний пункт может показаться спорным. Мы считаем его крайне важным и оказывающим непосредственное положительное влияние на уровень качества образования.

Наш опыт, накопленный в течение проведения массовых мероприятий в 1994 -2004 гг. в Приморском крае, свидетельствует о следующем: публикация результатов тестирования в виде рейтинговых списков приводит к возникновению феномена состязательности, побуждает учащихся проходить снова и снова сеансы тестирования, читать учебную литературу, консультироваться друг с другом и учителями по вопросам учебных дисциплин.

Разрешение учащимся беспрепятственно проходить тестирование вновь и вновь, в совокупности с публикацией результатов, приводит к:

- росту результатов;
- росту популярности тестирований;
- росту заинтересованности учителей;
- росту заинтересованности родителей.

Состязательность может быть поддержана локальными мерами. Например, если громко похвалить одного из участников за только что полученный высокий рейтинг, и попросить всех присутствующих поаплодировать ему, происходит удивительное: претенденты начинают конкурировать за место в компьютерном классе, придумывая невероятные причины, чтобы повторить тестирование. Такие «локальные» воздействия на аудиторию, кроме того, снимают часть сиюминутной усталости, повышают внимание.

Не менее интересен эффект воздействия на испытуемых интерфейса тестирующей программы. Если интерфейс скучен и однообразен, внимание участников падает быстрее, нежели в случае игрового интерфейса, сочетающего элементы сюрпризности, игры, новизны. Так же, распределяется и степень желания претендентов проходить тестирование с помощью того или иного комплекса. Игровой интерфейс побуждает к повторению тестирования.

Перечисленные здесь меры не только устраняют негативные факторы, но и положительно влияют на уровень качества образования.

04.03.Десять каналов утечки конфиденциальной информации. Обеспечение секретности БТЗ

*Генерация случайных чисел – слишком важный вопрос, чтобы оставлять его на волю случая.
Роберт КОВЗЮ, Окриджская лаборатория.*

Очевидно, учащийся, получивший несанкционированный доступ к экзаменационным тестовым заданиям до экзамена, имеет немалые преимущества перед остальными. Наличие таких учащихся приводит не только к появлению ложных оценок ЗУН, но и

- влияет на объективность результатов региональных мониторинговых исследований;
- делает необъективным анализ свойств тестовых заданий в экзаменационной базе.

Недоступность тестовых заданий учащимся до экзамена – один из важных компонентов технологии контроля ЗУН.

Распространенные в России компьютерные тестовые комплексы, применяемые для массовых тестирований школьников и студентов, не удовлетворяют требованиям секретности. Это связано со следующими обстоятельствами:

- ТЗ составляются преподавателями, работающими в момент составления заданий в окружении нежелательных свидетелей – коллег, друзей;
- Круг составителей ТЗ мал (обозрим);
- Количество ТЗ, включаемых в тест, невелико. Уже прошедшему тестирование учащемуся не составляет труда передать все увиденную и запомнившуюся информацию учителям, и друзьям, которые будут тестироваться позже;
- ТЗ предъявляются сидящим в компьютерном классе учащимся в едином пронумерованном порядке. Это упрощает создание шпаргалок;
- В компьютерных сетевых классах, как правило, расставлены компьютеры так, что соседи видят сразу несколько экранов. Если тест представлен одним вариантом, учащиеся, видящие действия соседей, имеют возможность «списывать»;
- Невариативность. Учащийся, который станет тестироваться повторно, получит тот же вариант заданий, с которым он познакомился несколько часов назад и который он имел возможность обсудить с репетитором;
- Кроме авторов ТЗ, к ним имеют доступ оформители (сканирование и включение иллюстраций, таблиц), программисты и др. – несколько десятков специалистов. «Утечку» информации от такого огромного коллектива трудно контролировать;
- Россия – большая страна, время начала тестирований в разных регионах различается на много часов. Телефон и электронная почта в течение этих часов обеспечивают и передачу на запад, и обработку рассекреченной на востоке страны информации. Наверное, это одна из причин того, что средний уровень знаний в западных регионах России оказывается неизменно выше, чем в восточных;
- Страна «не знает своих героев». Договора, которые заключаются центрами тестирования со своими сотрудниками и между собой, как правило, не включают конкретных условий по обеспечению конфиденциальности информации;

- Нет известий о том, что кого-то где-то когда-то как-то наказали за рассекречивание информации о школьных или студенческих тестах.

Перечисленные обстоятельства снижают уровень секретности (конфиденциальности) тестов. Информация о заданиях предстоящих тестирований, несмотря на титанические усилия Министерства образования, задолго до их начала мероприятий становится доступной всей стране, появляется на сайтах в Интернет. Покупатель не видит продавца: информация оплачивается желающими путем электронных переводов и рассылается им по электронной почте. Как говорится, не покупает заблаговременно ключи школьных тестов только ленивый.

В общем-то, считается, что продавец тестовых ключей – преступник. Только не всегда удается понять не только то, как его найти, но и за что и как его наказывать. Нам не известны случаи серьезных наказаний лиц, допускающих утечку конфиденциальной информации из центров тестирования и торгующих ключами школьных тестов.

Продажа информации о заданиях предстоящих школьных тестирований превратилась в своеобразную российскую индустрию. Продажа неверной, подложной информации о тестах стала точно такой же индустрией, и это, как ни парадоксально, несколько компенсирует потерю конфиденциальности.

*Неразбериха в обществе постоянно возрастает.
Только очень упорным трудом можно ее несколько уменьшить.
Однако сама эта попытка приведет к росту совокупной неразберихи.
Артур БЛОХ. Второй закон термодинамики Эверитта.*

В других странах мира ситуация с секретностью тестов считается иной. Пакеты с тестовыми заданиями распространяются по стране под охраной вооруженных полицейских. Личности составителей тестовых заданий засекречены, и секрет этот охраняется законом. Рассекречивание тестовой информации – преступление, подлежащее суровому наказанию в судебном порядке.

Более того, все эти меры – верхушка айсберга. Есть еще невидимая часть общественных взаимоотношений, защищающая секретность тестов не хуже вооруженной полиции. В США, например, давно воспитан среднестатистический добропорядочный гражданин, который, заметив преступные действия кого-либо, немедленно сообщает об этом «куда следует». Добропорядочный американский школьник никогда не будет списывать на виду у кого-либо контрольную работу. И никому не даст списать. И сам первым расскажет учителю или полицейскому о том, что кто-то из его друзей списывает.

Эта «честность» имеет, конечно же, подоплеку – систему общественно признанных стимулов. Рассмотрим ее с точек зрения заинтересованных сторон.

04.04. Полдюжины стимулов сохранения конфиденциальности тестовых материалов в США

*Каждый убийца, вероятно, чей-то хороший знакомый.
Агата КРИСТИ*

Стимул №1

Так сложилось, что возможность получения высокой оценки в американских школах и вузах – вопрос конкурса. Высших оценок не может быть больше нормы. Ученик знает: если кто-то рядом с тобой списывает, значит, он стремится попасть в список лучших вместе с тобой, а количество мест ограничено. Каждый ученик лично заинтересован в том, чтобы указать учителю на списывающего коллегу.

Стимул №2

В университетах существует ограниченное количество вакансий для бесплатного или льготного обучения. Есть система престижных именных стипендий, количество которых ограничено. Занятие льготной вакансии или получение престижной стипендии – не только предмет гордости, но и вопрос облегчения жизни сейчас и удачного старта карьеры в будущем. Следовательно, каждый сокурсник – твой конкурент, а самый простой способ избавиться от конкурента – найти и предоставить преподавателю или администрации доказательства его нечистоплотности в учебе.

Стимул №3

Администрация учебного заведения заинтересована в качестве обучения. Это не только вопрос престижа, это вопрос государственных и частных субсидий. Поэтому администрация заинтересована в поощрении учащихся, выявляющих экзаменационные подлоги, совершающиеся их друзьями и преподавателями. В отличие от российской школы, такой «шпион» в американской школе – герой!

Стимул №4

Администрация штата заинтересована в повышении конкурентоспособности своих университетов на уровне государства. Это и дополнительная реклама, и опосредованный источник поступления средств в бюджет (например, оплата обучения для нерезидентов штата всегда значительно выше). Администрация жестко относится к подлогу, для лиц, в подлоге замеченных, значительно осложняется дальнейшая карьера – сведения о них заносятся в специализированную государственную базу данных. Администрация поощряет лучшие учебные заведения снижением налогов, дополнительным финансированием и пр.

Стимул №5

Родители стремятся подготовить лучший «плацдарм» для старта карьеры ребенка. Они заинтересованы в выявлении недобросовестных однокурсников ребенка, которые потенциально могут его в будущем «обойти» и в сообщении о них администрации.

Стимул №6

Родители заинтересованы не только в поступлении ребенка в более престижный (и более дорогой по оплате обучения) вуз, но и в сохранении безупречности репутации ребенка. «Ставки» высоки. Поэтому родители сами контролируют процесс его обучения с точки зрения недопущения нечистоплотности.

04.05. Стимулы сохранения конфиденциальности тестовых материалов в России

*Из хорошего железа не делают гвоздей –
умный не пойдёт служить солдатом.
КОНФУЦИЙ*

Выше мы познакомились с работой системы обеспечения конфиденциальности тестовых материалов и, соответственно, защиты качества образования в США.

Россия – другой мир. Российский студент и школьник всегда готов помочь коллеге. Рискую собственной репутацией и будущей карьерой, он поможет сидящему за соседней партой во время тестирования конкуренту – решит его задачи и даст списать свои.

Иными словами, поднимая качество образования, россиянам приходится решать еще и задачу ломки менталитета, шире вводить в школьный и вузовский обиход элементы соревнования и конкуренции.

Традиционные для России «утечки» информации препятствуют успешному внедрению централизованных тестирований. Министерство образования предпринимает отчаянные меры по обеспечению секретности тестов, даже пакетики для хранения тестовых бланков заказывает в Германии. Однако на практике все эти меры оказываются затратными и неэффективными.

*Если четыре причины возможных неприятностей
заранее устранены, то всегда найдется пятая.
Артур БЛОХ, 4-е следствие закона Мерфи*

Меры по обеспечении секретности могут привести и к снижению качества тестовых заданий. Будучи созданными единоличным автором и хранящиеся в тайне, они оказываются закрытыми для рецензирования, редактирования и экспериментального контроля в референтных группах.

Где же выход? Где находится гордиев узел и как его разрубить?

Выход есть. Политики говорят, что для того, чтобы прекратить утечку информации на уровне государства, надо либо все засекретить, либо – все рассекретить. Проще и дешевле пойти вторым путем.

*Сам погибай – а знамя спасай.
Левицкая Н. В., Луганская Л. И., Лаврова К. И. Русский язык.
Учебное пособие для солдат, не владеющих или
слабо владеющих русским языком. Воениздат, 1982*

Оптимальная, на наш взгляд, технология обеспечения секретности (конфиденциальности) тестовых заданий может быть изложена на 1-2 страничках нормативного документа. Опишем ее в виде алгоритмической последовательности действий.

04.06. Обеспечение конфиденциальности теста в масштабе региона и страны

Не пытайтесь устранить зло, лежащее за пределами вашей досягаемости, иначе и зло не устраните, и авторитет потеряете.
Теодор РУЗВЕЛЬТ

На основании опыта участия в организации всероссийских централизованных тестирований в Приморском крае («Телетестинг» и ЦКТ), а также опыта участия в инспектировании таких тестирований в Республике Саха (Якутия) и Хабаровском крае, нам удалось выделить и систематизировать встреченные технологические недостатки и разработать достаточно безупречную и экономичную технологию обеспечения конфиденциальности.

В этой схеме совмещены два варианта сценария процедуры тестирования, разработанные для полностью и частично компьютеризованной формы тестирования.

Десять принципов безопасной технологии создания БТЗ

1. Список коллектива авторов, рецензентов и редакторов тестовых заданий является открытым и может быть опубликован.
2. В электронную базу вводится не менее 2-3 тысяч заданий по каждой дисциплине, включающих, например, материал всех опубликованных с грифом Министерства образования учебников и задачников. Это число обусловлено следующим:
 - Именно столько заданий (вместе с вариантами) по каждой учебной дисциплине можно найти в открытой печати, и что-то новое придумать довольно сложно.
 - Если средний ученик просто выучит хотя бы четверть верных вариантов выполнения такого количества заданий – он достоин высокой оценки.
 - Создать такую базу вполне возможно вдвоем – вчетвером за 1-2 месяца.
 - Такое количество заданий, сопровождаемых иллюстрациями и таблицами, вполне поместится на одном компакт-диске.
3. Задания сортированы по принципу подобия (равной сложности, равной трудоемкости, соответствия одной теме, равной важности). Групп подобных заданий для каждой дисциплины получится около сотни. В каждую группу войдет два-три десятка равных по сложности заданий. Процедура создания вариантов основывается на принципе: по одному заданию из каждой группы. *Математическое пояснение: Пусть тест включает 100 заданий, по одному из каждой группы. В каждой группе пусть будет 20 заданий. Если пересчитать все возможные варианты сочетаний заданий, которые создаст компьютер, получится 20^{100} . Это астрономическое количество, гарантирующее неповторимость вариантов.*
4. Нет традиционного засекречивания заданий. Задания публикуются в виде сборников для подготовки. Объем такого пособия составит 500-800 страниц, т. е. может превысить объем учебника. Страшного ничего не случится, если школьник выучит значительную часть заключенной в пособии информации. В этом реализуется главная дидактическая цель. Публикация сборников принесет доход центрам тестирования и, в то же время, устранил почву для нежелательных теневых явлений.

5. «Веса» заданий вычисляются заблаговременно, путем экспериментальных тестирований. Тестируются все желающие, эксперименты проходят многократно в течение года во многих территориях. Это возможно в условиях отсутствия секретности, поскольку все возможные варианты задач и вопросов заранее опубликованы. Выучить их все – равносильно повторному неоднократному прохождению курса обучения.
 - Заблаговременное определение весов ускорит процедуру статистической обработки результатов реальных тестирований, устранил этот общеизвестный объект многолетней критики.
 - Экономическую поддержку эксперимента можно осуществить введением оплаты участия. Желающих потренироваться будет много.
6. Алгоритм формирования единственного и неповторимого варианта теста (списка заданий) для каждого претендента основан на коде, включающем
 - паспортные данные;
 - точные дату и время создания кода;
 - четырехзначное число – ключ, меняющееся на едином Интернет-сайте через определенные промежутки времени и транслируемое по радио или телевидению одновременно на территории всей России в течение всего периода тестирования;
 - параметр (со значениями, например, 0 и 1) отражающий желание либо нежелание претендента опубликовать результат. Это необходимо в соответствии с законодательством РФ. В соответствии с выбранной цифрой, претендент узнает о своих результатах на следующий после тестирования день в общих списках, либо – по завершении всех мероприятий из присланного ему именного сертификата.

Блок-генератор вариантов ориентируется на эти данные, и дает возможность разрешить претендентам повторные тестирования по любой дисциплине, с гарантией неповторимости вариантов, в том числе и платные предварительные тренировки. Вариант теста может быть сгенерирован вновь и в другом месте при повторном введении перечисленных данных, например, при работе апелляционной комиссии.
7. Алгоритм формирования варианта теста включает выбор формы представления варианта – бланковой (если центр тестирования не оборудован достаточным количеством компьютеров) либо компьютерной. Выбор осуществляет руководитель центра тестирования. Блок выбора формы позволяет распечатать тест на принтере для каждого претендента. По окончании сеанса тестирования претенденты должны ввести результаты выполнения заданий в компьютер.
8. Нормативные документы и инструкции включают официальное разрешение претендентам тестироваться и тренироваться произвольное количество раз по любой дисциплине на протяжении периода тестирования (если это не ущемляет прав доступа других претендентов). Первый шанс (сеанс тестирования) предоставляется претендентам бесплатно, последующие – после оплаты.
9. Тестирования проводятся на базе специализированных центров, ответственные лица которых сертифицированы.
10. Руководители специализированных центров несут персональную ответственность за соблюдение правил тестирования – обеспечение доступности тестовых материалов, работоспособность печатных устройств и пр. Законодательством определена административная и пр. ответственность за подлог.

Десять шагов безопасной технологии тестирования

1. Утверждается и инструктируется персонал территориального Центра тестирования. Руководит тестированием специалист, сертифицированный в рамках централизованных мероприятий МОРФ.
2. Программное обеспечение, поступившее заблаговременно из МОРФ, устанавливается в компьютерном классе территориального Центра тестирования. Оно используется как для тренировочных, так и для аттестационных тестирований. Доступ претендентов на тренировки оговорен инструкцией. Тренировочные и аттестационные тестирования проходят по одному и тому же, описанному ниже, технологическому плану.
3. Администратор сети Центра тестирования регулярно, по установленному расписанию, получает по телевидению, по радио или из Интернет ключ, необходимый для формирования кода претендентов. Ввод числа синхронизован со временем начала и окончания сеансов.
4. Начало регистрации синхронизировано по времени с получением ключа. *Контроль документов претендентов при регистрации проводится независимо представителем милиции. Независимый контроль перекрывает один из вероятных путей подлога – подмену претендента.*
5. Приступающие к тестированию претенденты в течение нормативного промежутка времени вводят в компьютеры свои паспортные данные. На основе этой информации формируется код претендента и соответствующий вариант заданий теста, которые можно выполнять сидя перед экраном компьютера либо, если время сеанса тестирования велико и компьютеров мало – в аудитории, имея на руках распечатанный вариант и бланки для работы. Решение о форме тестирования принимает руководитель центра на основе сведений о количестве претендентов, наличия техники и пр.
6. Объявляется начало сеанса. Выполнение заданий в аудиториях центра тестирования и в компьютерном классе контролируется дежурными и представителями вузов региона. *Контроль в рассматриваемом случае упрощен, поскольку*
 - варианты заданий у всех участников разные;
 - вес и объем шпаргалки, имеющейся в продаже в ближайшем магазине, достаточен для того, чтобы ее вовремя заметить.
7. Объявляется окончание сеанса.
8. Учащиеся, выполнявшие задания письменно, возвращаются в компьютерный класс и в течение нескольких минут вводят в компьютер свои коды и свои результаты. Документы претендентов контролируется представителем милиции. Если результат введен в компьютер после окончания нормативного времени, например, после получения очередного ключа, задание считается невыполненным.
9. Результаты тестирования отправляются на специализированный сайт МОРФ для обработки. Они остаются неизвестными в течение оговоренного инструкцией периода. Учащиеся имеют право пройти сеанс вновь в тот же день либо в другой день периода тестирований.
10. По истечении технологического периода обработки, публикуются результаты. Претенденты с ними знакомятся на стендах. По окончании периода тестирований на всей территории России, претенденты получают сертификаты, изготовленные на бланках утвержденной МОРФ формы.

Десять шагов безопасной технологии обработки результатов и публикации

1. Результаты претендентов в виде зашифрованного файла из территориального Центра тестирования направляются на специализированный сайт тестирований для обработки. Обработка проводится в течение нормативного срока.
2. Обработанные результаты тестирований возвращаются в территориальный Центр тестирования в виде удобном для распечатки и публикации. Результаты претендентов, не пожелавших публикации, в публикуемых списках отсутствуют.
3. Принцип формирования зашифрованного файла результата и его обработки известен только двум специалистам – руководителю и заместителю руководителя специализированного сайта. Оба они являются и создателями сайта.
4. Ввод на сайт любой иной информации, кроме файлов-результатов блокируется специальными методами. Обработка производится автоматически без участия людей. Определена мера личной ответственности создателей сайта и методов защиты сайта за подлог результатов.
5. Вся входящая и исходящая информация специализированного сайта тестирований дублируется независимым Центром хранения в утвержденном порядке, исключающем возможность подлога. Дублированная информация может быть использована при работе федеральной апелляционной комиссии.
6. Результаты тестирований оперативно публикуются на стенде территориального ЦТ.
7. Претенденты имеют право повторить попытку тестирования произвольное количество раз при условии оплаты и наличия возможности. В новом сеансе код, который будет сформирован для претендента компьютером, будет другим, и вариант заданий – тоже.
8. Таблицы рейтингов учащихся, прошедших тестирования, без промедления параллельно публикуются на специализированном сайте, защищенном от посягательств. Это мероприятие не только повышает популярность и массовость тренировочных тестирований, но и дополнительно «перекрывает» пути фальсификации результатов.
9. После окончания тестирований все учащиеся получают заверенный печатью сертификат, включающий его рейтинги в 100- или 1000-бальной шкале по всем дисциплинам.
10. Именной сертификат претендент может представить в вуз при поступлении. Оценки по пятибалльной шкале абитуриенты получают в конкретном вузе на основании
 - его рейтинга;
 - рекомендаций МОРФ;
 - приказа руководителя учреждения или органа управления образованием, устанавливающего шкалу пересчета.

Десять направлений снижения общей себестоимости и себестоимости мер безопасности

Приведенная схема оптимальна для применения в регионах как обеспеченных техникой, так и не имеющих технику в достаточном количестве. Здесь нет необходимости:

- использования дорогостоящих сканеров;

- привлечения большого персонала для обработки результатов;
- привлечения персонала для вторичного контроля обработки результатов;
- тиражирования и распределения больших количеств стандартных бланков;
- авиационной пересылки бумажных материалов;
- одновременного использования площадей и технических средств;
- одновременной доставки больших количеств претендентов;
- охраны хранящихся тестовых материалов;
- исключены неэффективные потери времени работы компьютеров.

Расчеты, проведенные для Приморского края в 2003 году, показали, что на тогдашнем уровне развития компьютеризации и ресурсной базы, можно по такой схеме провести тестирование знаний выпускников всех школ в течение трех-четырех недель. Себестоимость мероприятий с учетом оплаты труда и транспортных расходов, расходов на инспекторские проверки и инструктивные мероприятия, составила около 100 руб./сеанс. Эта сумма вполне окупается за счет оплаты тренировок самими претендентами. Однократное же аттестационное тестирование должно быть бесплатным.

Десять известных в практике тестирования способов подлога

1. Подмена претендента при регистрации в ЦТ путем подделки документов или сговора.
2. Подмена претендента при тестировании в ЦТ путем изменения места или времени выполнения заданий.
3. Подмена бланков с результатами работы претендента в процессе транспортировки.
4. Подмена бланков с результатами работы претендента в ЦТ путем сговора.
5. Подмена файла результатов в территориальном ЦТ путем сговора.
6. Взлом и исправление файла результатов в течение времени хранения в ЦТ путем сговора. Для взлома нужно довольно много времени.
7. Подмена файла результата на сервере ЦТ или в момент пересылки между серверами путем хакерской атаки.
8. Шпаргалка. Покупка, пронос и использование шпаргалки в течение сеанса тестирования. При этом если вариантов теста достаточно много, используется сговор.
9. Списывание. Происходит, если претенденты сидят в аудитории так, что видят результаты действий друг друга.
10. Обмен файлами. Происходит, когда один претендент регистрируется с паспортными данными и паролем другого претендента.

Глава 05. Количество и качество ЗУН

В нашем мире, когда человеку есть что сказать, трудность заключается не в том, чтобы заставить его сказать это, а в том, чтобы не дать ему повторять это слишком часто....

Бернард Шоу

Многие термины и понятия, применяющиеся педагогами на практике, не определены как измеримые, т. е. не конструктивны. Знания, умения и навыки – именно такие понятия.

Оперировать неизмеримыми свойствами, когда речь идет о реализации ИТ, нельзя. Для того чтобы измерить свойство (или даже явление в целом, если его свойства не определены), т. е. приписать ему конкретное число, необходимо уметь делать одно из двух:

1. Разлагать свойство (явление) на составляющие части, которые можно перечислить (пересчитать), каждая из которых может быть отдельно описана и сравнена с другими;
2. Упорядочивать свойство (явление) в ряду ему подобных; ранжировать свойства (явления) по формальному алгоритму, позволяющему однозначно определить, почему данное конкретное свойство (явление) должно быть помещено между данными двумя подобными (впереди или позади, слева или справа, внизу либо сверху).

Сложно однозначно сказать – знание какой математической теоремы важнее знания другой, знание биографии какого писателя важнее знания биографии другого. Сегодня специалисты однозначно утверждают, что «так», завтра – что «этак».

Люди пока не определили другого общепризнанного способа измерения (соотнесения) объема и качества своих ЗУН, как экзамен. Экзамены принято делить на формальные (процедура описана нормативными документами) и неформальные.

Экзамен называют неформальным, если он происходит в реальной ситуации, например, когда нужно срочно найти лекарство или пищу в лесу, устранить аварию, решить сложную математическую задачу и т. п. Такой экзамен позволяет выделить **лучшего, на сей момент**, специалиста или знатока в группе. Результат неформального экзамена не гарантирует, что в другое время, в другой подобной ситуации лучшим не окажется другой претендент (может быть, в момент экзамена у него болела голова, или на него влияли действия других членов группы). В этом смысле, неформальный экзамен не объективен.

К неформальным можно отнести и такой экзамен, когда экзаменатор задает вопрос «на вскидку» и ставит оценку, ориентируясь по нескольким словам, не ожидая полного ответа.

Публичная защита выпускной работы учащимся также может считаться неформальным экзаменом, поскольку в зачетной книжке оценка выпускной работы отделена от других, и процедура защиты предусматривает проверку знаний только по одной узкой теме.

Формальный экзамен проводят в форме длительного собеседования или тестирования. Во время собеседования претенденту задают несколько основных вопросов (чаще два или три) из разных областей учебной дисциплины и несколько дополнительных. Выслушивает ответы комиссия либо, чаще, один преподаватель. На основе выслушанных ответов преподаватель, ориентирующийся на свой личный опыт, делает свою оценку количества знаний. Эта формальная оценка – всего лишь одно число, варьирующееся в рамках оценочной шкалы. Опытные преподаватели не останавливаются на одном числе, а заносят в свой блокнотик несколько. Больше некуда заносить. Оценка в блокнотике – объективнее формальной, она имеет несколько компонент, поэтому ее можно назвать векторной, а формальную оценку – скалярной.

05.01. Является ли традиционная оценка мерой качества и количества? Сепарабельность оценки

*В философских спорах выигрывает побежденный,
ибо приобретает новую мудрость.
ЭПИКУР, aforizm.kaminplus.ru*

Шкала оценок чаще бывает 3-х, 4-х, 5-ти и 10-ти балльной. Для более точного оценивания количества знаний преподаватель часто придумывает свой личный алгоритм оценивания, самостоятельно зауживает или расширяет формально установленную шкалу, урезая ее снизу или добавляя знаки – плюсы и минусы. Часто такая расширенная шкала, при ближайшем рассмотрении, просто является двойной (дуальной) – число становится мерой количества знаний, а плюсы и минусы – мерой умений и активности учащегося, т. е. качества.

Оценка, т. е. одно-единственное число, часто считается мерой гарантируемого преподавателем или экзаменационной комиссией количества и качества ЗУН претендента. По-видимому, это не вполне приемлемо.

Нельзя оценить количество и качество одной мерой. Скорее, и многие преподаватели с этим согласны, оценка – не мера, а пропуск (мандат) учащегося в одну из неформальных (формальных) групп:

- Учащиеся, усвоившие материал и умеющие учиться. Таким следует открыть путь для дальнейшей учебы, и которых можно привлекать к ответственным мероприятиям (выполнению работ, заказов и пр.);
- Учащиеся, которых можно считать первыми претендентами на продолжение учебы (предоставление рабочего места), если оно не занято представителями первой группы;
- Учащиеся, которые учатся слабо и обладают небольшим объемом знаний. Но есть еще надежда, что положение исправится;
- Учащиеся, не проявляющие рвения к учебе, которых лучше дольше не учить (либо – которым лучше не предоставлять рабочее место).

Такой подход к оценке хорош в рамках одного учебного заведения. Однако когда речь идет о переходах учащихся между учебными заведениями, выпускных и вступительных экзаменах, он не применим. Просто потому, что в разных учебных заведениях присутствуют разные критерии получения этого «пропуска».

Оценка, представляемая как мера количества и качества ЗУН, должна быть:

- объективной;
- унифицированной;
- объяснимой;
- сепарабельной.

Под сепарабельностью в разных науках понимают способность объекта к разделению на составляющие компоненты без потери свойств. Эта способность очень важна, например, для биологии, где часто нельзя разделить клетку без уничтожения свойств.

Здесь под сепарабельностью мы понимаем способность экзаменационной оценки описывать разные свойства претендента не только в совокупности, но и по отдельности.

05.02. Что измеряют в процессе тестирования?

*Все силы отдай, чтобы выполнить точно расчет,
Но знай: небеса не приемлют научный подход.
Уильям ВОРДСВОРТ (перевод Л. Владимирова)*

В настоящее время используют три метода обработки и анализа результатов тестирования:

- классический;
- IRT (Item Response Theory);
- наглядный.

Два первых подхода базируются на статистической обработке так называемого «сырого» балла, то есть балла, набранного испытуемым в результате тестирования. Обработка включает

- определение трудности или «веса» каждого задания, т. е. числа, отражающего количество участников с ним справившихся;
- пересчет результатов – «сушка сырых баллов».

Наглядные методы, как правило, не включают промежуточной обработки и позволяют участнику увидеть результат сразу после сеанса.

Люди, применяющие наглядные методы, экономят свое и чужое время, и опираются при этом на свой опыт, говорящий:

- если задания составлены и «взвешены» опытным преподавателем, то никакая статистическая обработка не повлияет существенно на результаты тестирования;
- какие бы системы пересчета баллов не применялись, все равно вверху рейтинговой таблицы фигурируют, слегка меняясь местами, одни и те же учащиеся.

Нельзя с ними не согласиться. «Сушка сырых баллов» после завершения тестирования с добротой созданным тестом действительно требует больших трудозатрат и редко приводит к существенным изменениям в рейтинговых таблицах. Пример тому – российский опыт проведения тестирований «Телетестинг», ЦКТ, ЕГЭ. Опытные члены жюри традиционных олимпиад, которым приходится иметь дело с разными системами пересчета баллов, придерживаются часто такого же мнения.

Математический аппарат классической теории проще, чем аппарат IRT, там возможен обсчет даже с помощью калькулятора, хотя удобнее воспользоваться специальными программами.

Считается, что классическая теория имеет ряд недостатков, главный из которых – большая, чем в IRT, зависимость результатов измерения от инструмента измерения (теста).

Качественный анализ хорошо представленных результатов наглядно демонстрирует успешность учебного процесса, дает возможность оптимизировать корректировку содержания и методики обучения. Используя результаты каждого испытуемого, можно индивидуализировать обучение.

*Икс с игреком не путай!
А. И. ПРИГОРНЕВ*

От процедуры тестирования ожидают, что после ее завершения все участники будут достаточно надежно распределены (дифференцированы) по определенным группам. Количе-

ство групп определяется разрешающей способностью применяемой шкалы оценивания – 5, 10, 20, 100, 1000.

Попавшие в одну группу претенденты должны иметь одинаковые характеристики – знания, умения, навыки. Всем, попавшим в одну группу, претендентам присваивается одинаковый рейтинг (число) или выставляются одинаковые оценки в рамках выбранной шкалы.

Давайте рассмотрим наиболее вероятную реальную ситуацию, возникающую при применении тестового контроля ЗУН к группе учащихся. Поставим эксперимент над группой учащихся, совершенно не знакомых с материалом дисциплины, для которого составлены тестовые задания. Подвергнем их тестированию и рассмотрим результат.

Количество заданий, включенных в тестовый комплекс, обычно не превышает сотни. Среди них есть простые и сложные, требующие от учащегося смекалки, памяти, знаний определений и умений решать задачи.

Если создатели тестовых заданий не затруднили себя привлечением современных тестологических технологий и ограничились (как обычно) лишь составлением заданий закрытого типа «с выбором одного правильного ответа из четырех», произойдет следующее.

Согласно теории вероятностей, методом «случайного тыка» усредненный претендент правильно выполнит около 25% заданий. Претенденты «продвинутые», т. е. могущие логически мыслить и знающие слабые места создателей тестовых заданий, найдут правильные ответы в 50%-80% случаев. Участники, не знающие слабых мест технологии и не знающие дисциплины, но попытавшиеся привлечь свой интеллект к поиску правильных ответов, покажут результат около 30%-40%.

Таким образом, даже с помощью непрофессионального тестового контроля можно добиться определенного результата, т. е. распределить учащихся на три группы – «интеллектуалов-хитрецов», «старательных – ищущих» и «хвостистов», результаты которых попали в интервал 0%-35%.

Правда, разбить учащихся по таким группам можно будет только приблизительно, ибо их результаты распределятся по интервалу возможных значений, как правило, непрерывно. Хотя, может быть, вы и обнаружите на кривой распределения три соответствующих серединам групп пика. Тогда задача разграничения групп, конечно же, упростится.

В действительности все не так, как на самом деле.

Станислав Ежи ЛЕЦ

Теперь рассмотрим случай, когда:

- все учащиеся с учебной дисциплиной знакомы;
- знания, умения и навыки всех учащихся не имеют резких различий;
- все учащиеся учились в одинаковой обстановке;
- все учащиеся имеют стимул к получению высокого результата при тестировании.

Тогда кривая распределения результатов учащихся будет гладкой, похожей на известные в теории вероятностей распределения типа Гаусса, Пуассона или Максвелла – с одной вершиной, поднимающуюся из нуля в начале координат и монотонно спускающуюся к нулю на границе 100%.

Если же группа учащихся неоднородна, т. е. сформирована, например, из учащихся обычной и специализированной школ, ситуация может измениться и кривая распределения результатов превратится в двугорбую. Вершины ее будут ориентировочно соответствовать средним знаниям учащихся той и другой подгрупп.

А может стать и так, что вершин у кривой распределения окажется так много, что они станут незаметными на фоне одной, объединяющей их, суммирующей вершины.

*Гораздо легче узнать человека вообще,
чем какого – либо человека в частности.
Франсуа де ЛАРОШФУКО, aforizm.kaminplus.ru*

Воспитанные на незыблемых законах классической теории вероятностей тестологи обычно рассматривают только кривые распределений результатов с одной вершиной. Более того, они считают, что только такие кривые и могут получаться, если тестовые задания и процедура тестирования правильные. Все остальное, следовательно, – ошибка измерений.

Более того, тестологи говорят, что тестовые задания и процедура тестирования правильные, если в результате тестирований получается именно одногорбая, плавная, узнаваемая математиками кривая.

На практике такая ситуация, когда кривая получается плавной и одногорбой, достаточно редка, и методики ее исправления нельзя всегда признать однозначно объективными.

Считается, что хорошим методом исправления ситуации, борьбы с «многогорбостью» является приписывание заданиям весов (относительной сложности и пр.), выражаемых числами. Результат каждого претендента взвешивают, т. е. суммируют не очки (плюс очко за каждое верно выполненное задание), а веса.

Существуют разные подходы к вычислению весов заданий, нацеленные на «исправление» кривой распределения. Задача такого исправления математически сложна, и не каждый из разработанных способов исправления приводит к желанному результату.

Так и должно быть, поскольку исходные положения теории основаны на предположении об однородности групп претендентов, т. е. статистически «гладком» распределении среди них ЗУН. А это случается не всегда. Например, возьмите группу, где четверть претендентов умеет в совершенстве решать задачи только по одной теме физики, три четверти – только по двум. Тестирование такой группы с помощью усредненного по темам теста приведет к удивительным результатам, из которых трудно сделать монотонную одногорбую кривую.

Вычисленные веса могут сильно зависеть от того, как сформирована группа претендентов. Это означает, что каждое новое тестирование, с новой группой претендентов, должно сопровождаться решением задачи вычисления весов вновь. А результаты этих двух тестирований нельзя будет однозначно сопоставить.

Какой смысл имеют веса, можно ли использовать их значения для решения иных задач, более значимых, чем подгонка кривой распределения под установленную волюнтаристски форму?

Вернемся к традиционной задаче объективного оценивания подготовки учащихся и распределения их по трем-четырем группам в соответствии с их уровнями. Кривая распределения результатов с одной вершиной, как бы она не была математически и эстетически красива, не дает оснований для объективного ответа на вопрос о том, где же объективно провести границы групп.

Опытный тестолог поделит учащихся с помощью простого алгоритма, давно придуманного и повсеместно применяемого:

- протестируйте учащихся и получите кривую распределения результатов;
- присвойте всем учащимся рейтинги и составьте список учащихся в соответствии с рейтингами - от высшего к низшему;
- определите, сколько учащихся может учиться в группе лучших (допустим – 20);

- первых 20 учащихся в рейтинговом списке определите в первую группу, вторую двадцатку во вторую, третью – в третью и т.д.;
- если необходимо поставить оценки, поставьте первой группе «пятерки», второй – «четверки», третьей – «тройки», остальным – что осталось.

Задача распределения, на первый взгляд, успешно решена.

Там, где светит солнце, всегда есть тени.

Казахская пословица

Правда, здесь все же допущена необъективность: число 20 было включено Вами в расчет волонтаристским путем, без надежного обоснования. Установленная Вами граница «20» стала барьером просто потому, что столько стульев стоит у Вас в учебном классе. Может быть, этим Вы ломаете чью-то судьбу. А может, наоборот, включите в группу нежеланных учащихся, которые разрушат дисциплину и учебный процесс.

Неощутимая, с точки зрения теории вероятностей, разница в 1-2% может привести к тому, что кто-то получит «четверку» и навсегда потеряет возможность поступить в престижный вуз, а кто-то получит «пятерку» не вполне обоснованно, и в группу «лучших» попадут люди, на обучение которых будет впустую затрачено много времени и средств.

Можно каким-то образом вычислить доверительный интервал, окружить им граничную точку и провести собеседование со всеми учащимися, результаты которых туда попали. И потом уже провести окончательное распределение, которое наверняка станет более точным за счет использования человеческого фактора, опыта преподавателя. Но это уже – не автоматическая процедура.

*Основная цель познания –
открытие первопричин и тайн движения материи.*

Френсис БЭКОН

Как добиться объективности автоматической дифференциации учащихся, т. е. объективности оценивания их ЗУН? Очевидно, одного знания величин их рейтингов для этого недостаточно и необходимо привлекать методы математического моделирования и статистики. Например, методы факторного анализа.

Как измерить количество ЗУН?

*И ракет, как говорится, Есть у нас не пять, не шесть...
Да к чему считать-трудиться, Сколько надо – столько есть!
Из песни «Сколько надо – столько есть!»
Стихи П. ГРАДОВА, Музыка В. Букина*

Хороший тест состоит из систематизированного множества заданий, наполнение которых равномерно «покрывает» материал дисциплины, для выполнения которых требуется привлечь примерно одинаковые массивы информации и примерно одинаковые умения. В тех случаях, когда «примерной одинаковости» авторам достичь не удалось, задания «взвешены», т. е. им сопоставлены числовые коэффициенты, характеризующие относительную наполненность работы претендента информацией и умениями.

*Мы приносим свои извинения читателю за использование здесь
не вполне определенной педагогической терминологии.*

Таким образом, правильное выполнение каждого задания теста свидетельствует о наличии у претендента конкретного знания и умения. Если разные задания позволяют выявить наличие разных ЗУН, то результат тестирования – оценка – будет арифметической суммой полученных баллов или просто количеством выполненных заданий.

Если сеанс достаточно длителен, чтобы усредненный претендент смог справиться со всеми заданиями, а задания теста полностью покрывают материал, то вычисленная, как указано выше, оценка вполне может служить мерилom количества знаний.

А как же измерить качество?

Как измерить качество ЗУН?

*Мой жизненный опыт убедил меня, что люди,
не имеющие недостатков, имеют очень мало достоинств.
Авраам ЛИНКОЛЬН aforizm.kaminplus.ru*

Количественная оценка ЗУН, несомненно, должна быть отражена в оценке качества. Однако оценка качества затрагивает значительно больший пласт характеристик претендента.

Качество ЗУН проявляется в конкретной производственной обстановке. Это проявления выражаются в:

- скорости реакции на ситуацию;
- точности отдаваемых распоряжений;
- объективности создаваемых отчетов;
- качестве создаваемой продукции;
- скорости производства работ;
- способности быстро переучиваться;
- способности и желании повышать свою квалификацию;
- широте кругозора среди производственной и научной информации;

и др. Нельзя объективно оценить качество образования путем прямого тестового опроса. Однако такую оценку можно произвести путем косвенного опроса. В косвенном опросе вместо самого претендента участвуют его руководители, ученики, потребители продукции.

Проведением таких опросов занимаются (должны заниматься) территориальные Центры мониторинга качества образования.

Качество в образовании важнее количества. Специалист, умеющий учиться, в стратегическом плане более ценен для фирмы, чем тот, который просто обладает узким комплексом ЗУН и прекрасно справился с тестированием. Энциклопедисты достигают в науке и изобретательстве больших успехов, чем узкие специалисты.

Оценка качества не может быть выражена одним числом. Сколько параметров – столько и показателей. Оценка качества – вектор. Одной из его компонент является оценка количества знаний. Количество ЗУН (т. е. параметры личности, которые можно измерить тестированием) являются неотъемлемой компонентой оценки качества.

Оценка качества образования с помощью компьютерного тестирования – сложная задача, которую нельзя сегодня считать решенной. Возможные пути решения этой задачи могут быть связаны со следующим:

- тестирование умения объяснить принятое решение;
- тестирование умения объяснить ход решения предложенной задачи и выбор пути ее решения;
- тестирование навыков ассоциативного мышления, умения найти связи между явлениями и объяснить их;
- тестирование умения отделить стандартное от нестандартного, нужного от ненужного (в решении данной проблемы), известное от неизвестного;
- тестирование умения объяснить ход развития конкретной учебной дисциплины в историческом, логическом, философском аспекте;
- ситуационное тестирование, когда выявляется умение найти нестандартное решение за достаточно малое время, а затем объяснить его и провести сравнение с другими решениями, выявить достоинства и недостатки.

Это вполне можно сделать на сегодняшнем уровне развития тестологии. В частности, с применением технологии тестирования «Диалог» (см. выше, а также Приложения).

*Мало обладать выдающимися качествами,
надо еще уметь ими пользоваться.
Франсуа де ЛАРОШФУКО, fraza.net.ua*

Как оценить качество оценивания качества?

*Качество – нравственность изделия.
Геннадий МАЛКИН, fraza.net.ua*

Оценка качества обучения, выставленная опытным и признанным преподавателем, считается объективной. Хотя, сами учащиеся часто с этим не согласны, поскольку знают слабости преподавателя и умеют ими пользоваться. Преподаватель может «не любить» некоторые разделы дисциплины и не задавать вопросы оттуда. Или, наоборот, преподаватель может питать пристрастие к какой-то теме, и студент, обмолвившийся о любви к ней, гарантированно повышает свою оценку.

Есть три лучших способа «оценки объективности оценки» преподавателя:

- Опрос выпускников. Студенты не питают уважения к не объективным экзаменаторам, даже если получали от них букеты пятерок;
- Опрос работодателей, у которых работают выпускники;
- Опрос преподавателей учебных заведений более высокого звена или опрос преподавателей, работающих на более высоких курсах, куда переходят выпускники для продолжения обучения.

Иными словами, качество и количество ЗУН оценивается преподавателем верно, если:

- он пользуется уважением своих выпускников;

- его рекомендательное письмо является гарантией получения выпускником престижного рабочего места;
- результаты его работы пользуются заслуженным уважением среди коллег.

Ясно, что такая «оценка объективности оценки» часто неприемлема по причинам долговременности, трудозатратности и пр. Особенно в России, где отсутствует культура рекомендательных писем, преемственности обучения и приема на работу.

В России отсутствует и культура опросов о качестве обучения: психологи придумывают анкеты, службы трудоустройства их распространяют, но анкеты исчезают, оставаясь не заполненными, а заполненные анкеты складываются в мешки и также исчезают. Публикуемые службами трудоустройств результаты опросов часто оказываются основанными на «умозрительных представлениях» как бы они могли выглядеть.

Хотя, кое-где в России давно, более тридцати лет, предпринимаются попытки построения систем непрерывного обучения, где решение о приеме абитуриентов на очередную ступень обучения или о приеме специалиста на работу основываются на авторитетной рекомендации преподавателя и договорах о совместной образовательной деятельности. В масштабе страны это – пока диковинка. Трудно, на основании всего лишь газетных публикаций, гарантированно указать адрес таких событий.

*В доверии, конечно, необходима осторожность,
но далеко необходимее еще более быть осторожным в недоверии.
Этвеш ЙОЖЕФ, чешский писатель, aforizm.kaminplus.ru*

05.03. Тестирование – средство объективизации оценки качества и количества. Пять предпосылок

Есть люди, которые полагают, что все, что делается с серьезным видом, разумно.
Георг Кристоф ЛИХТЕНБЕРГ, aforizm.kaminplus.ru

Если экзамен, как предположено выше, – это лишь мероприятие по выдаче мандатов, и мы к этому привыкли, надо ли менять ситуацию? Надо ли вводить новые шкалы оценивания, придумывать тонкие и сложные методы различения качества ЗУН? Нужны ли специальные компьютерные информационные технологии?

Можно предположить, что ответ большинства педагогов, имеющих дело с большими аудиториями и для которых проблема оценивания является насущной, будет утвердительным. Приведем предпосылки такого ответа.

1. Оценка выставляется преподавателями всегда на основе собственного опыта. Предыстория, внешний вид и поведение претендента, сведения, полученные ранее от коллег и других людей, играют при этом определенную роль.
2. Даже имея алгоритм оценки, трудно с его помощью «развести» претендентов, результаты которых попали в область близости оценок. Нельзя убедительно объяснить человеку, почему 72 балла – это «пятерка», а 71 балл – это «четверка». Здесь помогут только более сложные оценочные средства подвластные компьютеру.
3. Невозможно за краткий промежуток времени гарантированно оценить знания претендента в пределах всего материала дисциплины. Тем более, если речь идет о группе. Вопросы преподаватель часто задает в некоем случайном порядке (опытный студент может даже «дирижировать» этой «случайностью»), и оценивает ответы, ориентируясь лишь на наличие ключевых слов и выражений в ответе студента.
4. Претендентов иногда бывает очень много, а количество вакансий, на которые они претендуют – мало. Такая ситуация постоянна при поступлении абитуриентов на престижные специальности вуза. Тогда «драка» за каждое «очко» весьма драматична, доказательства правоты и неправоты членов экзаменационных комиссий могут быть самыми разными. Опорой в экзаменационных разбирательствах здесь могут стать многоуровневые и многомерные шкалы оценивания, а лучшим судьей – компьютер.
5. Традиционный экзамен – всегда лотерея, прочные знания необязательны. Таково отношение к традиционной технологии педагогических измерений подавляющего количества претендентов. Только ОИТ смогут помочь изменить такое отношение.

Впрочем, среди педагогов немало сторонников сохранения и традиционных способов оценивания ЗУН. Традиционный способ действительно часто является оптимальным в смысле малозатратности и объективности. Например, в случаях, когда:

- высшая оценка ЗУН не всегда является для учащихся желанной путевкой в жизнь; например – в российских ПТУ, где большинство выпускников, вне зависимости от оценок, попадают после выпуска в равные производственные условия;
- оценка выставляется формально по двузначной шкале «зачет – незачет», скажем, при сдаче абитуриентами математического факультета университета вступительного экзамена по русскому языку;
- отношение к данной дисциплине в учебном заведении не внимательное, дисциплина считается второстепенной, занятия отменялись, не проводились и пр.

Далее мы оставим такие случаи за пределами своего рассмотрения.

В тестологии созданы методы, позволяющие значительно повысить объективность измерения качества и количества ЗУН. Это так непривычно для традиционной педагогики, что многие преподаватели и руководители этому не верят. И правильно делают. Именно такое отношение к новшествам отличает российскую систему образования от зарубежных систем.

*Мужчина, который умно говорит о любви, не очень любит.
Жорж САНД, aforizm.kaminplus.ru*

Как правило, руководителей учреждений «мучают» следующие вопросы:

- А вдруг после тестирования окажется, что явные двоечники получили положительные оценки?
- А вдруг пятерок будет слишком мало?
- А вдруг пятерок будет слишком много?
- А вдруг пятерку получит нежеланный «абитуриент», т. е. не обладающий нужными качествами, обладание которыми можно выявить только на устном экзамене?
- А вдруг двойку получит «желанный» абитуриент, посещавший все подготовительные курсы и кружки?
- А вдруг больше половины учеников получают двойки, а преподаватели им ставили в процессе обучения положительные оценки?

Судя по нашему десятилетнему опыту, российский преподаватель признает технологию тестирования только

- когда он сам протестирует свои знания и получит тройку;

а российский руководитель учебного заведения признает технологию тестирования только при выполнении трех условий:

- когда он узнает, что преподаватели заведения уже тестировались и получили «разные» оценки;
- когда он получит право управления количеством итоговых пятерок и двоек;
- когда ему «на цифрах» докажут, что тестировать дешевле, чем устраивать экзамены по полной схеме.

Тестирование является одной из наиболее технологичных форм проведения автоматизированного контроля с управляемыми параметрами качества.

Российская система образования обретает новое качество. Современные компьютерные средства, корпоративные сети, телекоммуникационные системы позволяют создать объединенный (на уровне организации, региона, страны) ресурс для диагностики качества образования и собирать статистические данные результатов тестирования дистанционными методами. Это – мощный инструмент управления качеством, возможности которого пока не изведаны.

05.04. Разрешающая способность и объективность измерительных процедур

Чертеж должен работать!
В. Ф. ЕФИМЕНКО

Под «разрешающей способностью» технологии оценки понимают степень возможности адекватного распределения (дифференцирования, разбиения) с ее помощью всех претендентов на критериально-однородные группы в соответствии с их уровнями знаний, личностными качествами, практической подготовкой и пр.

Обычно подразумевается, что границы между группами учащихся, проведенные процедурой оценки с высокой «разрешающей способностью», легко объяснимы и не вызывают споров. Однако многие из нас сталкивались со случаями логически необъяснимого проведения границ между «пятеркой» и «четверкой» или «четверкой» и «тройкой» не только на устных экзаменах, но и при компьютерных тестированиях.

Чаше всего, когда технологии тестирования разрабатываются неспециалистами, границы определяются волюнтаристски, «с потолка» (здесь очень просится слово «всегда»).

Иногда работодатели предъявляют специфические претензии к вузовской оценке качества, полагая, что принимаемый на работу специалист должен, в первую очередь, грамотно проектировать изделия, а не знать назубок определения и математические теоремы. Требования вузовских приемных комиссий также нередко расходятся с тем, что требуется от выпускников в школах. Эта разница в точках зрения на качество исходящей и входящей «продукции» вполне объяснима и должна учитываться при проведении тестирований.

Разрешающую способность технологии можно и повысить и понизить. Она может быть повышена:

- во-первых, путем перехода к многобалльным шкалам (например, 100-балльные или 1000-балльные);
- во-вторых, путем использования многомерных оценок (например, когда оценивание каждой составляющей образованности специалиста ведется по-отдельности и вычисление интегральной оценки осуществляется с применением методик факторного анализа).

Разрешающая способность зависит от свойств не только каждого из применяемых тестовых заданий, но и всей совокупности ТЗ в целом.

Лучшим следует признавать тот массив ТЗ, который:

- во-первых, «равномерно покрывает» весь учебный материал (т. е. в заданиях нашла равное отражение учебная информация из подобных по информационному весу и важности разделов дисциплины);
- во-вторых, вопросы и ответы самих заданий не вызывают нареканий с точки зрения грамотности, взаимного соответствия и т. п.;
- в-третьих, веса оценок, которые учащиеся получают в свою «копилку» в процессе тестирования, объективно соответствуют соотношениям сложности и важности ТЗ;
- в-четвертых, задания нельзя выполнить путем простого логического заключения либо простого узнавания исходя из каких-либо примет.

Тестовые задания описывают многими характеристиками, например;

величиной приписываемых им эмпирических параметров:

- весом;
- сложностью;
- трудностью;
- трудоемкостью;

содержанием:

- информационно-отражающие;
- проблемно-эвристические;
- развивающие.

шкалой (логикой) оценивания:

- ТЗ с двузначной шкалой оценивания (верно – не верно);
- ТЗ с трехзначной шкалой оценивания (верно – не совсем верно – не верно);
- ТЗ с четырехзначной шкалой оценивания (верно – не полно – не точно – не верно).

Большей различающей способностью обладают тесты с большим разбросом весов заданий, с многозначной шкалой оценивания, построенные на основе развивающего метода.

Принципиального повышения уровня объективности оценки **знаний** (именно – знаний, а не умений) можно достичь лишь при полном устранении человеческого фактора из процедуры оценки и переходе к автоматизированному контролю. Все чаще и чаще педагоги выражают положительное отношение к автоматизации рутинных опросов, ибо понятно, что только компьютеру под силу задать несколько десятков вопросов типа «Назовите дату сражения при Ватерлоо» каждому из сотни учащихся и оценить их ответы за 15 минут. Такие «зачетные» тестирования для нас уже стали привычны.

Саша, никогда не оправдывайтесь!

Это вызывает подозрения

Эммануил ВИТОРГАН (в одном из сериалов)

Считается, что процедуры измерения и оценивания **умений** учащихся, за исключением простых и поддающихся формализации случаев, пока автоматизировать не представляется возможным. Эта область деятельности пока подвластна только человеку, и педагоги здесь правы, отстаивая здесь свой человеческий приоритет перед машиной.

05.05. Систематические ошибки в тестировании

Не верю!
К. С. СТАНИСЛАВСКИЙ

Систематическая ошибка называется систематической потому, что она присутствует в результате всегда, систематически, при наступлении определенных условий. Она не является случайной, вызванной неконтролируемыми изменениями внешних параметров. Это – характеристика теста, которую можно контролировать и учитывать.

Иногда бывает так, что проще определить и учесть величину систематической ошибки, возникающей в конкретных ситуациях, чем исправить тест. Ведь исправленный тест тоже может характеризоваться систематическими ошибками, которые еще не определены.

Систематическая ошибка теста – это важная его характеристика, связанная, например, с:

- различием валидности теста, измеренной для разных групп претендентов (ошибка наклона) и
- соотношением между групповыми значениями измеренных характеристик по тесту и по критерию (ошибка интерцепта).

Эти понятия используются для акцентирования того, что заложено в тесте, – в противоположность возможной случайной ошибке. Это в значительной степени связано со спецификой выборки, на которой адаптируется тест. Рассмотрим эти понятия более подробно.

- **Систематическая ошибка наклона.** Различие коэффициента валидности теста для двух групп обследуемых может быть связано с использованием в роли критерия субъективных оценок. Систематическая ошибка наклона теста может возникать и при сравнении выводов об эффективности его применения на представителях разных этнокультурных, социальных и профессиональных групп. Зависит она и от величины их выборки. Необходимо устанавливать различия между коэффициентами валидности для разных выборок. Если такие различия существенны (например, в тесте по химии в русскоязычных и национальных группах), то лучше всего проверить тест еще раз на двух независимых выборках, для того чтобы выяснить, действительно ли эти различия столь существенны. Если отклонения вызваны спецификой данного контингента обследуемых, то следует, просчитав эту систематическую ошибку наклона, учитывать ее при анализе полученных тестовых данных и при принятии педагогами решений об отсеве, распределении учащихся.
- **Систематическая ошибка интерцепта** означает, что тест систематически завышает или занижает значения критерия для части претендентов. Обнаружить ее можно, когда результаты анализируются по подгруппам обследованного контингента, что особенно необходимо на стадии выверки нового теста. Возможно, что она будет обнаружена даже при одинаковой валидности теста для двух групп. Следствием бывают разные прогностические значения теста для этих групп (например, этнического большинства и меньшинства). Если к этим группам мы начнем применять единый нормальный показатель, одна из них может оказаться в более, а другая в менее благоприятных условиях. Требуется расчет систематической ошибки интерцепта для тестов, ориентированных на обследование групп с существенными различиями по образовательному и социоэкономическому уровням.

Причиной систематической ошибки интерцепта может быть не только содержание знаний и умений, заложенных в заданиях теста, но и семантика теста – степень привычности или сложности формулировки его заданий. Для различий в национальных группах могут быть

существенны этнопсихологические факторы, особенно в тестах, связанных с гуманитарными ценностями и знаниями.

Могут сказаться и демографические, биографические различия учащихся. Так, для выпускников городских школ, в которых тестирование применялось в старших классах, сама форма предъявления теста при вступительных экзаменах в вуз привычна. Это делает тестирование для них операцией знакомой, рутинной. В то же время для абитуриентов из отдаленного села, которые тесты ранее не видели никогда, уже сама ситуация тестирования связана со стрессовыми дополнительными факторами.

Возможна и другая, семантическая ситуация: нечётко сформулированное задание во вступительном тесте. Не исключено, что в худшем положении окажутся как раз лучшие выпускники престижных школ, привыкшие к более строгим и точным формулировкам задания. Подумав, что за «простотой» в задании скрываются ловушки, они могут начать искать в нём его «второй смысл», особенно в ситуации отбора при высоком конкурсе. Потратив на выполнение задания много времени и сил, не найдя явного подвоха, эти хорошо подготовленные выпускники в ситуации, когда задания построены с выборочным вариантом ответа (а так строится большинство стандартизированных тестов), могут остановиться не на более очевидном и правильном варианте, а на том, что «похитрее», – и ошибиться. А троечники думать не будут – выберут более очевидный вариант и получат свой балл.

При составлении текста параграфа использованы материалы: Михайлычев Е. А. Дидактическая тестология. – М.: Народное образование, 2001. – 432 с.

05.06.Рейтинг и оценка. Их определение и назначение

*Никогда не выходи в море с двумя хронометрами:
всегда бери один или три.
(пословица)*

Вычисляемый по итогам проведения тестирования рейтинг учащегося является числом, по величине которого пытаются судить о целом спектре достоинств и недостатков учащегося.

Рейтинг, выраженный всего одним числом (будем называть его, вслед за математиками, **скалярным рейтингом**) можно назвать объективным, но не всеобъемлющим показателем. Подобно тому, как победный результат многоборца не говорит о том, что он умеет лучше других – плавать или бегать.

Кардинально изменить ситуацию можно лишь перейдя к рейтинговой модели, где рейтинг состоит из нескольких показателей, является **векторным**.

Давайте попытаемся это сделать. Все задания теста разобьем на группы однотипных, с точки зрения преподавателя, заданий. Например, это могут быть задачи на применение стандартной формулы или вопросы, касающиеся ограниченной литературной темы. Такие группы назовем **субтестами**.

Весь тест, таким образом, будет состоять из нескольких равноправных **субтестов**. Количество заданий субтеста Z_s , предъявляемых учащемуся, должно быть настолько большим, чтобы законы случайности накладывали малое влияние на результат.

*Проецируя глобус, мы пропускаем одну его полусферу
сквозь другую и соединяем их проекции...
Карл ЛЕВИТИН*

Пусть результат выполнения заданий субтеста будет однозначно характеризовать только одно из качеств учащегося, например, умение решать арифметические задачи на проценты. Или, например, умение быстро ориентироваться в массивах исторических или химических данных.

Значение результата выполнения субтеста S можно выразить, например, в виде отношения количества верно выполненных заданий к количеству предъявленных:

$$r_s = \frac{Z_+}{Z_s} \cdot (1)$$

Выполнение заданий часто оценивают по-разному. Например, выполнение одного задания или группы заданий можно оценивать по такой шкале:

- Задание (задания) выполнено верно (+1 балл);
- Задание (задания) выполнено неверно (-1 балл);
- Решение (задания) содержит погрешность (от -1 до +1 баллов).

Не обсуждая достоинства разных шкал, примем за интервал возможных значений r_s отрезок от -1 до +1.

Полный результат прохождения теста выразится строкой чисел или **вектором рейтинга**

$$\bar{R} = \{r_1, r_2, \dots, r_n\}, (2)$$

где n – количество субтестов в тесте, $1 \leq S \leq n$. Поскольку абсолютные значения r_s , согласно (1), могут принимать значения от нуля до единицы, вектор $\bar{R}$ указывает только точки внутри куба или части куба в n -мерном пространстве результатов субтестов. Назовем это пространство **пространством результатов**. Точку, которую указал вектор рейтинга, будем называть результатом, а куб – область нахождения всех возможных результатов – кубом результатов.

Можно попытаться представить, как плотно заполнен куб результатов. Пусть субтестов будет всего два. Тогда куб результатов будет двумерным и превратится во внутреннюю часть единичного квадрата.

В зависимости от выбранного теста, выбранной группы участников и их мотивации, разные области квадрата будут заполнены с разной плотностью. Например, совсем мало результатов окажется вблизи «идеального» результата; очень много результатов может сконцентрироваться в области «наиболее вероятного» результата, который можно получить методом «случайного тыка», и т.д.

Теперь, после рассмотрения этого наглядного примера, можно по-другому взглянуть и на задачу процедуры тестирования. Она состоит в том, чтобы, после проведения тестирования, точки-результаты учащихся оказались сгруппированными, распределенными внутри нескольких (немногих) малых областей внутри куба результатов. Учащихся, чьи результаты попали в одну такую область, будут считаться близкими по уровню (качеству) знаний и получают одинаковые оценки.

Оси координат в рассматриваемом пространстве результатов лучше представить перпендикулярными друг другу, а соответствующие им единицы измерений – неравными. Хотя, некоторые авторы считают, что более наглядно будет, если представить оси координат наклонными, а неравенство единиц измерений выразить через величины синусов и косинусов углов взаимного наклона осей координат. Принятие одного из этих представлений – дело вкуса.

*Когда математик слышит о четырехмерных вещах,
его охватывает священный трепет.
Альберт ЭЙНШТЕЙН*

Короче говоря, результаты прохождения разных субтестов r_s могут (и должны) иметь разные относительные веса p_s , отражающие как относительную важность данного субтеста среди остальных, так и сложность представленных там заданий. Веса субтестов применяют, например, когда вычисляют скалярный рейтинг путем простого суммирования результатов прохождения субтестов по формуле:

$$R = \sum_{s=1}^n p_s r_s. \quad (3)$$

Если автор теста не считает нужным различать важность и сложность заданий, он принимает все веса равными единице. Чаще всего, авторы приписывают заданиям веса необъективно, волюнтаристски, опираясь только на свой опыт и здравый смысл.

В пространстве результатов можно определить т.н. метрику, или формализованное расстояние между результатами разных участников тестирования:

$$M = |\bar{R}_2 - \bar{R}_1|. \quad (4)$$

В отличие от случая работы со скалярным рейтингом (3), где расстояние между результатами участников является простой разностью рейтингов, в n -мерном случае расстояния можно определять по-разному, в соответствии с целями тестирования.

Нетрудно понять, что при подстановке в (4) вместо одного из векторов нулевого вектора, мы получим величину (длину) другого вектора. Смысл этой длины может быть разным.

Длина вектора рейтинга не имеет такой же роли, как величина скалярного рейтинга. В векторном случае значительно важнее знать не длину вектора рейтинга учащегося, а расстояние от него до вектора «идеального» рейтинга. Идеальный рейтинг – это рейтинг «идеального» учащегося, который выполнил абсолютно верно все задания теста.

Чем меньше расстояние от вектора рейтинга учащегося до вектора «идеального» рейтинга, и чем больше расстояние от вектора рейтинга учащегося до области «наиболее вероятного» результата, тем более высоко следует оценить знания учащегося.

Поэтому для того, чтобы распределить учащихся на группы в соответствии с их выявленными качествами, необходимо, в первую очередь, вычислить расстояния между всеми вычисленными векторными рейтингами.

*Спорим, что тринадцать одинаковых шаров,
как их не расположи – не могут касаться еще одного шара!
Исаак НЬЮТОН*

После вычисления расстояний между рейтингами всех участников, можно попытаться сгруппировать результаты участников так, чтобы расстояния между результатами каждой группы были меньше, чем расстояния до результатов других групп. Это делается с помощью методов факторного анализа. С помощью факторного анализа можно подобрать метрику (4) и величины весов p_s так, чтобы группы результатов были более отчетливы, более разделены между собой. Поскольку такое разделение можно сделать строго математически, без привлечения дополнительной эмпирической информации, результаты его можно признать объективными.

Методам факторного анализа посвящено немало книг. Долгое время они не получали практического развития в связи с низким уровнем применявшейся вычислительной техники. Сейчас, в начале XXI века, ситуация изменилась, и применение методов факторного анализа в тестологии превращается в обыденность.

Системы кумулятивного индексирования

*Все мои работы это игры, серьезные игры.
Мауриц Корнелис ЭСХЕР*

Системы кумулятивного индексирования качества и количества приобретенных ЗУН (системы ИКИ – Индивидуального Кумулятивного Индексирования) тесным образом связаны с рейтинговыми системами тестирования. Они являются действенным инструментом:

- повышения уровня качества ЗУН;
- активизации учебного процесса;
- контроля качества.

Как правило, разнообразные системы оценки ЗУН носят констатирующий характер, не влияют прямо на качество и редко доставляют удовольствие ученикам и педагогам.

Кумулятивные системы способны положительно влиять как на количество, так и на уровень качества образования. Идея кумулятивных систем проста:

- каждое действие (учебное) учащегося имеет назначенную цену в очках;
- получаемые каждым учащимся за свои действия очки суммируются (аккумулируются) в течение определенного учебного периода;
- неуспехи учащихся никак не отмечаются (при неверных ответах, пропусках занятий и пр. суммы очков не снижаются, остаются теми же);
- суммы очков зависят от даты сдачи контрольных работ (опоздал со сдачей контрольной работы – получи понижение прибавляемого количества очков);
- итоговые семестровые оценки выставляются педагогом на основании сравнения сумм очков (кумулянтов) в учебной группе. Скажем, первые 10% учащихся, имеющих наивысшие кумулянты, получают «пятерки», затем те, кто набрал от 70% до 90%, получают «четверки» и т.д.;
- учебные «долги», а также самостоятельно изученный дополнительный учебный материал также можно «сдавать», увеличивая свои кумулянты.

В 70-80-х годах прошлого века на ряде образовательных конференций обсуждалась информация об использовании в вузах СССР систем ИКИ. Такие системы интенсивно развивались в 50-60-х годах за рубежом. В СССР среди первопроходцев были преподаватели ФЕН НГУ – факультета Естественных наук Новосибирского государственного университета.

В 1982-1992 гг., при поддержке коллег из Ленинграда и Новосибирска, аналогичные эксперименты проводились преподавателями ДВГУ на химическом, математическом, физическом, историческом факультетах, а также в специализированных классах нескольких государственных и негосударственных средних школ Владивостока. При этом система ИКИ в ДВГУ отличалась от своих предшественниц.

Компьютерные технологии тогда только появились в СССР, были труднодоступны, и о том, чтобы применить компьютер на занятиях, приходилось только мечтать.

Системы ИКИ обладают рядом положительных и отрицательных черт. Соотношение их количества подвержено динамике во времени в пользу положительных.

Перечислим отрицательные черты системы ИКИ, проявляющиеся на этапе внедрения:

- В группах, где введена система ИКИ в учебных занятиях по какой-либо дисциплине, учащиеся начинают посвящать данной дисциплине значительное количество сил и времени, в ущерб другим дисциплинам. Это порождает отрицательные эмоции среди коллег-преподавателей;
- Система ИКИ противоречит традиционной пятибалльной системе (нарушается график отчетности и пр.), что вызывает негативное отношение к введению кумулятивной системы со стороны руководителей учебного процесса;
- Система ИКИ резко дифференцирует успехи учащихся, их суммы очков часто различаются в десятки раз. Это порождает нервозность родителей учащихся, т. к. они привыкли к традиционной пятибалльной шкале, «нивелирующей» различия в активности учащихся;
- После внедрения системы ИКИ возникает необходимость дополнительной работы с учащимися, не воспринявшими систему на начальном этапе и значительно отставшими (можно с ними работать и традиционно, однако через некоторое время они все равно захотят быть «как все» и воспринимают систему как должное);

- Система ИКИ может вызвать «неспортивные» явления среди учащихся (исчезновение у «передовиков» тетрадей, обуви), т.к. не всем желающим иметь пятерку в четверти выгодно присутствие на занятиях лидеров.

Перечислим положительные качества кумулятивных систем:

- у учащихся значительно возрастает интерес к дополнительной литературе и обучению опережающим образом;
- велика активность учащихся на консультациях и коллоквиумах;
- публикация кумулянтов приводит к возникновению духа состязательности, появляется стремление учащихся к овладению дополнительными знаниями;
- исчезают текущие «долги», а если появляются, то сдаются оперативно;
- на занятиях исчезает «негатив», связанный с боязнью учащихся неправильно решить задачу, не выполнить задание, получить двойку – учащиеся всегда имеют право и возможность улучшить свои результаты.

В целом, положительные качества систем ИКИ с лихвой перекрывает отрицательные. Последние же исчезают по прошествии периода адаптации.

По-видимому, наилучшим «ареалом» для внедрения кумулятивных систем является среда дистанционного обучения. Именно здесь знание оценивается компьютером и, следовательно, есть все предпосылки для отработки количественных критериев.

Текущие значения кумулянтов учащихся разных учебных заведений, будучи вычисленными по единообразной системе и собранными в компьютерной БД, вполне могут служить основой для проведения работ по сравнительному мониторингу качества образования.

Определение скалярного рейтинга

*Учение Ньютона-Лейбница всесильно,
потому что оно верно.
БЕЛОКОНЬ В. И.*

Результаты тестирований используются для вычисления рейтинга учащихся. В основном используют два вида рейтинга: текущий и итоговый.

1. *Текущий рейтинг* – метод учета достижений студентов в процессе выполнения заранее определенной совокупности заданий, по каждой учебной дисциплине. Текущий рейтинг призван обеспечить ритмичную учебную работу. Задания различной трудности заранее оценены зачетными баллами. Выполнение каждого задания увеличивает суммы баллов учащихся. Всегда известно – сколько баллов набрано каждым учащимся. Используется два варианта текущего рейтинга:
 - определяется минимально необходимая сумма зачетных баллов для получения зачета без сдачи экзамена или какого-либо дополнительного собеседования;
 - определяется количество баллов, необходимое для оценки по пятибалльной шкале.
2. *Итоговый рейтинг* – метод определения ранга (места) учащегося в группе по результатам обучения по одной или нескольким учебным дисциплинам.

Если бы все учащиеся имели разные тестовые баллы, то ранжирование не представляло бы труда: чем выше балл, тем выше и занимаемое место. Но обычно один и тот же тестовый балл получают несколько человек, и тогда возникает вопрос – как распределить всех испытуемых на все имеющиеся места?

Если рейтинг вычисляется в малой учебной группе, то количество мест равно количеству учащихся группы. После тестирования всем присуждают соответствующие места. Имеющим одинаковые баллы, даются и одинаковые или, связанные ранги, с последующим пропуском стольких мест, сколько имеется человек с одинаковым баллом. Этот позволяет равномерно и справедливо распределить испытуемых на все места рейтинга. Например, в группе 13 человек, то результаты ранжируются так:

РАНЖИРОВАНИЕ

Баллы	Частоты	Cum. F	Ранг
1	1	13	13
2	1	12	12
3	1	11	11
4	2	10	9-10
5	3	8	6-8
6	2	5	4-5
7	1	3	3
8	1	2	2
9	1	1	1

Иногда проводится рейтинг в группе с очень большим количеством учащихся. Тогда нет особого смысла расставлять их на все имеющиеся места. Если количество учащихся больше ста, то лучше применять так называемый процентный рейтинг, где в качестве нормы берётся 100 процентных мест. Суть процентного рейтинга состоит в распределении всех учащихся именно на эти сто процентных мест. Это удобно при сравнении массивов данных, полученных, например, по различным регионам в процессе аттестации выпускников.

Технология определения процентного рейтинга отражена в приведенной ниже таблице. Там использованы обозначения:

- 1 столбец – баллы испытуемых;
- 2 столбец – частоты;
- 3 столбец – кумулированные частоты (cum.f);
- 4 столбец – усредненные частоты, получают сложением значения cum.f в строке со значением cum.f ниже этой строки и делением на 2. Например; $1+0=1$, $\frac{1}{2}=0,5$; $1+1=2$, $\frac{2}{2}=1$; $3+1=4$, $\frac{4}{2}=2$;
- 5 столбец – процентные доли, получают делением усредненных f (4 столбец) на N. В данном примере N = 120
- 6 столбец – процент успешности испытуемых, в зависимости от полученного тестового балла (процентный рейтинг);
- 7 столбец – процентный ранг, который получают вычитанием $100 - R$, с последующим округлением до целого ранга; данный ранг присваивается испытуемому, имеющим соответствующий тестовый балл

ОПРЕДЕЛЕНИЕ ПРОЦЕНТНОГО РЕЙТИНГА

Балл X	Наср.f	Cum. F	Уср. F	Доли	% рейт.	Место
20	1	120	119,5	.996	99,6	1
19	2	119	118	.983	98,3	2
18	3	117	115,5	.962	96,2	4
17	6	114	111	.925	92,5	8
16	7	108	104,5	.871	87,1	13
15	6	101	98	.817	81,7	18
14	8	95	91	.758	75,8	24
13	10	87	82	.683	68,3	32
12	12	77	71	.592	59,2	41
11	13	65	58,5	.487	48,7	51
10	10	52	47	.392	39,2	61
9	9	42	37,5	.312	31,2	69
8	9	33	28,5	.237	23,7	76
7	8	24	20	.167	16,7	83
6	4	16	14	.117	11,7	88
5	3	12	10,5	.087	8,7	91
4	6	9	6	.050	5,0	95
3	2	3	2	.017	1,7	98
2	0	1	1	.008	0,8	99
1	1	1	0,5	0,004	0,4	100

Наполнение таблиц и описание технологии в данном параграфе соответствуют (в сокр. и адапт. виде) тексту четвертого раздела книги: *Аванесов В.С «Композиция тестовых заданий». Учебная книга. 3 изд., доп. М.. Центр тестирования, 2002 г. -240 с.*

Глава 06. Основные сведения из области тестологии

*Чем мельче становятся мысли и чувства,
тем вычурнее и красивее подбираются для них названия....
ПИСАРЕВ Дмитрий Иванович aforizm.kaminplus.ru*

Тестология – (от англ. – test – проба, испытание) – наука об измерении психофизиологических и личностных характеристик, а также объема и качества ЗУН.

Тестологи изучают и создают способы, методы, технологии измерений психофизиологических и личностных характеристик, а также объема и качества ЗУН. Тестологи создают тестовые комплексы, где реализованы достижения тестологии в виде совокупности технологий, рекомендаций, тестов, автоматизированных систем, устройств.

Тестовые комплексы применяют для аттестации ЗУН учащихся и абитуриентов, для измерения объема и качества навыков и умений при приеме людей на работу и пр.

Теорию тестовых педагогических измерений называют кратко IRT – Item Response Theory.

В IRT не ставятся и не решаются фундаментальные проблемы валидности и надёжности теста: тест там заранее считается надежным и валидным. Вычисления IRT сводятся к получению оценок параметров трудности задания и к измерению уровня ЗУН испытуемых. К достижениям IRT относят использование одной шкалы в измерениях значений параметров испытуемых и заданий теста. Это позволяет соотносить уровни ЗУН испытуемых с мерой трудности тестовых заданий. На этой возможности совместных оценок параметров испытуемых и заданий основан применяемый большинством тестологов математический аппарат.

В России пока не сложилось представление о педагогической тестологии как выделенной науке или научной области:

- Нет должности тестолога в реестрах отделов кадров;
- Нет вакансий для академиков – тестологов в Академии Наук и пр.;
- Педагоги не воспринимают тестологию, так как в ней слишком много математики;
- Математики, физики, инженеры не воспринимают тестологию из-за присущих ей традиционных для педагогики «туманных сентенций»;
- Программисты считают, что тестология – пройденный этап, и там уже нечего делать творчески мыслящей личности.

В настоящем пособии предпринята одна из попыток исправления этого положения путем популяризации: здесь основные положения и достижения тестологии изложены на языке понятном педагогу без ужасающих математических выкладок и терминологии. Иногда Вы встретите здесь иностранное слово, но лишь там, где тестологи не применяют русскоязычного аналога.

06.01. Тридцать три принципа конструирования теста, создания тестовых заданий и выбора технологии тестирования

*Никто не несет наказания за мысли.
(положение Римского права)*

Педагогическое тестирование имеет довольно хорошо разработанную принципиальную базу. Она очень хорошо описана в научной и научно-методической литературе. Но, к сожалению, нам не удалось встретить ни одного учебного пособия, где бы была систематически изложена вся совокупность принципов, поэтому здесь мы ее публикуем впервые.

1. **Валидность теста.** Результаты тестирования группы претендентов должны соответствовать объективным характеристикам, данным претендентам их руководителями, коллегами, преподавателями. Это важнейший принцип, которому должны быть подчинены усилия составителя теста. В соответствии с множеством указанных характеристик, принято различать виды валидности.
2. **Надежность теста и технологии тестирования.** Результаты тестирований подобных групп претендентов с помощью одного теста должны быть одинаковыми (подобными) и не зависеть от времени. Это второй по важности принцип, которому должны быть подчинены усилия составителя теста. Чаще всего, надежность можно охарактеризовать лишь умозрительно, поскольку:
 - повторное тестирование одной группы претендентов по одному тесту приводит к росту результатов за счет эффектов памяти, самообучаемости и пр.;
 - очень трудно найти подобные группы претендентов и доказать, что это так;
 - на результаты оказывает влияние предыстория групп претендентов, технические сложности исключения предварительного общения между ними и пр.

Надежность – принцип, пришедший в педагогику из техники. Это принцип, предъявляемый к инструменту технического измерения. Но педагогика – иная область с иными закономерностями, и поэтому принцип надежности часто обоснованно подвергается критике (см. релевантность теста).

3. **Релевантность теста и технологии тестирования.** Выберем два одинаковых по валидности теста (или, точнее, две технологии тестирования). С точки зрения дидактики, науки подчиняющей всё цели росту качества обучения, если повторное тестирование с использованием первого теста (первой технологии) приводит к более выраженному улучшению результатов, чем с использованием второго теста (второй технологии), то первый тест (первая технология) – лучше. Пусть результаты в обоих случаях улучшаются по-разному за счет разной выраженности эффекта самообучаемости претендентов, привлекательности процедуры и др. С точки зрения принципа надежности, оба теста (обе технологии) – не надежны. Чем лучше тест (технология тестирования) с точки зрения дидактики, тем он (она) хуже, исходя из принципа надежности. Поэтому, в педагогике, в отличие от техники, лучше говорить не о надежности теста (технологии), а о релевантности теста (технологии), т. е. степени соответствия принципам дидактики. Релевантность теста (технологии тестирования) – мера его способности к инициации процессов самообразования. Релевантность можно выразить численно, взяв отношения прироста усредненных результатов при повторном тестировании с использованием разных тестов (технологий).

4. **Научная достоверность.** В тест включаются только те элементы знания и связи между ними, которые являются объективно истинными. Соответственно, спорные точки зрения, вполне нормальные в науке, не включают в тест.
5. **Обобщенность.** Тема тестирования не должны быть излишне обобщенной.
6. **Значимость.** Значимое ТЗ отражает структурный информационный элемент дисциплины либо связь между структурными элементами, без которых знания становятся неполными, с пробелами.
7. **Соответствие современному знанию.** В ТЗ включаются только современные элементы знания. Исключением могут быть ТЗ, составленные специально для ведения учебных занятий процесса в рамках концепции «диалога культур».
8. **Соответствие источникам знания.** Все ТЗ должны включать адрес источника, который может быть включен и в демонстрируемую часть текста ТЗ в случаях, когда имеются разночтения в рекомендованной литературе.
9. **Репрезентативность.** Репрезентативный тест содержит совокупность ТЗ, отражающую все структурные элементы содержания дисциплины и их связи. Репрезентативная (представительная) БТЗ содержит совокупность ТЗ, достаточную для ведения вариативного контроля по любой из имеющихся в структуре дисциплины учебных тем. Репрезентативная БТЗ может включать ТЗ по дополнительным учебным материалам для проведения олимпиад и пр.
10. **Вариативность.** Содержание теста может и должно варьироваться по мере:
 - изменения содержания дисциплины;
 - изменения и появления новых научных знаний;
 - изменения цели тестирования;
 - изменения направленности на особенные выборки испытуемых.Для достижения достаточной вариативности теста в рамках учебной дисциплины необходимо, чтобы БТЗ содержала тестовые задания соответствующие разным учебным пособиям и разным способам изложения учебного материала.
11. **Сложность.** Для каждого ТЗ в процессе контроля определяют усредненную, выраженную числом, сложность. Сложность задания № j определяют двумя способами. Например, так: $q_j = 1 - p_j$, где p_j – относительное количество претендентов, выполнивших данное задание. Либо, как «логит», равный $\ln(q_j/p_j)$. Претендент, правильно выполняющий сложные задания, с большой вероятностью правильно выполнит и легкие. Сложность теста соответствует содержанию обучения и поддерживает высокий уровень мотивации претендентов. Не следует пренебрегать заданиями, которые выполняются большим количеством претендентов (от 90 до 100%). Иногда может показаться, что задание лишнее, поскольку слишком лёгкое. Надо проконтролировать выборку – вошли ли в нее слабо подготовленные претенденты. Они могут сознательно не допускаться к тестированиям руководством учреждения под благовидным предлогом. Аналогичная, но обратная ситуация может случиться и с наиболее сложными заданиями. Прежде, чем отвергать задания, которые никто не выполнил, необходимо проверять, проходили ли тестирование наиболее подготовленные учащиеся. Они могли в это время отсутствовать по причине участия в олимпиаде или в работе физико-математического кружка.
12. **Трудоемкость.** Для каждого ТЗ в процессе контроля определяют усредненную, выраженную числом, трудоемкость. Она характеризует количество элементарных операций (и мыслительных, и физических, и расчетных), которые необходимо выполнить в процессе работы над заданием. Отношение трудоемкостей заданий можно определить через усредненное отношение времен их выполнения. Относительное количество труда,

который тестируемые затратят (в среднем) на выполнение каждого тестового задания, должно быть пропорционально относительной значимости отраженного в задании элемента курса.

13. **Взвешенность.** Для придания результату тестирования большей объективности, все ТЗ после сеанса «взвешивают». Затем общий результат претендента вычисляют как сумму весов выполненных заданий. Иногда за вес ТЗ принимают его относительную сложность, т. е. долю претендентов, не сумевших его выполнить. Иногда вес приписывают заданиям волонтаристски, опираясь на преподавательский опыт. Оба подхода к вычислению весов не лишены недостатков. Например, при первом подходе возможен вариант, когда наивысший вес приобретет задание, которое показалось всем претендентам слишком легким, недостойным сиюминутного внимания, которое они, в погоне за рейтинговыми очками, оставили «на потом» и не успели к нему вернуться.
14. **Системность.** Задания в БТЗ связываются между собой общей структурой знаний. Такая связь может быть определена умозрительно либо методами факторного анализа.
15. **Комплексность и сбалансированность.** Следует гармонично сочетать в тесте задания на проверку знаний теоретического материала (понятия, законы, закономерности, гипотезы, факты, структурные компоненты теории), методов научной и практической деятельности, умений решать типовые задания. Отношение количеств ТЗ перечисленных типов должно соответствовать отношениям значимостей и объемов рекомендованной учебной информации.
16. **Соответствие содержания и формы.** Педагогический тест характеризуют как результат единения содержания заданий с наиболее подходящей формой.
17. **Разрешающая способность теста.** Сложность ТЗ не может быть только легкой (все выполнили), средней (половина группы выполнила) или трудной (никто не выполнил):
 - Выраженные численно сложности совокупности ТЗ в тесте должны равномерно заполнять тот интервал, который соответствует обученности претендентов;
 - Трудоемкость ТЗ должна быть такой, чтобы во временные рамки теста могли уложиться все претенденты, способные их выполнить.От сбалансированности ТЗ по сложности и трудоемкости зависит способность теста дифференцировать претендентов в соответствии с их свойствами. Эта способность теста, выраженная числом, называется разрешающей способностью.
18. **Дидактическая направленность теста и технологии тестирования.** Одних претендентов трудные ТЗ могут подтолкнуть к учебе, других – оттолкнуть от нее. Скучные, однообразные, излишне формальные совокупности ТЗ вредоносны в учении. ТЗ, в своей совокупности, должны включать элементы:
 - Привлекательности (юмористическая окраска условий и пр.);
 - Сюрпризности (разные и неожиданные словарные обороты, редко встречаемые в учебнике, обращение в условии заданий непосредственно к исполнителю с использованием данных регистрации и пр.);
 - Практической направленности (примеры из жизни, возможность применить результаты выполненного задания непосредственно в деятельности и пр.).

Сюрпризные задания включаются в тест для противодействия возникновению в процессе работы претендентов эффекта «ровной дороги». Трудоемкость и сложность ТЗ, в совокупности, должна быть такой, что хотя бы пару-тройку заданий должны правильно выполнить абсолютно все претенденты. Это заставит слабых претендентов поверить в свои силы, увлечет, инициирует соревнование, даст основания для похвалы.

Тест (технология тестирования), в соответствии с принципами дидактики, должен (должна) не только дифференцировать и измерять свойства претендентов, но и обладать свойством инициирования их самообучения и проявлять их стремление к повышению качества знаний, умений, навыков.

19. **Оптимальность теста.** В тест, как правило, включаются только те ТЗ, содержание которых не дублируется и равномерно покрывает, в своей совокупности, учебный материал. Часть ТЗ исключают из теста, когда известно, что:

- претенденты не знают соответствующей части материала или, наоборот, знают ее гарантированно хорошо;
- претенденты не имеют навыков работы и поэтому не справятся с трудоемкими заданиями.

Тест считают более оптимальным, если с его помощью можно измерить такие же характеристики и провести дифференцирование претендентов за меньшее время.

20. **Темперированность сложности.** Последовательность заданий теста такова, что каждое последующее ТЗ, по мнению составителя, труднее предыдущего. Этот принцип, применяющийся в бланковом тестировании, нередко подвергается критике, поскольку действует лишь «в среднем». Если претенденты учились у разных преподавателей и по разным учебникам, предсказать ряд возрастающей трудности для большинства из них – большая проблема. А «усредненный» ряд не приносит желаемого эффекта повышения разрешающей способности. Поэтому в компьютерном тестировании чаще применяют иной принцип: претендентам позволяют самостоятельно выбирать порядок выполнения заданий. (см. «Свобода выбора ТЗ»)

21. **Темперированность трудоемкости.** Последовательность заданий теста такова, что каждое последующее ТЗ, по мнению составителя, более трудоемко, чем предыдущее. Применяется, когда оптимизируют тесты для измерения работоспособности, внимательности и подобных характеристик претендентов.

22. **Свобода выбора.** Принцип составления теста, пришедший на смену принципу «возрастающей сложности». Принцип свободы выбора в технологиях компьютерного тестирования проявляется в том, что претендентов не ограничивают в выборе последовательности выполнения заданий. Практика показывает, что при этом претенденты сами ищут и интуитивно находят для себя индивидуальный ряд ТЗ возрастающей сложности и трудоемкости.

23. **Вариативность теста.** Каждое обращение к БТЗ приводит к генерации нового варианта теста в соответствии с алгоритмом, ключ которого может формироваться в соответствии с регистрационными данными, датой и временем сеанса. Алгоритм, ориентирующийся при составлении варианта теста на личность претендента, может исключать возможность повторения ТЗ при повторном тестировании.

24. **Адаптивность теста.** Содержание теста адаптируют для придания тесту большей разрешающей способности в известной группе претендентов. Для адаптации теста к свойствам группы претендентов и каждого претендента в отдельности, необходимо наличие в БТЗ такого набора классифицированных заданий, что возможно построение вариантов теста нацеленных на измерение конкретных свойств:

- Тест дробят на субтесты, а сеанс – на ступени;
- Перед построением субтеста очередной ступени проводится анализ результатов выполнения предыдущих субтестов;

- При анализе конкретизируются результаты произведенных измерений свойств претендентов, затем, если объективная дифференциация группы претендентов еще не возможна, составляется очередной субтест;
 - Действия повторяются на очередной ступени сеанса до тех пор, пока не будет достигнута требуемая степень дифференциации.
25. **Полнота.** В банке ТЗ должны быть отражены абсолютно все структурные элементы содержания дисциплины и связи между ними.
26. **Целесообразность.** Содержание теста зависит от цели тестирования. Например, возможны ситуации:
- Если нужно отобрать самых подготовленных учащихся на олимпиаду, то задания должны быть сложные, ибо только с помощью таких заданий можно отобрать лучших;
 - Если претенденты учились по разным учебникам, то формулировки отобранных для теста ТЗ должны соответствовать каждому.
27. **Верифицируемость результата.** Тест может содержать равномерно распределенные среди всех заданий такие ТЗ, которые характерны разной формой и равным содержанием. Установив корреляцию результатов их выполнения можно установить, как претендент выполнял задания – вдумчиво либо методом «случайного тыка».
28. **Сбалансированность.** Количество разделов БТЗ и включенных в них заданий должны коррелировать с содержанием, относительными объемами информации разделов и формой дисциплины.
29. **Стилистическая непротиворечивость.** ТЗ формулируются в привычной для претендентов форме, с использованием словарных оборотов из рекомендованных им учебных пособий. Исключения – сюрпризные задания, включаемые в тест для противодействия возникновению в поведении претендентов эффекта «ровной дороги».
30. **Логическая и семантическая непротиворечивость.** Среди логических требований – определенность предмета измерения. Близость дисциплин затрудняет определение предметной принадлежности ТЗ. Чем меньше пересечений дисциплин и их разделов, тем чище, определеннее выражается в тесте их содержание. Логически ТЗ можно ассоциировать с утверждением. В случае правильной подстановки (Форма ТЗ №2) утверждение превращается в истинное, в случае неправильной – в ложное. Привычные для педагога вопросы не являются утверждениями, а потому содержание вопроса не может определяться как истинное или ложное. Вопросы применяются в ТЗ закрытой формы. Следует избегать несоответствий рода, числа, склонения в формулировках заданий и вариантов их выполнения. В ТЗ нет места двусмысленностям, например:
- на вопрос «Где находится Москва?» можно ответить «на берегу Москвы-реки», «в Европе» и пр.;
 - на вопрос «Кто первым полетел в космос?» можно ответить «советский человек», «Юрий Гагарин», «Лайка» и пр.;
 - подлежащее в именительном падеже легко путается с прямым дополнением в винительном падеже, например: «мать любит дочь» (пример В. С. Аванесова);
 - на вопрос «Когда началась Великая отечественная война?» можно ответить; «в июне», «в XX веке» и пр.;
 - в заданиях открытой формы следует указывать род, падеж и пр. ожидаемых слов. Например, на вопрос «Как называется судно, колющее лед?» можно ответить и «ледокол», и «ледоколом»;

- среди заданий на установление соответствия или на установление последовательности встречаются варианты, трактуемые в разных источниках по-разному (причины и последовательности исторических событий).

31. **Дистрактивность.** Варианты выполнения заданий (Форма ТЗ №1) формулируются так, чтобы правильные варианты были похожи на неправильные, и наоборот. Формулировки правильных вариантов должны точно соответствовать материалу учебного курса. Следует избегать характерной ошибки начинающих авторов, когда формулировки верных вариантов, в большинстве, длиннее формулировок неверных вариантов. Следование принципу дистрактивности значительно повышает валидность теста.

32. **Избыточность.** Часто случается так, что в группу претендентов попадают учащиеся, учившихся по разным программам и учебникам. Тест составляется один для всех и поэтому, чтобы «сравнять шансы», предпринимаются следующие шаги:

- в тест включаются ТЗ, составленные в формулировках разных учебников, примерно в равных пропорциях;
- эквивалентные по информации задания группируются, и участники информируются, что необходимо и достаточно правильно выполнить только одно задание группы;
- количество ТЗ в тесте превышает норму, которую мог бы выполнить усредненный участник сеанса.

Иногда завышают количество заданий теста выше нормы и из других соображений:

- чтобы увеличить шанс слабо успевающим учащимся хоть в чем-то проявить себя и получить похвалу;
- чтобы исключить при проведении образовательной олимпиады накладки связанной с тем, что несколько участников выполняют правильно все предложенные ТЗ.

33. **Деловая игра.** Технологии тестирования конструируются исходя из рекомендаций дидактики, утверждающей, что наилучших успехов можно достичь при игровом обучении. Игровая технология позволяет сделать процесс тестирования привлекательным и достичь высоких успехов за счет:

- естественной индивидуализации;
- возникновения соревнования;
- коллективности;
- публичности.

Ниже мы подробнее опишем смысл и практику применения перечисленных принципов.

06.02. Педагогический тест

Эволюция теста от создания до применения на практике проходит через три этапа.

На первом этапе создается задание в тестовой форме.

На втором – создаются тестовые задания.

На третьем этапе моделируются тесты.

В. С. АВАНЕСОВ

Ключевым понятием тестологии является понятие «педагогический тест».

Педагогический тест –

- система заданий такая, что результат их выполнения группой претендентов позволяет достаточно надежно ранжировать их (присвоить им порядковые номера) по качеству обученности, количеству имеющихся знаний;
- система стандартизированных заданий, результат выполнения которых позволяет с заданной степенью точности измерить знания, навыки и умения испытуемого.

Иногда под понятием «тест» понимают средство педагогического измерения; или стандартизированный метод исследования, предназначенный для количественных и качественных оценок характеристик человека. Путаница понятий (метод – средство – совокупность – система) здесь происходит от разночтений переводной литературы, а также от существования множества сокращений, разночтений и жаргонных выражений в специальной литературе по психологии, компьютерной технике, педагогике.

Форма приведенного определения педагогического теста предполагает некоторую неопределенность, невозможность точного выражения, поиск пути (Дао) построения идеала. Тестологи говорят, что идеального педагогического теста еще не создано. В этом смысле, в смысле поиска Дао, и следует понимать многие из определений тестологии.

Почему тестологи не довели определения до чисел? Почему определения тестологов, в то же время, так похожи на определения из новейших математических направлений – теории нечетких множеств и пр.? Наверное, потому, что место тестологии в общей картине наук «лежит» как раз на стыке педагогики и математики. Тестология еще только становится на ноги, и излишне математизированный педагогический язык здесь отпугивает педагогов и смешит математиков.

Педагогическим тестам, а также результатам их применения (результатам тестирования) приписывают характеристики и свойства. Важнейшие свойства теста – **валидность** и **надежность**.

Надежность теста – степень устойчивости (неизменности) результатов при повторном тестировании той же (такой же) группы претендентов (испытуемых).

Валидность теста – степень пригодности теста для измерения именно тех качеств претендентов (испытуемых), которые он по замыслу должен измерять.

Валидности и надежности обычно не придают точных количественных значений, используя при сравнении их, для различных тестов, лишь отношения порядка:

- **валидность** педагогического теста тем выше, чем более результаты его применения соответствуют решению проблемы ранжирования претендентов по качеству обученности;
- **надежность** педагогического теста тем выше, чем более устойчивы (одинаковы) результаты его применения при тестировании однородных групп претендентов.

Специалистами наиболее часто обсуждаются следующие характеристики тестов:

- **время измерения**, т. е. сколько времени необходимо тестировать группу претендентов для того, чтобы результаты тестирования стали достаточно валидными и надежными;
- **привлекательность**, т. е. насколько охотно претенденты избирают для своего тестирования тот или иной тест;
- **простота создания** (очень важная характеристика для условий поточного производства и для обеспечения экономичности производства);
- **простота применения** (иными словами, насколько велики требования к квалификации специалистов его применяющих или обслуживающих, а также к технике и оборудованию).

Этим характеристикам так же, как и приведенным выше свойствам, трудно сопоставить точные количественные критерии. Однако их можно сравнить, основываясь на мнениях экспертов, и, соответственно, ранжировать.

Дифференцирующая способность теста оценивается с точки зрения соответствия уровня его трудности уровню подготовленности претендентов. Бесполезно давать слабым учащимся трудные задания и знающим студентам легкие задания. В обоих случаях претенденты не будут различаться результатами. Измерение не состоится по причине несоответствия уровня трудности теста уровню подготовленности. Эффективный тест соответствует по трудности заданий уровню подготовленности испытуемых. Если включить в тест задания для измерения на всем диапазоне изменения трудности, то снижается точность измерения на отдельном участке. И наоборот, для точного измерения знаний претендентов одинакового уровня подготовленности, требуется иметь больше заданий соответствующего уровня трудности. Эффективный тест (с минимизированным количеством заданий) не может быть эффективным на всем диапазоне подготовленности претендентов.

Релевантность тестового комплекса. Редко обсуждается очень важная, на наш взгляд, характеристика тестовых комплексов (а также технологий тестирования и учебных пособий) – степень оптимальности соответствия учебным целям. Такую характеристику можно назвать релевантностью. Понятие «релевантность» по своему смыслу близко понятию «валидность», но охватывает больший смысловой круг, поскольку ранжирование претендентов является лишь малой (безусловно, важной) частью учебного процесса.

Педагогический тестовый комплекс создается с учетом единства семи систем:

- системы знаний проверяемой учебной дисциплины;
- системы заданий;
- системы предъявления заданий испытуемым (алгоритм, интерфейс);
- сценария тестирования;
- системы методического обеспечения;
- системы технического обеспечения;
- системы статистической обработки результатов испытуемых.

Информация, которую содержит педагогический тестовый комплекс, меняется со временем: включаются новые задания, меняются статистические веса заданий, меняются сценарии и пр.

06.03. Семь оснований для классификации и двадцать видов компьютерных педагогических тестов

*Тьма дорог туда низводит,
Ни одной оттуда нет,
И отшедший не приходит
Никогда опять на свет.*

ШИЛЛЕР Иоганн Фридрих, aforizm.kaminplus.ru

Педагогический тест можно определить не только относительно группы испытуемых, как это сделано выше. Но и относительно одного испытуемого, взятого в отдельности.

Педагогические тесты классифицируются в соответствии с выделенными группами.

*Сперва любовь, потом брак.
Сперва пламя, потом дым.
Никола ШАМФОР*

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №1

(по тематике заданий)

Тест гомогенный – совокупность стандартизованных заданий по одной учебной дисциплине (разделу дисциплины). Система заданий возрастающей трудности для объективной и эффективной оценки уровня подготовленности учащихся по одной из учебных дисциплин. Гомогенные тесты распространены больше других. Они создаются для контроля знаний по одной учебной дисциплине или по одному разделу объемной учебной дисциплины.

Тест гетерогенный – совокупность стандартизованных заданий, созданная с целью измерения знаний по нескольким учебным дисциплинам (разделам дисциплин), но не требующая одновременного привлечения знаний из различных дисциплин (разделов дисциплин). Гетерогенные тесты используются для комплексной оценки выпускника школ, для отбора абитуриентов при приеме в вузы. Гетерогенный тест состоит из гомогенных. Вычисление результата ведется для каждого гомогенного теста в отдельности с последующей интеграцией в единый рейтинг.

Тест интегративный – система заданий для общей диагностики подготовленности выпускника образовательного учреждения. Задания таковы, что для их выполнения требуются синтетические знания по нескольким учебным дисциплинам.

Тест адаптивный – система заданий с известными свойствами и параметрами. Задания предъявляются, посредством компьютера, в порядке, зависящем от ответа испытуемого на предыдущее задание. С помощью разных стратегий предъявления заданий удастся значительно сократить время компьютерных тестирований без потери точности измерений.

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №2

(по процедуре формирования очереди представления)

Тест смешанный – совокупность стандартизованных заданий разной сложности, предъявляемых в случайном порядке.

Тест возрастающей сложности – совокупность стандартизованных заданий, предъявляемых в порядке возрастающей сложности.

Тест адаптивный – система стандартизованных заданий с известными (заданными) параметрами трудности и дифференцирующей способности. Очередность предъявления каждого задания зависит от успешности выполнения испытуемым предыдущего задания. При верном выполнении предыдущего задания – степень трудности последующего возрастает, и наоборот.

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №3

(по количеству претендентов)

Тест индивидуальный – тест, созданный для отдельного испытуемого или отдельного типа испытуемых.

Тест групповой – тест, который группа испытуемых выполняет совместно (коллективно, сообща).

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №4

(по форме представления заданий)

Тест стандартный – тест для применения в наиболее распространенной методике диагностики испытуемых, согласно которой они выполняют одну или схожие совокупности заданий (варианты гомогенного теста) за одинаковый ограниченный период времени и в одинаковых условиях. При этом содержание заданий испытуемым заранее не объявляется, а условия проведения тестирования исключают возможность общения с целью подсказки. Проверка результатов производится вручную или с привлечением средств автоматизации. Результатом диагностики (измерения) для каждого испытуемого является число – рейтинг, позволяющее упорядочить испытуемых в рамках группы в соответствии с выявленным уровнем знаний (умений, свойств, характеристик).

Тест скрытый – тест созданный для выявления не актуализированных способностей и оценивания, насколько успешно испытуемый мог бы осваивать знания, умения и пр. При скрытом динамическом тестировании преподаватель активно взаимодействует с испытуемым, консультирует его, выдаёт конкретные рекомендации по оптимизации действий.

Тест игровой – тест предназначенный для выявления ЗУН испытуемых в процессе компьютерной игры. Форма деловой игры применяется на стандартных учебных занятиях в разных видах

Тест мультимедийный – тест, в котором информация разной природы присутствует равноправно и взаимосвязано – тексты, изображения, звуки, анимация, фрагменты фильмов. В качестве примера тестирующих и обучающих компьютерных игр можно назвать тренажеры для лиц военных профессий – летчиков, танкистов.

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №5

(по процедуре анализа результатов выполнения)

Тест бланковый – тест, проводящийся с использованием бланков – стандартизованных бумажных, картонных, пластмассовых листов, на которых нанесена тестовая информация и присутствуют места для занесения информации испытуемого. Информация испытуемого, символизирующая результаты выполнения им тестовых заданий, может наноситься на бланк путем перфорации отверстий, нанесения знаков пишущей ручкой, вырезания фрагментов ножницами. Информация анализируется путем механического сравнения изме-

нений материала бланка с образцом. Механическое сравнение может производиться путем продевания спиц в технологические отверстия бланка и пр.

Тест машинный – тест, результаты которого обрабатываются автоматически при помощи компьютерных программ и специальных технических устройств

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №6

(по процедуре информационной защиты)

Тест безвариантный – тест, состоящий из фиксированного количества тестовых заданий. Все претенденты получают для выполнения единственный тест.

Тест многовариантный – совокупность безвариантных тестов (вариантов). Задания с одинаковыми номерами близки между собой во всех вариантах. Вариантов составляется так много, чтобы все соседи каждого участника тестирований имели разные варианты.

Тест со случайной выборкой – тест, формирующийся из базы тестовых заданий непосредственно перед началом тестирования. Это делается с помощью какой-либо процедуры, генерирующей случайный список номеров заданий.

ОСНОВАНИЕ КЛАССИФИКАЦИИ ТЕСТОВ №7

(по цели)

Тест достижений – тест контроля успешности учебной деятельности испытуемого. Тесты достижений подразделяются по целям и задачам измерения:

- общей результативности;
- степени усвоения знаний;
- общего развития;
- интеллектуальных способностей;
- функциональных возможностей;
- профессиональной пригодности.

Тест усвоения знаний – тест для определения уровня усвоения учебного материала учащимися по конкретной учебной дисциплине, теме. Тестирование проводится сразу после изучения или спустя некоторый большой промежуток времени (тест остаточных знаний).

Тест остаточных знаний – тест для определения уровня усвоения учебного материала учащимися по конкретной учебной дисциплине, теме. Тестирование проводится сразу после изучения или спустя некоторый большой промежуток времени (тест остаточных знаний).

06.04.Десять недостатков педагогических тестов

*Если женщина сердится,
значит она не только не права, но и понимает это.
(источник – Интернет)*

Тест – это совокупность или система нескольких ТЗ. Характерные встречающиеся недостатки тестовых заданий мы рассмотрим ниже, а здесь сосредоточимся на недостатках именно тестов. Перечислим основные недостатки:

- Неравномерность «покрытия» учебного материала, отсутствие равенства отражения подобных по информационному весу и важности разделов дисциплины;
- Необъективность весов ТЗ, т. е. количеств баллов, которые претенденты получают в свою «копилку» за выполнение каждого задания; вес не отражает среднестатистическую сложность и важность задания;
- Малое количество ТЗ в тесте или одновариантность теста, предполагающие простоту пользования шпаргалкой и тестирование не знаний, а способности запоминать последовательности букв и цифр. Задания для составления варианта теста должны выбираться из БТЗ случайным образом, а БТЗ должна быть столь велика, чтобы меры по обеспечению ее секретности были не нужны (более 300 заданий для этого достаточно);
- Отсутствие ориентации заданий на учет присущих человеку свойств и качеств (настроение, усталость, темперамент, возраст, пол, национальность), т. е. отсутствие в тесте элементов гуманистичности или толерантности;
- Если в тестирующую программу не заложены методики случайного выбора порядка предъявления ТЗ, и тест создавался одним человеком, номера правильных вариантов иногда можно достаточно надежно определить с помощью калькулятора и простых алгоритмов, созданных кибернетиками еще в середине прошлого века. Человеку только кажется, что он распределил варианты в случайном порядке;
- Присутствие связи между последовательными заданиями, когда для верного выполнения одного задания учащийся пользуется или вынужден пользоваться информацией из других заданий. Такая, заложенная автором, связь легко прослеживается в бланковом варианте теста, но исчезает в компьютерном;
- Вопросы и ответы однообразны, нудны. Это отрицательно воздействует на учащихся, вызывают так называемый эффект «ровной дороги» – известный в практике водителей-дальнобойщиков эффект засыпания за рулем, отвлечение и рассеяние внимания;
- Инструкция для участников дана только в одном ТЗ в предположении, что участник выполняет их последовательно. Это приводит к неразберихе и ошибкам при отклонениях;
- Отсутствие элементов адаптивности;
- Отсутствие элементов психофизической разгрузки.

Часто эти недостатки взаимосвязаны, это легко увидеть при некотором опыте составления заданий. Есть и другие недостатки тестов, выявляемые уже в процессе математической обработки результатов тестирования. Недостатки, связанные с валидностью и надежностью теста, обсуждены в специальном разделе настоящего пособия.

06.05. Валидность и надежность теста

*Думать – самая трудная из работ.
Видимо, поэтому так мало людей ею занимаются
Генри ФОРД, aforizm.kaminplus.ru*

Основными свойствами и параметрами качества тестов считаются валидность и надёжность. Эти свойства, в отличие от многих других, обсуждаемых в педагогике, выражаются числами.

Валидность отражает пригодность теста для измерения того, что он по замыслу должен измерять. Измерение валидности теста может включать ряд процедур, главной из которых является валидация.

Валидация производится путем математического сравнения результатов тестирования с успешностью выполнения соответствующей практической деятельности испытуемыми. При этом учитывается шкалированное мнение руководителей группы прошедших тестирование испытуемых об их конкретных исполнительских качествах.

Валидность считается достаточно высокой, если коэффициент корреляции будет более 0,6. При значении коэффициента корреляции 0,45 – 0,65 валидность считается вполне удовлетворительной.

Считается, что стопроцентно валидных тестов нет. Различают следующие виды валидности:

- содержательную – степень соответствия теста программам обучения и образовательным стандартам;
- критериальную – степень соответствия результатов тестирования внешнему, не относящемуся к тесту критерию;
- квалитетическую – степень связи результатов математической обработки результатов тестирования и их интерпретации;
- прогностическую – степень полноты достижения цели тестирования;

и др. Наибольший «вес» здесь приходится на содержательную валидность. В Приложениях содержится обзор и более полное изложение разных видов валидности и методов ее обеспечения.

Надёжность теста. Контролируемые и неконтролируемые факторы влияния

*Чтобы произвести впечатление на окружающих,
дети стремятся выглядеть старше,
мужчины – умнее,
женщины – моложе и глупее.
(неизвестный автор)*

Надёжность теста определяется как устойчивость результатов при повторном тестировании на той же (такой же) выборке испытуемых. Считается, что при коэффициенте корреляции

ляции результатов повторного тестирования более 0,75 уровень надежности теста приемлем.

Надёжность теста проверяется относительно

- временных изменений;
- выбора конкретных заданий;
- конкретных индикаторов;
- роли индивидуальности персонала при накоплении и обработке данных;
- аспектов процедуры тестирования.

Вычисление ошибки измерений – вероятных пределов колебаний измеряемой величины – основано на понятии надёжности.

Ни один тест не является абсолютно надёжным. Дисперсия ошибки (мера отклонения результатов сеансов тестирования) отражает случайные колебания, вызываемые неконтролируемыми факторами:

- тренинг и дообучение претендентов в течение периода измерения надёжности;
- изменениями погоды;
- случайными отвлекающими моментами;
- обучаемостью претендентов в течение сеанса и в перерывах между сеансами;
- мотивированностью претендентов и ее изменениями;
- адаптацией, привыканием претендентов к форме проведения тестирования;
- изменениями состояния претендентов.

Первый из этих факторов, наиболее влияющий на результаты и «путающий все карты», – обычное явление, создаваемое родителями, руководством учебных заведений, добросовестными педагогами, не желающими, чтобы их воспитанники предстали перед экспериментаторами в худшем виде.

Стандартизация и компьютеризация тестов уравнивают условия тестирования и способствуют повышению надежности.

Стандартный набор данных о тесте, предназначенном для широкого употребления, обязательно должен включать сведения о мере его надёжности.

Надёжность обычно вычисляется с помощью коэффициента корреляции произведения моментов К. Пирсона (его можно найти в статистических справочниках).

При измерении надежности следует обращать внимание на критерий целесообразности, помнить об отношении «затраты – польза». Следует контролировать, в каком интервале времени измерялась надежность, произошли ли за этот период какие-либо события, способные повлиять на результаты.

Существуют различные типы надёжности и подходы к их вычислению. Педагогу, апробирующему тест на своих классах, важнее всего владеть техникой измерения:

- ретестовой надёжности;
- надёжности эквивалентных форм.

Надежность не обязательно предполагает высокую валидность. Среди тестологов распространено поверье, что на практике всегда выполняется следующее выражение: валидность

< надежность. Иными словами, значение валидности теста не может превышать значение его надежности, какие бы процедуры их определения не использовались. Трудно этому поверить, пока не проведено математическое сравнение процедур вычисления этих значений.

Три типа надежности теста и три способа ее определения

Тестовые методы

- это стандартизированные инструментальные технологии оценки знаний, которые обязательно базируются на массовых статистических исследованиях и оказываются тем более объективными, чем шире охват этих статистических исследований

А. Г. ШМЕЛЕВ

Надежность и валидность тестов могут быть значительно повышены, если их качественные и содержательные характеристики будут связаны со статистическими данными, полученными при обработке больших массивов результатов тестирования испытуемых.

Разработано несколько способов определения и повышения надежности теста.

- **Метод повторного тестирования** (метод ретеста). Двукратное или многократное использование одного и того же теста в одной группе испытуемых. Достоинство метода заключается в простоте его использования, ясности основных посылок, простоте сравнений и расчетов. К недостаткам относят неопределенность выбора временного интервала между опросами. Этот интервал может колебаться от нескольких минут до нескольких дней, месяцев и даже лет. Естественно, что при этом по-разному проявляются факторы: запоминание или, наоборот, забывание, влияние опыта, полученного в первом опросе на второй, влияние общения испытуемых между собой после первого опроса.
- **Метод параллельного тестирования** (метод эквивалентных форм). Одной и той же группе испытуемых дается вначале одна форма теста, и после перерыва – другая. Затем вычисляется величина коэффициента корреляции верности выполнения заданий, которая и принимается за значение коэффициента надежности. Если между предъявлением обоих форм имеется значительный временной интервал, то коэффициент надежности называют по-другому: коэффициентом эквивалентности или коэффициентом стабильности.
- **Метод раздельного коррелирования.** В основе лежит допущение о параллельности не только отдельных форм, но и частей внутри одной формы теста. Для получения величины коэффициента надежности сравниваются результаты выполнения частей теста. В зависимости от способа деления теста могут меняться значения коэффициента. Часто применяемая процедура деления теста – это сведение в одну часть результатов респондентов в нечетных номерах высказываний и в другую – четных. Суммирование баллов в этих половинах теста дает два вектора, коэффициент корреляции между которыми и служит коэффициентом надежности теста. Его называют еще коэффициентом внутренней состоятельности теста. Этот метод имеет преимущество перед другими, поскольку позволяет оценить надежность при однократном тестировании.

Чаще всего рассматривают три типа надежности:

- **Ретестовая надежность.** Измеряется при повторном проведении того же самого теста на том же контингенте и в тех же условиях. Ее аналогом является проверочная кон-

трольная работа, проводимая, когда результаты первой контрольной кажутся сомнительными. Ретестовая надёжность обычно подсчитывается с использованием коэффициента корреляции моментов Пирсона. Чтобы повысить ретестовую надёжность, надо отбирать из первоначального, апробируемого варианта теста те задания, на которые испытуемые дают устойчивые ответы. Чем выше ретестовая надёжность, тем менее чувствительны результаты к влиянию неконтролируемых факторов.

- **Надёжность эквивалентных форм.** Измерение надёжности эквивалентных форм проще измерения ретестовой надёжности. Это вычисление корреляции результатов выполнения двух форм одного и того же теста, считающимися равноценными. Их называют также параллельными, взаимозаменяемыми, сопоставимыми, подобными. Здесь коэффициент надёжности одновременно отражает временную стабильность теста и согласованность результатов сеансов по двум формам. Только если варианты сеансов следуют один за другим, можно точно измерить надёжность эквивалентных форм теста. Для снижения влияния эффекта самообучения претендентов, меняют способ предъявления эквивалентных форм сходных заданий. На эквивалентность также проверяют:
 - инструкции;
 - персонал;
 - временные рамки работы;
 - формат бланков.
- **Надёжность теста на скорость.** Для проверки надёжности тестов на скорость считаются наиболее эффективными метод ретеста и метод эквивалентных форм. Есть приёмы разделения полного времени выполнения теста на четыре части с регистрацией результатов отдельно для каждой четверти.

При составлении текста параграфа использованы материалы:

1. Кречетников К. Г. *Задания в тестовой форме и методика их разработки: Учебно-методическое пособие* – Владивосток: Изд-во Дальневост. ун-та, 2002. – 36 с.
2. Михайлычев Е. А. *Дидактическая тестология*. – М.: Народное образование, 2001. – 432 с.

Развитие понятия валидности и надёжности теста

*Я полагаю, что мы пришли после других для того,
чтобы делать лучшие их, чтобы не впасть в их ошибки,
в их заблуждения и суеверия.
П. Я. ЧААДАЕВ*

Свойства валидности и надёжности обычно приписывают тесту, набору тестовых заданий. Это правильно, пока используется одна-единственная процедура тестирования. Но таких процедур уже много: есть адаптивные процедуры, есть тестирующие деловые игры. Они отличаются от стандартов пятилетней давности и непрерывно развиваются. Это направление деятельности развивается и нельзя забывать об этом, строя понятийный аппарат тестологии.

Сама процедура тестирования, то есть способ представления и организации теста, сильно и по-разному влияет на степень решения образовательных задач, на измеряемые величины

валидности и надежности теста. Поэтому целесообразно приписывать свойства валидности и надежности не тесту как таковому, а всей применяемой тестовой технологии, включающей и тест, и интерфейс программного комплекса, и процедуру публикации результатов и др. В этом случае результаты измерения валидности и надежности станут более объективными.

Кроме этого, в связи с развитием технологий, следует ввести в рассмотрение новые свойства, характеризующие степень соответствия образовательным задачам не теста, как такового, а всей технологии тестирования.

Часто считают, что тестирование предназначено для выполнения единственно значимой функции: измерения уровня (количества, качества) ЗУН учащегося (группы учащихся). В связи с этим, самой важной характеристикой процедуры тестирования считают надежность. Иными словами, считается, что тестирующий комплекс хорош тогда, когда результаты тестирования при повторении не меняются.

Требование воспроизводимости (понимается как синоним надежности) обусловлено отнюдь не образовательными потребностями или нормами, а спецификой мышления (образования) создателей тестирующих комплексов – технарей, часто подсознательно отождествляющих человека с каким-либо устройством.

На самом же деле, образовательные потребности и дидактические принципы направлены как раз на то, чтобы результаты тестирований не воспроизводились, а повышались. С этой точки зрения, процедура тестирования и тестирующий комплекс хороши лишь тогда, когда они выполняют главную образовательную задачу: побуждают учащегося к самостоятельному повышению уровня качества ЗУН и соответствующей самопроверке. Рецепт известен: процедура тестирования должна включать не только механизм измерения, но и элементы состязательности и привлекательности (деловой игры, сюрпризности и пр.). А каждое прохождение этой процедуры должно побуждать у учащегося стремление к самосовершенствованию, т. е. к самообразованию, и приводить к повышению результата измерения. И мера этого роста – коэффициент релевантности (соответствия образовательным целям и принципам) – одна из важнейших характеристик процедуры.

Эта точка зрения вызывает немало споров, в том числе и в среде профессиональных педагогов, иногда считающих, что в таком серьезном деле, как образовательный процесс, нет места игре. Есть авторы, серьезно утверждающие, что в образовании нет места соревнованию, человек должен самосовершенствоваться сам, без оглядки на коллег (мы не называем авторов). Якобы, это – высшая степень развития учащегося. Существование таких споров удивительно – ведь классики педагогики и психологии в один голос высказываются за важность и необходимость применения именно игровых элементов в учении (см. напр.: Хуторской В. А. Современная дидактика, Спб: Питер, 2001.), а значит, и сюрпризности и соревновательности.

Проводимые нами с 1994 года эксперименты среди школьников (2-11 классы), абитуриентов, студентов (1-5 курс), учителей, преподавателей вузов, показали следующее:

- Использование в процедуре тестирования элементов игры, таких как полуинтерактивный диалог учащегося с компьютерными персонажами, вместо сухой демонстрации табличек с вопросами и списками ответов, побуждает желание учащихся к неоднократному тестированию;
- Включение в процедуру тестирования элементов состязательности (напр., публикация результатов, символический приз) с параллельным позволением неоднократного повтора попыток приводит к росту популярности тестирования и росту уровня результатов участников;

- Учителя и преподаватели относятся к тестированию своих ЗУН, как правило, настороженно. Они опасаются процедуры регистрации, они опасаются прикасаться к клавиатуре и мыши. Раскрепощение (чудесное преображение) наступает, когда оказывается, что «тестируются знания не самого участника, а руководимого им компьютерного персонажа». При такой постановке педагоги часами готовы сидеть за компьютером, обсуждая с коллегами нюансы возникающих на экране эпизодов, стремясь «заработать» как можно больше баллов;
- Допущение сценарием игры компьютерной «шпаргалки» приводит к интенсификации запоминания учебной информации. Информация получаемая «обходным» путем, оказывается, лучше запоминается. Учащиеся (и педагоги) с удовольствием делятся друг с другом почерпнутыми знаниями, помогают друг другу их найти, раскрепощаются при коллективном общении на учебные темы.

Повторим: описанные явления происходят как среди детей, так и среди взрослых. Более того, в игровом тестировании с интересом участвуют даже присутствующие родители и школьные «двоечники»: они с интересом обсуждают с одноклассниками и учителем перипетии своих компьютерных диалогов в рамках учебной дисциплины, обращаются к учебнику.

Последнее из перечисленных выше наблюдений, вероятно, может привести в будущем к необычной технологии образовательного процесса, когда учащиеся самостоятельно, группой и в одиночку, «добывают» учебную и дополнительную информацию, становящуюся коллективным достоянием.

Проведенные нами исследования попыток тестирований школьников показали, что, в случае снятия количественных ограничений на набор тестовых заданий, «кривая роста» результатов участников в среднем близка экспоненте. При этом вычисляемый экспоненциальный параметр (коэффициент) устойчив и, по-видимому, может использоваться в качестве характеристик как участника, так и коллектива участников тестирований. На форму кривой «роста» должны оказывать влияние такие характеристики учащегося как скорость адаптации к ситуации, способность быстро усваивать учебный материал (обучаемость), психическое состояние во время сеанса тестирования, осведомленность. Важным показателем, который можно извлечь из результатов повторных тестирований, вместе с суммарным количеством набираемых участником баллов, является разность количеств баллов, набранных при повторных тестированиях. В зависимости от времени, прошедшего между попытками, эту разность можно интерпретировать и как показатель привыкания, и как показатель обучаемости (в перерыве участник мог по своей инициативе общаться с коллегами, читать учебник и пр.). Отклонения этого показателя в положительную (отрицательную) сторону вполне могут характеризовать учащегося с положительной (отрицательной) стороны.

На основании сказанного, целесообразно обсудить введение в научный обиход такой численной характеристики процедуры тестирования (мы ее здесь называем «релевантность»), которая прямо связана с валидностью, но характеризует не степень устойчивости результата измерения, а степень его роста при повторных измерениях.

Уточним специально для физиков: описанная характеристика имеет интересный аналог в статистической физике – производство энтропии.

06.06. Информационный граф

У меня сложная и красивая фигура...

Технология создания электронного учебного пособия и педагогического теста – это основанный на достижениях тестологии сложный и трудоемкий процесс. Дополнительные сложности здесь создают своеобразные педагогические традиции. Цитируем:

Методологической основой разработки и использования тестовых средств В. С. Идиатулин считает «представления когнитивной психологии о пропозициональной репрезентации усвоенной информации в человеческой памяти, о существовании ее наименьших значимых единиц, которые могут быть выражены отдельным утверждением, имеющим самостоятельное смысловое содержание, описывающее объекты и отношения. При этом выделение достаточно однородных элементов знания становится предпосылкой измерения его совокупности». (Кречетников К. Г. Задания в тестовой форме и методика их разработки: Учебно-методическое пособие – Владивосток: Изд-во Дальневост. ун-та, 2002.- 36с).

К сожалению, многие педагоги любят излагать свои мысли именно так.

Опытные составители сначала всегда «конвертируют» материал дисциплины, т. е. переводят его в последовательность кратких утверждений – текстовых блоков. При этом «выливается вода» и, в качестве блоков, остаются:

- положения и понятия;
- определения и формулировки;
- факты;
- классификации;
- законы и закономерности;
- выводы и следствия;
- тезаурус;
- важные иллюстративные материалы.

Информация блоков должна быть достаточной для контроля ЗУН. «Рядом на полях» отмечаются внутренние и внешние связи блоков между собой и со смежными дисциплинами. Блоки связаны между собой в общей структуре материала дисциплины, включены один в другой, иерархически соподчинены. Отношения блоков могут быть представлены в виде ориентированного графа. В вершинах графа располагаются блоки, а их взаимосвязи обозначаются направленными ребрами. Такой граф, оснащенный некоторыми дополнительными атрибутами, иногда называют моделью знаний. В таком виде проще:

- искать и устанавливать новые связи;
- производить внедрение нового материала;
- находить и устранять дублирование материала;
- производить деление информации на крупные модули – подграфы;
- проектировать интеллектуальные электронные системы.

Граф знаний удобно использовать при составлении тестовых заданий и гипертекста.

06.07. Два типа ошибок измерений в тестологии и их источники

*Архитекторы прячут свои ошибки под плющом,
врачи под землей, а хозяйки под майонезом.
Бернард ШОУ*

Так же, как и в любой технической области, тестологические измерения сопровождаются ошибками. Эти ошибки могут быть

- систематическими, возникающими, например, по вине создателей тестовых заданий, тестирующей программы, учебного пособия, и
- случайными, возникающими из-за того, что участники пытаются «схитрить», выполнить часть заданий методом «случайного тыка».

Систематические ошибки возникают из-за:

- разницы толкования определений в ТЗ и в учебнике;
- неясностей в интерфейсе;
- логических неувязок в текстах ТЗ;
- неправильных объяснений, данных сопровождающим тестирования оператором, и др.

На величину ошибок влияет человеческий фактор. Наиболее вероятные величины случайных ошибок определения рейтинга, если их отложить на графике, образуют небольшой интервал значений или, говоря языком статистики, – доверительный интервал. Величина этого интервала зависит от того, в какую область рейтингов попал исследуемый рейтинг. Проще говоря, – если участник правильно выполнил много заданий, вероятность того, что он нажимал клавиши «на авось», мала. И наоборот, эта вероятность велика для участников, результаты которых попали в область «наиболее вероятного» результата.

В общем случае, величину доверительного интервала можно представить функцией двух переменных:

- расстояния от исследуемого результата до «идеального» и
- расстояния от исследуемого результата до центра области «наиболее вероятного» результата.

Очевидно, с изменением этих переменных, величина области «доверительного интервала» будет изменяться в определенных пределах. Эти пределы заключены между величиной области «наиболее вероятного» результата и нулем.

Знание величин доверительных интервалов поможет на практике при разборе экзаменационных апелляций, а также очень важно для правильного применения методов факторного анализа.

*Детей нужно учить врать.
Г. ОСТЕР*

06.08. Педагогическое тестовое задание

*Воображение рисует, разум сравнивает, вкус отбирает, талант исполняет.
Гастон де ЛЕВИС, aforizm.kaminplus.ru*

Не каждое задание, созданное для дальнейшего включения в тест, можно назвать тестовым или педагогическим. ТЗ, составляющие педагогический тест, называют педагогическими.

Ключевое задание – нацелено на проверку знания конкретных блоков учебной дисциплины. В случае адаптивного тестирования, тест должен содержать не менее 3-х ключевых заданий разных уровней сложности для каждого блока.

Педагогическое задание – средство интеллектуального развития, образования и обучения, способствующее активизации учения, повышению качества знаний, а также повышению эффективности педагогического труда (В. С. Аванесов).

Конкретное педагогическое задание является результатом труда по отбору (редактированию, изменению) совокупности тестовых заданий для педагогического теста. До того, как задания будут апробированы и включены в **педагогический тест**, во избежание путаницы, их называют **заданиями в тестовой форме**.

Задание в тестовой форме – это единица учебно-контрольного материала. Будучи включенным в педагогический тест, задание в тестовой форме становится педагогическим тестовым заданием.

Тестовое задание – самостоятельный элемент теста (т. е. совокупности тестовых заданий), состоящий из восьми компонент:

1. инструкции;
2. постановки задачи (собственно, задания);
3. демонстрационного материала;
4. эталонов результата выполнения задания, среди которых есть, как минимум, один абсолютно верный;
5. специально организованного места или области для ввода результата выполнения;
6. эмпирических параметров – веса, трудности задания;
7. номер и указания принадлежности задания к конкретной теме учебной дисциплины;
8. шкалы оценивания результата выполнения.

Приказом МОРФ № 1122 от 17 апреля 2000 г. рекомендовано использовать для оценки знаний методом тестирования сертифицированные педагогические тестовые материалы. Целью сертификации материалов является повышение качества образования за счет совершенствования педагогических тестовых материалов. Предметом и содержанием сертификации является установление степени соответствия характеристик педагогических тестовых материалов требованиям к их качеству, утвержденным Координационным советом МОРФ. Экспертиза педагогических тестовых материалов предусматривает оценку качества их содержания, правильности форм заданий и статистических характеристик.

В инструкции ТЗ пишут, что испытуемый должен сделать, каким образом, где и как делать пометки и записи. Иногда, для экономии времени и ресурсов, дается только одна инструкция для группы однотипных заданий, которая помещается в начале группы. Инструкцию принято снабжать примерами.

Текст постановки задачи обычно состоит из вопроса, на который следует ответить, или утверждения, которое следует подтвердить, исправить или опровергнуть.

Демонстрационный материал задания должен быть таким, чтобы для знакомства с ним претендент затратил не более 3-5 минут.

Эталоны выполнения ТЗ определяются явно (перечнем) либо неявно (указанием возможностей). Эталоны соответствуют выбранной автором качественной либо количественной шкале оценивания – верные, неверные, неполные, неточные и пр. Чаще, для простоты, авторы выбирают двузначную шкалу: верно – неверно. Эталонов выполнения должно быть не меньше двух. Оптимальным считается количество – 4-6. Эталоны не демонстрируются в случае ТЗ открытой формы, где претендент самостоятельно вводит знаки в специально выделенных местах. Обычно считается, что ТЗ закрытой формы должно содержать хотя бы один правильный эталон выполнения. Их может быть и несколько.

Наличие определенного места для ответов обязательно как при бланковом, так и при компьютерном тестировании:

- В ТЗ закрытой формы – это код (цифры или буквы) ответа из числа демонстрируемых. В ТЗ открытой формы ответ пишется вместо прочерка;
- В ТЗ на установлении соответствия ответы или пишутся в специально отведенной для этого строке ниже текста задания, или фиксируется с помощью мыши;
- В ТЗ на установлении правильной последовательности испытуемый ставит ранги в специально отведенном для этого месте.

Каждое ТЗ взаимосвязано с остальными. Эта связь бывает явной и неявной. ТЗ часто неявно связано с другими просто в силу структуры теста, подобно, атомам в кристалле.

Задания в рамках теста **связаны явно**, если (приведены примеры – образы):

- все ТЗ посвящены одной, резко очерченной теме, и любое дополнительное задание будет лишним (тест – кристалл);
- все ТЗ имеют веса, веса являются составляющими формулы расчета результата тестирования, и любое дополнительное ТЗ разрушит равновесие (тест – весы);
- важна последовательность изложения и проверки знаний (тест – конвейер);
- важна последовательность возрастания сложности ТЗ (тест – эволюция);
- каждое ТЗ занимает определенную нишу в структуре теста, как архитектурный элемент в сооружении, и без его выполнения нельзя судить о каком-либо качестве претендента (тест – дворец).

ТЗ в рамках теста **связаны неявно**, если:

- каждое ТЗ посвящено отдельному блоку информации, не связанному с другими отношениями порядка, а сам тест формируется из множества ТЗ случайным образом;
- ТЗ составлялись без учета того, сколько и каких заданий уже вошло в состав теста.

ТЗ могут быть **разграниченными**, т. е. результаты их выполнения свидетельствуют о разных составляющих знаний претендента и **взаимопроникающими**, когда результат выполнения одного ТЗ подтверждает результат выполнения другого.

Если из какого-либо теста сделать выборку меньшего количества ТЗ, то может образоваться система, не уступающая заметно по своим свойствам первоначальному тесту. Тест с меньшим количеством ТЗ тогда можно называть более эффективным.

06.09. Четыре формы тестовых заданий

*Само понятие нормы не существует без ее нарушения.
М. ГОРБАНЕВСКИЙ*

В тестологической практике принято считать, что существуют всего четыре формы тестовых заданий. Все остальное разнообразие можно классифицировать как их вариации.

Вместе с тем, существует мнение, что четырьмя типами все описать нельзя. Тестология развивается. Например, не так давно появилась методика дифференцированной оценки выполнения заданий. Она известна под названием «grading» (градуирование, распределение). В этой методике предполагается, что оценка выполнения заданий может быть не только двузначной «правильно – неправильно», но и принимать целый спектр значений. Считается, что более всего эта методика применима в математике.

Вот названия (характеристики) всех четырех традиционных форм тестового задания:

- задания с выбором ответов (испытуемый выбирает правильные ответы из числа готовых, предлагаемых в задании теста) – задание в закрытой форме;
- задания на дополнение (испытуемый сам дает краткий или развернутый ответ) – задание в открытой форме;
- задания на установление соответствия между элементами двух множеств;
- задания на установление правильной последовательности в ряду элементов.

Выбор формы зависит от:

- цели тестирования;
- содержания теста;
- технических возможностей;
- уровня подготовленности преподавателей и персонала.

Приведем (с некоторыми сокращениями) объемную цитату, проясняющую взаимоотношение формы и содержания тестовых заданий:

Овладение формой является необходимым, но недостаточным условием создания полноценных тестов. Форма придает заданиям лишь структурную целостность и определенность, внешнюю организованность. Задания в тестовой форме только внешне похожи на тестовые задания, но это недостаточно для их включения в тест; нужна еще проверка свойств, позволяющих включить их в тест. Вместе с тем, меняющееся от дисциплины к дисциплине содержание позволяет абстрагировать форму в качестве самостоятельного предмета и рассматривать ее как подобие инварианта, независимого, в значительной степени, от содержания конкретной учебной дисциплины. Тем самым здесь подтверждается общий философский тезис о сравнительном постоянстве формы, остающейся устойчивой при изменении содержания.

Начиная с Аристотеля, форма понимается как идеальный принцип расположения элементов. Другое истолкование выдвинул И. Кант, согласно которому форма есть принцип упорядочения. Эти две идеи – организации и упорядочения, плюс идея существования содержания в каких-либо формах легли в основу нашего определения формы тестовых заданий: это способ организации, упорядочения и существования содержания теста.

Исследованием научных основ разработки тестов занимается педагогическая тестология. В западных странах получили развитие эконометрика, биометрика, социометрия и

другие науки, имеющие своим предметом изучение количественных свойств и отношений, разработку объективных показателей состояния интересующих явлений
Общие принципы разработки формы тестовых заданий рассматриваются в тестологии как важный предмет исследования, в то время как содержание, будучи не менее важным компонентом хорошего теста, относится все же не столько к тестологии, сколько к той науке, знание которой проверяется.

В. С. АВАНЕСОВ

Первая форма – закрытая форма задания

Основой этой формы является закон исключенного третьего, сформулированный Аристотелем. Выбор правильного варианта порождает истину, выбор неправильного – ложь. Третьего не дано. Из этого закона следует: каждое задание должно иметь один верный вариант выполнения.

Чаще всего в литературе встречаются задания в закрытой форме, имеющие вопрос и четыре варианта ответа, из которых один – верный. Это – уже классика.

Для выбора правильного варианта испытуемый должен совершить одно из действий:

- ввести код (цифры или буквы) требуемого ответа в специальном окне;
- щелкнуть мышью по «радиокнопке» с номером варианта;
- щелкнуть мышью по тексту варианта.

Главными недостатками классической закрытой формы являются:

- высокая вероятность угадывания верного варианта;
- непроизводительные затраты времени на прочтение всех вариантов;
- возможное непроизвольное подсознательное запоминание испытуемым неправильных вариантов как правильных.

Форма налагает логический запрет на применение таких вариантов, как:

- «правильного ответа нет»;
- «все ответы правильные»;
- «все ответы неправильные».

Он часто нарушается как в зарубежных, так и в российских тестах. Нарушения эти свидетельствуют о невнимательности либо об отсутствии культуры логического мышления среди составителей тестовых заданий.

В публикуемых в печати заданиях закрытой формы часто встречаются логические неувязки, несоответствия, многозначности. Это происходит в том случае, когда:

- вопрос содержит скрытый вопрос, который некоторые, особо искушенные в тестированиях, испытуемые могут принять за основной;
- вопрос содержит дополнительную контекстную информацию, могущую показаться неверной части испытуемых (особенно это касается истории, политологии и т. п.);
- вопрос многословен, содержит придаточные предложения;

- в задании отсутствует задание или вопрос, т. е. есть только текст в области задания и четыре варианта текста в области ответов;
- задание содержит утверждение с многоточием и варианты замещения многоточия.

Чаще всего, логические неувязки встречаются в четвертом из этих случаев.

Указанные и другие недостатки есть во всех тестах. Есть три пути борьбы с ними:

- повторное авторское редактирование;
- самотестирование автора;
- проведение тестирования в референтных группах, состоящих не только из учащихся, но и из преподавателей.

Наиболее быстрым, экономичным и эффективным следует признать третий путь.

Есть два достаточно известных пути развития закрытой формы, восполняющие указанный недостаток:

- задание с оцениванием верности, т. е. выбором одного, наиболее верного варианта из нескольких верных в разной степени;
- задание с выбором нескольких верных вариантов из нескольких предложенных.

Оба они так же считаются заданиями в закрытой форме.

Есть третий и четвертый пути развития задания в закрытой форме, еще более снижающие вероятность угадывания, описанию которых в данном пособии уделено особое внимание:

- задание с оцениванием степени верности каждого из вариантов;
- задание с оцениванием степени верности одного из вариантов, где испытуемый не может видеть все варианты одновременно.

В последнем случае испытуемому предоставляется возможность просматривать варианты, выпадающие в случайном порядке.

Третий и четвертый пути, кроме наличия очевидных преимуществ в оценивании степени владения знаниями, весьма перспективны для:

- развития эвристичности мышления;
- обучения студентов педагогических специальностей.

Ситуации, в которых необходимо выбор, особенно множественный, как в последних случаях, часто вызывают у испытуемых чувство неуверенности. Источником сомнений является:

- отсутствие информации об альтернативах;
- субъективной ценностью альтернатив;
- конфликтный (бесконфликтный) характер альтернатив.

Опыт показывает, что неуверенность испытуемых исчезает после первых успехов.

Вторая форма – открытая форма задания

Задание в открытой форме является обычно утверждением или большим текстом. В тексте есть специальные места, куда следует вносить информацию, оговоренную инструкцией. Обычно это «клетки», в каждую из которых вписываются буквы, цифры или пробелы. После заполнения клеток, текст превращается в истинное или ложное высказывание.

Обычно используется три вида тестовых заданий открытой формы:

- задания дополнения, где испытуемый должен сформулировать дополнения к предъявленным текстам;
- задания – кроссворды, ориентированные на выяснение знания претендентами определений, могущие содержать подсказки на пересечениях слов и игровой элемент;
- задания свободного изложения, где требуется изложить полное решение задачи, сочинение или перевод текста, где почти никакие ограничения на выполнение задания не накладываются.

Правильно сконструированные задания открытой формы полностью исключают догадку – главный недостаток заданий закрытой формы. Это – главное достоинство таких заданий.

Задания открытой формы часто вызывают недопонимания среди педагогов в связи с частотой одинаковых случаев: претенденты, выполняющие задания открытой формы по географии (истории, литературе, математике...) делают орфографические ошибки. Считать ли варианты с орфографическими ошибками ошибочными? Ставить ли учащемуся, превосходно знающему теоремы и умеющему решать задачи, двойку по математике за то, что он неправильно пишет слова «прямая» и «плоскость»?

Однако задания свободного изложения имеют существенный недостаток – трудно предсказуемая многовариантность. Результаты их выполнения невозможно полноценно проверить автоматически. По крайней мере, на сегодняшнем этапе развития компьютерных интеллектуальных систем. Можно с этим бороться:

- уделять больше внимания формулировке заданий, чтобы существенно сократить количество вариантов;
- указывать количество букв в допустимом варианте выполнения;
- предусмотреть все возможные верные варианты.

Все эти попытки приводят к резкому ограничению длины возможных текстов, вводимых претендентами в качестве вариантов ответа.

Учитывая, что количество цифр и букв ограничено, можно считать задание в открытой форме эквивалентным заданию в закрытой форме, только с очень большим количеством (миллионы, миллиарды) неправильных вариантов выполнения, скрытых от претендента.

Третья форма – задания на установление соответствия

Это задания, где элементам одного множества нужно верно сопоставить элементы другого. Вариантом выполнения задания является перечень связей (соответствий) элементов. Для обозначения этих связей поступают так. Элементы одного множества обозначаются цифрами, другого – буквами. Ответ конструируется как последовательность связанных

цифр и букв. Например, для двух пятиэлементных множеств, нужно ввести в качестве результата выполнения задания кодовое слово: 1P2M3B4E5H6J.

Результаты выполнения задания на установлении соответствия:

- или пишутся в отведенной для этого строке по оговоренному инструкцией правилу;
- или указываются мышью графические изображения связей (соответствий).

Задания на установление соответствия рекомендуются для проверки ассоциативных и алгоритмических навыков пользования знаниями. Поиск аналогий на основе учебного материала, позволяет делать выводы не только о владении обучаемым теми или иными понятиями, но и о других характеристиках структуры знаний.

Задания на установление соответствия позволяют вести проверку усвоения сразу нескольких дисциплин. Поэтому они могут обладать большой проверочной емкостью.

Выполнение таких заданий легко проверять путем сравнения с единственным верным вариантом. Это жесткий подход. Но ведь претендент может сделать одну ошибку – неужели из-за этого все задание он выполнил неправильно? Проверить выполнение таких заданий «мягко», дифференцированно – вручную нельзя, а с компьютером – можно.

Часто встречаемый недостаток заданий третьей формы – «особо начитанные» претенденты усматривают в таких заданиях не только изоморфные (однозначные) связи, но и гомеоморфные (многозначные). Это приводит к необоснованным снижениям оценок и справедливым апелляциям (см., напр., раздел «ЕГЭ» в Приложениях).

Поскольку результат выполнения задания третьей формы – кодовое слово с четко оговоренной длиной, в этом смысле эта форма совпадает со второй (открытой).

Четвертая форма – задания на установление последовательности

В этих заданиях требуется установить правильную последовательность вычислений, действий, шагов, операций, терминов.

Установив правильную, со своей точки зрения, последовательность, испытуемый вводит свой вариант в специально отведенном для этого месте путем

- или щелканья мышью по последовательности мнемонически понятных элементов (цифр и пр.);
- или ввода последовательности цифр с клавиатуры.

Эта форма заданий вызывает часто нарекания не из-за себя, а из-за недостатков интерфейса и инструкции, приводящим к разночтениям алгоритма ввода варианта.

Эту форму применяют для проверки верности понимания испытуемыми хода решения.

Четвертая форма заданий становится эквивалентной третьей, если представить одно из элементарных множеств задания третьей формы простой последовательностью цифр.

06.10. Вес тестового задания. Полдюжины факторов снижающих ценность понятия веса ТЗ

Независимо от единиц измерения, используемых поставщиком или покупателем, производитель будет использовать свои собственные произвольные единицы измерения, переводимые в единицы поставщика или покупателя с помощью странных и неестественных коэффициентов пересчета.

Теорема Вышковского (Артур Блох. Мерфология)

Можно «взвесить» задания относительно друг друга внутри каждого уровня, если это необходимо. Однако здесь следует иметь в виду следующие соображения.

Вычисление относительного веса заданий, т. е. их относительной трудности для учащихся – довольно сложная задача. Она требует проведения большого количества тестирований с участием многих учащихся. При этом на величины вычисляемых Вами весов могут непредсказуемо повлиять следующие факторы:

- Разные учащиеся в разное время могут пользоваться при подготовке к тестированию разными учебниками, где один и тот же вопрос освещен с разной степенью подробности и ясности;
- Смена (или временная подмена) преподавателя дисциплины приводит к освещению разных вопросов в разных учебных группах с разной степенью подробности и ясности. Поэтому при взвешивании можно вес одного и того же вопроса для одних избрать чересчур высоким, для других – низким;
- Учащиеся, прошедшие тестирование первыми неизбежно передадут почерпнутую информацию следующим, и те придут тестироваться более подготовленными. Неодновременность сеансов тестирования и неконтролируемость распространения информации среди учащихся о тестовых заданиях сделают Ваши попытки «скрупулезного взвешивания» бессмысленными;
- Учащиеся, лучше всех знающие учебный материал, лучше всех прогнозируют для себя время, необходимое для выполнения заданий. Поэтому они выполняют в первую очередь простые задания, а сложные оставляют на «потом». Это приведет к ошибочному пониманию сложности заданий – за «сложные» Вы ошибочно примете «долгие»;
- Рабочие учебные программы по дисциплинам иногда меняются, в соответствии со взглядами руководства. Каждая такая смена будет разрушать Ваши результаты «взвешиваний» и заставлять Вас приниматься за эту задачу вновь.
- Величины весов ТЗ, вычисленные на разных выборках претендентов, получатся разные. Следовательно, претендент, результаты тестирования которого учитывались в совокупности с разными выборками, будет иметь разный результат.

Таким образом, для учащихся, учившихся в разных учебных группах и в разные периоды времени, веса заданий, вычисленные по результатам тестирований других групп, будут относительно необъективны. Кроме того, опыт работы жюри олимпиад говорит, что, по какой бы системе веса не вычислялись, дополнительной объективности решению жюри это никогда не придает и список призеров от этого меняется мало.

Поэтому «сложность» заданий надо учитывать по-другому. Например, путем перехода к многомерной векторной оценке, векторному рейтингу. Компоненты вектора – рейтинга характеризуют относительный успех претендента в выполнении конкретных групп ТЗ.

06.11.Трудоемкость и сложность тестового задания

На узкой дороге тон задает черепаха.

В классической тестологии понятия «трудность» и «сложность» использовали как синонимы. Разные авторы предлагали различать их. В современной тестологии термин «трудность» заменяют на более подходящий – «трудоемкость» (предложено В. П. Беспалько).

- Трудоемкость задания – мера количества и длительности предполагаемых операций, необходимых для успешного выполнения задания.
- Сложность задания – мера объема и сложности информации, которую должен освоить претендент для того, чтобы выполнить задание.

Если претенденты имеют достаточно времени, чтобы продумать варианты выполнения и не соревноваться в скорости, определяющим фактором становится сложность заданий. Поэтому сложность – наиболее часто используемое слово для характеристики заданий.

Сложность задания зависит от пути изучения учебной темы претендентами (по какому учебнику), от их подготовки и сообразительности. Сложность принимает разные значения для разных выборок претендентов. Одно и то же задание для студента одного вуза может стать неразрешимым, для студента другого – «семечками». Чем больше претендентов, правильно выполнивших задание, тем меньше его сложность для данной группы. На величину сложности могут влиять артефакты, например – накапливающаяся в процессе длительного сеанса усталость или недостаточность времени сеанса. Простое задание могут не выполнить большинство претендентов только потому, что пока они до него добрались, они устали. Или, вообще не успели добраться. Так простое задание с большим номером автоматически формально становится очень сложным.

И в школе, и в вузе часто бывает ситуация, когда:

- учащиеся одной параллели учатся у разных преподавателей и по разным учебникам;
- преподаватели меняются в учебной группе в течение учебного года.

А если так, то усилия по эмпирическому определению сложности заданий теста не приведут к результату, и сложность заданий будет принимать разные значения.

Вычисленная сложность заданий может реально использоваться в дальнейших тестированиях при условии сохранения трех внешних параметров:

- личность преподавателя и материал лекций, контрольных работ и пр.;
- комплект учебных пособий;
- параметры претендентов (т. е. группы учащихся, подвергающиеся тестированию, должны быть статистически неотличимы по характеристикам).

В противном случае, сложность заданий придется пересчитывать после каждого сеанса тестирований перед определением рейтингов. А тогда может произойти казус:

- претенденты, тестировавшиеся в составе разных групп и получившие одинаково высокий рейтинг, будут значительно различаться своими характеристиками;
- претендент, прошедший тестирование по одному тесту, но в разных группах, получит разный рейтинг.

Один из путей преодоления этих проблем – переход к векторному рейтингу.

06.12. Две дюжины оснований типологии тестовых заданий

Создаваемые в настоящее время в России БТЗ, на первый взгляд, так различны, что трудно их сопоставить для дальнейшего совместного использования. Ряд исследователей, занимающиеся этой проблемой, предложили способы классификации БТЗ.

ПЕРВАЯ ДЮЖИНА

оснований определена комбинированной типологией ТЗ предложенной Е. А. Михайлычевым (Михайлычев Е. А. Дидактическая тестология. М.: Народное образование, 2001. - 432 с.) на основе типологий Г.А. Балла, В.И. Андреева, В.П. Беспалько. Мы здесь внесли в нее некоторые не принципиальные коррективы, следуя лишь соображениям лаконичности и однообразия.

1. Ориентация на диагностируемый уровень усвоения (по концепции В.П. Беспалько):
 - I уровень;
 - II уровень;
 - III уровень (не типовая задача, связанная с графическими и модельными построениями, не имеет эталона ответа);
 - IV уровень (эвристическая задача, не имеет эталона ответа).
2. Характер обобщённости предмета задачи:
 - индивидуальный;
 - родовой.
3. Характер предмета решения задачи:
 - материальный;
 - информационный.
4. Характер разрешимости тестовой задачи:
 - принципиально неразрешимая;
 - принципиально разрешимая.
5. Субъективная подготовленность претендента к решению тестовой задачи:
 - рутинная;
 - не рутинная.
6. Чёткость построения тестовой задачи:
 - чёткая (алгоритмическая);
 - нечёткая.
7. Нацеленность формулировки на адекватность восприятия:
 - ориентированная на адекватное восприятие;
 - ориентированная на неадекватное восприятие.
8. Возможность изменения предмета задачи:

- теоретическая;
- практическая.

9. Особенности формулировки задания:

- задачи на исполнение;
- задачи на преобразование;
- задачи на восстановление;
- задачи на построение;
- задачи на использование процедуры;
- задачи на использование имеющегося состояния, условия.

10. Логическая схема построения задания:

- индуктивная;
- дедуктивная;
- смешанная (индуктивно-дедуктивная).

11. Доминирующие приемы метода познания, требуемого для решения задачи:

- логическая;
- аналитическая;
- экспериментальная;
- графическая.

12. Форма организации выполнения задачи:

- индивидуальная;
- групповая;
- фронтальная.

На основе этой схемы можно описывать основные аспекты ориентации ТЗ. Она же может помочь в планировании структуры вновь создаваемой БТЗ.

В некоторых случаях удастся добиться еще более детальной классификации заданий. Для этого дополним приведенную первую дюжину оснований еще одной.

ВТОРАЯ ДЮЖИНА ОСНОВАНИЙ

может быть построена на основании работы (Васильев В. И., Тягунова Т. Н. *Теория и практика формирования программно-дидактических тестов*. М.: Издательство МЭСИ, 2001. - 130 с.) и ряда сделанных нами дополнений:

13. Форма:

- на указание верных вариантов (закрытая форма);
- на исправление или ввод информации (открытая форма);
- на указание соответствия;
- на указание последовательности.

14. Тип оценивания:

- с конструктивной оценкой
- с неконструктивной оценкой

15. Шкала (логика) оценивания

- с двузначной шкалой оценивания (верно – не верно)
- с трехзначной шкалой оценивания (верно – не совсем верно – не верно)
- с четырехзначной шкалой оценивания (верно – не полно – не точно – не верно)

16. Форма демонстрационных материалов

- не мультимедийное
- мультимедийное
- графическое
- текстовое

17. Тип носителя

- бланковое;
- компьютерное.

18. Величина эмпирических параметров:

- вес;
- трудоемкость;
- сложность.

19. Содержание:

- информационно-отражающие;
- проблемно-эвристические;
- развивающие.

20. Направленность выяснения характеристик претендента:

- знаний;
- умений;
- навыков.

21. Общие требования:

- технологичности;
- формы;
- содержания

22. Статистические требования:

- известной трудности;
- достаточной вариации тестовых баллов;
- положительной корреляции баллов задания с баллами по всему тесту.

23. Ориентация:

- нормативные тесты – ориентированные на статистические нормы, основанием для сравнения в которых служат статистически полученные данные о выполнении теста репрезентативной выборкой претендентов;
- критериальные тесты – ориентированные на критерии, предназначенные для определения уровня достижений претендента относительно заданного критерия;
- прогностические тесты – ориентированные на прогноз успешности деятельности.

24. Вид тестирования:

- статическое тестирование – направлено на измерение результатов обучения, но не может оценить потенциал претендента;
- динамическое тестирование – направлено на выявление скрытых способностей претендента, а также оценку – насколько успешно он может далее обучаться.

Чаще авторы тестов не пользуются типологиями, составляя тесты так, как они считают нужным. Их оправдывает прагматичная цель – не повышение качества измерения, а надежная дифференциация групп претендентов. Тестирования с такими тестами похожи на обычную экзаменационную лотерею, в которой гарантированно выигрывает только подготовленный претендент. Чаще всего, свойств таких тестов вполне достаточно для практического применения.

В упомянутой книге Е. А. Михайлычева приведен обширный обзор различных типологий тестовых заданий. В настоящее время наиболее распространенной типологией ТЗ является дифференциация по четырем формам.

06.13. Четыре группы из восемнадцати характеристик ЗУН, выявляемых путем тестирования

*Когда я был маленьким, у меня тоже была бабушка ...
Тов. Дынин (к/ф «Добро пожаловать или посторонним вход воспрещен»)*

Приведем классификацию параметров претендентов, выделенных разными авторами. Мы разбили их на четыре больших группы.

1. Кругозор:

- знание основных фактов;
- знание истории науки;
- методологическая подготовка (классификации, внутридисциплинарные и междисциплинарные связи);
- умение применить философские методы, понимание противоречий, различение формы и содержания;
- знакомство с основной литературой;
- словарный запас.

2. Знания основ дисциплины (темы):

- знание законов;
- знание определений и ключевых понятий, номенклатуры;
- знание правил;
- знание основных концепций.

3. Умение решать стандартные задачи в рамках дисциплины (темы):

- умение применять знания в решении простых задач;
- умение защищать свои взгляды;
- умение логичного изложения;
- умение приложить теорию к практике, описать и объяснить простое явление;
- наличие стандартных умений и навыков (построение таблиц и графиков, формул, проведение экспериментов, анализ текстов и изображений);
- умение перевести стандартный текст на иной язык.

4. Умение разрешать проблемы:

- понимание взаимосвязанности тем и разделов курса, умение применить взаимосвязи;
- знакомство с периодической и не основной литературой;
- наблюдательность, умение вычленить главное;
- умение распознать структуру текста (проблемы, изображения, формулы, явления);
- умение применить межпредметные знания, «сквозные» учебные умения и навыки.

06.14. Два десятка требований к тестовым заданиям

*Было бы ошибкой думать...
(классический пример неверно составленного тестового задания)*

Тестовые задания должны отвечать следующим требованиям (в алфавитном порядке):

1. Адекватность – соответствие инструкции, формы и содержания задания;
2. Взвешенность – наличие веса, т. е. определенного, выраженного в виде числа или отношения порядка уровня сложности или трудоемкости задания;
3. Дистрактивность – степень похожести неправильных вариантов на правильные и наоборот;
4. Дифференцирующая способность – свойство отдельных заданий и теста дифференцировать претендентов на группы в соответствии со степенью их подготовки;
5. Доступность – потенциальная возможность выполнения для большинства претендентов;
6. Интегрированность, синтетичность – наличие или отсутствие межпредметных связей;
7. Корректность – отсутствие малопонятных и редко употребляемых выражений, а также неизвестных для претендентов символов и слов, затрудняющих восприятие смысла;
8. Краткость – тщательный подбор слов, символов, графиков, позволяющая минимумом средств добиваться максимума ясности смыслового содержания задания;
9. Критериальность – наличие возможности предъявить претендентам не только двузначную шкалу оценки вариантов выполнения («верно – не верно») но и иные, в соответствии с критериями (напр., «верно – не точно – не полно – не верно»);
10. Многофункциональность – наличие возможности диагностики и памяти, и мышления;
11. Оригинальность – минимизирующая вероятность пользования шпаргалкой;
12. Открытость – претенденты имеют возможность познакомиться с подобными ТЗ заранее;
13. Понятность – отсутствие двойных отрицаний, придаточных предложений, скрытых вопросов, подразумеваемого смысла и т. п.;
14. Привлекательность – присутствие доли юмора и аналогичных элементов, способствующие развитию желания претендентов к самообразованию;
15. Технологичность – простота составления и корректировки, наличие четких правил составления и оценивания весов;
16. Толерантность – отсутствие фрагментов, вызывающих непонимание и возмущение, связанные с различием людей по объективным признакам (национальность, пол и пр.);
17. Универсальность – возможность использования в разных тестовых комплексах;
18. Формальность – соответствие избранной форме;
19. Эргономичность выполнения и обработки – минимальность затрачиваемого времени на прочтение, понимание, ответ; минимальность затрат ресурсов и времени на проверку и анализ результатов.
20. Целесообразность.

06.15. Две дюжины недостатков тестовых заданий и полдюжины их причин

*Мы шли по тропе. Тропа поднималась в гору.
Левицкая Н. В., Луганская Л. И., Лаврова К. И. Русский язык.
Учебное пособие для солдат, не владеющих или
слабо владеющих русским языком. Воениздат, 1982*

В отечественной и зарубежной научной тестологической литературе неоднократно отмечалось, что большинство создателей тестовых заданий совершают ряд одинаковых ошибок. Перечислим основные ошибки в порядке частотности появления.

1. В списке ответов правильный ответ, как правило, самый длинный. Поэтому его трудно найти. Это основная ошибка начинающих составителей тестовых заданий. Часто учащиеся об этом знают, и умело пользуются этим знанием.
2. Неправильные варианты ответов бывают не согласованными с вопросом задания в падежах, родах, числах и т.д. Авторы и редакторы обращают на них внимания меньше, чем на правильные варианты.
3. В вариантах выполнения заданий встречаются перекрестные логические и информационные ссылки, что усложняет нахождение верного варианта, если ответы расставлены компьютером по случайному алгоритму.
4. Верный вариант выполнения задания можно с большой степенью вероятности «вычислить» логически, путем исключения явно неверных.
5. Вопрос задания содержит скрытый вопрос, который, особо искусственные в тестированиях («особо одаренные»), испытуемые, ища логический подвох, могут принять за основную.
6. Вопрос содержит дополнительную контекстную информацию, могущую показаться неверной части испытуемых (особенно это касается истории, политологии и т. п.).
7. Вопрос задания многословен, содержит придаточные предложения, двусмыслен.
8. В задании отсутствует задание или вопрос, т. е. есть только текст в области задания и четыре варианта текста в области ответов.
9. Задание содержит утверждение с многоточием и варианты замещения многоточия.
10. Задание предполагает непропорционально высокие затраты времени на знакомство испытуемого со всеми вариантами выполнения.
11. Задание предполагает возможное произвольное подсознательное запоминание испытуемым неправильных вариантов как правильных.
12. Задание содержит логически невозможные варианты выполнения. Например, в литературе часто обсуждается известный логический запрет на применение таких вариантов ответа в заданиях закрытого типа, как
 - «правильного ответа нет»;
 - «все ответы правильные»;
 - «все ответы неправильные».
13. В задании отсутствует инструкция и трудно определить, что же требуется от испытуемого на этот раз. Например, в заданиях третьей и четвертой форм, когда на экране компьютера присутствуют два-три одинаково нумерованных элементов, а в задании

предлагается «указать» один из них, испытуемый вынужден обращаться за толкованием к дежурному в классе, который может не иметь нужной информации.

14. Нарушение изоморфизма. В заданиях третьей и четвертой форм, где требуется перенумеровать или сопоставить друг другу элементы, встречаются элементы, которым можно сопоставить разный порядок или которые можно одновременно сопоставить нескольким элементам.
15. Присутствие логической связи между разными вариантами выполнения задания, не позволяющей рассматривать их по-отдельности, например, если компьютер демонстрирует их по очереди или в случайном порядке.
16. Вопрос задания не однозначен, содержит в себе более одного вопроса (т. е. несколько подвопросов). Техника составления таких вопросов используется в социологических опросниках для «подталкивания» человека к нужному ответу, но в школьных тестовых заданиях они появляются исключительно по неопытности составителей. Подвопросы часто логически скрыты и видны только специалисту. Поэтому иногда трудно однозначно определить, какой ответ авторы считают полным и верным.
17. Вопрос задания содержит двойное отрицание. В этом случае часто можно построить не менее двух верных и совершенно противоположных по смыслу ответов на такой вопрос.
18. Вопрос задания не подразумевает полный однозначный ответ, не очерчивает круг обязательной информации, которую должен содержать ответ. Такая ошибка составителей часто встречается в заданиях всех четырех форм.
 - Например, на вопрос: «Где произошло Бородинское сражение?» можно абсолютно точно ответить «Под Москвой», «В России», «У деревни Бородино», «В поле».
 - Например, в заданиях с вопросами типа «Кто первым изобрел...?» не поясняется, должен ли ответ содержать фамилию, инициалы, национальность ученого, а также один или несколько ученых имеются в виду. Не учитывается, что вопрос «Кто?» применяется в русском языке и для случая одного человека, и для случая группы.
 - Например, вопрос сформулированный так: «Назовите клички лучших английских скакунов...», не содержит пояснения, сколько кличек нужно назвать.
 - Подлежащее в именительном падеже легко путается с прямым дополнением в винительном падеже, например: «мать любит дочь» (пример В. С. Аванесова);
 - На вопрос «Когда началась Великая отечественная война?» можно ответить; «в июне», «в XX веке» и пр.;
 - В заданиях открытой формы следует указывать род, падеж и пр. ожидаемых слов. Например, на вопрос «Как называется судно, колющее лед?» можно ответить и «ледокол», и «ледоколом».
 - среди заданий на установление соответствия или на установление последовательности встречаются варианты, трактуемые в разных источниках по-разному (причины и последовательности исторических событий).
19. Вопрос неоднозначен либо варианты ответа не настолько полны и однозначны, как подразумевается вопросом задания. Например, нельзя в задании задавать вопрос «Когда началась Великая отечественная война?». Следует задание формулировать так: «Назовите дату начала Великой отечественной войны». Ведь правильный ответ для первой формулировки неоднозначен: «В июне», «В XX веке» и пр.
20. Варианты ответа не согласованы с вопросом в числе и падеже, дают повод для противоречивого толкования.
21. ТЗ содержат орфографические и иные языковые ошибки. Эти ошибки, ставят в логический тупик учащегося, поскольку ему не известно, случайна или намеренна ошибка.

Особенно такие ситуации драматичны в тестированиях по русскому языку.

22. В инструкции задания открытого типа не указывается, в каком числе и падеже должен быть написан вариант словосочетания. В итоге – введенный принципиально правильный вариант может быть принят при автоматической проверке за неверный.
23. В задании открытого типа возможно несколько формулировок ответа, а составителем бывает предусмотрена только одна. В итоге – введенный принципиально правильный вариант может быть принят при автоматической проверке за неверный
24. Задания однообразны, нудны. Последовательность таких заданий отрицательно воздействует на учащихся, вызывают так называемый эффект «ровной дороги» – известный в практике водителей-дальнобойщиков эффект засыпания за рулем, отвлечения и рассеяния внимания, когда дорога слишком ровна и пустынна.

Часто учащиеся знают о типовых ошибках, и умело пользуются этими знаниями, получая высокие отметки. Известны случаи, когда шутник, осведомленный о таких ошибках, уверенно сдавал экзамен, например, по специальной медицинской дисциплине.

*Бесполезно придумывать защиту от дурака –
ведь дураки так гениальны....
Э. Мэрфи, DEC*

Применяемые в системе образования БТЗ часто далеки от идеала. Приведем самые веские причины этого объективного положения:

1. Составлять тексты так, чтобы они были удобочитаемы и удовлетворяли правилам русского языка, не так просто, как представляют себе многие. Не всякий учитель способен сформулировать вопрос так, чтобы он однозначно воспринимался всеми;
2. В специальных науках укоренились жаргонные словосочетания, которые могут быть применяемы в кругу единомышленников для оптимизации времени общения, но не в печати. Часто такие словосочетания противоречат языковым нормам, и смысл их трудно воспринимается неспециалистами;
3. Объективно невозможно полностью исключить вероятность угадывания правильного ответа, хотя и существует целый ряд методик, позволяющих ее минимизировать. Сравните два задания (пример А. Анастаси):
 - Плохо составленное задание: «Нелегальный: несекретный, незаконный, сексуальный, безграмотный, дерзкий».
 - Хорошо составленное задание: «Нелегальный: секретный, противозаконный, сексуальный, неграмотный, дерзкий».

В первом случае слово «незаконный» предполагает правильный ответ с приставкой «не», а во втором – слово «неграмотный» наводит на ложный след;

4. Можно утверждать, что практически в любом тесте можно найти «дикие» с точки зрения русской языковой культуры фразы, несогласованности падежей, родов;
5. Отсутствие культуры критики. В вузовских тестах (так же, как и в учебных пособиях) значительно меньше «дикостей», чем в школьных. Причина этого культурно-языкового превосходства, по нашему мнению, заключена в следующем. В вузовской среде, в отличие от школьной, поощряется критика: критические замечания по поводу методических пособий публикуются. Рассмотрите критически сборники тестов для школьников и студентов, рекомендованные для подготовки к тестированиям;
6. Нет предела совершенству. Идеал трудно достигим еще и по причине несогласованности спорности разрабатываемых требований к тесту.

Глава 07. ОИТ для защиты качества и управления качеством образования

*Говорят, что, лучшее правило политики – не слишком управлять.
Это правило столько же верно и в воспитании.
Жан ПОЛЬ. <http://frazza.net.ua>*

Защита качества образования и управление качеством образования – пока непривычные понятия для российского педагога. Они пришли в образование из производственной сферы. Появление их здесь обусловлено массовым внедрением ОИТ и появлением новых эффективных взглядов на управление развитием общественных отношений и производства.

ОИТ и управление качеством. Антураж компьютерных тестирований

*Первый миф науки управления состоит в том,
что она существует.
Закон Хеллера (Артур БЛОХ, Прикладная Мерфология)*

Эффективное управление системой образования страны невозможно без наличия обратной связи – надежной информации о состоянии системы, включающей показатели, характеризующие:

- кадровое и материально-техническое обеспечение образования;
- учебный процесс;
- результаты обучения.

Получение такой информации в полном объеме невозможно в настоящее время в России по ряду причин, например, отсутствие:

- надежной и научно обоснованной системы показателей, характеризующих состояние системы образования;
- федеральной системы мониторинга образования.

Данные, полученные в ходе исследовательских проектов, представляют научную и практическую ценность, при условии доказанности их валидности и надежности, а также представительности выборки обследуемых объектов.

Управлению качеством посвящено немало научной и методической литературы. Есть стандарты качества учреждений, например популярный стандарт – ISO-9000. К сожалению, для образовательной области многие описанные там положения мало подходят.

Большинство авторов этих книг и теорий рассматривают, в основном, качество управления производством, считая качество продукции вторичным понятием. Действительно, они по-своему правы. Там, где производят овеществленный товар, всегда есть много написанных на бумаге инструкций, технологий. Нет проблем с исполнителями, кто их не исполняет – их исполнителей просто меняют, благо, что безработица еще существует. Проблемы есть с руководителями, которые умеют взбираться на верхние этажи руководящей пирамиды и не умеют руководить. Для них и создают десятки томов – стандартов качества, где

расписаны алгоритмы управленческой деятельности на все случаи жизни. Потрясающее облегчение руководящей жизни! Трудовой коллектив работает строго по книгам, сам себя подбирает, воспитывает и контролирует. Все по стандарту. Чем строже выполняются инструкции – тем качественнее руководство и производство.

В образовании – не так. Особенно – в российском. Наличие подробной обязательной инструкции, как учить, – это «погибель» качества в образовании. Слишком непростой товар – образование. Никто не знает, в точности, не только, что это такое, но и как его производить. Зато есть старые добрые педагогические рецепты – деловая игра, публичность успеха, соревнование и др.

Много есть педагогов, превосходно знающих, как обеспечить методичность и дидактичность, и как потом рассказать об этом с трибуны, но мало таких, которые бы действительно применяли в учебном процессе рецепты. Почему мало, понятно – это трудно и часто невозможно. Вот именно здесь – в деле нового внедрения тысячелетиями проверенных технологий обучения – управленцу и могут помочь компьютерные технологии.

Для повышения уровня качества образования недостаточно только методично и дидактично вести учебный процесс. Грамотно организованное многоэтапное соревнование между образовательными учреждениями в виде, например, образовательной компьютерной олимпиады может привести к значительному росту успешности обучения без особого увеличения нагрузки на учебный процесс. В такой олимпиаде могут померяться знаниями и учителя, претендующие на переход в новую категорию.

Результаты массовых компьютерных олимпиад могут служить:

- основанием для аттестации специалистов и учреждений;
- основанием для выставления итоговых оценок учащимся;
- неиссякаемым источником данных для мониторинга качества образования в учреждении, регионе, стране.

Сами такие олимпиады – не что иное, как великолепный инструмент управления качеством образования. Приведем главные характеристики олимпиад:

- Массовость (всеохватность);
- Непрерывность (многостадийность);
- Привлекательность;
- Соревновательность;
- Значимость (результаты влияют на оценки в журнале, размеры премий для участников учебного процесса, оценки аттестационных комиссий, повышение категорий и зарплат и пр.);
- Объективность;
- Публичность («страна знает своих героев»).

*Все люди зануды, кроме тех случаев,
когда мы в них нуждаемся.
Оливер Уэнделл Холмс (старший)*

ОИТ и организация управления качеством. Пять важнейших характеристик антуража массовых компьютерных тестирований

*У всякого безумия есть своя логика.
Вильям ШЕКСПИР*

Есть много методических пособий о новых образовательных информационных технологиях, наполненных призывами и таблицами, которые должны помогать повышать качество образовательной продукции. Толку от большинства из них не много. С другой стороны, каждый действующий педагог может привести рецепты успеха в этой области.

Здесь мы опишем один из таких рецептов применения ОИТ, кажущийся нам вполне достойным для освоения и массового применения. Он известен, и ОИТ играют здесь не ведущую, а вспомогательную, хотя и очень важную роль. Речь пойдет о внедрении в учебный процесс старого доброго соревнования.

Наш десятилетний опыт позволяет утверждать, что грамотно организованное, под видом турнира или олимпиады, массовое компьютерное тестирование способно творить чудеса.

Известны компьютерные комплексы, с помощью которых можно организовать массовое компьютерное тестирование. Многие из них можно найти по ссылкам в Приложениях.

Лучше, если интерфейс комплекса удовлетворяет изложенным здесь требованиям.

Не менее чем интерфейс компьютерного комплекса, важен антураж тестирования. Приведем описание антуража, который мы считаем наилучшим:

1. Турнир является многостадийным, ведущимся в течение учебного года;
2. В турнире участвуют все учащиеся соревнующихся образовательных учреждений, без исключения. Результат каждого должен войти в копилку команды;
3. Результаты каждого турнирного сражения немедленно публикуются с указанием вклада каждого участника;
4. Вклады участников всегда положительны. Вклады только накапливаются;
5. Результаты побед немедленно сказываются в жизни: оплаченные экскурсии, походы, пикники, личные и командные медали, публичные поздравления и вознаграждения для педагогов, выставление оценок в журнал по итогам сражений.

Все это трудно было бы организовать, не имея компьютерных классов. С ними все иначе: Имея в распоряжении двух инженеров и ресурс школьного класса из 12 ПК, можно в течение светового дня провести тестирование 300-400 человек. Публикация результатов, с учетом автоматизации не займет более получаса.

Вот, вкратце, и все основное. Можно предложить и иные пути использования ОИТ в благородном деле развития системы образования и управления качеством ее продукции.

ОИТ и кадровое обеспечение

*Слишком послушные сыновья никогда не достигают многого.
Абрахам БРИЛ. <http://frazza.net.ua>*

Повышение квалификации и переподготовка кадров сферы образования в области ОИТ являются неотъемлемой частью мероприятий, направленных на поддержание и повышение качества и количества ЗУН выпускников образовательных учреждений. Новые пути повышения эффективности переподготовки и повышения квалификации связаны с появлением и внедрением новых компьютерных и телекоммуникационных технологий обучения и оценки качества ЗУН.

Не каждая районная образовательная структура укомплектована штатом методистов, способных вести эту работу в полном объеме. Проведение же мероприятий по переподготовке через краевую структуру столь же значимо, как и затратно: отрывать от учебного процесса на длительное время большие группы учителей, обеспечивать их транспортом, жильем, оборудованием – часто становится непозволительной роскошью.

Компьютерные классы есть уже во всех районных структурах. Организовать на их базе кратковременные курсы переподготовки – не проблема. При наличии программного обеспечения справиться с задачей дистанционной групповой переподготовки в области профессиональных знаний справится один подготовленный методист.

Именно таким путем пошли, в рамках договора о совместной деятельности, Департамент образования и науки администрации Приморского края и ДВГУ.

В течение 2000-2002 гг. созданы 17 электронных учебных пособий с электронными тренажерами для реализации полнообъемного дистанционного обучения в рамках Программы ДПО «Менеджмент в образовании», созданы коллектив преподавателей и документационное обеспечение, инициирован и успешно развивается учебный процесс на базе территориальных представительств и филиалов ДВГУ.

Важной особенностью является ориентированность Программы на переподготовку кадров в области ОИТ, мониторинговых технологий (массовых тестирований знаний, сбора и обработки данных, прогноза) и технологий менеджмента качества.

Развитие Программы позволит подготовить достаточный контингент специалистов для успешного полнообъемного внедрения на территории Приморского края новых технологий контроля и управления качеством образования.

ОИТ и защита качества образования

*Проблема – в том, что все, кто здесь работал, и я в том числе, хотели
сделать действительно чистую работу,
только они не сделали чистой работы, они хотели много – и быстро.
(Рик БЭЙКЕР, пример в фильмах: «Кинг-Конг», «Звездные войны»)*

Создание национальной системы открытого образования (СОО) является приоритетным направлением деятельности Министерства образования РФ. Решение этой задачи неразрывно связано с процессом общего реформирования системы образования. Формирование СОО опирается на новые, непривычные для России принципы – технология и структура создается по инициативе «снизу». В качестве инициаторов создания СОО в настоящее время лидируют несколько столичных и региональных университетов, в частности, Даль-

невосточный государственный университет, становящийся форпостом национальной СОО в Азиатско-Тихоокеанском регионе, и др.

Образование уже воспринимается нами не только как некое понятие, но и как товар. В России строится рынок образовательных услуг. А раз это так, то сам товар, производители и потребители товара, организаторы производства и потребления товара должны быть соответствующим образом защищены правовым механизмом.

Действующий Закон «Об образовании» не включает четкого, доведенного до численных характеристик, определения качества и уровня качества образования; нет в Законе понятия конкуренции и монополизма в сфере образования, равно как нет и многих важных понятий, связанных с рынком, отсутствуют непротиворечивое описание механизма защиты качества.

Система образования, как производственный механизм, состоит из (зависит от) многих компонентов и факторов, влияющих на качество и зависящих от него: преподаватели и учителя; производственный персонал учебных заведений и научно-методических центров; авторы учебных пособий и образовательных технологий; абитуриенты и учащиеся; администраторы учебных заведений и управлений образованием; работодатели; грантодатели, вкладывающие средства в развитие учебных заведений; зарубежные организаторы внедрения российских образовательных технологий; эксперты и гаранты качества; государство.

Все они должны быть защищены не только от воздействий извне, но и взаимно. Например, вполне логичны такие меры защиты: решения о реформировании, занятии должностей и др. не должны приниматься администратором без учета объективных показателей качества; выпускник вуза с более высоким рейтингом качества должен иметь приоритет при трудоустройстве; работодатель должен иметь четкие государственные либо иные гарантии качества принимаемого на работу выпускника вуза.

Производственные принципы требуют, чтобы качество образовательных услуг выражалось количественно во взаимном сравнении. Причем эти выражения – напр., рейтинговые таблицы – должны быть защищены достаточными гарантиями.

Мониторинг качества образования не менее важен, чем медицинский, химический, радиационный, экологический и социальный. Это основа для построения в федеральном масштабе механизма защиты, гарантирования и повышения качества образования. Этот механизм должен опираться не только на материальную и методическую базу сети ЦМКО, не только на результаты ее деятельности, но и на некоторые нетрадиционные для нашей системы образования аспекты. Обсудим некоторые из них.

Работодатель, как правило, выбирает будущего сотрудника из нескольких претендентов. Его выбор будет более объективен, если он станет опираться на результаты деятельности регионального ЦМКО – таблицу рейтингов учебных заведений и оценки независимого тестирования остаточных знаний. Так реализуется один из аспектов механизма защиты качества. Если же, при этом, претенденты законодательно защищены друг перед другом и перед работодателем на основе тех же рейтинговых таблиц, то реализуется второй аспект защиты. Одновременно с первыми двумя, реализуется третий аспект – взаимная защита образовательных учреждений.

Современный работодатель заинтересован в упреждающем привлечении будущих сотрудников, до окончания ими обучения. Его выбор претендентов будет объективным, если он получит доступ к данным независимого тестирования ЗУН студентов, например, в сети Интернет. Студенты, зная, что за их успехами следят будущие работодатели, несомненно, более серьезно отнесутся к своему образованию. Таким образом, повышается индивидуальная образовательная мотивация и реализуется четвертый аспект – поддержка повышения уровня качества.

Вузы заинтересованы в заблаговременном формировании контингента абитуриентов. Данные ЦМКО являются основой для:

- построения межвузовской политики такого формирования;
- проведения профориентационных мероприятий.

Публикация результатов деятельности ЦМКО – рейтинговых таблиц – неизбежно приводит к появлению конкуренции среди учащихся и образовательных учреждений. Учреждения стремятся привлечь лучших преподавателей, внедрить лучшие образовательные программы, повысить оснащенность и др., что приводит к общему подъему уровня образованности.

Компьютеризированный мониторинг позволяет анализировать деятельность не только вузов в целом, но и конкретных преподавателей и управленцев. Эти данные позволяют объективно оценить деятельность и стать основой защиты прав:

- преподавателей и учителей как производителей образования;
- абитуриентов и учащихся, как покупателей и носителей образования;
- учебных заведений, как организаторов производства образования;
- работодателей, как потребителей образования;
- государства, как организатора рынка образовательных услуг.

Ведение сравнительного мониторинга качества образования на уровне регионов и государств не только станет основой стимулирования конкуренции в этой сфере, но и является абсолютно необходимым шагом на пути к интеграции образовательных систем и построению глобальной системы открытого образования.

Следует внимательно отнестись к решению о конкретном виде механизма защиты и гарантирования качества. Мы должны быть уверены, как это не парадоксально, что качество образования будет защищено и от самого механизма своей защиты.

Каждое нововведение приводит не только к позитивным, но и к негативным последствиям. В частности, одним из возможных негативных результатов внедрения нового механизма, может оказаться появление общего административного стремления к унификации образования. Это может привести, как это случилось в некоторых странах, где такие механизмы давно сложились, не к развитию образованности, а к падению уровня качества образования. О глубинной связи этих, на первый взгляд, несвязанных явлений – унификации и устойчивости – свидетельствует наука.

Свойства стабильности и самовосстановления систем прямо связаны со степенью их неоднородности. Установлено, что наименее устойчивыми являются унифицированные, однородные системы. Этот факт, кроме прочего, дает ключ к пониманию одного из российских «чудес»: именно в неунифицированности и традиционной разнородности образовательной системы России заложена ее магическая способность к самостоятельному восстановлению и поддержанию уровня качества.

ОИТ и проблемы финансирования

*Советы принимай от всех дающих,
но собственное мнение береги.
Вильям ШЕКСПИР*

Возникающая в России система ДО изначально была поставлена в жесткие условия. На деньги государственного бюджета были дополнительно оборудованы несколько столичных вузов, объявивших раньше других о своем желании строить СДО и победившие в конкуренции. Остальные вузы проявляют финансовую самостоятельность, заключающуюся часто в том, что дистанционным просто называют традиционное заочное обучение для привлечения абитуриентов.

Есть мнение, что российская система образования сама «породила» свое недофинансирование. Здесь есть доля истины, ибо система давно уже страдает известным пороком – неэффективный менеджмент. Сапожник без сапог: вузы производят тысячи вполне успешных менеджеров и маркетологов, и забывают о них, принимая на соответствующие должности в штат не специалистов, а остепененных научных сотрудников и преподавателей. Сотрудники вузов всегда работают на два фронта – кандидат наук не человек, пока не стал руководителем (менеджером), а хозяйственник – не человек, пока не стал кандидатом наук. Эффективность производственной и научной деятельности при этом, естественно, невысока.

Наше дворянство – самая невежественная часть общества
Павел Строганов, генерал-лейтенант
Зам. министра внутренних дел России, герой 1812 года

Есть и другая известная сторона «медали» недофинансирования. Поясним. Вузы «ташат» с собой из недавнего прошлого традицию повышения оценок в студенческой зачетке не за знания, а за «пользу» – организацию мероприятия, ремонта и пр. Совершить организационный «подвиг» для некоторых студентов значительно проще, чем понять материал сложной дисциплины. Организация отнимает довольно много сил у студентов, пошедших по такой тропе, на учебу времени не остается. Знаний у них нет, но вуз «закрывает на это глаза». В итоге они гарантированно защищают диплом и становятся молодыми специалистами. Далее они, обойдя прилежно учившихся коллег, традиционно быстро делают карьеру, получают степени, не являясь специалистами, и, в конце концов, решающий голос при распределении государственных средств.

Это хорошо, что у нас в обществе есть такие активные самобытные менеджеры. Наша беда в другом – эти самобытные менеджеры не обладают достаточным объемом знаний ни в своей специальности, ни в менеджменте. Со студенчества они несут в себе неприязнь не только к преподавателям, кривившим душой при выставлении оценок, но и к системе образования в целом. Они опасаются профессионалов и неохотно допускают их в свои ряды. Они по-своему честны, и у них просто рука не поднимется проголосовать за направление достаточно больших средств на образовательные проекты, т. е., по существу, в сферу ОИТ. Покопайтесь в дипломных вкладышах депутатов и министров – и Вы найдете среди них множество «махровых» троечников.

Россией правят троечники
(из газетных цитат)

Мы уверены, что строящаяся на основе ОИТ федеральная Система открытого образования способна избавить российскую образовательную структуру от этих и других недостатков.

Глава 08. Применение компьютерных технологий массового тестирования

*Рассказывать что-нибудь без особенной цели даже своим знакомым –
свойственно только праздному болтуну.*

ПИСАРЕВ Дмитрий Иванович, aforizm.kaminplus.ru

Автоматизированные технологии массового тестирования ЗУН в ряде стран используются уже несколько десятков лет. Есть сведения, что профессиональное тестирование государственных служащих проводилось в Китае уже в XXI веке до нашей эры. Согласно историческим источникам, определенные технологии тестирования на проф.пригодность (например, определение качеств воинов) разрабатывались и использовались в государствах древнего Средиземноморья.

В России, США, Великобритании, Франции, Германии, Австралии, Канаде, Японии, Голландии, Австрии, Китае и др. странах существуют государственные и частные организации, занимающиеся разработкой тестовых комплексов, их распространением и организацией массового тестирования. Есть организации, занимающиеся регулярным сбором информации тестирований в целях ведения мониторинга качества образования. Результаты такой деятельности позволяют планировать развитие систем управления качеством.

Развита индустрия тестового контроля знаний в США. В каждом штате США сегодня функционируют десятки специализированных центров тестирования. Многие такие центры имеют отделения за рубежом, занимающиеся «рекрутированием» абитуриентов и дешевой рабочей силы. В США есть общенациональные тестовые системы (SAT, GRE и др.) для:

- оценки знаний учащихся;
- аттестации образовательных учреждений;
- отбора перспективных абитуриентов в университеты;
- оценивания характеристик людей при приеме на работу и на службу в армии;
- мониторинга образовательной среды;
- глобальных социологических исследований.

Массовые тестирования, результаты которых прямо влияют на уровень государственных дотаций или «определяют судьбу» абитуриентов университетов, как правило, проводятся независимыми организациями. Личности экспертов и авторов тестовых заданий при этом содержатся в строгом секрете. Это делается для повышения уровня объективности (и еще, как говорят американцы, во имя укрепления демократии). В результате проведения таких массовых тестирований:

- учащиеся – получают аттестационные оценки за периоды обучения;
- абитуриенты – получают информацию о том, в какие университеты они могут быть зачислены автоматически либо с льготами оплаты обучения, либо – насколько успешно они могут участвовать в абитуриентском конкурсе;
- педагоги – получают объективную характеристику качества своей работы;
- учебные заведения – получают официальное место (рейтинг) в рейтинговых таблицах штата, страны, содружества (это самое важное основание для повышения или

снижения уровня оплаты обучения, привлечения спонсоров, получения льгот в налогообложении и пр.);

- родители – получают информацию об образовательных учреждениях;
- выпускники учебных заведений – получают возможность претендовать на более престижные рабочие места при приеме на работу (во многих штатах США это право выпускников защищено законодательно);
- государственные органы управления – получают основания для распределения дотационных средств между образовательными учреждениями на следующий год (это важнейший момент управления качеством образования в США).

Существуют договорные отношения между многими европейскими и скандинавскими странами в области проведения международных сравнительных исследований качества образования на основе тестовых технологий. Такие отношения имеют хорошую основу – историю развития международных физико-математических олимпиад, зародившихся в середине XX века в СССР и странах социалистического лагеря.

Тестирования ЗУН могли бы быть более популярными не только среди абитуриентов и безработных, но и среди специалистов и пенсионеров, если бы создатели их сценариев не пренебрегали (относились как к чему-то не серьезному) важными чертами человеческой натуры – спортивным азартом, стремлением к лидерству и т. п. Люди, стремящиеся показать лучший результат в тестировании, обязательно будут изучать учебники, энциклопедии. Велика часть российского населения, считающая, что победа в сравнении знаний важнее, спортивной. Азартных людей много, и это означает, что хороший сценарий тестирования может привести к заметному росту образованности населения.

Массовые тестирования знаний в России

*Чем раньше солнце встает – тем выше оно поднимается
(Удмуртская пословица)*

В России работы в области массового компьютерного тестирования знаний стали реально проводиться только в последние два-три десятилетия. К несомненным «старейшинам» и лидерам в этой области следует отнести Новосибирский государственный университет (НИИ МИОО НГУ, разработка и реализация технологий образовательных тестирований студентов и школьников) и Московский государственный университет (Центр «Гуманитарные технологии», разработка и реализация проекта Всероссийской компьютерной олимпиады «Телетестинг», массовое компьютерное тестирование на профессиональную пригодность).

Много лет организует массовые всероссийские тестирования на основе бланковой, и частично компьютеризированной технологии Центр тестирования (ЦТ) МОРФ. В последние три года ЦТ МОРФ развернул массовые ежегодные тестирования и на основе полностью компьютерной технологии.

Подробное описание этих проектов можно найти, например, на сайтах

1. Проект «Единый Государственный Экзамен»

<http://www.ege.ru/>

<http://ege.ru/>

2. Телетестинг: всероссийское компьютерное тестирование для старшеклассников

<http://www.teletesting.ru/>

3. Центр тестирования МО РФ, проект «Централизованное тестирование»

<http://www.rustest.ru/>

<http://rustest.ru/>

4. Центр «Гуманитарные технологии», проект «Телетестинг»

<http://www.ht.ru/>

В России появилось немало специалистов – тестологов (далее – в алфавитном порядке): В. С. Аванесов, А. Н. Майоров, Л. В. Макарова, В. И. Нардюжев, А. И. Субетто, А. О. Татур, В. А. Хлебников, М. Б. Чельшкова, А. Г. Шмелев и др. Защищено много диссертаций, посвященных тестовым технологиям. Можно говорить о складывающихся в разных регионах страны научных школах тестологов.

*Взрослые всегда должны помнить о том;
чтобы не показать детям дурного примера
ЮБЕНАЛ Децим Юний, aforizm.kaminplus.ru*

С 1995 года массовые компьютерные тестирования школьников ведутся в Приморском крае (ЦНОТ ДВГУ, Приморский краевой ЦНИТ, Приморский краевой ЦМКО). Усилиями ДВГУ, районных и городских управлений образованием, на территории Приморского края созданы и создаются около двух десятков территориальных Центров мониторинга качества образования (ЦМКО). Среди них крупнейшие – Владивостокский ЦМКО и Дальнегорский ЦМКО.

К несомненным успехам созданной ДВГУ структуры отнесем шестикратное общероссийское лидерство по количеству проведенных сеансов в период централизованных тестирований.

В 2002 году ДВГУ первым в России предпринял и осуществил проект массового круглогодичного компьютерного тестирования знаний всех студентов младших курсов очных отделений университета (более 8 тыс. человек) по 10 общим дисциплинам (работы ведутся специализированным подразделением ДВГУ – Открытым университетом).

Более подробные материалы о мониторинговой деятельности в Приморье содержится в электронных Приложениях.

Полдюжины средств профилактики негативных ситуаций при компьютерном тестировании

*Когда исходят лишь из выгоды, то множат злобу.
КОНФУЦИЙ*

В процессе тестирования не все претенденты стремятся добросовестно выполнять задания. Возможны негативные ситуации, связанные с широким спектром мотивов: от принципиального неприятия тестирования как метода контроля знаний до чисто детского баловства даже со стороны взрослых.

Это бывает следствием сговора группы, не заинтересованной по каким-либо причинам проходить тестирование и желающей его сорвать. Отказ искажает результаты, особенно если количество отказавшихся превышает 5%. Декларированные отказы (независимо от обоснованности) снижают у остальных претендентов мотивацию работы.

Особенно негативно публичные отказы сказываются на репрезентативности выборки теста, если инициаторами являются лидеры групп. Попытки убеждать и уговаривать ни к че-

му обычно не приводят. Запуганная группа при дидактическом тестировании даёт заведомо заниженные результаты.

Способами профилактики ситуации отказов является, как показывает практика:

- продуманный выбор времени тестирования;
- определение разумной продолжительности теста (с учётом того, что после 30– 35 мин работы с тестом продуктивность решений снижается);
- комфортность помещения, температуры, отсутствие отвлекающих факторов (шум);
- предварительное разъяснение цели и задач тестирования, значимости его результатов для объективного оценивания ЗУН учащихся;
- внешний вид и стиль общения персонала, должны не отвлекать претендентов, а стимулировать добросовестную работу, создавая доброжелательную атмосферу.

Второй вид искажений – стремление претендента, по возможности, прятаться за нейтральными ответами типа «затрудняюсь ответить». Считается, что если в тесте у кого-либо из обследованных до 30% ответов такого типа, то его следует исключить из выборки (при оценке репрезентативности, надёжности, валидности теста).

Опыт показывает, что если более 50% претендентов затрудняются выполнить задание теста – это задание неудачно, имеет низкую дискриминантную способность.

Третьим видом являются преднамеренные искажения ответов со стороны тех, кто знает правильный ответ, но по каким-то причинам решает пошутить. Среди них встречаются любители «системного» подхода. Они отмечают как верный каждый первый (третий, четвертый, поочередно) вариант выполнения, либо на первую половину теста отвечают утвердительно, а на вторую – отрицательно, и т. п.

Проверку искренности отвечающего лучше проводить методами, которые не вызывают возникновения ощущения подозрения. Психологически опросники включают группы вопросов, контролирующих добросовестность претендента – «шкалы лжи», а в профессионально составленные анкеты – специальные контрольные вопросы.

В дидактические тесты включают такие вопросы также. Этот приём целесообразно применять не в первичном варианте тестов, а в том, где после отбора заданий проведена стандартизация. Если при тестировании диагностом были замечены отвлекающие факторы – он обязан их зафиксировать в отчёте и попытаться дать прогноз возможного влияния этих факторов на результаты. К таким факторам относятся: появление в аудитории постороннего или представителя администрации, шум около аудитории, яркие внешние проявления болезни кого-либо из претендентов (тошнота, громкий кашель).

Для того чтобы предупредить попытки тестируемых найти ответ с помощью подсказки или других тестируемых, необходимо заранее в инструкции по проведению тестирования чётко оговорить порядок действий педагога-диагноста:

- замечания с фиксацией нарушителя и задания, по которому запрашивалась помощь;
- исключение для нарушителей этого задания из числа оцениваемых;
- повторное тестирование нарушителя в индивидуальном порядке;
- прекращение тестирования всей группы и т. п.

При создании параграфа использованы материалы книги: Михайлычев Е. А. *Дидактическая тестология*. – М.: Народное образование, 2001. – 432 с.

Дидактический принцип индуктивной мотивации в организации массовых компьютерных тестирований

*Доброе слово и кошке приятно.
(пословица)*

Мотивация участников – важнейшее условие не только массовости, но и дидактической целесообразности массовых тестирований. Неправильно мотивированные участники сохраняют в своей памяти произвольно возникшие у них в течение сеанса тестирования ощущения усталости и «принудиловки». Это может негативно сказаться на последующей образовательной траектории личности. Позитивная же мотивация может непосредственно «подтолкнуть» рост качества образования в школах территории.

В конечном итоге, проводниками мер мотивации являются руководители и персонал территориальных центров тестирования (мониторинга качества образования).

Можно создать и развесить плакаты, можно провести разъяснительные передачи для школьников и их родителей по телевидению. Но весь созданный при этом настрой участников тестирований и их родителей улетучится, когда кто-либо из персонала в компьютерном классе выразит сомнение:

- в качестве теста;
- в целесообразности участия в тестированиях;
- в объективности подведения итогов тестирования;
- в искренности заявлений руководителей системы образования относительно судьбы лучших участников;
- в честности руководства территориального центра и отсутствии подлога;
- в истинности целевого расходования средств (денег), полученных от испытуемых за участие в тестированиях, и пр.

Людей, не способных к самоконтролю, не способных держать при себе свои чувства и сомнения, не должно быть в штате центра.

Позитивная мотивация участников, создаваемая централизованными мерами, подкрепленная общим положительным настроем сотрудников центра (опосредованно переданной мотивацией), может творить чудеса. Об этом свидетельствует описанный в настоящем пособии опыт ДВГУ массовых тестирований на территории Приморского края.

Организуя тестирования в масштабах региона, следует позаботиться о мотивации сотрудников территориальных центров. Свою, созданную внешними организаторами, мотивацию они передадут (**индуцируют**) участникам тестирований. В ДВГУ это делается следующим образом:

- создана прозрачная система выплат вознаграждений сотрудникам центров, ставящая суммы в прямую зависимость от количества проведенных сеансов;
- ежедневно сотрудники центров получают неформальные письма, где выражается исключительно позитивное отношение организаторов к ежедневным успехам, называются имена лучших, проводятся сравнения и обсуждения;
- организовано соревнование между центрами за получение призов по итогам мероприятий.

Действенность этих мероприятий очевидна: ДВГУ шесть раз стал российским лидером по массовости участия школьников в централизованных тестированиях.

08.01. Мониторинг качества образования в России. Международное сотрудничество

*Люди только тогда сообщают нам интересные сведения,
когда мы им противоречим.
Бернард Шоу, aforizm.kaminplus.ru*

Качество образования школьников, работы учителей и школ России в недалеком будущем будет контролироваться независимыми экспертными центрами, координирующими набор абитуриентов в вузы, строящими таблицы рейтингов образовательных учреждений, ведущими мониторинг образовательных систем. Новые для России рыночные термины – менеджмент, маркетинг – прочно укоренятся в лексиконе работников образования.

Работы по внедрению мониторинговых технологий в настоящее время координируются Министерством образования РФ (<http://www.informika.ru/>) и Институтом общего среднего образования Российской Академии образования (ИОСО РАО) (<http://www.ioso.ru/>)

Россия осуществляет ряд международных проектов в области тестирования. Информацию об этом можно найти на официальном сайте www.informika.ru. Результаты этих мероприятий используются для построения международных рейтингов образовательных систем. Ежегодно публикуются рейтинговые таблицы, где можно почерпнуть сведения о том, школьники какой страны лучше знают, например, математику или химию. Объективность рейтингов обеспечивается специальными стохастическими процедурами, на основании которых выбираются конкретные школы и классы на территории стран, где проводятся тестирования.

В последние годы отмечается повышенный интерес к проблемам качества образования во всем мире. Страны объединяют усилия в разработке подходов к оценке и управлению качеством образования. Создается система мониторинга качества образования в мире. В ней участвуют более 50 стран. Ее организаторы:

- Международная ассоциация по оценке учебных достижений IEA (International Association for the Evaluation of Educational Achievement);
- Организация экономического сотрудничества и развития – OECD (Organisation for Economic Co-operation and Development).

Примерами международного сотрудничества в данной области являются международные сравнительные исследования, среди которых следует выделить: IAEP-II-1991, TIMSS-1995, TIMSS-R 1999, CIVIC -1999, CiViC-2000, PISA-2000.

Полный цикл подобного исследования в настоящее время, в связи с быстрым развитием теории и практики педагогических измерений, а также технологии проведения тестирования и обработки результатов, составляет 3-4 года. Это позволяет организовать эффективный мониторинг образовательных систем по ключевым показателям качества и эффективности образования. Одного срезового исследования, проводимого один раз в десять лет, недостаточно для выявления тенденций развития или последствий реформ.

Россия участвовала в международных исследованиях учебных достижений в области

- математики;
- естественнонаучных предметов;
- граждановедения;
- чтения.

(IAEP-II, TIMSS-1995, TIMSS-1999, SITES, CIVIC-1999, CIVIC-2000, PISA-2000, PIRLS-2000).

Наиболее важным международным исследованием по оценке качества образования в мире в 90-х годах, в котором участвовала Россия, было III Международное исследование по оценке качества математического и естественнонаучного образования – TIMSS. Цель исследования – сравнительная оценка естественно-математической подготовки учащихся средней школы в странах с различными системами образования и выявление факторов, влияющих на уровень этой подготовки.

В исследовании TIMSS участвуют многие страны: Австралия, Австрия, Англия, Аргентина, Бельгия (языки обучения: французский и фламандский), Болгария, Венгрия, Германия, Греция, Гонконг, Дания, Индонезия, Иран, Ирландия, Исландия, Израиль, Индонезия, Италия, Испания, Канада, Кипр, Колумбия, Корея, Кувейт, Латвия, Литва, Мексика, Нидерланды, Новая Зеландия, Норвегия, Португалия, Россия, Румыния, Сингапур, Словацкая Республика, Словения, США, Таиланд, Филиппины, Франция, Чешская Республика, Швеция, Швейцария, Шотландия, ЮАР, Япония.

Задачи исследования TIMSS:

- Сбор, анализ и обобщение информации о системах образования.
- Анализ и сравнение систем естественно-математического образования на уровне определения целей, планирования содержания образования и требований к подготовке учащихся по конкретным предметам.
- Анализ учебного процесса обучения математике и естественнонаучным предметам.

Результаты проведенных в России международных исследований по оценке качества образования публикуются в сборниках материалов. Их готовит Центр оценки качества образования (ЦОКО) ИОСО РАО. Например:

- ВЫПУСК 1. Ковалева Г. С. Основные подходы к сравнительной оценке качества математического и естественнонаучного образования в странах мира (по материалам международного исследования – TIMSS)
- ВЫПУСК 2. Сравнительный анализ математической и естественнонаучной подготовки учащихся основной школы России. Денищева Л. О., Дюкова С. Е., Ковалева Г. С., Корощенко А. С., Краснянская К. А., Мягкова А. Н., Найденова Н. Н., Резникова В. З., Суравегина И. Е. / Под ред. Г. С. Ковалевой.
- ВЫПУСК 3. Что отвечают учащиеся, учителя и директора школ России на вопросы международных анкет? (Альманах международного исследования TIMSS /Часть I/) Составители: Гуссак Н. А., Ковалева Г. С.
- ВЫПУСК 4. Сравнительная оценка естественно-математической подготовки выпускников средних школ России. Денищева Л. О., Ковалева Г. С., Краснянская К. А., Лошаков А. А., Найденова Н. Н., Нурминский И. И., Тарасов Ю. П. / Под ред. Ковалевой Г. С.
- ВЫПУСК 5. Что отвечают выпускники и директора средних школ России на вопросы международных анкет? (Альманах международного исследования TIMSS /Часть II/) Составители: Гуссак Н. А., Ковалева Г. С.
- ВЫПУСК 6. Сравнительный анализ программ и учебников по математике и естественнонаучным предметам. Ивченкова Г. Г., Краснянская К. А., Ковалева Г. С. и др. / Под ред. Ковалевой Г. С.

- ВЫПУСК 7. Сравнительная оценка математической и естественнонаучной подготовки учащихся начальной школы России. Ивченкова Г. Г., Ковалева Г. С., Краснянская К. А. и др. / Под ред. Ковалевой Г. С.
- ВЫПУСК 8. Модели взаимодействия факторов, влияющих на уровень математической и естественнонаучной подготовки школьников России. Ковалева Г. С., Краснянская К. А., Найденова Н. Н. и др. / Под ред. Найденовой Н. Н.
- ВЫПУСК 9. Технический отчет по проведению в России международного исследования – TIMSS. Каменщикова Е. К., Ковалева Г. С., Крайнова Н. А., Найденова Н. Н. / Под ред. Ковалевой Г. С.

Проект TIMSS является одним из самых крупномасштабных образовательных исследований по охвату исследуемых проблем и по количеству стран-участниц.

TIMSS объединил усилия специалистов более 50 стран мира. В процессе исследования изучаются тысячи программ и учебников различных стран по математике и естественнонаучным предметам, проводится тестирование и анкетирование сотен тысяч школьников, тысяч учителей и директоров школ.

В проведение исследования TIMSS в России вовлекается большое количество специалистов: представители РАО и МО РФ, работники образования десятков регионов России, директора и учителя сотен школ, десятки тысяч учащихся. Формирование обследуемой выборки осуществляется в соответствии с международными требованиями, анализируется и утверждается международными экспертами.

Участие нашей страны в TIMSS позволило также получить информацию о системе образования России, которой она раньше не располагала.

Особенностью TIMSS является его значительная ориентация не на ранжирование стран по уровню учебных достижений учащихся, а на объяснение полученных данных.

ОРГАНИЗАЦИЯ И ПРОВЕДЕНИЕ ИССЛЕДОВАНИЯ TIMSS-R (1999) В РОССИИ

В TIMSS-R (1999) участвовало 38 стран: Австралия, Англия, Бельгия, Болгария, Венгрия, Гонконг, Иордания, Иран, Израиль, Индонезия, Италия, Канада, Кипр, Южная Корея, Латвия, Литва, Македония, Малайзия, Молдова, Марокко, Нидерланды, Новая Зеландия, Россия, Румыния, Сингапур, Словацкая Республика, Словения, США, Таиланд, Тайбэй (китайский), Тунис, Турция, Филиппины, Финляндия, Чешская Республика, Чили, ЮАР, Япония. Из них 26 стран участвовали в исследованиях 1995 г. и 1999 г., 12 стран присоединились к исследованию на втором цикле.

Отличительной особенностью исследования TIMSS является научное сотрудничество специалистов всех участвующих в исследовании стран. Это позволяет включить в работу ведущие коллективы мира и таким образом поддерживать высокое качество проводимого исследования. Так, в проведении исследования и разработке его инструментария принимали участие многие научно-исследовательские центры или профессиональные организации мира: Служба педагогического тестирования (ETS, США), Канадский Центр Статистики (Statistics Canada, Канада), Секретариат IEA (Нидерланды), Центр обработки данных (DPC, Германия) и др. Для координации усилий специалистов различных стран были созданы совещательные комитеты, которые состояли из ведущих специалистов мира. Координация всего исследования осуществлялась Международным координационным центром, который располагается в Бостонском колледже в США (Бостон). Экспериментальная проверка инструментария TIMSS осуществлялась во всех странах.

Исследование TIMSS в России проводилось специалистами ЦОКО ИОСО РАО при участии Департамента общего среднего образования Министерства образования РФ. В задачи российских специалистов входили организация и проведение исследования в стране, участие в разработке и адаптация инструментария к условиям России, формирование представительной выборки учащихся России, подготовка материалов исследования (их перевод, адаптация, подготовка к печати, тиражирование), организация и проведение тестирования и анкетирования по школам страны, обеспечение контроля качества проведения исследования в России, ввод и обработка данных, анализ результатов исследования, подготовка научных и технических отчетов, создание банка данных исследования. Организация и проведение исследования TIMSS в регионах России осуществлялись при участии МОРФ через министерства образования, управления или департаменты образования отобранных регионов России, педагогические институты, институты повышения квалификации работников образования и другие учреждения. Около 1500 специалистов приняли участие в проведении исследования TIMSS в России в 1999 г.

Выборка учащихся 8 класса основной школы России строилась из 89 регионов, объединенных в 10 экономико-географических зон на основе вероятностно-стратифицированного подхода. При выборе регионов России расчет производился на основе федеральной статистики Министерства образования. На федеральном уровне отсутствует соответствующая международным требованиям статистика, необходимая для формирования представительной выборки учащихся. Поэтому процедура выборки учащихся в России содержит дополнительную стадию – выбор регионов. Это позволяет собственными силами собирать необходимую региональную статистику в соответствии с международными требованиями.

Всего в исследовании было отобрано 47 регионов. Отбор школ проводился для каждого выбранного региона в отдельности. Всего в исследовании участвовало 189 школ. Из них 75% городских, 25% поселковых и сельских школ. В каждой школе выбирался один класс, все учащиеся которого принимали участие в тестировании. Всего в исследовании участвовало 4332 учащихся 8 класса. Учителя, ведущие в отобранных классах математику и все естественнонаучные предметы, участвовали в анкетном опросе. Всего было охвачено 189 учителей математики и 756 учителей естественнонаучных предметов.

При создании инструментария исследования TIMSS-R (1999) использовались те же подходы, что и в исследовании TIMSS (1995). Полный инструментарий включал тесты достижений (8 вариантов по 75-80 заданий закрытого и открытого типа по математике и естественнонаучным предметам, выполнение одного варианта было рассчитано на 90 мин); анкеты для учащихся, учителей математики и естественнонаучных предметов, администрации школы, наблюдателей за проведением тестирования в школе, экспертов-предметников по математике и естественнонаучным предметам; руководства по формированию выборки, проведению апробации инструментария и основного исследования, проверке заданий с открытыми ответами, вводу и обработке результатов, обеспечению контроля качества проведения тестирования в школах.

При разработке тестов дополнительно учитывалось, что на основе их выполнения будут сравниваться результаты двух срезовых исследований 1995 г. и 1999 г. В связи с этим в них частично использовались задания 1995 г., которые не были опубликованы в открытой печати. Опубликованные задания были заменены на другие из тех же разделов. Особое внимание уделялось соответствию статистических параметров. Во всех анкетах по ключевым показателям использовались те же вопросы.

При интерпретации результатов международных исследований математической подготовки учащихся 8 класса и выпускников средней школы, в которых участвовала Россия в 1995 г. и 1999 г., следует иметь в виду присущие им особенности. В подобных исследованиях для проверки, как правило, выделяются вопросы содержания обучения, которые

являются общими для большинства стран-участниц. Тем не менее, и при таком подходе содержание проверочных заданий в той или иной степени не отвечает содержанию программы обучения математике соответствующей параллели классов в каждой из стран.

Для России несоответствие проверочных заданий содержанию программы 8 класса значительно больше, чем для многих стран. Это связано с тем, что она сильно отличается от программ других стран по номенклатуре вопросов и требованиям к подготовке учащихся. По отношению к особенностям содержания курсов математики в России проверочные задания распределились следующим образом. Более половины (60%) заданий составлены на материале, который традиционно изучается в курсе математики 5-6 классов, а в большинстве стран продолжает изучаться в 7-8 классах. Только 23% заданий связано с материалом, изучаемым в курсе 7-8 классов. При этом основные темы курсов алгебры и геометрии этих классов весьма незначительно или вовсе не нашли отражения в содержании проверки. Поэтому наши учащиеся не получили возможности продемонстрировать уровень усвоения значительного по объему материала, который был ими изучен к моменту проведения тестирования.

Форма проверки международных исследований оказалась непривычной для россиян. За 90 минут ученик должен был ответить на 75-80 тестовых заданий различного типа, составленных на материале 6-ти учебных предметов естественно-математического цикла. В этих условиях явно проявились как положительные качества, так и недочеты нашего математического образования.

Анализ результатов исследований в 1995 г. и 1999 г. показал, что состояние математической подготовки восьмиклассников России не претерпело существенных изменений в сравнении с собственными результатами 1995 г., а также в сравнении с другими странами.

По отношению к результату по России (средний балл по тесту) страны-участницы можно распределить на три группы: результаты существенно выше российских, результаты статистически не отличаются от российских, результаты существенно ниже российских. Состав стран-участниц исследования в 1999 г. изменился по сравнению с 1995 г., изменился и состав этих групп.

Интересно отметить, что в 1995 и в 1999 г.г. первая группа включает 6 стран. Это страны Азии (Сингапур, Корея, Тайбэй, Гонконг, Япония) и две европейских страны – Чешская Республика (только 1995 г.) и Фламандская часть Бельгии. Они показали более высокие результаты в сравнении не только с Россией, но и со всеми странами-участницами.

В 1995 г. среди стран, результаты которых не отличаются от российских, входят страны, имеющие хороший уровень математического образования: Франция, Венгрия, Нидерланды, Австралия, Канада. Результаты этих же стран (кроме Франции, не принимавшей участия в исследовании) не отличаются от российских и в 1999 г.

В 1995 г. среди стран с результатами ниже российских – США, Англия, Германия, Италия. Результаты этих же стран (кроме Германии, не принимавшей участия в исследовании) остались ниже российских и в 1999 г.

При написании параграфа использованы (в сокращении) материалы перечисленных здесь выпусков ЦОКО ИОСО РАО, любезно предоставленные нам Ковалевой Г.С.

*Обращайся всегда к чужим богам.
Уж они-то точно выслушают.
Станислав Ежи ЛЕЦ*

08.02.Единый государственный экзамен в России. Чертова дюжина препятствий

*Когда страна отступит от закона,
тогда много в ней начальников
СОЛОМОН, царь иудейский aforizm.kaminplus.ru*

Среди новых информационных технологий, внедряемых МО РФ в системе образования, выделяется технология Единого государственного экзамена (ЕГЭ). Масштабное экспериментальное внедрение ЕГЭ в регионах стало возможно только благодаря всесторонней поддержке территориальных органов власти.

Первоначально в ЕГЭ применялась только бланковая технология, но, в последнее время обозначился «крен» в сторону внедрения компьютерного тестирования.

Основа ЕГЭ – объединение школьного и вузовского экзамена – пока не находит всеобщей поддержки во многих крупных государственных вузах России. Причин тому много, а главная – это недоверие руководства вузов к положениям технологии ЕГЭ предложенной МО РФ:

- фактически ответственность за качество контингента принимаемых по эксперименту абитуриентов перекладывается с вузовских приемных комиссий на плечи неизвестных вузам контролеров;
- нет ясности с технологией планирования набора студентов;
- нет ясности с ГИФО – государственными именными финансовыми обязательствами – которые государство предполагает закреплять за выпускниками школ.

Особенности технологии ЕГЭ позволяют автоматизировать обнаружение недочетов и недобросовестности в пунктах проведения ЕГЭ. Статистический анализ таких явлений показал, что более всего они проявляются в «переселенческих» территориях, где население проживает временно либо прибыло ранее из других регионов и стран.

К «тонким» моментам технологии следует отнести, например, неотработанность формы бланков и связанную с этим многократную перепроверку сканированных данных. В частности:

- На бланке регистрации в названиях полей используются разные типографские шрифты, что отвлекает внимание школьников от образцов и приводит к многовариантному написанию цифр и букв.
- В перечне образцов отсутствуют написания латинских букв, они могут потребоваться.
- Написание инструкций «красным по красному» и «розовым по розовому» приводит к сложностям для детей с пониженным уровнем зрения.
- Надписи – пояснения к полям непонятны школьникам, вызывают споры и требуют дополнительных устных пояснений.
- Боязнь школьников написать неправильную букву или неверно нарисовать крестик, поддерживаемая педагогами, играет негативную роль и может отрицательно сказываться на итоговой оценке знаний.
- Низкая скорость обработки результатов и доставки сертификатов к местам проживания.

Многих из отмеченных здесь, а также других «тонких» моментов используемой технологии можно было бы избежать, если видоизменить технологический план, более широко включить в него использование возможностей компьютерной техники. В настоящем пособии этому посвящен соответствующий раздел.

Препятствиями к внедрению компьютерных технологий ЕГЭ могут стать, например:

- слабое владение учащимися компьютерной клавиатурой;
- малая обеспеченность компьютерной техникой и отсутствие соответствующих центров тестирований;
- наличие тестовых заданий выполняемых в свободной форме, контроль которых пока невозможно автоматизировать с помощью компьютера.

Первые два из этих препятствий временные и, скорее всего, будут разрешены в большинстве территорий в течение ближайшего года – двух. Третье – более сложное, и связано оно с предполагаемой необъективностью оценки с использованием тестовой системы. Предложенная организаторами ЕГЭ система оценки «свободных» заданий не лишена недостатков, поскольку проверяющие и перепроверяющие выполненные задания специалисты неизбежно приносят в оценку «человеческий» фактор.

При разработке технологий ЕГЭ и ЦКТ, специалистами министерства несправедливо забыты положительные моменты, отработанные и внедренные Центром «Гуманитарные технологии» под руководством профессора А. Г. Шмелева при проведении всероссийских мероприятий «Телестестинг». Это, в частности, – соревновательность. Вряд ли есть особая нужда в запрещении повторной сдачи ЕГЭ, тем более, при переходе на компьютерную технологию, когда варианты комплектов заданий не повторяются. Если технические средства позволяют сделать возможной многократную сдачу теста, система образования территории и региона только выиграет, когда школьники в свободном режиме, в порядке живой очереди станут повторно (многократно) проходить тренировочное тестирование с итоговым зачетом лучшего результата. Показатели качества знаний от этого улучшатся.

ЕГЭ не может оставаться одноразовым мероприятием, проводимым всего лишь раз в год. Такая дорогостоящая, имеющая государственный размер, отлаженная машина не может стоять и ждать своего часа в течение года. Технология ЕГЭ должна поселиться в каждой школе на постоянной основе, перерасти из инструмента для измерений в инструмент для поддержки, повышения, гарантирования и защиты качества образования.

Мониторинговые исследования позволяют диагностировать образовательную среду и прогнозировать ее развитие. Такие работы, имеющие широкое распространение в ряде стран, не являются обыденностью в России. Благим пожеланием для нас остается применение результатов мониторинга для упорядочения формирования контингента абитуриентов. И вузам, и абитуриентам было бы неплохо определяться заранее с показателями конкурса и величинами проходных баллов на специальности. Пока полученные на ЕГЭ баллы не являются гарантией поступления абитуриента на выбранную специальность в выбранном вузе. Это – одна из проблем, удерживающих вузы от участия в эксперименте ЕГЭ.

Внедрение ЕГЭ требует многих предварительных мероприятий:

- обучение специалистов;
- создание и оборудование специализированных центров;
- отработка путей финансирования;
- разъяснительная работа с населением.

08.03. Мониторинг, аттестация, сертификация. Опыт США

*Перед Судьбой склоняясь – молчим:
Баланс не в нашей власти.
Расчет последствий и причин
Возможен лишь отчасти.
Роберт БЕРНС (перевод Е. Ереминой)*

Высшей возможной степени объективности оценки качества образования и обучения претендента, закончившего курс обучения в учебном заведении, можно достичь, если складывать (вычислять) ее, исходя из трех компонент:

- Результат тестирования ЗУН претендента;
- Результат опроса потребителей выпускников о качестве выпускников конкретного подразделения учебного заведения;
- Рекомендации преподавателей выпускника.

Во многих странах развита культура рекомендательных писем, а упомянутые опросы проводятся независимыми мониторинговыми центрами. Деятельность территориальных мониторинговых центров обычно финансируется либо государством, либо крупными образовательными ассоциациями. Статистические данные пользуются спросом:

- Работодатели – определяют, выпускников каких учебных заведений следует брать на работу в первую очередь;
- Руководители учебных заведений – определяют, каких преподавателей следует пригласить на работу;
- Абитуриенты – выбирают учебное заведение для поступления;
- Родители учащихся – планируют дальнейшее обучение своих детей;
- Студенты – определяют, к какому преподавателю следует пойти для выполнения дипломной работы и производственных практик;
- Органы управления образованием – регулируют ценовую политику и определяют планы распределения субсидий.

Абитуриенты, учащиеся, претенденты на занятие вакантных должностей и действующие специалисты регулярно проходят аттестацию. Как правило, все лица, успешно прошедшие аттестацию получают документ, удостоверяющий измеренный уровень их ЗУН.

Успешно прошедшие аттестацию претенденты получают **сертификат**. Специалистов, или выпускников образовательных учреждений, имеющих сертификат о качестве и количестве своих ЗУН, называют **сертифицированными**.

Аттестации или сертификации подвергаются учебные заведения. В отличие от России, где эта процедура упрощена и состоит в одновременном заполнении и проверке «красивости» заполнения огромного количества таблиц, за рубежом к аттестации часто подходят как к перманентному, длительному мероприятию. Лозунг здесь был бы таким: «Аттестация через всю жизнь!». Мониторинговыми центрами ведутся опросы, заполняются базы данных о выпускниках и пр. Внимание уделяется и газетным скандалам, и журналистским расследованиям об экзаменационных подлогах и нецелевых расходах средств в системе образования.

Специальные учреждения, где проводится аттестация учреждений, ЗУН специалистов, обычно называют **сертификационными центрами**.

Сертификаты могут иметь разные степени (рейтинги), в соответствии с удостоверяемым ими уровнем качества. Уровень сертификатов (дипломов, аттестатов и пр.) можно сравнить по ежегодно публикуемым рейтинговым таблицам.

Служба трудоустройства не имеет права отдать рабочее место специалисту с сертификатом низшего рейтинга, если в очереди есть специалист с сертификатом более высокого рейтинга. В случае нарушений прав специалиста, за него успешно вступаются судебные органы.

Аналогично, сертификат охраняет право образовательного учреждения на получение субсидий определенного рейтингом уровня.

Как правило, процедура аттестации включает обязательное бланковое либо компьютерное тестирование. Сертификационный центр ведет свою деятельность от имени

- государства;
- крупного образовательного учреждения;
- негосударственного учреждения.

В зависимости от своего статуса, сертификационный центр выдает успешно прошедшим аттестацию претендентам (испытуемым)

- государственные сертификаты;
- сертификаты государственного образца;
- удостоверения;
- иные документы.

Считается, что качество деятельности сертификационного центра будет выше, если он независим, т. е. изучаемые им структуры никак не влияют ни на оснащенность, ни на уровень вознаграждений сотрудников центра.

Патриотизм – последний довод подлеца.
(старая истина)

Трудно обеспечить независимость образовательного сертификационного (или мониторингового) центра в России. Сотрудники центра – такие же люди, как все окружающие, с такими же мало защищенными и обеспеченными правами. У них есть дети, которым нужно поступать в вуз или переводиться из школы в школу. Следовательно, не смотря на всеобщее стремление к принципиальности, здесь есть все предпосылки к традиционному российскому бартеру «ты мне – я тебе». В настоящем пособии приведены перечни способов, позволяющих превзойти эту российскую особенность.

*Не врет только сумасшедший
А. МИХАЛКОВ-КОНЧАЛОВСКИЙ*

08.04.К развитию технологий тестирования

Тезис о том, что поспешность, мол, необходима при ловле блох, поддерживают и развивают именно те, кому и блоху поймать лень.

Виталий КОРОТИЧ

Существующие теории тестовой оценки качества и количества пока далеки до завершения. Тем не менее, построенные на их основе технологии получили широкое распространение в мире и России.

Наблюдающееся в России бурное внедрение существующих (часто критикуемых профессионалами) технологий тестирования и сдержанное отношение к развитию теории объясняется, на наш взгляд;

- традиционным здравым недоверием педагогов к любой автоматизации учебного процесса;
- прагматичностью целей, ставящихся руководителями образовательных учреждений на первый план.

Прагматичный подход к оценке ЗУН, пришедший в Россию вместе с реформами последних лет на смену бытовавшему совсем недавно бережному отношению, снизил рейтинг российского образования в мире. Об этом говорят известные результаты международных исследований.

Качество среднего выпускника российских образовательных учреждений теперь ниже, чем два-три десятилетия назад. Среди работодателей укрепилось мнение, что на вкладыш диплома специалиста можно теперь не обращать внимания. Работодатели стали сами экзаменовать, по своему разумению, принимаемых на работу специалистов. Ответственность образовательного учреждения перед государством за качество образования постепенно подменяется личной ответственностью работодателя за правильный выбор работника.

Но не все потеряно. Ответственное, государственное отношение к развитию теории оценки и внедрению новых компьютерных технологий аттестации ЗУН помогут вернуть российской системе образования былую славу.

Человечество, смеясь, расстаётся со своим прошлым.

К. МАРКС

Вряд ли возможно добиться заметного повышения объективности измерений в рамках существующих технологий тестирования. В этом убеждают многие профессиональные исследования, проведенные самими создателями технологий и их коллегами. Будущие успехи в деле объективизации связывают с созданием:

- новых подходов к конструированию тестовых заданий и тестов;
- новых форм представления тестовых заданий;
- новых технологий оценивания качества и количества ЗУН.

Среди перспективных направлений развития форм тестовых заданий можно назвать т.н. грейдинг (от англ. – grading – сортировка, классификация) – методику, в которой оценка выполнения тестовых заданий является дифференцированной (ступенчатой, множественной). При применении методики grading, спектр оценки шире, чем двузначный (верно – неверно).

Применяется grading чаще при тестировании знаний по математике. Признанного русскоязычного аналога пока нет (первое известное нам упоминание об этой технологии в русскоязычной литературе принадлежит В. С. Аванесову).

... тестовое задание часто представляют в вопросной форме, в то время как к этим структурам мышления предъявляются разные требования. Известно, что вопрос не истинен и не ложен, в то время как тестовые утверждения либо истинны, либо ложны.

Васильев В. И., Тягунова Т. Н.

Другим вариантом развития технологии тестирования является переход от наиболее распространенной, уже привычной и понятной линейной схемы, использующей задания закрытой формы, к другой – схеме «мягкого тестирования» (перевести этот термин можно как «Soft testing» или «Soft grading»).

Технология «мягкого тестирования» предложена и реализована в ДВГУ, реализованные варианты ее многократно представлялись и обсуждались на научных конференциях всероссийского уровня. Подробно технология «мягкого тестирования» описана ниже, здесь же мы кратко остановимся на основных ее отличиях от других технологий.

Сам термин «мягкое тестирование» означает, что из технологии устранены чересчур резкие границы оценки вариантов выполнения тестовых заданий. Вместо двухступенчатой шкалы «верно – неверно», здесь используется четырехступенчатая: «верно – неполно – неточно – неверно». Такая шкала более соответствует человеческой логике, чем двузначная.

Здесь можно использовать уже имеющиеся базы заданий закрытой и открытой формы. Однако, часто они требуют «косметических» корректив. Коррективы внести несложно, проведя в каждом задании сортировку неверных вариантов ответов на неверные, неточные и неполные. Ниже приведены примеры, как это можно делать для разных учебных дисциплин.

Изменена здесь и форма представления тестовых заданий. Испытуемому предъявляется выборка тестовых заданий закрытой формы, причем, выполняя задание, он видит на экране только один вариант ответа на вопрос. Именно этот, единственный, вариант и должен оценить испытуемый. Остальные варианты скрыты, и могут открываться испытуемым по очереди, не давая возможности воспользоваться простым логическим сравнением. Такой подход позволяет значительно повысить достоверность измерения.

Процедура предъявления заданий испытуемому оформлена здесь в виде деловой ролевой игры. Испытуемый играет роль учителя, проводящего устный опрос в виртуальном классе. Интерфейс оформлен соответственно – ученики, доска, парты. Учеников можно переспрашивать и исправлять уже выставленные оценки.

Описанный вариант сценария реализован коллективом Центра новых образовательных технологий в 1995-1996 гг. (ЦНОТ ДВГУ) в электронном тренажере «STeacher». Этот тренажер был создан для работы в среде DOS. Он долгое время эксплуатировался в ДВГУ и школах Приморья, но, с развитием компьютерной техники, потребовалась его обновление для работы в Windows.

Модернизированный тренажер, реализующий технологию «мягкого тестирования», созданный коллективом Тихоокеанского института дистанционного обучения и технологий (ТИДОТ ДВГУ) и применяемый для тестирования знаний студентов и школьников в Приморье в настоящее время, получил название «Дидактор».

В 2004 году создана версия тренажера «Дидактор-6», где реализованы мультимедийные возможности, необходимые, например, при изучении языков. «Дидактор-6» позволяет вести тестирование с использованием открытых тестовых заданий и в нетрадиционной форме, например, как диктант.

Адаптивное тестирование. Три варианта и семь возможностей

Ecce spectaculum dignum, ad quod respiciat intentus operi suo deus
- Вот зрелище, достойное того,
чтобы на него оглянулся Бог, созерцая свое творение)
СЕНЕКА Луций Аней, fraza.net.ua

Адаптивное обучение и адаптивный тестовый контроль, в своем единстве, представляют собой современный компьютерный вариант реализации принципа индивидуализации обучения. Этот принцип не мог быть реализован эффективно в условиях традиционных классно-урочных форм занятий. Начало адаптивного обучения можно отнести ко времени возникновения педагогических трудов Коменского, Песталоцци и Дистервега.

Адаптивное обучение:

- помогает отойти от классно – урочной формы обучения, открыть возможности свободного выбора траектории учения;
- потребует определения перечня учебных задач и тестовых заданий для проверки подготовленности не только по минимальным, но и по более высоким требованиям;
- основано на использовании оценок, получаемых при входном адаптивном тестовом контроле. По итогам контроля выбирается первый учебный материал, с которого и начинается адаптивное обучение. Следующий за этим контроль выявляет меру усвоения предложенного материала, и т.д.;
- позволяет обеспечить выдачу учебного материала на оптимальном уровне трудности. Известно, что легкие материалы не обладают заметным развивающим потенциалом, а трудные задания снижают учебную мотивацию у большинства учащихся;
- обеспечивает знания без привычных, для массового образования, многочисленных пробелов в индивидуальной подготовленности учащихся;
- достигает требуемой структуры и желаемого уровня знаний;
- позволяет регулировать трудоемкость и количество предъявляемых заданий в зависимости от ответа учащегося на текущее задание. В случае правильного ответа следующее задание он получит труднее, в случае неправильного – легче.

Целесообразность адаптивного контроля вытекает из соображений рационализации традиционного процесса тестирования, где всем претендентам дается одинаковый набор заданий. Знающему претенденту нет необходимости давать легкие задания, из-за высокой вероятности их правильного решения. Из-за высокой вероятности неправильного решения нет смысла давать трудные задания слабому претенденту. Использование заданий, соответствующих уровню подготовленности, существенно повышает точность измерений и минимизирует время индивидуального тестирования.

Три варианта адаптивного тестирования:

1. Пирамидальное тестирование – всем претендентам дается задание средней трудности и уже затем, в зависимости от ответа, каждому дается задание легче или труднее. На каждом шаге применяется правило деления шкалы трудности заданий пополам. Например, если в тесте имеется двадцать одно упорядоченное по трудности задание, тестирование начинается с одиннадцатого. Если претендент действовал правильно, ему предъявляется шестнадцатое задание и, в случае успеха, девятнадцатое. Аналогично, слабо подготовленный претендент после неудачи на одиннадцатом задании пробует

силы на шестом, и далее, по тому же принципу, до момента стабилизации на близких по трудности заданиях;

2. Flexilevel – когда контроль начинается с любого уровня трудности, а потом происходит постепенное приближение к реальному уровню подготовленности;
3. Stradaptive – когда тестирование проводится с БТЗ, где задания разделены по уровням трудности. После верного выполнения, следующее задание берется из более высокого уровня трудности, после неверного – наоборот. Это требует предварительного определения трудности всех заданий.

В адаптивном тесте заранее известны параметры трудности и дифференцирующей способности каждого ТЗ. Он создаётся в виде БТЗ, где задания упорядочены в соответствии с их характеристиками. Главная характеристика заданий адаптивного теста – это их сложность, полученная эмпирически. Это означает: каждое задание адаптивного теста прошло эмпирическую апробацию на заданной выборке, которая соответствует предполагаемым характеристикам претендента.

Адаптивный контроль позволяет рационализировать традиционное тестирование. В рамках адаптивного контроля эффективно выбираются и используются те задания БТЗ, которые соответствуют уровню подготовленности претендента. Это позволяет существенно повысить точность измерений и минимизировать время сеансов до, примерно, 5-10 минут.

Литература:

1. Weiss D. J.(Ed.) New Horizons in Testing: Latent Trait. Test Theory and Computerised Adaptive Testing. N.Y., Academic Press, 1983. – 345 pp.
2. Lord P. M. Application of Item Response Theory to Practical Testing Problems. Hillsdale N – J. Lawrence Erlbaum Ass., Publ. 1980, – 266 pp.
3. Бобков А. И., Далматов С. Б., Преснякова Г. В., Шашин Г. В. Принципы построения адаптивных аналоговых систем обучения и контроля знаний. Учебное пособие. – Л.: Лен. инст. авиац. приборостроения, 1987. – 80 с.
4. Аванесов В. С Математические модели педагогического измерения. – М.: Исследовательский центр проблем качества подготовки специалистов, 1994. – 26 с.
5. Аванесов В. С. Научные проблемы тестового контроля знаний. Моногр. – М.: Исследовательский центр проблем качества подготовки специалистов, 1994. – 135 с.

Текст настоящего параграфа основан на материале четвертой части книги: Аванесов В. С. Композиция тестовых заданий: Учебная книга. 3 изд., доп. М.: ЦТ, 2002. – 240 с.

*Если учитель имеет только любовь к делу, он будет хороший учитель.
Если учитель имеет только любовь к ученику, как отец, мать,
он будет лучше того учителя,
который прочел все книги, но не имеет любви ни к делу, ни к ученикам.
Если учитель соединяет в себе любовь к делу и к ученикам,
он - совершенный учитель.
Лев Николаевич ТОЛСТОЙ, fraza/net/ua*

Необычные формы тестирования

О, сколько нам открытий чудных...
А. С. ПУШКИН

Ниже помещен, с сокращениями, поучительный текст, попавшийся нам в новостях Интернет. Этот текст полон рекламы, поэтому мы полагаем, что ни автор, ни компания Google ни администрация CNET News.com на нас не обидятся за эту перепечатку. Для тех, кто не знает: Кремниевая долина – район США с наивысшей концентрацией программистов на единицу площади.

Google привлекает яйцеголовых при помощи таинственного рекламного щита

Стефани Олсен (Stefanie Olsen), CNET News.com, 12 июля, 2004, 11:14

На прошлой неделе в самом сердце Кремниевой долины (Маунтин-Вью, штат Калифорния) на выезде Ралстон, ведущем в Санта-Клару, появился рекламный щит с математической головоломкой. Чтобы решить ее, большинству водителей, проезжавших по хайвею No 101, пришлось прибегнуть к помощи Google.

Поисковый гигант, стоявший за анонимной вызывающе пустой рекламой с короткой черной надписью, хотел лишь привлечь внимание людей с математическим складом ума и предложить им работу. Этот рекрутский ход определенно замешан на известном афоризме математика Пола Ардоса: «Математик – это устройство для переработки кофе в теоремы».

Надпись на рекламном щите гласит: {first 10-digit prime found in consecutive digits e}.com. Ответ, 7427466391.com, приводит математических гениев на веб-страницу, где без всякого намека на то, что за этой игрой стоит Google, предлагается решить еще одно уравнение.

Победа над этим уравнением ведет на страницу Google Labs, отделения исследований и разработок компании, где говорится: «Одна из вещей, которые мы усвоили, создавая Google: то, что ищешь, легче найти, если оно само ищет тебя. Мы ищем лучших в мире инженеров. И вот вы здесь. К нам каждый день поступает множество резюме, и мы придумали этот нехитрый процесс, чтобы улучшить отношение сигнал/шум».

Google методично подходит к найму людей. Компания, в которой по состоянию на 31 марта работало 1907 сотрудников, с первого дня существования очень серьезно относится к своим методам. Один из первых сотрудников фирмы, работающий в ней еще с тех времен, когда та размещалась в гараже в Пало-Альто, рассказывает, что во время его интервью с соучредителями Google Сергеем Брином и Ларри Пейджем на соседнем столе лежало не меньше восьми руководств по найму людей. Рекрутеры компании тоже говорят, что прежде чем принять кого-то, они проводят множество интервью, иногда с привлечением тестов по математике или бизнес-стратегии.

Еще один остроумный способ рекрутинга Google заключается в том, что для выявления потенциальных работников компания размещает спонсируемые ссылки на страницах результатов поиска. Например, одно время такая ссылка была связана с ключевым словом/именем Udi Manber – это шеф A9, нового отделения поисковой технологии Amazon.com.

Этот пример поучителен, его нетрудно адаптировать в повседневной образовательной практике. Объявления «загадки» могут помещаться на школьном сервере, на сайте школы в Интернет, на стенах школьных коридоров.

СПИСОК ЛИТЕРАТУРЫ

1. Аванесов В. С. Композиция тестовых заданий. Учебная книга. 3 изд., доп. – М.: Центр тестирования, 2002 г. – 240 с.
2. Башмаков А. И., Башмаков И. А. Разработка компьютерных учебников и обучающих систем. – М.: Информационно-издательский дом «Филинъ», 2003. – 616 с.
3. Васильев В. И., Тягунова Т. Н. Теория и практика формирования программно-дидактических тестов. – М.: Издательство МЭСИ, 2001. – 130 с.
4. Вопросы тестирования в образовании / Под ред. Хлебникова В. А., Неймана Ю.В. М.: «Век книги», 2001. – 115 с.
5. Гузеев В. В. Оценка, рейтинг, тест. Школьные технологии. 1998, №3. – С. 3-40.
6. Дзюбенко А. А. Новые информационные технологии в образовании. М.: 2000. – 104с.
7. Дистанционное обучение: Учеб. пособие / Под ред. Е . С. Полат. – М.: Владос, 1998. – 192 с.
8. Кречетников К. Г. Задания в тестовой форме и методика их разработки: Учебно – методическое пособие. – Владивосток: Изд-во Дальневост. ун-та, 2002. – 36 с.
9. Кречетников К. Г. Теоретические основы создания креативной обучающей среды на базе информационных технологий для подготовки офицеров флота.: Моногр. – Владивосток: Изд-во Дальневост. ун-та, 2001. – 360 с.
10. Макарова Т. Д. Тестирование в системе мониторинга качества образования // Стандарты и мониторинг в образовании. 1998, №1. – С. 60-61.
11. Машбиц Е. И. Психолого-педагогические проблемы компьютеризации обучения: (Педагогическая наука – реформе школы).– М.: Педагогика, 1988.– 192 с.
12. Михайлычев Е. А. Дидактическая тестология. – М.: Народное образование, 2001. – 432 с.
13. Нардюжев В. И., Нардюжев И. В. Модели и алгоритмы информационно-вычислительной системы компьютерного тестирования. – М.: Прометей, 2000. – 148 с.
14. Нейман Ю. М., Хлебников В. А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000. – 168 с.
15. Нестеров А.В., Тимченко В.В., Трапицын С.Ю. Информационные педагогические технологии. Учебно-методическое пособие. – СПб.: Издательство ООО "Книжный дом", 2003. – 340 с.
16. Осин А.В., www.rnmc.ru/ideas/ideas.html
17. Основы открытого образования. Т. 1 / Ответ. редактор В.И. Солдаткин. – РГ ИОО. – М.: 2002. – 676 с.
18. Основы открытого образования. Т. 2 / Ответ. редактор В.И. Солдаткин. – РГ ИОО. – М.: 2002. – 680 с.
19. Роберт И.В. Современные информационные технологии в образовании: дидактические проблемы; перспективы использования. – М.: "Школа–Пресс", 1994. – 205 с.
20. Софронова Н.В. Программно-методические средства в учебном процессе общеобразовательной школы. – М.: ИИО РАО, 1998. – 178 с.
21. Тесты для старшеклассников и абитуриентов. Оценка знаний / Под ред. Шмелева А. Г. – М.: «Первое сентября», 2000. – 132 с.
22. Хуторской А. В. Современная дидактика: Учебник для вузов. – СПб: Питер, 2001. – 544 с.
23. Шеншев Л. В. Основы адаптивного обучения языку (семиотические аспекты развития речи с помощью автомата). – М.: Наука, 1995. – 113 с.
24. Шмелев А. Г., Похилько В. И. Анализ пунктов при конструировании и применении тест-опросников: ручные и компьютерные алгоритмы. – Вопросы психологии, № 4, 1985. – С. 126-134.

Учебное издание

Морев Игорь Авенирович

**ОБРАЗОВАТЕЛЬНЫЕ
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
Часть 2. Педагогические измерения**

В авторской редакции

Технический редактор: М. И. Морева

Компьютерная верстка: И. А. Кугуенко

Подписано в печать 14.09.2004.

Формат 60х84 1/8

Усл. печ. л. 20,00 . Уч.-изд. л. 18,50 Тираж 500 экз.

Издательство Дальневосточного университета
690950, г. Владивосток, ул. Октябрьская, 27

Отпечатано на множительной технике ТИДОТ ДВГУ
690950, г. Владивосток, ул. Октябрьская, 27