

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Пензенский государственный педагогический университет
имени В. Г. Белинского

Кафедра английского языка

**УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС
ПО ДИСЦИПЛИНЕ**

СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

0503030 Иностранный язык

2007

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
Пензенский государственный педагогический университет
имени В. Г. Белинского

ПРИНЯТО
на заседании Ученого совета
факультета иностранных языков

Протокол заседания совета факультета
№ 8 от «2» мая 2007 г.

УТВЕРЖДАЮ
Проректор по учебной работе


подпись Ф.И.О.

Декан
факультета 
подпись Ф.И.О.

ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

специальность 050303 Иностранный язык

Факультет иностранных языков

Кафедра английского языка

Пенза – 2007



Квалификационные требования

Выпускник, получивший квалификацию учителя иностранного языка, должен быть готовым осуществлять обучение и воспитание обучающихся с учетом специфики преподаваемого предмета; способствовать социализации, формированию общей культуры личности, осознанному выбору и последующему освоению профессиональных образовательных программ; использовать разнообразные приемы, методы и средства обучения; обеспечивать уровень подготовки обучающихся, соответствующий требованиям Государственного образовательного стандарта; осознавать необходимость соблюдения прав и свобод учащихся, предусмотренных Законом Российской Федерации "Об образовании", Конвенцией о правах ребенка, систематически повышать свою профессиональную квалификацию, участвовать в деятельности методических объединений и в других формах методической работы, осуществлять связь с родителями (лицами, их заменяющими), выполнять правила и нормы охраны труда, техники безопасности и противопожарной защиты, обеспечивать охрану жизни и здоровья учащихся в образовательном процессе.

Требования ГОС по дисциплине «Современные средства оценивания результатов обучения».

Понятие о качестве образования. Оценка как элемент управления качеством. Традиционные и современные средства оценки (рейтинг; мониторинг; накопительная оценка («портфолио»). Виды, формы и организация контроля качества обучения. Оценка, её функции.

Этапы развития тестирования в России и за рубежом. Психолого-педагогические аспекты тестирования. Понятие теста. Виды тестов. Формы тестовых заданий. Компьютерное тестирование и обработка результатов. Интерпретация результатов тестирования.

Единый государственный экзамен, его содержание и организационно-технологическое обеспечение. Контрольно-измерительные материалы.

Цель дисциплины – познакомить студентов с методологическими и теоретическими основами тестового контроля, современными средствами оценки результатов обучения, методическим обеспечением дисциплины, а также порядком организации и проведения единого государственного экзамена (ЕГЭ).

При этом реализуются следующие **задачи**:

- изучение методов конструирования и использования педагогических тестов;
- рассмотрение методов шкалирования и интерпретации полученных результатов;
- знакомство с компьютерными технологиями, используемыми в тестировании;
- определение психологических и педагогических аспектов использования тестов для контроля знаний учащихся;
- развитие умения составлять тесты и оценивать результаты тестовых заданий по различным аспектам языка.

Место дисциплины в профессиональной подготовке студентов

Дисциплина «Современные средства оценивания результатов обучения» является системообразующим компонентом в процессе подготовки учителя иностранного языка в ВУЗе. Изучение данной дисциплины обеспечивает профессиональную подготовку молодых специалистов – будущих учителей иностранного языка. Данному курсу отводится 70 часов в 7 семестре. Таким образом, студенты готовятся к самостоятельной творческой работе учителя в период прохождения педагогической практики.

В ходе изучения дисциплины студенты приобретают необходимые знания, умения и навыки для оценивания результатов обучения; и так называемую «диагностическую компетенцию учителя иностранного языка», которая включает готовность проверять понимание материала учащимися, применять стандартизированные тесты и интерпретировать результаты, проверять самостоятельную работу, отслеживать систематически допускаемые ошибки и другие виды деятельности учителя иностранного языка.

Данная дисциплина входит в цикл общепрофессиональных дисциплин. Специальность «Иностранный язык» 050303.

Распределение времени, отведённого на изучение дисциплины по учебному плану:

Форма учебной работы	Форма обучения				
	очная	з/о 3,5 года	з/о 6 лет		ДВО
	VII	V	VI	VII	V
Общая трудоёмкость	70	72	72		72
Аудиторные занятия АЗ	34	8	8		10
Лекции Л	17	4	8		6
Практические занятия ПЗ					
Семинары					
Лабораторные занятия ЛЗ	17	4			4
Другие виды АЗ					
Самостоятельная работа	36	64	64		62
Контрольная работа					
Компьютер.тестирование					
Форма итогового контроля (зачёт)	+	+		+	+

Тематические планы для разных форм обучения

№	Наименование разделов и тем	Форма обучения											
		очная			з/о 3,5 г.			з/о 6 лет			ДВО		
		Л	ЛЗ	СР	Л	ЛЗ	СР	Л	ЛЗ	СР	Л	ЛЗ	СР
1	Понятие о качестве образования.	2	2	4			8	1		8			8
2	История развития системы тестирования в России и за рубежом.	2	2	4	1		8	1		8	1		8
3	Психологические и педагогические тесты.	2	2	4	1	1	8	1		8	1	1	8
4	Виды тестов и формы тестовых заданий.	2	2	4	1		8	1		8	1	1	8
5	Оценка качества тестовых заданий.	2	2	4			8	1		8	1		8
6	Интерпретация результатов тестирования	2	2	4		1	8	1		8		1	7
7	ЕГЭ. Организационно-технологическое обеспечение ЕГЭ.	2	2	4	1	1	8	1		8	1	1	8
8	Содержание и структура тестовых заданий по разным аспектам языка	3	3	8		1	8	1		8	1		7
Всего часов		17	17	36	4	4	64	8		64	6	4	62

Содержание дисциплины

1. Качество образования. Система контроля и управления качеством образования. Показатели качества образования. Оценка как элемент управления качеством. Традиционные и новые средства оценки результатов обучения. Виды контроля. Формы и организации контроля. Оценка и её функции. Связь оценки и самооценки.
2. Возникновение тестирования. Родоначальники тестового движения. Развитие тестологии в Америке и странах Европы. Развитие тестирования в России. Современные центры тестирования.

3. Деление тестов на педагогические и психологические. Педагогические тесты Э. Торндайка.
4. Классификация тестов. Зависимость видов и форм тестов от специфики учебной дисциплины. Виды педагогических тестов. Диагностическое тестирование. Тематические тесты, рубежные тесты, итоговая аттестация. Тестовые задания открытой и закрытой формы. Структура тестового задания. Принципы отбора содержания тестов. Определение целей тестирования. Проверка и обработка результатов.
5. Оценка качества тестовых заданий. Шкалирование результатов тестирования. Представление и интерпретация результатов тестирования.
6. Пакеты прикладных программ обработки и конструирования тестов. «Портфолио» как средство оценки результатов обучения. Мониторинг как средство оценки результатов обучения.
7. ЕГЭ и качество образования. ЕГЭ как одно из средств повышения качества общего и педагогического образования. Задачи ЕГЭ: расширение доступности высшего образования, снижение психологической нагрузки на выпускников общеобразовательных учреждений, унификация требований к подготовке поступающих в вузы. Организационные основы ЕГЭ. Структура заданий ЕГЭ.
8. Составление типовых тестовых заданий по разным аспектам языка. Обобщённые способы выполнения тестовых заданий. Виды контроля.

Требования к уровню освоения программы

В результате изучения дисциплины **студент должен:**

знать:

- историю и современное состояние системы тестирования в России и за рубежом;
- традиционные и современные подходы к оценке учебных достижений;
- особенности тестовых технологий, типы и виды тестовых заданий, оценку их качества;
- методы оценивания результатов тестирования;
- документы, регламентирующие проведение ЕГЭ;
- процедуру проведения ЕГЭ;

уметь:

- давать оценку тестовым заданиям;
- использовать на практике разные виды тестов;
- проводить тестирование и анализировать полученные результаты;

владеть:

- методами применения методического обеспечения дисциплины;
- навыками работы с компьютерными пакетами программ по обработке результатов тестирования.

Примерный перечень вопросов для зачёта

1. Расскажите о понятии “качество образования”. Охарактеризуйте оценку как элемент управления качеством.
2. Сравните традиционные и новые средства оценки результатов обучения. Опишите их достоинства и недостатки.
3. Как оценивается качество российского образования отечественными и зарубежными экспертами?
4. Расскажите историю возникновения тестирования в России.
5. Назовите современные центры тестирования.
6. Назовите функции контроля в современном учебном процессе.
7. Назовите традиционные формы контроля. Укажите их достоинства и недостатки.
8. Какие современные средства контроля выделяются в учебном процессе? В чем их преимущество по сравнению с традиционными формами контроля?
9. Укажите место психологических и педагогических измерений в современном образовании.
10. Охарактеризуйте цели и задачи педагогического и психологического тестирования.
11. Сходство и различие педагогических и психологических тестов в учебном процессе.
12. Назовите психологические тесты, применяемые в учебном процессе.
13. Расскажите о видах тестов.
14. Охарактеризуйте основные положения классической теории тестов.
15. Опишите виды педагогического контроля (текущий, тематический, рубежный, итоговый контроль).
16. Дайте классификацию тестов по разным основаниям.
17. Понятие гомогенных и гетерогенных тестов.
18. Расскажите о применении компьютерного тестирования.
19. Назовите основные виды заданий в тестовой форме.
20. Как определяются цели тестов?
21. Расскажите о структуре тестового задания.
22. На какие принципы необходимо опираться при отборе содержания тестового задания?
23. Как производится экспертиза качества содержания теста?
24. Расскажите об основных подходах к качеству знаний.
25. Охарактеризуйте мониторинг как средство оценки результатов обучения.
26. Назовите и охарактеризуйте основные модели педагогических измерений.
27. Расскажите о видах оценочных шкал. Дайте характеристику одному из пакетов прикладных программ по обработке или конструированию тестов.

28. Опишите достоинства и недостатки “портфолио” как средства оценки результатов обучения.
29. Расскажите о задачах ЕГЭ. Каковы преимущества и недостатки ЕГЭ перед другими формами контроля?
30. Опишите организацию проведения ЕГЭ.
31. Опишите структуру заданий ЕГЭ.
32. Расскажите о порядке создания контрольно-измерительных материалов (КИМ-ов) для ЕГЭ.
33. Каков порядок проверки тестовых заданий ЕГЭ?

Литература

Основная

1. Аванесов В.С. Композиция тестовых заданий. М., 2002.
2. Анастаси А., Урбина С. Психологическое тестирование. СПб., 2002.
3. ЕГЭ. Сборник нормативных документов. М., 2002.
4. Майоров А.Н. Теория и практика создания тестов для системы образования. М., 2000.
5. Мельникова М.Б. Теория и практика конструирования педагогических тестов. М., 2002.
6. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. М., 2002.
7. Беспалько В.П. Критерии оценки знаний учащихся и пути оптимизации процесса обучения// Теория поэтапного формирования умственных действий и управление процессом обучения. М., 1960.
8. Поддубная Л. М., Татур А. О., Чельшкова М. Б. Задания в тестовой форме для автоматизированного контроля знаний студентов. М., 1995.
9. Чельшкова М. Б., Савельев Б. А. Методические рекомендации по разработке педагогических тестов для комплексной оценки подготовленности студентов в вузе. М., 1995.

Дополнительная

1. Амонашвили Ш.А. Воспитательная и образовательная функции оценки умения. М., 1984.
2. Ананьев В.Г. Психология педагогической оценки// Избранные педагогические труды. М., 1982.
3. Боголюбов Л. Н., Дик Ю.И., Иванова Е.О. и др. О подходах к разработке требований к обязательному уровню подготовки выпускников основной школы // Перспективы развития общего среднего образования: Сб. науч. тр. М., 1998.

4. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии/ Пер. с англ. Л.И. Хайрусовой. М., 1976.
5. Клайн П. Введение в психометрическое программирование: Справочное руководство по конструированию тестов. Киев, 1994.
6. Кларин М. В. Инновационные модели обучения в зарубежных педагогических поисках. М., 1994.
7. Концепция оценки достижения учащимися требований общеобразовательного стандарта / Колл. авторов под рук. В. С. Леднева. М., 1993.
8. Нардюжев В.И., Нардюжев И.В. Модели и алгоритмы информационно-вычислительной системы компьютерного тестирования. М., 2000.
9. Родионов Б.У., Татур А.О. Стандарты и тесты в образовании. М., 1995.
10. Селевко Г.К. Современные образовательные технологии. Уч. пособие. М., 1998.
11. Симонов В. П. Педагогический менеджмент: Учеб. пособие. М., 1997.
12. Селезнева Н. А., Байденко В. И. Проблема качества образования: актуальные аспекты пути решения // Проблемы качества, его нормирования и стандартов в образовании: Сб. науч. статей. М., 1998.
13. Чельшкова М. Б., Ковалева Г. С., Татур А. О., Хлебников В.А. Концепция аттестационного тестирования выпускников общеобразовательных учреждений на этапе перехода от школы к вузу // Проблемы качества, его нормирования и стандартов в образовании: Сб. науч. ст. М., 1998.
14. Чельшкова М. Б., Хлебников В. А. Основные подходы к оценке качества подготовки обучаемых // Проблемы качества, его нормирования и стандартов в образовании: Сб. науч. ст. М., 1999.

Лекция 1. Понятие о качестве образования. Оценка как элемент управления качеством.

1. Показатели качества образования. Оценка эффективности и качества образования.
 2. Традиционные и новые средства оценки результатов обучения.
 3. Виды контроля (входной, текущий и итоговый). Формы и организация контроля.
- Оценка, ее функции.

1. Показатели качества образования. Оценка эффективности и качества образования.

Как известно, отношение общества к системе образования неоднозначно: периоды удовлетворенности сменяются периодами критики, когда поднимаются вопросы качества преподавания, содержания образования, уровня профессионального мастерства имеющих специалистов и качества их подготовки (1), (5). На сегодняшний день имеются существенные расхождения в понимании и интерпретации качества подготовки специалистов, что можно продемонстрировать на примере оценок социума, начинающих учителей, выпускников и т.д. Оценкой качества образования могут быть: академические успехи студентов, результаты государственных экзаменов, достижения учеников и т.д. В результате мы получаем разные оценки одного и того же феномена. Тем не менее, при оценке качества образования необходимо учесть, по крайней мере, три точки зрения. Точку зрения личности, точку зрения государства, точку зрения социума. Данные таблиц говорят о том, что они не всегда совпадают (6), (7). К примеру из таблицы 3 следует, что в результате итоговой аттестации (а это точка зрения государства) подготовка в психолого-педагогической области оценивается средним баллом 4,27. Результат неплохой, даже выше, чем по специальным дисциплинам. А вот, согласно данным таблицы 1, отражающей оценку качества образования с точки зрения социума, около 80% родителей считают реальную подготовку учителей в этой области недостаточной. Результаты разные, поэтому необходим некоторый общий взгляд на проблему оценки качества образования.

Качество – это сложная философская, системная, экономическая и социальная категория, раскрываемая через систему определений, отражающих единство системно-структурного и ценностно-прагматического аспектов (9).

Качество – это философская категория, выражающая существенную определенность объекта, благодаря которой он является именно этим, а не иным, это характеристика объектов, обнаруживающаяся в совокупности их свойств (2).

Качество знаний выявляется в качестве многоаспектного анализа усвоения и применения знаний человеком в различных видах деятельности. Основными параметрами

качества знаний в системе являются: полнота, глубина, оперативность, гибкость, осознанность (8).

Качество образования – социальная категория, определяющая состояние и результативность процесса образования в обществе, его соответствие потребностям и ожиданиям общества (различных социальных групп) в развитии и формировании гражданских, бытовых и профессиональных компетенций личности (9).

Под **качеством образования** понимается "степень удовлетворения ожиданий различных участников образовательного процесса от предоставляемых образовательным учреждением образовательных услуг" или "степень достижения поставленных в образовании целей и задач" (9).

Из последних приведенных определений вытекает аналогия качества образования с рыночным понятием цены, если рассматривать образование как товар или услугу.

Мы будем придерживаться следующего взгляда на понятие качества образования.

Качество образования определяется соответствием принятой доктрине образования, социальным нормам, требованиям общества, государства и личности.

В этом смысле существуют три направления требований, предъявляемых к качеству образования: требования личности, требования общества, требования государства.

При этом требования личности к качеству образования, в соответствии с вышесказанным, определяются отсутствием существенных диссонансов между целями и результатами образования, соответствующим состоянием внутренней комфортности. Учитывая вышеприведенные оценки качества первичной подготовки, данной молодыми учителями, можно считать, что с точки зрения личности качество высшего педагогического образования достаточное.

Требования государства трансформируются в освоение на необходимом уровне соответствующих образовательных и профессиональных программ, проведение необходимых квалификационных процедур (в виде сдачи государственных экзаменов или защиты дипломного проекта и т.д.) и получение соответствующего государственного сертификата (диплома). При этом в данную процедуру явно или неявно обязательно вмешивается посредник, которым выступает соответствующее государственное (или негосударственное) образовательное учреждение, непосредственно осуществляющее образовательный процесс. Его интерпретация государственных требований находит свое отражение в итогах текущего и промежуточного контролей в виде экзаменов, зачетов, семинаров, лабораторных, различных практик и т.д. Учитывая оценки, полученные выпускниками на государственных испытаниях, и их соответствие промежуточным результатам (таблица 3), можно сказать, что с точки зрения государства с качеством

образования все нормально и образовательное учреждение не слишком вольно интерпретирует требования государственных стандартов.

Требования общества трансформируются в оценку социума, в успешность вхождения индивида в профессию и самостоятельную жизнь. Из таблицы 1, следует, что оценка социумом качества подготовки (особенно в психолого-педагогической области), не очень высока и необходимы меры по его коррекции.

В силу выше принятого толкования понятия процесс формирования качества образования определяется следующими факторами (7):

- качеством образовательных целей;
- качеством образовательных стандартов и эталонов;
- качеством образовательных программ;
- качеством кадрового и научного потенциала;
- качеством абитуриентов на входе;
- качеством выпускников на выходе;
- качеством средств образовательного процесса;
- качеством образовательных технологий;
- эффективностью системы контроля достижений;
- наличием обратной связи по течению и результатам образовательного процесса;
- системой традиций;
- успешностью вхождения в социум;
- согласованностью потребностей личности общества и государства в уровне и

качестве образования.

В достаточно общем виде цели высшего образования на сегодняшний день можно было бы определить следующим образом:

1. Удовлетворение образовательных потребностей личности.
2. Удовлетворение потребностей общества.
3. Удовлетворение потребностей государства.
4. Создание условий для овладения личностью профессиональной деятельностью и

получения соответствующей квалификации.

Причем, профессиональная деятельность должна рассматриваться, по крайней мере, в двух аспектах: как средство самореализации и самоутверждения личности и как средство ее устойчивой социальной защиты и адаптации.

Для того, чтобы иметь возможность оценить качество образования, необходимо ввести понятие **нормы** или **эталона**.

Норма качества образования – выявленная и зафиксированная документально система требований к качеству образования, соответствующих потребностям личности, общества, государства (11).

Оценка качества образования есть мера качества (числовая или семантическая), выражающая собой соотнесенность измерений свойств (функций) с базой, которая фиксирует эталонный уровень, норму качества (11).

Следует отметить, что оценка качества может быть внутренней или внешней, где под внутренней оценкой мы понимаем оценку, данную самой личностью или учреждением, осуществляющим образовательный процесс. Последняя имеет смысл, по-видимому, только в качестве промежуточной оценки, необходимой для осуществления обратной связи и коррекции в проведении образовательного процесса. Под внешней – понимаем оценку, данную качеству образования государством через государственную аттестационную комиссию (ГАК), социумом, профессиональной средой и др.

Естественно, что определение нормы качества образования со стороны государства (в силу сложности и многоплановости проблемы) требует введения системы эталонов, которая нашла свое отражение в различных концепциях стандарта образования (9), (10), (11). Кроме того, она отражает точку зрения на качество образования только одной стороны - государства. Все оценки на экзаменах, зачетах и прочие, полученные в процессе учебы, можно рассматривать лишь как внутренние и промежуточные, которые, в лучшем случае, могут дать, хотя и очень полезную и важную (отражающую интерпретацию образовательным учреждением государственного стандарта образования), но все-таки косвенную и достаточно условную, предметно ориентированную оценку качества профессиональной подготовки студента и выпускника.

2. Традиционные и новые средства оценки результатов обучения.

Разработка общих принципов методологии оценивания до сих пор остается проблемой потому, что исключительно трудно осуществить последовательное сопоставление целей образования с достигаемыми результатами обучения. Цели выражены, как правило, в весьма общих и потому абстрактных категориях: «подготовить квалифицированного специалиста», «сформировать научное мировоззрение», «вооружить научной методологией» и т. п. Измерение же результатов обучения проводится на совсем ином уровне – более узком, более конкретном, более осязаемом. Получается, что конечные цели образования и результаты обучения, проверяемые в конкретной оценочной ситуации, формулируются на разных языках. Цели – на языке интегральных, общих категорий, а результаты – на языке конкретных знаний, умений, навыков, то есть языке действий. Для выработки эффективных и достаточно строгих критериев оценивания

необходимо стараться излагать цели и результаты обучения на одном и том же языке, в одних и тех же понятиях и терминах, что не всегда представляется возможным.

Результаты контроля учебно-познавательной деятельности учащихся выражаются в ее оценке. *Оценка* – это определение степени усвоенности знаний, умений и навыков.

Количественным выражением оценки является *отметка*.

До настоящего времени не только в педагогической практике, но и в дидактической и методической литературе термины «оценка» и «отметка» иногда понимаются как синонимы. Часто, когда имеется в виду отметка, говорят оценка (выставление оценок, система оценок и т.д.). Между тем «оценка» и «отметка» – понятия хотя и близкие, но далеко не идентичные.

Отметка – это условное выражение количественной оценки знаний, умений и навыков обучаемых в цифрах или Галлах.

Оцениваться результаты обучения могут не только отметкой, но и другими средствами. Например, это может быть словесное одобрение и неодобрение, вынесение благодарности, награждение грамотами и медалями и т.д.

Балльная система оценки знаний учащихся существовала в России еще до революции 1917 г. Оценка знаний осуществлялась по шестибалльной системе с баллами от нуля до пяти. Затем нулевая оценка была упразднена, система превратилась в пятибалльную.

Отношение к оцениванию результатов обучения в баллах в истории школы было различным. После 1917г. получила развитие идея обучения без отметок. Согласно концепции советской трудовой школы учебная деятельность должна строиться на интересе учащихся, ориентироваться на самостоятельность, инициативу, творческий характер учения. Непригодными были признаны прежние методы дисциплинирования учащихся с помощью отметок. В мае 1918 г. вышло Постановление Наркомпроса РСФСР «Об отмене отметок», в котором определялось следующее:

- применение балльной системы для оценки познания поведения учащихся отменяется во всех без исключения в случаях школьной практики;
- перевод из класса в класс и выдача свидетельств производятся на основании успехов учащихся по отзывам педагогического совета об исполнении учебной работы.

Запрещались любые виды экзаменов: вступительные, переводные и выпускные. Отменялась индивидуальная: проверка учащихся на уроке. Лишь в крайних случаях пускались фронтальные устные проверки и письменные работы зачетного характера. Рекомендовались периодические беседы с учащимися по пройденной теме, их выступления о прочитанных книгах, статьях. Приветствовались отчеты учеников о

самостоятельно выполненных работах по личному выбору. Вместо традиционной системы контроля основной формой стал самоконтроль, выявление достижений не отдельного учащегося, а школьного коллектива. Получили широкое распространение тестовые задания как одна из самых пригодных форм самоконтроля.

Обучение без отметок имело определенные положительные стороны, так как развивало у части обучаемых самостоятельность. Но в большинстве своем такое обучение приводило к снижению качества знаний, ухудшению дисциплины. Школьники перестали регулярно заниматься в школе и дома, посещать занятия. Поэтому сначала стихийно стали вводиться различные формы контроля, а в 1932 г. официально восстановлен принцип систематического учета знаний каждого ученика на основе индивидуального подхода.

В сентябре 1935 г. были введены пять словесных (вербальных) оценок: «отлично», «хорошо», «удовлетворительно», «плохо», «очень плохо», которые просуществовали до конца 1943 г. В январе 1944 г. было принято решение заменить применявшиеся словесные оценки на цифровую пятибалльную систему оценки успеваемости и поведения учащихся. В вышедшей вслед за этим постановлением Инструкции о применении цифровой пятибалльной системы оценки были сформулированы параметры, за что выставляется та или иная отметка.

Введение регулярного учета знаний оправдало себя, учебная подготовка и дисциплина учащихся повысились.

Сложившаяся система оценивания результатов учебной деятельности в виде баллов несмотря на недостатки до сих пор не нашла достойной замены. Хотя некоторые педагоги, ученые выступают против существующей системы оценки степени обученности. Их главные аргументы:

- обученность фактически оценивается по трехбалльной шкале (отметки «1» и «2» фиксируют незнание);
- недостоверность оценки обученности (отметки «5», «4», «3» ставятся и в гимназических классах, и в классах для одаренных детей, и в классах общеобразовательных школ, и в классах коррекционно-развивающего обучения, поэтому они не равнозначны);
- трехбалльная шкала недостаточна, чтобы оценить знания учащихся (поэтому преподаватели используют суррогатную шкалу: к баллам добавляют «плюсы» и «минусы»). Ученые предлагают новые оценочные шкалы: десятибалльные, двенадцатибалльные, стобалльные и др.

Наряду с традиционными методиками оценивания все шире используется тестовый контроль знаний. Происходит тщательная эмпирическая проверка и шлифовка

стандартных предметных тестов, и постепенно меняется психологическая привычка учителей и учащихся к самой процедуре тестирования.

Педагогическое тестирование как один из методов контроля усвоения обучаемыми знаний, умений и навыков обладает важными преимуществами перед традиционными методами контроля знаний. Вот некоторые из них:

1. Более высокая, чем в традиционных методах, объективность контроля. Обычно на оценку, получаемую студентами, влияют, помимо уровня его учебных достижений, многие другие факторы: личность преподавателя и самого студента, их взаимоотношения, строгость или, наоборот, либеральность преподавателя и т.п. В оценке, выставляемой на основе традиционных методов контроля, оказывается существенным субъективный компонент. Это подтверждается тем известным фактом, что за один и тот же ответ разные преподаватели могут поставить разные отметки. В тесте такое влияние субъективных факторов исключается. Если тест достаточно качественный, то получаемая оценка может рассматриваться как объективная.

2. Оценка, получаемая с помощью теста, более дифференцирована. В традиционных методах контроля пользуются четырехбалльной шкалой ("отлично", "хорошо", "удовлетворительно", "неудовлетворительно").

Результаты тестирования, благодаря особой организации тестов, могут быть представлены, если необходимо, в более дифференцированных шкалах, содержащих больше градаций оценки. При этом обеспечивается более высокая точность измерений учебных достижений.

3. Тестирование обладает заметно более высокой эффективностью, чем традиционные методы контроля. Тесты можно одновременно проводить на больших группах студентов. Обработка результатов для получения окончательных оценок проводится легче, быстрее, чем, скажем, проверка контрольных работ. Особенно велика экономия времени, когда в форме тестирования проводятся экзамены.

Конечно, у тестирования как метода контроля есть и свои ограничения. Легче всего с помощью тестов проверять овладение просто организованным учебным материалом. Проверка глубинного понимания предмета, овладения стилем мышления, свойственным изучаемой дисциплине, с помощью тестов затруднена, хотя в принципе возможна. Отсутствие непосредственного контакта со студентом, с одной стороны, делает контроль более объективным, но, с другой стороны, повышает вероятность влияния на результат других случайных факторов. Например, невозможно проконтролировать случайные ошибки студента, вызванные невниманием или неправильным пониманием задания.

Существуют довольно изощренные методы, позволяющие в значительной степени преодолеть эти недостатки.

Ограничения тестирования как метода контроля необходимо хорошо осознавать, чтобы правильно применять тесты. В ряде случаев наилучший эффект дает сочетание тестов с традиционными методами контроля. Достоинство тестов заключается еще в том, что можно оценить степень вероятности влияния на результат случайных факторов и, следовательно, знать, насколько точен полученный результат.

3. Виды контроля (входной, текущий и итоговый). Формы и организация контроля. Оценка, ее функции.

Контроль бывает разных видов, форм и может составляться с помощью разнообразных методов. В педагогической практике применяется несколько *видов контроля* предварительный, текущий, периодический, тематический, итоговый и отсроченный. **Предварительный контроль**, как правило, имеет диагностические задачи. Он проводится с целью выявления имеющихся знаний, умений и навыков учащихся (к началу) обучения. Применяется обычно в начале учебного года или перед изучением новой темы. Предварительный контроль позволяет обучающему выбрать наиболее эффективные методы и формы работы.

Текущий контроль осуществляется по ходу обучения и дает возможность определить степень сформированности знаний, умений, навыков, а также их глубину и прочность. Этот контроль позволяет своевременно выявить пробелы в знаниях учащихся и оказать им помощь в усвоении программного материала. Текущий контроль стимулирует ответственность ученика за подготовку к каждому занятию.

Периодический контроль подводит итоги работы за определенный период времени. Он осуществляется в конце четверти или полугодия.

Тематический контроль проводится после изучения темы, раздела для определения степени усвоенности данного материала.

Итоговый контроль призван определить конечные результаты обучения. Он охватывает всю систему знаний, умений и навыков по предмету.

Отсроченный контроль - определение остаточных знаний и умений спустя какое-то время после изучения темы, раздела, курса (этот срок может колебаться от трех месяцев до полугодия и более). Отсроченный контроль как вид контроля соответствует требованию судить об эффективности процесса по конечному результату.

Контроль осуществляется в различных *формах*. По форме контроль подразделяется на индивидуальный, групповой и фронтальный.

При контроле используются различные методы. *Методы контроля* – это способы, с помощью которых определяется результативность учебно-познавательной деятельности.

обучаемых и педагогической работы обучающихся. В педагогической практике используются методы устного, письменного, практического, машинного контроля и самоконтроля.

Устный контроль осуществляется в процессе устного опроса обучаемых. Он позволяет выявить знания обучаемых, проследить логику изложения ими материала, умение использовать знания для описания или объяснения процессов и происходящих событий, для выражения и доказательства своей точки зрения, для опровержения неверного мнения и т.д.

Письменный контроль предполагает выполнение письменных заданий (упражнений, контрольных работ, сочинений, отчетов и т.д.). Такой метод контроля позволяет проверять знания всех обучаемых одновременно, но требует больших временных затрат на проверку письменных заданий.

Практический контроль применяется для выявления сформированности умений и навыков практической работы или сформированности двигательных навыков.

С развитием информационных технологий распространение получил контроль с использованием компьютеров. *Машинный контроль* экономит время учащихся и учителя. С помощью контролирующих машин легко установить единые требования к измерению и оцениванию знаний. Результаты контроля легко поддаются статистической обработке. Устраняется субъективизм учителя при оценивании знаний.

Применение контролирующих машин позволяет успешно осуществлять *самоконтроль*. Самоконтроль возможен и без применения машин. Но для этого необходимо научить обучаемых самостоятельно находить ошибки, анализировать причины неправильного решения познавательных задач, искать способы их устранения.

Сочетание различных методов контроля называется *комбинированным (уплотненным) контролем*. Обычно сочетание устного и письменного опроса.

К контролю в процессе обучения предъявляются следующие *педагогические требования*:

- индивидуальный характер контроля. Контроль должен осуществляться за работой каждого ученика, за его личной учебной деятельностью. Нельзя допускать подмены результатов учения отдельных учащихся итогами работы коллектива, и наоборот;
- систематичность, регулярность проведения контроля на всех этапах процесса обучения;
- разнообразие форм проведения контроля, что в большей степени обеспечивает выполнение обучающей, развивающей и воспитывающей функций контроля;
- всесторонность контроля. Контроль должен давать возможность проверки

теоретических знаний, интеллекту субъективные и ошибочные суждения и выводы; индивидуальные личностные качества обучаемых;

- единство требований со стороны обучающихся.

В последнее время все большее распространение получает *тестовый контроль*. Основным инструментом такого контроля является *тест*. В зависимости от предмета измерения выделяются тесты педагогические, психологические, социологические, социально-психологические, культурологические и др.

Рейтинговая сумма баллов с учетом перечисленных выше активов формируется по результатам таких видов контроля, как:

1. *Входной контроль* — контроль знаний и умений студентов при начале обучения очередной дисциплине.

2. *Текущий контроль* — непрерывно осуществляемое "отслеживание" уровня усвоения знаний и умений студентами на лекциях, лабораторно-практических занятиях; своевременного и аккуратного оформления отчета о лабораторной работе и его защите; быстроты и точности решения ситуационных задач. По всем видам деятельности ведется учет баллов, принятый кафедрой, исходя из важности и трудности выполняемых работ.

3. *Рубежный контроль* — контроль умений и знаний студентов по окончании изучения темы (раздела), модуля. Этот вид контроля заключается в написании контрольной работы, расчетно-графическом задании или системе заданий в виде тестов.

4. *Итоговый контроль* — контроль знаний, умений, навыков и поддающихся критериально-оценочной процедуре личностных качеств. Итоговый контроль оформляется в виде зачетного или экзаменационного теста по всему предмету.

5. *Отсроченный контроль* — контроль остаточных знаний и умений спустя какое-то время после изучения темы, раздела, курса (этот срок может колебаться от 3 месяцев до полугодия и более).

Поскольку рейтинг — это шкала достижений студентов, то совершенно очевидно, что должен быть стандартный инструмент измерения. Таким инструментом является правильно построенный и хорошо составленный тест, который соответствует не только предмету обучения, но и его задачам, является педагогическим тестом достижений. Обычно в тесте нет заданий, на которые можно найти прямой ответ в учебнике, поэтому возможность списывания практически исключена или бессмысленна.

Таким образом, рейтинг служит развитию и закреплению системного подхода к изучению дисциплины. Это, пожалуй, наиболее важная положительная сторона рейтинга, кроме тех, что отмечены выше.

Обобщая достоинства рейтинговой системы как одного из инновационных методов, можно констатировать, что рейтинговая система — это не только оценка уровня усвоения знаний, но и метод системного подхода к изучению дисциплины. При рациональном, с дидактической позиции, разделении дисциплины на блоки возможно ее усвоение каждым обучающимся самостоятельно. В этом случае самостоятельная работа студентов, ее реализация в вузе, создает условия для развития личности студента. Самостоятельная деятельность по приобретению студентами знаний и умений предполагает четкую регламентацию обучения в зависимости от содержания предмета, условий обучения, уровня подготовленности студентов к восприятию и усвоению материала.

Функции оценки, как известно, не ограничиваются только констатацией уровня обученности. Оценка – единственное в распоряжении педагога средство стимулирования учения, положительной мотивации, влияния на личность. Именно под влиянием объективного оценивания у студентов создается адекватная самооценка, критическое отношение к своим успехам. Поэтому значимость оценки, разнообразие ее функций требуют поиска таких показателей, которые отражали бы все стороны учебной деятельности студенток и обеспечивали их выявление. С этой точки зрения ныне действующая система оценивания знаний, умений требует пересмотра с целью повышения ее диагностической значимости и объективности.

Важнейшими принципами диагностирования и контролирования обученности (успеваемости) обучающихся являются объективность, систематичность, наглядность (гласность). *Объективность* заключается в научно-обоснованном содержании диагностических тестов (заданий, вопросов), диагностических процедур, равном, дружеском отношении педагога ко всем обучаемым, точном, адекватном установленным критериям оценивания знаний, умений. Практически объективность диагностирования означает, что выставленные оценки совпадают независимо от методов и средств контролирования и педагогов, осуществляющих диагностирование.

Требование *принципа систематичности* состоит в необходимости проведения диагностического контролирования на всех этапах дидактического процесса — от начального восприятия знаний и до их практического применения. Систематичность заключается и в том, что регулярному диагностированию подвергаются все обучаемые с первого и до последнего дня пребывания в учебном заведении. Контроль необходимо осуществлять с такой частотой, чтобы надежно проверить все то важное, что обучаемым надлежит знать и уметь. Принцип системности требует комплексного подхода к проведению диагностирования, при котором различные формы, методы и средства контролирования проверки, оценивания используются в тесной взаимосвязи и единстве,

подчиняются одной цели. Такой подход исключает универсальность отдельных методов и средств диагностирования.

Принцип наглядности (гласности) заключается прежде всего в проведении открытых испытаний всех обучаемых по одним и тем же критериям. Рейтинг каждого обучающегося, устанавливаемый в процессе диагностирования, носит наглядный, сравнимый характер. Принцип гласности требует также оглашения и мотивации оценок. Оценка — это ориентир, по которому обучаемые судят об эталонах требований к ним, а также об объективности педагога. Необходимым условием реализации принципа является также объявление результатов диагностических срезов, обсуждение и анализ их с участием заинтересованных людей, составление перспективных планов ликвидации пробелов.

Диагностировать, контролировать, проверять и оценивать знания, умения обучающихся нужно в той логической последовательности, в какой проводится их изучение.

Качество усвоения студентами подлежащего изучению материала, приобретенного (усвоенного) ими опыта и, следовательно, деятельности, которую они могут осуществлять в результате обучения, может характеризоваться уровнями усвоения (деятельности). Напомним их.

1-й уровень — уровень представления (знакомства). Студент, выведенный на этот уровень, способен узнавать объекты и процессы, если они представлены ему сами (в материальном виде) или даны их описание, изображение, характеристика. На этом уровне студент обладает знанием-знакомством и способен опознать, различить и соотнести эти объекты и процессы.

2-й уровень — уровень воспроизведения. Студент может воспроизвести (повторить) информацию, операции, действия, решить типовые задачи, рассмотренные при обучении. Он обладает знанием-копией.

3-й уровень — уровень умений и навыков. На этом уровне усвоения студент умеет выполнять действия, общая методика и последовательность (алгоритм) которых изучены на занятиях, но содержание и условия их выполнения новые. Здесь различают две разновидности усвоения: *умение*, когда студент выполняет действия после довольно продолжительного предварительного продумывания последовательности и способов их осуществления, *навык*, когда действие выполняется автоматически. Обдумывание каждой предстоящей операции резко "свернуто" во времени. Создается впечатление, что исполнитель работает "не думая".

4-й уровень – уровень творчества. Как известно, творчеством считают проявление продуктивной активности человеческого сознания. Например, рационализаторство и изобретательство, работа по реконструкции в ходе реального курсового проектирования, при участии в научно-исследовательской работе. Чтобы вывести студента на уровень творчества, недостаточно, чтобы он овладел знаниями, умениями и навыками по определенному, пусть даже весьма широкому набору учебных элементов. Необходимо обучить его умению самостоятельно "добывать" необходимые знания и умения. Нужно пробудить и развить в нем творческие склонности. А это возможно только при условии, что в процессе обучения будут применяться специальные творческие задачи научно-исследовательской, проектной, конструкторской, технологической деятельности, т.е. будут реализовываться мотивационные знания.

Для достижения любого уровня усвоения студент должен осуществить учебную деятельность (УД), состоящую из трех видов действий: ориентировочной основы действия (ООД), исполнительских действий (ИД) и контрольных действий (КД), выполняемых студентами обычно с помощью преподавателя:

$$\text{УД} = \text{ООД} + \text{ИД} + \text{КД}.$$

Здесь ООД заключается в получении студентами необходимой информации, осознании поставленной перед ними задачи усвоения. С помощью полученных от преподавателя инструктажа и ориентировки они выбирают пути, средства и методы (выбирают программу) ее решения; ИД состоят в интеллектуальной переработке полученной информации и выполнении упражнений с целью усвоения знаний, умений и навыков. Студент выполняет программу, разработанную в ходе ООД; КД – действия, с помощью которых проверяются полнота, правильность и качество выполнения ООД и ИД.

Приведенные выше характеристики контроля качества знаний и умений весьма условны. Качество обучения, по С.И.Архангельскому [4], рассматривается как способность студентов выполнять определенные требования, поставленные перед ними, с учетом целей и задач изучения того или иного предмета.

Лекция 2. История развития системы тестирования в России и за рубежом

1. Возникновение тестирования. Деление тестов на педагогические и психологические.

2. Современное развитие тестологии в Европе, Японии, Канаде, США. Современная теория тестов (IRT). История ее создания.

3. Развитие тестирования в России. Начало развития тестирования в рамках педологии. Период игнорирования тестов. Использование тестов в 70-е годы XX в. Современные центры тестирования.

1. Возникновение тестирования.

Деление тестов на педагогические и психологические.

Для успешной разработки педагогических заданий в тестовой форме полезно знать основы той науки, которая занимается изучением тестов. Тестология появилась на стыке психологии, социологии, педагогики и других, так называемых поведенческих наук, и призвана заниматься вопросами разработки тестов для объективного контроля знаний, умений, навыков и представлений.

Впервые тест возник как метод изучения индивидуальных различий. Временем его возникновения считается конец XIX – начало XX веков, когда для определения физических, физиологических и психических особенностей человека психологи попытались применить различные способы измерения.

Быстрое развитие естествознания способствовало проникновению точных измерительных методов в такие «неточные» сферы, как психология и педагогика, что, в свою очередь повлекло за собой возникновение и распространение теорий, объяснявших процесс развития человека как изменяющийся и усложняющийся во времени. Первым ученым, попытавшимся измерить различия между людьми в области элементарных психических процессов, был англичанин Френсис Гальтон (F. Gallon; 1882–1911).

Проводя исследования индивидуальных различий, Ф. Гальтон использовал набор методик на определение зрительной, слуховой, тактильной чувствительности, на мускульную силу и другие характеристики.

Ф. Гальтон называл испытания, проводившиеся в его лаборатории, умственными тестами. Важным вкладом Ф. Гальтона в развитие теории тестов было определение трех основных принципов:

1. Применение серии одинаковых испытаний к большому количеству испытуемых.

2. Статистическая обработка результатов.

3. Выделение эталонов оценки.

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

Термин «умственный тест» приобрел наибольшую популярность после выхода статьи Джеймса Мак-Кина Кеттелла (Cattell J. Mc-K.; 1860–1944) «Умственные тесты и измерения», опубликованной в 1890 году с послесловием Ф. Гальтона. Горячий сторонник и пропагандист тестового метода, Дж. Кеттелл считал тест средством для проведения научного эксперимента с соответствующими требованиями к чистоте научного эксперимента. Все эти идеи, выдвинутые Дж. Кеттеллом, в настоящее время составляют основу для современной тестологии. Одинаковость условий для всех испытуемых, одинаковые инструкции и четкое их понимание испытуемыми – фундаментальные принципы, положенные в основу стандартизации процедуры проведения тестирования. Ограничение времени, в настоящее время, после дополнительных исследований, реализуется в зависимости от возраста испытуемых и особенностей применяемого инструментария. Идеи статистической обработки результатов реализованы в достаточно сложных методах статистического анализа и моделирования. В 1895–96 годы в США были созданы два национальных комитета, призванных объединить усилия тестологов и придать общее направление появлявшимся тестологическим работам.

Большой вклад в развитие тестологии внес видный французский психолог Альфред Бине (Binet A.; 1857–1911). А. Бине может считаться родоначальником современных тестов, предназначенных для диагностики уровня развития интеллекта. А. Бине ввел в обиход понятие умственного уровня (позднее замененного на умственный возраст, а еще позже на показатель умственного развития IQ).

Длительное время к тестам относились как к инструменту для индивидуальных измерений индивидуумов. Массовый характер тестирования вызвал необходимость перейти от индивидуальных тестов к групповым. В 1917–18 годы в США появились

первые групповые тесты для нужд армии. Наибольшее применение нашли тесты Артура Синтона Отиса (Otis A.S.; 1866–1963). Им были предложены два набора тестов (Альфа- и Бета-тесты): для владеющих английским языком и «немые» тесты для не владеющих или плохо владеющих языком.

Основные принципы, использованные при составлении этих тестов, были систематизированы и впоследствии легли в основу всей методологии групповых тестов.

Одновременно широко развивались методики обработки результатов тестирования, создания тестовых систем. Ф. Гальтон первый разработал метод статистического сравнения двух рядов переменных и ввел индекс совместного отношения, названный коэффициентом корреляции. Также Ф. Гальтону принадлежит заслуга изучения взаимосвязи переменных и построения линий регрессии одной переменной на другую. В 1896 году Карл Пирсон (Pearson K.; 1857–1936) изложил основы теории корреляции. А в 1904 году Чарльз Эдвард Спирмен (Spearman Ch. E.; 1863–1945) показал, что корреляция между двумя переменными свидетельствует о наличии общего фактора (причины, определяющей величины этих переменных) и специфических факторов, присущих каждой переменной. Ч. Спирмен явился основателем двухфакторной теории.

Луис Левон Терстоун (Thurstone L.L.; 1887–1955) развил далее идею факторного анализа, исходя из понимания интеллекта как целого, состоящего из нескольких первичных способностей (вербальной, перцептивной, способности к восприятию пространства, памяти, беглой речи). Все эти факторы равноправны, не существует единого фактора интеллекта. С помощью факторного анализа не просто устанавливается связь изменения одной переменной с изменением другой, а должна определяться мера этой связи, и найдены основные причины, лежащие в основе указанных изменений. Достижения корреляционного и факторного анализа находят свое применение и в современных исследованиях.

С начала XX века определилось и педагогическое направление в развитии тестологии. Американец В.А. Макколл разделил тесты на педагогические (Educational Test) и психологические – по определению уровня умственного развития (Intelligence Test). Основной задачей педагогических тестов являлось измерение успешности обучения обучающихся по тем или иным дисциплинам за

определенный период обучения, а также успешность применения определенных методов и технологии преподавания.

Целью использования педагогических тестов, по мнению В. Макколла, является объединение в группы обучающихся, усваивающих равный по объему материал и усваивающих материал с одинаковой скоростью.

Разработка первого педагогического теста принадлежит американскому психологу Эдуарду Ли Торндайку. Он считается основоположником педагогических измерений. Результатом большой исследовательской работы в области измерений и использования метода тестов в педагогике явилась книга «Введение в теорию психологии и социальных измерений» (1904). Первый стандартизованный тест, вышедший под руководством Э. Торндайка был тест Стоуна на решение арифметических задач, впервые снабженный «нормами».

В 1915 году Йеркс создал свою серию тестов, главное отличие которой – изменение системы подсчета. Вместо возрастных долей, предложенных А. Бине, испытуемый получает за каждый правильно решенный тест известное количество баллов. Это повысило удобство проведения и подсчета результата теста. Количество полученных баллов переводилось по приложенным стандартам в коэффициент одаренности или успешности.

2. Современное развитие тестологии. Развитие тестирования в России.

Группа американских психологов работает над созданием «интеллектуальных» тестов для обучения. Эти тесты назывались «национальными тестами». Национальные тесты впервые были составлены в 2 сериях по 2 варианта в каждой серии. Совершенствуя методику, американские тестологи в этих тестах впервые перед каждым разделом теста дают примерные упражнения.

Разработкой и проверкой тестов занимаются специальные государственные службы. Еще в 1900 году в США был создан Совет по вступительным экзаменам. В 1926 году – Совет колледжей принял тест SAT, были разработаны «тесты для квалификационной и профессиональной оценки деятельности педагога». С 1947 года существует Служба тестирования (Educational Testing Service), которая считается наиболее представительным научно-исследовательским центром.

К 1961 году только в США были созданы 2126 стандартизированных тестов. Здесь существует множество тестовых издательств, общая продукция которых представляет несколько тысяч различных тестов. Естественно, они различаются по качеству и по спросу на них. Изданные тесты ежегодно аннотируются и наиболее значительные комментируются в ежегоднике «Mental Measurement Yearbook».

Интеграционные процессы, характеризующие современную цивилизацию, дали толчок развитию новых направлений тестирования. Оценка качества образования приобретает все большую значимость по мере увеличения количества стран, осознающих преимущества сопоставления собственных достижений с международным контекстом.

С 1959 года основные крупномасштабные международные исследования проводятся под руководством неправительственной организации Международная ассоциация по оценке успеваемости (ИЕА). Учреждения ИЕА созданы в более чем 45 странах мира и подключают к своей деятельности министерства образования, педагогические факультеты университетов, научно-исследовательские институты.

Исследования ИЕА строятся в определенной последовательности и базируются на объяснительных моделях, основанных на учебных программах. Предполагается обследование тысяч обучающихся в каждой стране с помощью академических тестов, составляемых на основе учебных программ и других занятий. Используемые тесты успеваемости направлены на измерение достижений обучаемых, которые можно оценить в баллах, скорректировать и сравнить с аналогичными данными по стране, а также соотнести с другими факторами, влияющими на успеваемость (например, с содержанием обучения, технической оснащенностью школы, полом обучающихся, местностью проживания и пр.). Цель такого исследования – создание достаточно гибкой модели, позволяющей измерить комплексное воздействие школы на процесс овладения знаниями.

Помимо ИЕА существуют другие организации, занимающиеся проблемами международного тестирования. Например, Институт образования ЮНЕСКО (Гамбург), Международный центр педагогических исследований в Париже провели совместно с ИЕА исследование навыков чтения в 35 странах (1990 год). При этом определение лучшей или худшей страны в изучаемой области не является целью

тестирования. Задачей исследования становится поиск наиболее эффективных условий обучения.

В России до 1917 года вопросам тестирования не уделялось достаточного внимания. Практическое значение тесты получили только после 1925 года, когда была создана особая тестовая комиссия. Тестовая комиссия существовала при педагогическом отделе Института методов школьной работы. В ее задачи входила разработка стандартизированных тестов для советской школы. И уже в 1926 году были разработаны тесты по некоторым предметам: природоведению, обществоведению, счету, решению задач, знанию географической карты, на понимание чтения и правописание, созданные на основе американских. К этим тестам прилагались инструкции и личная карточка для учета прогресса учащегося.

Проблемой разработки тестов вплотную занимались видные российские психологи и педагоги: С.Г. Геллерштейн, П.П. Блонский, А.П. Болтунов, М.С. Бернштейн, А.М. Шуберт, Г.И. Залкинд и др. В 1927 году вышла книга С.М. Василейского «Введение в теорию и технику психологического, педологического и психотехнического исследования», где излагались теоретические и практические подходы к созданию и использованию тестов.

В 20-е годы Центральная педологическая лаборатория МОНО под руководством педолога Е. Гурьянова разработала и выпустила в свет несколько тестов:

1. Шкала для измерения умственного развития испытуемых.
2. Тесты для учета навыков в чтении, счете и письме.
3. Тесты коллективного испытания умственной одаренности.

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

Однако вскоре положение кардинально изменилось. Вышедшее в 1936 году постановление ЦК ВКП(б) «О педологических извращениях в системе Наркомпросов» негативно сказалось на развитии педологии в целом и тестологии

в частности. Метод тестов был признан буржуазным орудием для дискриминации обучающихся и «изгнан» из советской школы. Одной из причин негативного отношения к тестам того времени было отсутствие заинтересованности системы управления в объективных результатах обследований педологов.

В настоящее время в нашей стране появилось несколько центров, в которых достаточно профессионально занимаются работой с тестовыми методиками. Среди наиболее активных следует назвать: Центр тестирования профессионального образования (ЦТПО, г. Москва), Центр оценки качества образования Института общего среднего образования РАО, Центр тестирования выпускников общеобразовательных учреждений Российской Федерации, Центр психологического и профессионального тестирования МГУ, Лаборатория аттестационных технологий Московского института повышения квалификации работников образования (МИПКРО), Лаборатория изучения образовательных систем Центра развития образования (г. Санкт-Петербург), Центр аттестации областного института повышения квалификации и переподготовки педагогических кадров (г. Вологда), Научно-информационный центр государственной аккредитации Минобрнауки России (г. Йошкар-Ола), Исследовательский центр проблем качества подготовки специалистов, Центр аттестации Института развития регионального образования (г. Екатеринбург) и целый ряд других.

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

Лекция 3. Психолого-педагогические аспекты тестирования

1. Роль психологической подготовки к тестированию.
2. Тестирование и мотивация.
3. Тестирование и личностно-психологические, национальные особенности тестируемых.

1. Роль психологической подготовки к тестированию

Введение в контекст обучения новой формы контроля – тестирования, приводит к изменению прежде всего восприятия и установок контроля, меняется вся обстановка, в том числе психологическая ситуация, связанная с формированием у участников тестирования положительного отношения к ним, с приятием, одобрением или неприятием, отторжением тестов как формы контроля.

Среди педагогов много таких, которые уже в течение ряда лет используют тестовые методы в обучении и в контроле. Но есть педагоги, которые опасаются открытых процедур.

Интересные данные об оценке учащимися методов традиционного и тестового контроля были получены Н.Ф.Ефремовой в ходе диагностического исследования такого рода предпочтений (опрошено около 500 респондентов). Наибольшим баллом при этом респондентами-участниками теста была оценена объективность процедуры тестирования и независимость оценки от мнения педагога, а также постулат неразглашение оценок. В качестве преимуществ контроля традиционными методами были отмечены: возможность подсказки, более низкий, уровень трудностей задания. Вместе с тем ни один из респондентов не удостоил традиционный контроль такими качествами, как объективность, глубина и полнота проверки (Современные тестовые технологии в образовании, 2003, с.133).

2. Тестирование и мотивация

Под мотивацией в педагогике понимается система целей, потребностей, мотивов, побуждающих индивида сознательно относиться к учению и к контролю как его составной части, быть активным в учебно-познавательной работе и т.д. Мотивация выполняет не только побуждающую, но и направляющую, организующую, смыслообразующую функции. В нашем

случае – при тестировании – мотивация служит одним из основных психологических факторов успешности контроля.

Как известно, на уровень мотивации учебно-познавательной деятельности влияют объективные факторы (статус, история изучаемого предмета и т.д.) и субъективные (жизненные планы человека, стремление к самореализации и т.д.), а также внешние мотивы (желание иметь хорошие оценки и др.).

Не будет преувеличения сказать, что процесс побуждения человека к тестированию нередко оказывается сложным социально-психологическим, интеллектуальным, эмоциональным и волевым актом анализа и оценки альтернатив, выбора и принятием решения. Этот процесс психологически осложняется тем, что далеко не всегда реальные мотивы осознаются субъектом актуально, (например, так называемая инструментальная мотивационная основа связана с потребностью проверить уровень владения русским языком для выполнения профессионально-деловых, учебно-образовательных обязанностей.) т.е. при подготовке и выполнении тестирования. Есть мотивы скрытые, могут быть материальные, семейные, туристические мотивы и т.д.).

Результаты теста, несомненно, зависят от привычности и искушенности испытуемых. Понятно, что человек, имеющий опыт выполнения тестов, обладает преимуществом перед теми, кто впервые участвует в тестировании. Отчасти эти преимущества складываются из преодоленного чувства неизвестности, развившейся уверенности в себе, в своих возможностях. Чувство напряженности, тревожности, сомнения в удачном исходе экзамена снимают разного рода предтестовые тренировки. Они помогают механизмам памяти, речи настраиваться на работу в жестко лимитированном по времени экзамене. Американский специалист в области тестологии, психолог Анна Анастаси считает, что у человека, искушенного в работе заданиями тестового формата, показатели на 11% выше, чем у «неискушенных» испытуемых (Анастаси А., 1982, с. 48).

Желательно, чтобы группа тестируемых (в том случае, если можно объединить участников процедуры) была выравнена по опыту прохождения теста.

Были выделены два важных компонента тревожности – эмоционального и озабоченного. Эмоциональная составляющая тестовой тревожности охватывает чувства, физиологические реакции, такие как напряжение. Озабоченность связана с негативными мыслями, с ожиданием неудачи при выполнении теста и с озабоченностью последствиями неуспеха. Эти мысли отвлекают внимание от выполнения заданий теста. Зарубежные психологи разработали в этой связи методики поведенческой терапии (G.S. Tryon, 1980). Вместе с тем, тестовая тревожность, как оказалось, способна убывать – и без «терапевтического вмешательства» – с каждым последующим тестированием (Анастаре А., Урбина С., 2001, с. 37,38).

Тестовая тревожность, тем не менее, – комплексный феномен, вызываемый множеством разнородных причин, специфических для конкретного человека. Однако, к числу причин, позволяющих преодолевать тревожность, относятся, кроме вышеназванных, совершенствование языковых, коммуникативно-речевых навыков и умений.

Немаловажны и руководства к тестам, которые могут предоставлять пользователю информацию о процедуре проведения тестирования, подсчете показателей, оценивающих качество выполнения теста.

3. Тестирование и личностно-психологические, национальные особенности тестируемых.

Исследователями в области психологии (см. Немов Р.С., 1994) установлено, что только около 40% людей в возрасте до 23 лет обладают зрительно-образной памятью. Она помогает запоминать материал через смысловую группировку, через его соотнесение с ранее известным. Большинство же людей нуждается в осмыслении материала, к примеру, в его повторном чтении. В результате «развивается» дефицит времени, растет чувство напряженности во время экзамена, снижается сосредоточенность.

Кроме того (и это обнаруживает тестирование), не все учащиеся обладают способностью к моментальному вводу информации (психологи именуют это состояние «ретроградной амнезией»). Сдающему устный экзамен бывает нужно «разговориться» - «внутренне» или «внешне». Это так называемый этап «вработывания», который следует учитывать при планировании временных рамок тестирования. Как правило, после фазы «вработывания» наступает этап устойчивой работоспособности, который при соблюдении условий тестирования (при отсутствии помех, шума, в атмосфере доброжелательности) может длиться до трех и более часов. Затем возможен этап снижения работоспособности, нередко завершающийся состоянием «конечного порыва», когда мобилизуются дополнительные силы через волевую установку, обязывающую закончить экзаменационную работу с самоотдачей. Волнообразный характер деятельности несложно учесть при комплектовании блоков тестовых заданий. Желательно не только располагать их в последовательности, но и устранять однообразные целевые установки типа «Выберите правильный вариант ответа». Экзаменуемый может мобилизовать внимание, силы, способности, если ему на помощь «придет» удачная словесная инструкция или средство, регулирующее внимание и настроение (знаки, символы, призывы), разнообразные формы предъявления стимула задания (вербальный тест может сменяться сюжетными картинками, видеорядом, аудиозаписью).

Лекция 4. Педагогические тесты. Термины и определения

1. Педагогический контроль. Функции, формы и принципы контроля.
2. Понятийный аппарат тестологии.
3. Валидность, надёжность, объективность теста. Гомогенность и гетерогенность.
4. Адаптивное тестирование.

1. Педагогический контроль. Функции, формы и принципы контроля.

Известно, что контроль стимулирует обучение и влияет на поведение студентов. Как показала практика, попытки исключить контроль частично или полностью из учебного процесса приводят к снижению качества обучения. Внедряемые в настоящее

время интенсивные методы обучения ведут неизбежно к новым поискам в области повышения качества и эффективности педагогического контроля и появлению его новых форм, например, таких как рейтинг.

В области контроля можно выделить три основные взаимосвязанные **функции**: диагностическую, обучающую и воспитательную.

Диагностическая функция: контроль – это процесс выявления уровня знаний, умений, навыков, оценка реального поведения студентов.

Обучающая функция контроля проявляется в активизации работы по усвоению учебного материала.

Воспитательная функция: наличие системы контроля дисциплинирует, организует и направляет деятельность студентов, помогает выявить пробелы в знаниях, особенности личности, устранить эти пробелы, формирует творческое отношение к предмету и стремление развить свои способности.

В учебно-воспитательном процессе все три функции тесно взаимосвязаны и переплетены, но есть и формы контроля, когда одна, ведущая функция превалирует над остальными. Так, на семинаре в основном проявляется обучающая функция: высказываются различные суждения, задаются наводящие вопросы, обсуждаются ошибки, но вместе с тем семинар выполняет диагностическую и воспитывающую функции.

Зачеты, экзамены, коллоквиумы, тестирование выполняют преимущественно диагностическую функцию контроля.

При применении программированного контроля проявляется его обучающая и контролирующая функции.

Формы педагогического контроля.

Систему контроля образуют экзамены, зачеты, устный опрос (собеседование), письменные контрольные, рефераты, коллоквиумы, семинары, курсовые, лабораторные контрольные работы, проектные работы, дневниковые записи, журналы наблюдений. Каждая из форм имеет свои особенности.

Во время устного опроса контролируются не только знания, но тренируется устная речь, развивается педагогическое общение. Письменные работы позволяют документально установить уровень знания материала, но требуют от преподавателя больших затрат времени. Экзамены создают дополнительную нагрузку на психику студента. Курсовые и дипломные работы способствуют формированию творческой личности будущего специалиста. Умелое сочетание разных видов контроля -

показатель уровня постановки учебного процесса в вузе и один из важных показателей педагогической квалификации преподавателя.

По времени педагогический контроль делится на текущий, тематический, рубежный, итоговый, заключительный.

В российской высшей школе выделяют три основных **принципа контроля**.

Воспитывающий принцип проявляется в том, что активизирует творческое и сознательное отношение студентов к учебе, стимулирует рост познавательных потребностей, интересов, организует учебную деятельность и воспитательную работу. Всякий контроль, принижающий личность студента, не может применяться в вузе.

Систематичность. Систематический контроль упорядочивает процесс обучения, стимулирует мотивацию, дает возможность получить достаточное количество оценок, по которым можно более объективно судить об итогах учебы.

Всесторонность. Круг вопросов, подлежащих контролю и оцениванию, должен быть широк настолько, чтобы охватить все основные темы и разделы предмета.

При оценке личности следует учитывать деловые, моральные качества, ценностные ориентации, общественную работу, учебные достижения. Всесторонность контроля – одна из составляющих комплексного подхода к воспитанию.

2. Понятийный аппарат тестологии.

Ядро научного аппарата науки о тестах образуют понятия.

Адаптивное (последовательное) тестирование - вид компьютерного тестирования, при котором тестовые задания с известными характеристиками последовательно изображаются на экране компьютера, а уровень подготовленности испытуемого со все возрастающей точностью оценивается сразу же после каждого его ответа. Очередное задание в зависимости от ранее данных ответов испытуемых подбирается так, чтобы его уровень трудности позволял наилучшим образом оценить уровень подготовленности тестируемого. Количество заданий теста заранее не фиксируется, а процесс тестирования заканчивается по достижении заданной точности оценки уровня подготовленности испытуемого.

Апробация теста - предварительное тестирование стратифицированной выборки испытуемых для определения соответствия теста его целям и априорным характеристикам. Апробация есть необходимый этап для создаваемого теста перед его широким использованием.

Ассоциации в тестовом задании - словесная подсказка в содержании задания, позволяющая тестируемому угадать правильный ответ; свидетельство некорректности задания.

Балл - условная единица для оценки по определенной шкале результатов выполнения теста или его задания.

Балл истинный - латентный объект измерений, представляющий собой тот балл по определенной шкале, которому объективно соответствует уровень подготовленности данного испытуемого при указанной методике оценивания (при этом ошибки измерений условно предполагаются отсутствующими).

Балл категории ответа на тестовое задание - балл, соответствующий определенной категории ответа на политомическое задание в зависимости от уровня полноты того ответа (например, категория ответа на уровне узнавания - один балл, на уровне репродукции - два балла, на уровне оперативного применения - 3 балла и т.д.).

Балл критериальный - граничное значение тестового балла, с помощью которого заданная выборка тестируемых разделяется на выполнивших данный тест удовлетворительно ("зачет") и неудовлетворительно ("незачет").

Балл первичный дихотомического задания - количество участников тестирования, верно выполнивших данное задание; дополнение до общего количества испытуемых, выполнявших задание, отражает некоторую меру трудности задания для данного контингента участников.

Балл первичный испытуемого - сумма баллов, приписанных тем категориям ответов на тестовые задания, которые испытуемый указал в качестве правильных. Для тестов, состоящих из дихотомических заданий - количество верно выполненных заданий, что отражает некоторую меру подготовленности данного испытуемого относительно данного теста.

Балл первичный категории задания - количество участников тестирования, указавших данную категорию ответа для определенного задания в качестве правильного ответа; в случае дихотомического задания совпадает с первичным баллом этого задания.

Балл тестовый - окончательное количественное выражение по определенной шкале индивидуальной оценки уровня подготовленности тестируемого, полученное на основании стандартизованной обработки результатов выполнения испытуемым тестовых заданий.

Банк тестовых материалов - совокупность систематизированных тестовых заданий и тестов, разработанных различными авторами, для различных целей, прошедших апробацию и имеющих известные характеристики.

Бланк ответа - стандартный бланк для записи ответов на предложенные в тесте задания; тестируемый отмечает или записывает номера выбранных ответов или сами ответы в определенные позиции бланка.

Валидизация - процедура улучшения валидности теста по результатам оценки валидности критериальной.

Валидность теста критериальная - характеристика теста, отражающая показатель соответствия диагноза и прогноза теста определенным внешним критериям, характеризующим объект измерения.

Валидность теста прогностическая - частный случай валидности критериальной. Отражает эффективность прогноза теста о возможностях испытуемых в будущем.

Валидность теста содержательная - характеристика теста, выражающая показатель охвата заданиями теста той области знания, подготовленность в которой этот тест оценивает.

Валидность сопоставительная (текущая) - частный случай валидности критериальной. Отражает соответствие текущего диагноза теста результатам другого измерения того же объекта.

Валидность теста - комплексная характеристика теста, отражающая его способность измерять именно то, для чего он предназначен. Характеризует возможности генеральной совокупности заданий в тестируемой области знаний несмещенно оценить объект измерений теста.

Валидность теста факторная - характеристика теста, выражающая корреляцию между данным тестом и каждым из выбранных факторов контроля знаний, умений и навыков.

Варианты теста - набор неидентичных тестов, созданных по единой спецификации и имеющих одинаковую структуру.

Время тестирования - интервал времени, отводимого на выполнение теста.

Выравнивание вариантов теста - определенный метод решения комплекса вопросов, связанных с отображением на определенной шкале латентных характеристик всех тестовых заданий (в том числе, составляющих его различные варианты).

Генеральная совокупность тестовых заданий - гипотетическое множество тестовых заданий, связанных общей целью и теоретически полностью отражающих определенную область знаний.

Дистрактор (отвлекающий ответ) - вариант ответа на тестовое задание закрытого типа, похожий на правильный, но не являющийся таковым.

Дихотомическое (альтернативное) тестовое задание - задание, выполнение которого оценивается только альтернативно : выполнено верно (обычно символизируется единицей) или выполнено неверно (обычно символизируется нулем).

Длина теста - количество тестовых заданий теста.

Задание закрытого типа - тестовое задание, содержание которого сопровождается несколькими номерованными вариантами ответа; испытуемому предлагается выбрать номер правильного ответа.

Задание открытого типа - тестовое задание без указания возможных вариантов ответа; испытуемому предлагается самостоятельно указать правильный ответ.

Задания узловые - тестовые задания, общие для всех или нескольких вариантов одного и того же теста. Используются для выравнивания результатов тестирования с использованием различных вариантов тестов.

Инструкция о проведении тестирования - документ, устанавливающий порядок и организацию тестирования, которые определяются используемой методикой, техническими и организационными средствами и запланированными способами обработки.

Инструкция тестового задания - словесные указания, испытуемому, связанные с выполнением тестового задания (выбором правильного ответа из нескольких вариантов; решением математической задачи и т.п.). Указывается способ записи правильного ответа (что, каким образом и где надо отметить, вписать и т.д.). Инструкция может быть единой для нескольких заданий теста, если эти задания однотипны по действиям.

Ключ тестового задания - правильный ответ для тестового задания.

Коэффициент дискриминации (дифференцирующая способность) тестового задания - количественная характеристика способности тестового задания дифференцировать испытуемых по уровню их подготовленности. Изменяется от -1 до +1.

Логит - единица измерения уровней подготовленности участников тестирования и трудностей тестовых заданий в рамках логистических моделей тестирования. Если

разность между упомянутыми двумя понятиями составляет 1 логит, то вероятность верного выполнения таким испытуемым такого задания равна 0,73.

Матрица ответов - прямоугольная таблица, в каждой позиции которой указываются ответы участника тестирования. Обычно номер строки соответствует номеру испытуемого, а номер столбца соответствует номеру задания теста.

Метод Кронбаха - обобщение метода Кьюдера-Ричардсона на случай, когда задания теста не являются дихотомическими.

Метод Кьюдера—Ричардсона - оценка надежности теста, основанная на вычислении по одноименной формуле среднего значения коэффициента надежности, оцененного методом Рюлона при всевозможных расщеплениях теста на две половины; при этом задания теста предполагаются дихотомическими, а само расщепление практически не требуется.

Метод расщепления - оценка надежности теста, основанная на сопоставлении результатов тестирования по двум или нескольким эквивалентным частям теста.

Метод Рюлона - метод оценивания числителя отношения, определяющего коэффициент надежности, по разностям результатов тестирования испытуемых по двум эквивалентным половинам одного и того же теста.

Модель двухпараметрическая - логистическая модель, в которой функция успеха зависит от разности между уровнем подготовленности тестируемого и уровнем трудности тестового задания и от коэффициента дискриминации тестового задания.

Модель однопараметрическая - логистическая модель, в которой функция успеха зависит только от разности между уровнем подготовленности тестируемого и уровнем трудности тестового задания.

Модель тестирования - одна или несколько функциональных зависимостей, гипотетически связывающих подлежащие определению характеристики (параметры) участников тестирования и тестовых заданий с такими величинами, которые реально проявляются в результате выполнения соответствующего теста (например, с вероятностью правильного выполнения испытуемым определенного уровня подготовленности тестового задания определенной трудности).

Модель частичного оценивания - обобщение логистических моделей для полиномических тестовых заданий, позволяющее дополнить альтернативное оценивание заданий (выполнено верно, выполнено неверно) оцениванием частично правильного ответа. Определяется количеством категорий ответа, их оценками и видом соответствующих функций успеха.

Модератор - комплексная характеристика контингента испытуемых (возраст, пол, регион и т. п.).

Надежности коэффициент - количественная характеристика надежности, изменяющаяся от 0 до 1; показывает, в какой мере результаты тестирования можно считать реальными, а в какой приписывать влиянию случайных ошибок. Представляет собой отношение дисперсии измеряемого объекта в выборке (обычно истинного балла) к реально полученной полной дисперсии с учетом неизбежных ошибок измерений (обычно тестового балла).

Надежность теста - показатель точности и устойчивости результатов измерения с помощью теста при его многократном применении. Характеризует степень адекватности отражения тестом соответствующей генеральной совокупности заданий.

Нормативная выборка стратифицированная - группа тестируемых, содержащая представителей всех наиболее значимых страт, реально существующих в генеральной совокупности потенциальных испытуемых, причем в той же пропорции.

Параллельные варианты теста - варианты теста, имеющие одинаковые характеристики.

Педагогические измерения - область педагогики, разрабатывающая и применяющая методы и средства измерений и оценки результатов учебной деятельности испытуемых.

План теста - таблица, в которой каждое тестовое задание соотносится с определенным элементом содержания учебного предмета, конкретным видом знаний или умений, позволяющим объективно судить об уровне подготовленности испытуемых. Указывается также планируемое время выполнения каждого тестового задания и всего теста в целом.

Политомическое тестовое задание - задание, выполнение которого допускает несколько категорий ответа, каждая из которых оценивается по-разному (например, полностью верный ответ - 2 балла, частично верный ответ - 1 балл, неверный ответ - 0 баллов).

Пользователь теста - юридическое или физическое лицо, использующее тест или результаты тестирования для оценки уровня подготовленности испытуемых в определенной области.

Предтест (предварительный, пилотный тест)- педагогический тест, априорно соответствующий спецификации, но подлежащий апробации и, возможно, последующей коррекции.

Протокол результатов тестирования - официальный документ о результатах тестирования определенного контингента испытуемых с указанием тестового балла для каждого участника тестирования.

Разрешающая способность теста (РСТ) - длина промежутка на шкале уровня подготовленности испытуемых, соответствующая изменению первичного балла участника тестирования на единицу.

Ретестовый метод - оценка надежности теста, основанная на повторном, желательно многократном его использовании при приближенно одинаковых условиях. Надежность теста характеризуется степенью соответствия полученных результатов друг другу.

Сертификат результатов тестирования - индивидуальный официальный документ государственного образца о результате централизованного педагогического тестирования определенного лица с указанием полученного тестового балла.

Ситуативные переменные - совокупность случайных факторов, влияющих на результаты тестирования (например, место и время проведения тестирования, особенности поведения тестирующего лица, волнение испытуемого и т.д.).

Спецификация теста - документ, в котором содержится информация о целях, задачах, плане и структуре теста, а также указаны основные требования к правилам проведения тестирования, обработки результатов тестирования и их интерпретации.

Стратификация - описание структуры определенного множества объектов (например, генеральной совокупности потенциальных участников тестирования), расслоение его на страты.

Страты - определенные слои в некотором множестве объектов, например, в генеральной совокупности потенциальных участников тестирования. Характеризуются различными программами обучения, региональным представительством, бытовыми условиями и т.п.

Структура теста - совокупность сведений о числе и типе тестовых заданий по каждому элементу содержания учебного предмета и по каждому виду знаний или умений, позволяющих объективно судить об уровне подготовленности испытуемых. Указывается также предполагаемый уровень трудности каждого тестового задания и, по возможности, его коэффициент дискриминации.

Субтест - определенное подмножество тестовых заданий некоторого теста, допускающее независимую обработку результатов тестирования и позволяющее,

таким образом, решать специфические частные задачи тестирования. Тест может содержать несколько субтестов, возможно перекрывающихся.

Теория моделирования и параметризации педагогических тестов (ТППТ) - теоретическая основа педагогических измерений, позволяющая ввести метрику шкалы, по которой фиксируются результаты тестирования. Основные задачи : разработка моделей тестирования; проверка адекватности различных предположений тестирования реальной действительности; измерение характеристик тестовых заданий и участников тестирования в виде оценки параметров модели; шкалирование результатов тестирования и оценивание их точности.

Тест аппаратурный - тест с использованием технических устройств для изучения внимания, восприятия, памяти и т.д.

Тест гетерогенный - педагогический тест, проверяющий уровень подготовленности по нескольким родственным дисциплинам.

Тест гомогенный - педагогический тест, проверяющий уровень подготовленности по конкретной дисциплине.

Тест достижений - тест, предназначенный для определения степени освоения испытуемым определенной области знаний или умений.

Тест критериально-ориентированный - частный случай теста, предназначенного для абсолютного тестирования, позволяющий оценить, преодолел ли испытуемый определенный порог усвоения учебного материала рассматриваемый учебной программы или ее части. При этом результаты тестирования сравниваются с некоторым критерием уровня подготовленности.

Тест мономорфный - тест, объединяющий задания одного типа (например, задания закрытого типа).

Тест мощности (тест на сложность) - тест, состоящий из заданий возрастающей трудности. Время выполнения лимитируется мягко. Оценивается в зависимости от уровня трудности верно выполненных заданий.

Тест нормативно-ориентированный - частный случай относительного теста, позволяющий сравнивать учебные достижения (уровень подготовленности) отдельных испытуемых друг с другом или относительно предварительно полученных тестовых норм.

Тест обучающий - педагогический тест, включающий актуальный учебный материал, предназначенный для формирования у обучаемого определенного уровня

знаний, навыков, умений и позволяющий осуществлять самоконтроль усвоения материала.

Тест педагогический - совокупность тестовых заданий, позволяющая по результатам их выполнения объективно измерить уровень подготовленности испытуемых по конкретным разделам определенной области знания. Тест педагогический представляет собой конечную репрезентативную выборку из генеральной совокупности калиброванных тестовых заданий.

Тест полиморфный - тест, в котором присутствуют задания различных типов.

Тест репетиционный - тест, позволяющий проверить степень готовности испытуемого к педагогическому тестированию, знакомящий с порядком работы, объемом и сложностью заданий и т.п.

Тест скорости - тест, состоящий из заданий приблизительно одинаковой трудности, оценка по которому определяется количеством выполненных заданий в строго ограниченное время.

Тест смешанный - тест, содержащий черты теста скорости и теста мощности. Наиболее распространен при оценке учебных достижений испытуемых.

Тест стандартизованный - педагогический тест, имеющий спецификацию и определенные характеристики, стабильно подтвержденные на представительной выборке испытуемых. Предназначен для многократного использования.

Тестирование абитуриентское - педагогическое тестирование с целью измерения уровня подготовленности абитуриентов ВУЗов.

Тестирование абсолютное - педагогическое тестирование, предназначенное для оценивания уровня подготовленности каждого испытуемого персонально (безотносительно к успехам других) в четко определенной области знаний.

Тестирование аттестационное - педагогическое тестирование по окончании обучения определенного уровня, как правило, основного и полного среднего образования с целью сопоставления уровня подготовленности учащихся с образовательными стандартами и аттестации испытуемых.

Тестирование компьютерное - педагогическое тестирование на компьютере под управлением специальной программы, обеспечивающей нужную презентацию тестовых заданий и обработку результатов тестирования.

Тестирование на бланках - педагогическое тестирование, при котором носителем содержания теста служит бланк определенного образца; дополняется соответствующим бланком ответов.

Тестирование относительное - педагогическое тестирование, предназначенное для сопоставления уровня подготовленности испытуемых в определенной области знаний относительно друг друга или относительно нормативной выборки.

Тестирование педагогическое - совокупность методических и организационных мероприятий, обеспечивающих разработку педагогических тестов, подготовку и проведение стандартизированной процедуры измерения уровня подготовленности испытуемых, а также обработку и анализ результатов.

Тестирование тематическое - педагогическое тестирование с целью измерения уровня подготовленности испытуемых в определенной области знаний, соответствующей законченной теме или разделу учебной дисциплины.

Тестирование централизованное - педагогическое тестирование, проводимое в разных регионах России по тестам и инструкциям, исходящим из единого центра. Обработка результатов тестирования и шкалирование также выполняется в том же центре.

Тестируемый (испытуемый) - человек, проходящий тестирование, с целью оценки его уровня подготовленности в определенной области знаний.

Тестирующее лицо - наблюдатель, инструктор, непосредственно проводящий тестирование и обеспечивающий самостоятельную работу испытуемых в соответствии с инструкцией по проведению тестирования.

Тестовая батарея - совокупность субтестов, объединенных единой методикой использования и направленных на измерение различных сторон сложного объекта.

Тестовое задание - задание специфической формы, элемент теста, минимальная законченная составляющая единица теста.

Тестовое задание в виде логические соотношения - тестовое задание, при выполнении которого предполагается установить логические соотношения между указанными в тестовом задании понятиями.

Тестовое задание на соответствие (тестовое задание перекрестного выбора) - тестовое задание, при выполнении которого необходимо установить соответствие между элементами двух множеств (двух списков).

Тестовое задание на установление правильной последовательности - тестовое задание, выполнение которого состоит в установлении правильной последовательности операций, действий, событий.

Тестовое задание на установление точки зрения - задание с несколькими ответами качественного характера, среди которых нет правильных и нет

неправильных; испытуемый выбирает ответ в соответствии со своей точкой зрения по сформулированному в задании вопросу, основываясь на своих склонностях, представлениях и т.п. (примером предлагаемых ответов могут быть следующие: полностью согласен, частично согласен, безразличен, не согласен, категорически не согласен).

Тестовое задание с несколькими правильными ответами - задание, в котором правильных ответов может быть несколько.

Тестовые нормы - наиболее важные статистические характеристики распределения результатов тестирования нормативной выборки испытуемых.

Тестолог - специалист, занимающийся теоретическими и практическими вопросами тестологии.

Тестология - наука, изучающая возможности оценить различные свойства личности с помощью специально разработанных тестов и методов их интерпретации.

Трудность тестового задания (уровень трудности) - основная количественная характеристика тестового задания, не зависящая от выборки испытуемых и отраженная на определенной шкале.

Уровень подготовленности испытуемого - количественная характеристика знаний, навыков и умений участника тестирования по определенному кругу вопросов.

Условия тестирования - окружающая обстановка (готовность помещения, состояние рабочих мест, поведение тестирующих лиц) и другие факторы, оказывающие влияние на процесс тестирования.

Функция успеха для категории ответа - функциональная зависимость, указывающая вероятность выбора данной категории ответа в зависимости от близости этой категории к правильному ответу и уровня подготовленности испытуемого. Определяет модель тестирования для политомических заданий.

Функция успеха для тестового задания - функциональная зависимость, указывающая вероятность правильного выполнения задания в зависимости от характеристик этого задания (уровня трудности и др.) с одной стороны и уровня подготовленности испытуемого с другой. Определяет модель тестирования для дихотомических заданий.

Характеристическая функция заданного уровня подготовленности испытуемых - функция характеристик заданий теста, полученная из функции успеха для тестового задания при фиксированном значении заданного уровня подготовленности испытуемых.

Характеристическая функция заданных параметров тестовых заданий - функция одного переменного уровня подготовленности участников тестирования, полученная из функции успеха для тестового задания при фиксированных значениях характеристик заданий.

Шкала - упорядоченное множество действительных чисел (индексов), соответствующих ряду возможных значений измеряемой величины.

Шкала нормализованная с постоянным шагом - порядковая шкала, индексы которой соответствуют равноудаленным значениям стандартной нормальной переменной $z \in N(0;1)$ с некоторым шагом z (например, десятибалльная шкала с $z=0,5$).

Шкала нормализованная стандартная - нормализованная шкала типа $N(0;1)$, тестовый балл по которой пропорционален стандартной нормальной переменной с нулевым математическим ожиданием и единичной дисперсией.

Шкала нормализованная типа N (математическое ожидание μ , дисперсия) - порядковая шкала, полученная в результате преобразования эмпирического распределения частот первичных баллов к модельному нормальному распределению N (математическое ожидание μ , дисперсия), где математическое ожидание μ , и дисперсия- параметры шкалы. Тестовый балл по такой шкале определяется отклонением от математического ожидания в единицах стандарта.

Шкала нормализованных процентилей - порядковая шкала, тестовый балл, по которой соответствует квантилям нормального распределения, в частности, квантилям стандартной нормальной переменной с нулевым математическим ожиданием и единичной дисперсией.

Шкала первичных процентилей - порядковая шкала, тестовый балл испытуемого, по которой определяется соответствующей квантилью эмпирического распределения частот первичных баллов участников тестирования.

Шкала процентная - порядковая шкала, тестовый балл испытуемого, по которой прямо пропорционален первичному баллу соответствующего участника тестирования (в частности, равен первичному баллу, выраженному в процентах).

Шкалирование - определенный метод решения комплекса вопросов, связанных с выставлением каждому испытуемому тестового балла по единой шкале вне зависимости от того, в какой группе и над каким вариантом теста он работал.

3. Развитие понятия валидности и надежности теста

При проведении тестирования учитываются три критерия качества теста: надежность, валидность, объективность.

Надежность - определение степени погрешностей в педагогической оценке, в вычислении истинного значения оценки. В последнее время получил распространение экспертный опрос, когда студента оценивают 2-3 и более преподавателей, и посредством коррелирования результатов появляется возможность надежности оценки.

Валидность теста - соответствие форм и методов контроля его цели.

Объективность - критерий, в котором сочетаются надежность, валидность плюс ряд аспектов психологического, педагогического, этического, ценностного характера.

Понятие валидности (перевод слова с английского valid \approx годный, пригодный) трактуется довольно широко.

В теории тестирования различают два типа валидности: внешнюю и внутреннюю. Внешняя валидность присуща тестам, не имеющим сходных компонентов с оцениваемой системой. Внутренняя валидность свойственна тем тестам, которые содержат в себе компоненты оцениваемой педагогической системы.

Мера валидности теста определяется несколькими способами, например, сравнением с тестом-эталонном и сопоставлением с более объективными показателями, полученными другими методами.

Сопоставление с объективными показателями предусматривает сравнение достижений испытуемого в тесте с результатами, полученными с помощью экспертов. В этом случае обычно берутся оценки, выставляемые испытуемому преподавателями-экспертами традиционным образом, без использования тестов, после чего результаты по заданиям теста и по оценкам экспертов коррелируются. Если получены однотипные изменения общих результатов, то считается, что тест обладает валидностью. Мера согласованности указывает на меру валидности. Следует заметить, что тест нельзя назвать валидным или невалидным без указания сферы его применения.

Тест считается содержательно валидным, если аспекты измерения соответствуют целям тестирования (аспектная полнота теста). При этом основное внимание в тесте (наибольшее число заданий) уделено наиболее важным из аспектов измерения (тест сбалансирован). Наконец, последовательность, в которой расположены задания и ответы к ним, логически обоснованы, такой тест представляет собой единое целое (конструктивно целостный). В педагогических тестах, разрабатываемых для контроля, диагностики знаний, валидность по содержанию играет решающую роль.

К традиционным тестам относятся тесты гомогенные и гетерогенные.

Гомогенный тест представляет собой систему заданий возрастающей трудности, специфической формы и определенного содержания - система, создаваемая с целью объективного, качественного, и эффективного метода оценки структуры и измерения уровня подготовленности учащихся по одной учебной дисциплине. Легко видеть, что в своей основе определение гомогенного теста совпадает с определением традиционного теста.

Гомогенные тесты распространены больше других. В педагогике они создаются для контроля знаний по одной учебной дисциплине или по одному разделу объемной учебной дисциплины (например, физики или информатики). В гомогенном педагогическом тесте не допускается использование заданий, выявляющих другие свойства. Наличие последних нарушает требование дисциплинарной чистоты педагогического теста. Ведь каждый тест измеряет что-то заранее определенное.

Например, тест по информатике измеряет знания, умения, навыки и представления испытуемых в данной науке. Одна из трудностей такого измерения заключается в том, что знания информатики изрядно сопряжены с математическими. Поэтому в тесте по информатике экспертно устанавливается уровень математических знаний, используемых при решении информативных заданий. Превышение принятого уровня приводит к смещению результатов; по мере превышения последние все больше начинают зависеть не столько от знания информатики, сколько от знания другой науки, математики. Другой важный аспект - стремление некоторых авторов включать в тесты не столько проверку знаний, сколько умение решать информативные задачи (читать программы и д.т.), вовлекая, тем самым, интеллектуальный компонент в измерение подготовленности по информатике.

Гетерогенный тест представляет собой систему заданий возрастающей трудности, специфической формы и определенного содержания - система, создаваемая с целью объективного, качественного, и эффективного метода оценки структуры и измерения уровня подготовленности учащихся по нескольким учебным дисциплинам. Нередко в такие тесты включаются и психологические задания для оценки уровня интеллектуального развития. Обычно гетерогенные тесты используются для комплексной оценки выпускника школ, оценки личности при приеме на работу и для отбора наиболее подготовленных абитуриентов при приеме в вузы. Поскольку каждый гетерогенный тест состоит из гомогенных тестов, интерпретация результатов тестирования ведется по ответам на задания каждого теста (здесь они называются

шкалами) и кроме того, посредством различных методов агрегирования баллов делаются попытки дать общую оценку подготовленности испытуемого.

4. Адаптивное тестирование.

Целесообразность адаптивного контроля вытекает из необходимости рационализации традиционного тестирования. Каждый учитель понимает, что хорошо подготовленному ученику нет необходимости давать легкие и очень легкие задания. Потому что слишком высока вероятность правильного решения. К тому же, легкие материалы не обладают заметным развивающим потенциалом. Симметрично, из-за высокой вероятности неправильного решения нет смысла давать трудные задания слабому ученику. Известно, что трудные и очень трудные задания снижают учебную мотивацию многих учащихся. Нужно было найти сопоставимую, в одной шкале, меру трудности заданий и меру уровня знаний. Эта мера была найдена в теории педагогических измерений. Датский математик Г. Раск назвал эту меру словом "логит". После появления компьютеров эта мера легла в основу методики адаптивного контроля знаний, где используются способы регулирования трудности и числа предъявляемых заданий, в зависимости от ответа учеников. При успешном ответе следующее задание ЭВМ подбирает более трудным, при неуспешном – легким. Естественно, этот алгоритм требует предварительного опробования всех заданий, определения их меры трудности, а также создания банка заданий и специальной программы. Использование заданий, соответствующих уровню подготовленности, существенно повышает точность измерений и минимизирует время индивидуального тестирования до, примерно, 5 - 10 минут. Адаптивное тестирование позволяет обеспечить компьютерную выдачу заданий на оптимальном, примерно 50%-ом уровне вероятности правильного ответа, для каждого ученика. В западной литературе выделяется три варианта адаптивного тестирования. Первый называется пирамидальным тестированием. При отсутствии предварительных оценок всем испытуемым дается задание средней трудности и уже затем, в зависимости от ответа, каждому испытуемому дается задание легче или труднее; на каждом шаге полезно использовать правило деления шкалы трудности пополам. При втором варианте контроль начинается с любого желаемого, испытуемым, уровня трудности, с постепенным приближением к реальному уровню знаний. Третий вариант - когда тестирование проводится посредством банка заданий, разделенных по уровням трудности. Таким образом, адаптивный тест представляет собой вариант автоматизированной системы тестирования, в которой заранее известны параметрами

трудности и дифференцирующей способности каждого задания. Эта система создана в виде компьютерного банка заданий, упорядоченных в соответствии с интересующими характеристиками заданий. Самая главная характеристика заданий адаптивного теста – это уровень их трудности, полученный опытным путем, что означает: прежде чем попасть в банк, каждое задание проходит эмпирическую апробацию на достаточно большом числе типичных учащихся интересующего контингента. Слова "интересующего контингента" призвано представлять здесь смысл известного в науке понятия более строгого понятия "генеральная совокупность". Распространенная у нас образовательная модель адаптивной школы Е.А. Ямбурга, исходит, по существу, из общих идей адаптивного обучения и адаптивного контроля знаний.

Истоки такого подхода можно проследить с момента возникновения педагогических трудов Коменского, Песталоцци и Дистервега, которых объединяют идеи природосообразности и гуманности обучения. В центре их педагогических систем был Ученик. Например, в малоизвестной у нас работе А. Дистервега "Дидактические правила" (Киев, 1870 г.) можно прочесть такие слова: "Преподавай сообразно природе... Учи без пробелов... Начинай преподавание с того, на чем остановился ученик... Прежде чем приступить к преподаванию, нужно исследовать точку исхода... Без знания того, на чем остановился ученик, невозможно порядочно обучить его".

Недостаточная информированность о реальном уровне знаний учеников и естественные различия в их способностях усвоить предлагаемые знания стали главной причиной появления адаптивных систем, основанных на принципе индивидуализации обучения. Этот принцип трудно реализуем в традиционной, классно-урочной форме.

Лекция 5. Тест как метод педагогического измерения

1. Классификации тестов
2. Процедура разработки тестов
3. Фактор времени в тестировании

1. Классификации тестов

Современное понимание тестов и тестирования можно представить в виде трех уровней.

Первый – «бытовой» уровень. Здесь тест понимается как набор вопросов с вариантами ответов, который стоит в одном ряду с кроссвордами,

головоломками и служит в большей степени для развлечения и удовлетворения познавательных интересов. Педагоги с таким пониманием тестирования считают тестирование очень ненадежным, ограниченным, а создание тестов простейшим делом.

Второй уровень понимания тестирования можно назвать «словарным». В этом понимании выделяются основные составляющие понятия тестирования. При этом не учитываются особенности процедуры создания, использования и анализа, специфичные для той или иной сферы применения. Для этого понимания характерны разночтения и противоречия в понятиях и определениях. Современное состояние развития тестологии находится именно на этом уровне: многие понятия до конца не определены, многие авторы трактуют по-разному одни и те же понятия и, в свою очередь, одно явление может иметь несколько названий. Это вполне естественный этап развития молодого научно-практического феномена, когда понятийный аппарат находится в стадии становления, а часть терминов и понятий иноязычны и в разных переводах приобретают отличающиеся значения.

Третий уровень понимания может быть назван научным. Он наиболее точен, учитывает особенности тестов и отражает требования к тестам, которые появляются в процессе развития и научного обоснования тестирования.

Все определения тестов подчеркивают то, что тест означает проверку, испытание, но это не простое установление факта наличия или отсутствия какого-либо качества или свойства, что в состав теста входят тестовые задания, что тест должен быть стандартизирован и что назначение теста – это выявление личностных особенностей или приращений.

Первое, на чем надо остановиться, рассматривая классификацию тестов – это два подхода, которые в настоящее время сложились в тестировании – тесты, ориентированные на критерий (критериально-ориентированные) и тесты, ориентированные на норму (нормативно-ориентированные). Появившись как разные подходы к анализу результатов тестирования,

отражающие разные основания для сравнения, сейчас эти два подхода определяют разницу на большинстве этапов создания теста.

При проверке минимально допустимого уровня знаний содержание заданий носит принципиально облегченный характер. Этот подход выработан для органов управления образованием, стоящих перед необходимостью в короткое время проверить состояние образования в большом количестве учебных заведений, и не позволить последним опускаться ниже предельно допустимого уровня требований (такими являются тесты на определение остаточных знаний). В таких случаях говорят о критериально-ориентированной интерпретации результатов.

При критериально-ориентированной ориентации требуется несколько меньшее число заданий, посредством которых определяется – что испытуемый знает, а что не знает из заданного стандарта. Иначе говоря, здесь ответы оцениваются относительно не всей области требуемых заданий, а только из области, ограниченной конкретным стандартом или уровнем (критерием) знаний.

При нормативно-ориентированной интерпретации результатов задача определения точного объема и структуры освоенного учебного материала отступает на задний план. В тест отбирается такое минимально достаточное количество заданий, которое позволяет сравнительно точно определить, образно говоря, не «кто что знает», а «кто знает больше».

Каждый из перечисленных видов выполняет свою функцию, используется для достижения определенной цели.

Понимание критериальной и нормативной ориентированности как разных подходов к интерпретации результатов можно найти у В.С. Аванесова.

С точки зрения интерпретации результатов: «Для критериально-ориентированной интерпретации вывод выстраивается вдоль логической цепочки: задания → ответы → выводы о соответствии испытуемого заданному критерию».

«Для нормативно-ориентированной ориентации вывод достраивается рейтингом: задания → ответы → выводы о знаниях испытуемого → рейтинг, понимаемый как вывод о месте или ранге испытуемого».

Описание этих двух подходов в своей схематичной форме вполне отражает смысл разницы в интерпретации.

Кроме приведенных подходов, тесты можно классифицировать по целому ряду оснований.

1. *По процедуре создания* могут быть выделены стандартизированные и не стандартизированные тесты.

Стандартизируются процедура и условия проведения тестирования, способы обработки и интерпретации результатов, которые должны привести к созданию равных условий для испытуемых и минимизировать случайные ошибки и погрешности, как на этапе проведения, так и на этапе обработки результатов и интерпретации данных.

2. *По средствам предъявления:*

– бланковые (тесты «бумага и карандаш») с использованием тестовых бланков или тетрадей, в которых находятся тестовые задания и в которых испытуемый фиксирует результаты;

– предметные – в которых необходимо манипулировать материальными объектами, результативность выполнения этих тестов зависит от скорости и правильности выполнения заданий;

– аппаратные – тесты с использованием устройств для изучения особенностей внимания, восприятия, памяти и мышления;

– практические – появившиеся относительно недавно, эти тесты сходны с известными у нас лабораторными работами, однако снабженные соответствующими инструкциями и имеющие тестовое оснащение;

– компьютерные.

Кроме этого в рамках компьютерного тестирования можно выделить еще адаптивные тесты – такие, в которых можно изменять параметры тестирования, адаптируя (приспосабливая) тест к разным условиям. В адаптивной системе тестирования (АСТ), разработанной Центром тестирования профессионального образования (ЦТПО), есть возможность изменять алгоритм предъявления заданий (строго последовательный, случайного выбора или по возрастанию сложности), изменять время, отводимое на каждое задание или весь тест, изменять способы оценивания и шкалу отметок в зависимости от процента выполненных заданий, изменять количество заданий в самом тесте. В этом есть смысл, поскольку

большинство тестов предполагает предъявление конечного набора вопросов испытуемому и не предполагает зависимости от ответа на предыдущие вопросы.

Каждый из способов предъявления имеет свои плюсы и минусы.

Для нужд образования подходит любой из этих способов, при этом нужно помнить о том, что предъявляя один и тот же тест в разных формах, мы получим разные результаты. Нельзя сравнивать результаты тестирования, полученные в результате разных способов предъявления.

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

– тесты интеллекта, выявляющие особенности последнего;

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

– тесты достижений.

4. *По характеру действий*:

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

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

5. *По ведущей ориентации*:

– тесты скорости, содержащие простые задачи, время решения которых ограничено настолько, что ни один испытуемый не успевает решить все задачи в заданное время;

– тесты мощности или результативности, включающие трудные задачи, время решения которых либо вовсе не ограничено, либо мягко лимитировано. Оценке подлежит успешность и способ решения задачи;

– смешанные тесты, которые объединяют в себе черты двух вышеперечисленных. В таких тестах представлены задачи различного уровня сложности, от самых простых до очень сложных. Время испытания в данном случае ограничено, но достаточное для решения большинства предлагаемых задач определенной группой обследуемых. Оценкой в данном случае служат как скорость выполнения заданий (количество выполненных заданий), так и правильность решения.

Эти тесты наиболее часто применяются на практике, к ним относится большинство тестов учебных достижений.

6. По степени однородности задач:

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

– гетерогенные (многомерные), имеющие несколько шкал, которые позволяют оценить разнообразные характеристики личности и включают задания, отличающиеся и по характеру, и по содержанию.

7. По широте использования (только для тестов в системе образования).

Приведенная ниже классификация является сугубо эмпирической, однако практически достаточно значимой, поскольку в зависимости от широты предполагаемого использования зависит и уровень требований к качеству инструмента и важность этой разработки, и время подготовки:

– для использования преподавателем;

– для использования группой преподавателей или администрацией образовательного учреждения;

– для целей отбора и формирования групп;

– для аттестации обучающихся.

2. Процедура разработки тестов

Для тестов учебных достижений существуют общие правила и этапы разработки.

В зависимости от тех целей, которые стоят перед разработчиками, процедура разработки тестов имеет некоторые особенности, однако перечень этапов создания тестового инструментария одинаков для всех видов тестов.

1. Определение целей тестирования.
2. Определение ресурсных возможностей разработчиков.
3. Отбор содержания учебного материала.
4. Конструирование технологической матрицы и ее экспертиза.
5. Составление тестовых заданий и их экспертиза.
6. Построение выборки для апробации заданий и тестов.
7. Компоновка заданий для апробации.
8. Апробация тестовых заданий.
9. Определение и расчет показателей качества тестовых заданий.
10. Отбраковка заданий и составление теста.
11. Апробация теста.
12. Определение и расчет показателей качества теста.
13. Составление окончательного варианта теста.
14. Стандартизация теста.
15. Нормирование теста.
16. Оснащение теста.

Этот список полный, поскольку в нем представлены все этапы создания тестов. Создание теста начинается с определения целей тестирования. Конечно, для тестов, которые предполагается использовать для сравнения результатов между несколькими классами в одном образовательном учреждении и тестами, предназначенными для итоговой аттестации обучающихся, существует значительная разница.

Лекция 6. Технология составления тестов

1. Технологическая матрица как модель педагогического тестирования
2. Состав и характеристика тестовых заданий
3. Проблемы составления тестовых заданий
4. Использование невербальных материалов в тестировании

1. Технологическая матрица как модель педагогического тестирования

Определив цели составления тестов, уточнив подходы и выбрав уровень использования, разработчик определяет необходимые этапы создания тестового инструментария. После этого составляется модель педагогического тестирования. Такая модель может быть представлена в виде технологической матрицы. Эта

матрица может носить разные названия: содержательная решетка, сетка, организационно-деятельностная матрица, тестовая решетка, таблица содержания и другие. Однако смысл у нее остается один. Технологическая матрица задает содержание, которое будет отобрано для проверки, и важность того или иного элемента содержания. Она может содержать уровни достижений, которые будут проверены, их соотношение, соответствие стандарту и некоторые другие компоненты.

При составлении матрицы для любого предмета тестирования разработчик обязан убедиться, что весь предмет охвачен предлагаемыми вопросами. Содержание предмета должно полностью покрываться матрицей по всем темам. Если же имеет место тестирование по отдельным подтемам, то и в этом случае необходимо, чтобы вся подтема была охвачена вопросами теста. В случае если вопрос или часть вопроса не соответствует теме, или не полностью ясен в рамках данной темы, от вопроса следует воздержаться.

Таким образом, фиксируется требование широты теста, полного учета всех разделов предмета, который находит выражение в матрице.

Для тестов, ориентированных на критерии (критериально-ориентированных), отбор содержания теста является самым важным этапом его создания, так как для принятия решения о достижении данной цели обучения, например, стандарта, необходимо достаточно точно и полно описать содержание стандарта и выразить его совокупностью заданий, которая была бы представительной для этой цели. Поэтому главной проблемой в разработке тестов, используемых для оценки достижения образовательных стандартов, является достижение соответствия содержания стандарта и содержания теста.

В самом простом случае технологическая матрица может описывать только предметы, предметные области или отдельные темы разного уровня обобщения, которые должны войти в тест, и определяет соотношение заданий в тесте.

Важно проследить, чтобы общее количество вопросов было распределено по темам таким образом, чтобы оно отражало их важность, объем, и количество учебного времени.

Основанием для распределения вопросов в технологической матрице могут служить три фактора – важность, объем, и количество учебного времени.

Например, для тестов, предназначенных для входного контроля и итоговой аттестации, важность темы будет носить разный характер. В первом случае материал будет отбираться, исходя из важности для продолжения образования в конкретном вузе или просто для продолжения образования, во втором случае важность будет определяться в рамках всех тем, которые изучались в процессе обучения. Соответственно, между технологическими решетками этих тестов может быть существенная разница. В ряде стран этой разницы не существует, поскольку итоговая аттестация выпускников проводится независимыми службами, организованными на базе отдельных вузов или их ассоциаций, но на построение такой системы им потребовались десятилетия согласований и уточнений.

Для практической работы распределение заданий между темами следует увязать с целями предстоящего тестирования и в том случае, если речь идет о промежуточной аттестации, следует преимущественно ориентироваться на время (объем) темы. В том случае, если тесты готовятся для итоговой аттестации, следует ориентироваться на важность темы для дальнейшего образования.

Особенностью матрицы является планирование не только заданий, но и количества баллов, которыми могут быть оценены эти задания. (Вопрос о количестве баллов на каждое задание требует отдельного обсуждения, так как в последнее время укрепилось мнение – присваивать каждому верно выполненному заданию только один балл.).

Более сложные технологические матрицы содержат две шкалы и оформляются в виде таблицы.

Содержание горизонтальной строки матрицы, как правило, определяют предметы, разделы, учебные темы, разделы учебных тем.

Вертикальная составляющая составляется, как правило, на основе той или иной таксономии учебных целей. Идея второй шкалы реализуется таким образом: устанавливается взаимно однозначное соответствие конкретного тестового задания и уровня (или свойства, умения и пр.) той графы матрицы, к которой отнесено это задание.

К первой относят знания, различные уровни его усвоения. Ко второй – умения со своей иерархией подцелей. И к третьей – отношения, интересы, склонности, ориентации.

Наиболее известны в мировой литературе несколько классификаций знания. С ними связаны категории учебных целей.

Авторы международных сравнительных исследований при создании тестов достижений используют содержательно-деятельностные матрицы, основанные на таксономии Блума.

Основные категории учебных целей (по Блуму).

1. ЗНАНИЕ. Эта категория означает запоминание и воспроизведение изученного материала. Речь может идти о различных видах содержания – от конкретных фактов до целостных теорий. Общая черта этой категории – припоминание соответствующих сведений.

СТУДЕНТ:

- знает употребляемые термины;
- конкретные факты;
- методы и процедуры;
- основные понятия;
- правила и принципы и др.

2. ПОНИМАНИЕ. Показателями способности понимать значение изучаемого материала могут служить преобразование (трансляция) материала из одной формы выражения в другую, перевод его с одного «языка» на другой (например, из словесной формы в математическую). В качестве показателя понимания может также выступать интерпретация студентом (объяснение, краткое изложение) или же предположение о дальнейшем ходе явлений, событий (предсказание последствий, результатов). Такие учебные результаты превосходят простое запоминание.

СТУДЕНТ:

- понимает факты, правила и приемы;
- интерпретирует словесный материал;
- интерпретирует схемы, графики, диаграммы;
- преобразует словесный материал в математическое выражение;

– предположительно описывает будущие последствия, вытекающие из имеющихся данных и др.

3. *ПРИМЕНЕНИЕ*. Эта категория обозначает умение использовать изученный материал в конкретных условиях и новых ситуациях. Сюда входит применение правил, методов, понятий, законов, принципов, теорий. Соответствующие результаты обучения требуют более высокого уровня владения материалом, чем понимание.

СТУДЕНТ:

- использует понятия и принципы в новых ситуациях;
- применяет законы, теории в конкретных практических ситуациях;
- демонстрирует правильное применение метода или процедуры и др.

4. *АНАЛИЗ*. Эта категория обозначает умение разбирать материал на составляющие так, чтобы ясно выступала его структура. Сюда относится вычленение частей целого, выявление взаимосвязей между ними, осознание принципов организации целого. Учебные результаты характеризуются при этом более высоким интеллектуальным уровнем, чем понимание и применение, поскольку требуют осознания как содержания учебного материала, так и его внутреннего строения.

СТУДЕНТ:

- выделяет скрытые (неявные) предположения;
- видит ошибки и упущения в логике рассуждений;
- проводит различия между фактами и следствиями;
- оценивает значимость данных и др.

5. *СИНТЕЗ*. Эта категория обозначает умение комбинировать элементы, чтобы получить целое, обладающее новизной (или заданными свойствами). Таким новым продуктом может быть сообщение (выступление, доклад), план действий или совокупность обобщенных связей (схемы для упорядочивания имеющихся сведений). Соответствующие учебные результаты предполагают деятельность творческого характера с акцентом на создание новых схем и структур.

СТУДЕНТ:

- пишет небольшие творческие сочинения;
- предлагает план проведения эксперимента;

– использует знания из различных областей, чтобы составить план решения той или иной проблемы и др.

6. *ОЦЕНКА*. Эта категория обозначает умение оценивать значение того или иного материала (утверждений, художественных произведений, исследовательских данных). Суждения студента должны основываться на четких критериях. Критерии могут быть как внутренними (структурными, логическими), так и внешними (соответствие намеченной цели). Критерии могут определяться самим студентом или же задаваться ему извне (преподавателем). Данная категория предполагает достижение учебных результатов по всем предшествующим категориям плюс оценочные суждения на ясно очерченных категориях.

СТУДЕНТ:

- оценивает логику построения материала в виде письменного текста;
- оценивает соответствие выводов имеющимся данным;
- оценивает значимость того или иного продукта деятельности, исходя из внутренних критериев;
- оценивает значимость того или иного продукта деятельности, исходя из внешних критериев и др.

Другие авторы предлагают иной перечень видов знаний, который сформулирован исключительно для задач педагогического измерения:

- знание названий, имен;
- знание смысла названий имен;
- фактуральные знания;
- знание определений;
- сравнительные, сопоставительные знания;
- классификационные знания;
- знание противоположностей, противоречий, синонимичных и антонимичных объектов;
- ассоциативные знания;
- причинные знания, знание причинно-следственных отношений, знание оснований и принципов классификации;
- процессуальные, алгоритмические, процедурные знания;
- технологические знания;

- обобщенные, системные знания;
- оценочные знания;
- вероятностные знания;
- абстрактные знания;
- структурные знания;
- методологические знания.

Перечисленные виды знаний не претендуют на полную классификационную систему и поэтому допускают возможность заметного расширения, замены одних видов другими, объединения их в различные группы.

В соответствии с данной системой каждое задание теста предназначено для проверки овладения обучающимися определенными умениями, характеризующими отдельные компоненты познавательной деятельности на материале различных разделов, тем конкретной учебной дисциплины.

Таксономия Б. Блума и ее развитие служит основой для разработки современных тестов учебных достижений, однако ее реально используемая часть определяется целями тестирования.

Учебная программа составлена на основании ГОС ВПО 2005 г. для
специальности 050303 Иностранный язык

Программу составил(и)

1. Разумова М.В. ассистент
Ф.И.О., учёная степень, ученое звание

**Настоящая программа не может быть воспроизведена ни в какой
форме без предварительного письменного разрешения кафедры-
разработчика программы.**

Программа одобрена на заседании кафедры английского языка

от «__» _____ 2007 года протокол № ____

Заведующий кафедрой
английского языка

Толмачева

Т.А.Полукарова

Программа одобрена учебно-методическим советом факультета

«__» _____ 2007 г.

Председатель учебно-методического совета *Шинин*

Программа одобрена учебно-методическим управлением университета

«31» ~~05~~ 2007 г.

Начальник учебно-методического управления *Шинин*

УМК разработал (и)

1. Разумова М.В.
2. _____
3. _____
4. _____
5. _____

УМК одобрен отделом менеджмента качества университета

« 28 » апреля 2008 г.

Начальник отдела менеджмента качества,
Д.э.н., профессор

Скворцова

В.А. Скворцова