

ПЕДАГОГИЧЕСКОЕ ОБРАЗОВАНИЕ

Н. Н. Самылкина

**СОВРЕМЕННЫЕ
СРЕДСТВА
ОЦЕНИВАНИЯ
РЕЗУЛЬТАТОВ ОБУЧЕНИЯ**



БИНОМ

ПЕДАГОГИЧЕСКОЕ ОБРАЗОВАНИЕ

Н. Н. Самылкина

СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

2-е издание
(электронное)



Москва
БИНОМ. Лаборатория знаний
2012

УДК 004.9
ББК 32.97
С17

Серия основана в 2007 г.

Самылкина Н. Н.

С17 Современные средства оценивания результатов обучения [Электронный ресурс] / Н. Н. Самылкина. — 2-е изд. (эл.). — М. : БИНОМ. Лаборатория знаний, 2012. — 172 с. : ил. — (Педагогическое образование).

ISBN 978-5-9963-1000-5

Книга представляет собой курс лекций для новой общепрофессиональной дисциплины «Современные средства оценивания результатов обучения» для подготовки будущих учителей информатики в области теории и практики педагогических измерений с использованием компьютерных технологий. Она также может быть полезна при изучении раздела «Использование современных информационных и коммуникационных технологий в учебном процессе» и курса «Психолого-педагогическая диагностика на основе компьютерного тестирования» для учителей информатики.

Для студентов педагогических вузов, учителей, методистов и администрации образовательных учреждений, а также для аспирантов и соискателей при подготовке ими диссертаций (при обработке результатов педагогического эксперимента).

**УДК 004.9
ББК 32.97**

**По вопросам приобретения обращаться:
«БИНОМ. Лаборатория знаний»**

Телефон: (499) 157-5272

e-mail: binom@Lbz.ru

<http://www.Lbz.ru>, <http://metodist.Lbz.ru>

ISBN 978-5-9963-1000-5

© БИНОМ. Лаборатория знаний,
2012

СОДЕРЖАНИЕ

Введение в проблему педагогических измерений	5
Актуальные вопросы оценки качества образования	11
Таксономия образовательных целей и результаты обучения	17
Педагогический тест как объективный способ оценивания	27
Основные этапы конструирования педагогического теста	37
Общие подходы к оценке достижения требований к уровню подготовки выпускников основной и средней школы	49
Классическая теория тестов. Достоинства и недостатки	66
Немного о современной теории создания тестов.	79
Методы шкалирования и интерпретации результатов тестирования	88
Использование программных средств автоматизации хода выполнения и обработки результатов тестирования	94
Инструментальные программные системы для разработки и использования тестовых заданий	108
Литература.	113
Приложение 1. Кодификатор элементов содержания по информатике для составления контрольных измерительных материалов (КИМ) единого государственного экзамена.	116
Приложение 2. Спецификация экзаменационной работы по информатике для выпускников XI (XII) классов общеобразовательных учреждений 2005 г.	122

Приложение 3. Единый государственный экзамен по информатике. Демонстрационный вариант 2005 г.	138
Приложение 4. Система оценивания заданий с развернутым ответом.	158

ВВЕДЕНИЕ В ПРОБЛЕМУ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

- *Основы теории педагогических измерений*
- *Переменные измерения и их взаимосвязь*
- *Систематические и случайные ошибки измерений, истинный балл*
- *Доверительный интервал*
- *Точечные и доверительные оценки*

Начиная разговор о педагогических измерениях в образовании, необходимо познакомиться с некоторыми понятиями, которые будут использоваться в дальнейшем.

Контроль — важнейший компонент практической деятельности педагога, позволяющий выявить достоинства и недостатки новых методов обучения, установить взаимосвязи между планируемыми, реализуемыми и достигнутыми уровнями образования, оценить достижения ученика, выявить пробелы в его знаниях и умениях, определить эффективность работы педагогического коллектива в целом (для принятия управленческих решений) и многое другое. С другой стороны, контроль — это объект теоретических исследований, в настоящее время тесно связанный с изменениями подходов к определению и оценке качества образования.

Понятие «*педагогический контроль*» можно рассматривать как единую дидактическую и методическую систему проверочной деятельности. Применительно к повседневному учебному процессу, проверочная деятельность направлена на выявление и оценку результатов учебной деятельности школьников или студентов. В действительности, несмотря на существующие общие критерии выставления, оценка остается субъективным мерилом учебных достижений.

Возникает проблема объективного измерения результатов учебной деятельности: «Измерение предполагает проведение объективного количественного сопоставления оцениваемого свойства ученика с некоторым эталоном, принятым в качестве единицы измерения. При педагогическом измерении роль оце-

ниваемого свойства отводится знаниям, умениям и навыкам ученика, так традиционно сложилось в отечественной практике, — а вместо единицы измерения используются контрольные задания или части заданий по проверяемому содержанию предмета» [20].

Существует и общее определение: «Измерение — процесс сопоставления оцениваемых характеристик объекта на числовую ось» [13] — настолько общее, что оно не отражает процедуру измерения и характеристики используемой шкалы. «Измерение — оценка (как процедура), проводимая с использованием понятия шкалы вполне определенных способностей для количественной оценки поведения или характеристики, например, математических способностей» — таково современное понимание *измерения* в международном сообществе специалистов [5].

Итак, проблема педагогических измерений ставит перед нами три вопроса:

- Для чего измерять?
- Что измерять?
- Чем измерять?

Проще всего ответить на первый вопрос, связанный с *поставкой целей контроля*. Для определения степени владения базовыми знаниями, умениями и навыками, необходимыми для начала обучения, служит *входной контроль*. Для выявления пробелов в знаниях в процессе изучения нового материала предназначен *текущий, или тематический, контроль*. Для оценки качества знаний после прохождения всего предметного курса или итоговой аттестации выпускников основной или средней (полной) школы служит *итоговый контроль*.

Второй же и третий вопросы взаимосвязаны и более сложны. В педагогическом контроле нет выделенного объекта измерения, нет точки отсчета («абсолютного нуля»), нет единиц измерения. Все это создает определенные трудности, которые специалисты пытаются решать, опираясь на *математическую теорию измерений*.

Понятие «измерения» специалистами трактуется следующим образом. **Измерения** — это конструирование числовой функции, осуществляющей *изоморфное* отображение эмпирической структуры в некую числовую шкалу, где изоморфизм определяет условия взаимно-однозначных соответствий при отображении. Отсюда главная задача измерения: поиск такой числовой структуры, которая будет изоморфна эмпирическим референтам (измеряемым характеристикам) [20].

Для педагогических измерений необходимо определить следующие параметры: *объект измерения (переменные измерения)*, *измерительная процедура*, *средства измерения* и *шкала*.

Главная проблема при этом — выделить **переменные измерения**, которые в педагогике не носят явного характера. Такие переменные называют *латентными* (неявными), т. е. скрытыми от возможности непосредственного измерения. Такая переменная может быть выражена в виде совокупности заданий и проявляться только в эмпирических результатах выполнения теста, путем анализа; точность такого измерения, разумеется, ниже, чем в точных науках. В качестве переменной измерения могут выступать дискретные и непрерывные характеристики: из них дискретные не имеют промежуточных значений между двумя точками, тогда как непрерывные могут принимать любые значения на всем заданном интервале.

При измерениях используют следующие обозначения (рис. 1). На некоторой числовой оси выделяют интервал, называемый *доверительным*, в котором находятся наблюдаемый балл (X) и истинный балл (T).

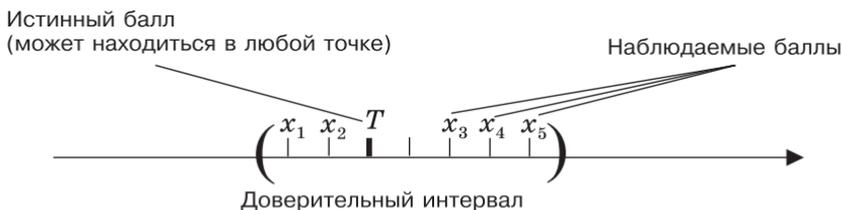


Рис. 1

Доверительным интервалом специалисты называют такие пределы выборочного распределения, в которых можно с заранее определенной вероятностью ожидать нахождения значения *генеральной совокупности* (т. е. результатов работы всех возможных тестируемых).

Следовательно, все точки, лежащие в доверительном интервале, можно считать *доверительными оценками*.

Интервальное исследование — это получение некоторого интервала вокруг наблюдаемого балла (X), границы которого определяются ошибкой измерения и внутри которого лежит истинный балл (T). Постепенно сужая интервал, мы приближаемся к истинному баллу. Математические способы позволяют от X

максимально близко подойти к T , но в классической и современной теории тестирования это реализуется по-разному.

Процедура измерения — совокупность операций, позволяющая перейти от эмпирических референтов (заданий) к числовым оценкам измеряемых характеристик.

Измерительный инструмент состоит из двух компонентов: *теста* и *шкалы* для фиксации результатов измерения. **Педагогический тест** — это инструмент, состоящий из квалиметрически (с использованием различных методов определения качества) выверенной системы тестовых заданий, стандартизированной процедуры проведения и заранее спроектированной технологии обработки и анализа результатов, предназначенный для измерения качеств и свойств личности, измерение которых возможно в процессе систематического обучения [9].

Оставим пока понятие «педагогический тест» для отдельного обсуждения и рассмотрим понятие **шкалы**.

«Шкала — числовая система, в которой отношения между различными свойствами изучаемых явлений, процессов переведены в свойства того или иного множества, как правило — множества чисел» [13].

Существуют шкалы отношений, интервальные шкалы, порядковые (ранговые) и номинальные (шкалы наименований). Здесь шкалы были перечислены в порядке убывания их *мощности*; в тестологии же, описывая использование тех или иных шкал, говорят об «*уровнях измерений*»:

1. **Шкала наименований** представляет собой *номинальный уровень*, не связана с понятием «величина» (т. е. является неметрической) и используется для того, чтобы отличать один объект от другого (фамилии учеников, номера телефонов и пр.). Над элементами такой шкалы не допускается никаких арифметических действий, — возможен лишь подсчет количества объектов с совпадающими признаками. Для школьной практики это означает возможность подсчета количества отличников, «хорошистов» или двоечников и сравнение этих групп по количеству учащихся.
2. **Порядковая шкала (ранговая, шкала упорядоченной классификации)** представляет собой *ординальный уровень*, когда объекты измерения разбиваются на группы, соответствующие определенной точке шкалы. Такая шкала только упорядочивает объекты, приписывая им те или

иные *ранги*. Относительно же значений этой шкалы нельзя говорить ни о том, во сколько раз измеряемая величина больше или меньше другой, ни о том, на сколько она меньше или больше; не допускается никаких арифметических действий со значениями, — допустима лишь замена чисел, характеризующих величину того или иного признака, другими, не нарушающими ранговый порядок объектов.

Для корректного использования этой шкалы в школе также следует помнить, что с ее помощью можно подсчитать лишь количество учащихся, получивших различные оценки. Сравнить же качество их подготовки путем вычисления среднего балла недопустимо!

3. **Интервальная шкала** (*интервальный уровень измерения*) основана на сравнении различий между объектами по величинам измеряемых признаков или свойств и характеризуется тем, что не имеет естественного начала отсчета и единиц измерения. Начало отсчета и интервалы здесь выбираются по договоренности относительно выбранного параметра, т. е. шкалу можно растягивать и сжимать. Интервальная шкала позволяет рассчитать среднее арифметическое и среднее квадратичное отклонение и коэффициенты корреляции. Современная теория тестов позволяет преобразовать результаты тестирования в интервальную шкалу.

Для школьной практики это новый формат оценивания, который может прижиться при активном внедрении тестовых форм контроля качества подготовки выпускников.

4. **Шкала отношений** — самая мощная шкала. Она позволяет оценивать, во сколько раз один измеряемый объект больше (меньше) другого объекта, принимаемого за *эталон*. Шкала отношений имеет естественное начало отсчета, но в ней нет естественной единицы измерений. Шкалами отношений измеряют почти все физические величины, но ею нельзя пользоваться в социальных измерениях.

Говоря о качестве педагогических измерений, прежде всего подразумевают их *объективность*. Существует несколько градаций объективности.

Процедурная объективность отождествляется с объективностью процедуры тестирования (отсутствует педагог, одинаковые условия для всех, единые критерии проверки и пр.).

Классическая объективность описывается формулой:

$$X \text{ (наблюдаемый балл)} = T \text{ (истинный балл)} + E \text{ (ошибка измерения)}.$$

Здесь T (истинный балл) — это оценка параметра испытуемого, не зависящего от средств и способа измерения, но меняющегося в процессе обучения. T постоянно в момент измерения. Чем меньше ошибка измерения, тем измерение объективнее, где T — абсолютно объективная оценка.

Таким образом, классическая объективность связана с величиной ошибки измерения, с содержанием и процедурными проблемами. При этом совокупность заданий должна работать на оценивание переменной измерения, которая должна быть доминирующей.

АКТУАЛЬНЫЕ ВОПРОСЫ ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАНИЯ

- *Содержание понятия «качество образования» в отечественной и зарубежной образовательной практике*
- *Первые шаги становления общероссийской системы оценки качества образования*
- *Представления о качестве подготовки учащихся общеобразовательных учреждений*
- *Показатели качества*
- *Принципы отбора показателей качества*
- *Международные программы по оценке образовательных достижений учащихся*

По мнению как ученых, так и политиков, качество образования и интеллектуальных ресурсов становится главным геополитическим фактором в мире [1].

В России повышенное внимание к проблеме качества образования обусловлено происходящими в последнее десятилетие политическими и социально-экономическими изменениями и связанными с ними процессами модернизации российского образования. Отечественная система образования призвана решить ряд принципиально новых задач:

- обеспечить новое качество образования за счет развития лично-ориентированного обучения;
- сохранить единство образовательного пространства, создавая при этом условия для развития системы вариативного образования;
- обеспечить равный доступ к получению качественного образования для всех социальных слоев населения.

Изменение целей и задач образования, а также условий его получения тесно связано с изменением подходов к определению и оценке качества образования. В настоящее время становится все более ясно, что образовательную ценность представляет не только система знаний человека, но и освоение способов их получения, способность выявлять потребность в новых знаниях,

умение быстро и эффективно, самостоятельно или во взаимодействии с другими людьми восполнять имеющиеся пробелы. Такой подход не позволяет отождествлять качество образования с качеством знаний (обученности): оценка качества образования подразумевает оценку качества образовательных достижений обучающихся и оценку качества образовательного процесса. Таким образом, под **качеством образования** понимается интегральная характеристика системы образования, отражающая степень соответствия реально достигаемых образовательных результатов нормативным требованиям, социальным и личностным ожиданиям [21].

В ближайшее время наиболее актуальной государственной задачей в России является создание *общероссийской системы оценки качества образования (ОСОКО)* как совокупности организационных и функциональных структур, обеспечивающей основанную на единой концептуально-методологической основе оценку образовательных достижений и определение личностных качеств детей и взрослых граждан, а также выявление факторов, влияющих на образовательные результаты.

Создание общероссийской системы оценки качества образования в России должно способствовать достижению следующих *целей*:

- обеспечение объективности и справедливости при приеме в образовательные учреждения;
- сохранение единого образовательного пространства;
- повышение уровня информированности потребителей образовательных услуг при принятии решения о продолжении образования или трудоустройстве;
- развитие государственной аттестационной службы;
- принятие обоснованных управленческих решений по повышению качества образования.

Для управления качеством образования в нашей стране, прогнозирования развития образования, определения политики в области образования необходимо иметь единую картину о результатах образования и о факторах, влияющих на эти результаты. Оценка качества образования должна быть нацелена на фиксацию не только некоторого «мгновенного» состояния системы, но и динамики ее развития.

При создании ОСОКО необходимо учесть, что оценка качества образования (как процесс) включает в себя несколько *этапов*, составляющих некий инвариант:

- выделение показателей качества (они могут различаться на различных этапах);
- определение критериев оценки (на основе чего будет оцениваться каждый показатель);
- формирование шкалы уровней достижения критериев качества;
- разработка инструментария для измерения показателей;
- организация сбора, обработки, анализа и интерпретации полученных данных;
- предоставление и распространение обобщенной информации для разных категорий пользователей (учащиеся и их родители, педагогические коллективы, органы управления образованием, работодатели);
- организация широкого общественного и профессионального обсуждения;
- подготовка рекомендаций по совершенствованию системы образования и принятие управленческих решений на различных уровнях.

Можно выделить следующие основные *принципы отбора показателей* для оценки качества образования:

- ориентация на требования внешних пользователей;
- учет потребностей системы образования;
- минимизация системы показателей с учетом потребностей разных уровней управления системой образования;
- надежность и технологичность используемых показателей;
- оптимальность использования источников первичных данных для определения показателей качества и эффективности образования (с учетом возможности их многократного использования и экономической обоснованности);
- иерархичность системы показателей (для удобства пользователей на различных уровнях системы образования);
- соответствие международным аналогам;
- соблюдение морально-этических норм в отборе показателей [21].

В настоящее время Россия участвует в нескольких международных сравнительных исследованиях, из которых наибольший интерес представляют исследования *PISA* и *TIMSS*. Участие России в международных сравнительных исследованиях качества образования на современном этапе развития общества имеет большое значение: они позволяют оценить состояние образова-

ния в международном контексте, выявить сильные и слабые стороны российского образования, наметить пути более эффективного достижения поставленных целей.

Получение объективной информации о качестве образования в стране позволяет принимать обоснованные решения с целью обеспечения эффективности образования, его эквивалентности не только в пределах России, но и в мире, конкурентоспособности российских специалистов на мировом рынке труда.

Исследование TIMSS (Trends in Mathematics and Science Study) — мониторинговое исследование качества математического и естественнонаучного образования.

Исследование PISA (Programme for International Student Assessment) — международная программа для оценки компетентности выпускников основной школы.

Сравнивая исследования TIMSS и PISA, можно сказать, что каждое из них имеет свои особенности, но вместе они дополняют друг друга. Для определения объективности и достоверности полученные результаты международных исследований сопоставляют с данными, полученными в ходе проведения единого государственного экзамена (ЕГЭ), а также многочисленных мониторинговых исследований. Одновременное проведение двух этих международных исследований дает более полную информацию о системе образования, оценивая разные его стороны — уровень освоения школьных программ (оценивается общеобразовательная подготовка по математике и естествознанию учащихся IV и VIII классов) и способность применять полученные в школе знания и умения в ситуациях, приближенных к реальной жизни (оценивается математическая и естественнонаучная грамотность, грамотность чтения и понимания текстов, умение решать проблемы непредметного содержания).

Исследование TIMSS позволяет получить информацию, необходимую для совершенствования преподавания математики и естественнонаучных предметов в школе.

Исследование PISA в большей степени оценивает готовность выпускников основной школы к жизни в современном обществе и сфокусировано на аспектах, связанных с непрерывным образованием.

Оба исследования отличаются подходами к формированию выборки (множества отобранных для исследования единиц). В исследовании TIMSS выборка учащихся формируется в соответствии с годами обучения в школе (IV и VIII класс), а в исследовании PISA — по возрастному признаку (учащиеся 15-летнего возраста, независимо от класса и программы обучения). Поэто-

[. . .]

Учебное электронное издание

Серия: «Педагогическое образование»

Самылкина Надежда Николаевна

**СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ
ОБУЧЕНИЯ**

Ведущий редактор *Д. Усенков*

Корректор *Е. Клитина*

Компьютерная верстка: *Л. Катуркиной*

Подписано 20.03.12. Формат 60×90/16.

Усл. печ. л. 11.

Издательство «БИНОМ. Лаборатория знаний»

125167, Москва, проезд Аэропорта, д. 3

Телефон: (499) 157-5272

e-mail: binom@Lbz.ru

<http://www.Lbz.ru>, <http://metodist.Lbz.ru>

Системные требования: процессор *Intel* с тактовой частотой от 1,3 ГГц и выше; операционная система *Microsoft Windows XP, Vista* или *Windows 7*; от 256 Мб оперативной памяти; от 260 Мб свободного пространства на жестком диске; разрешение экрана не ниже 1024×768; программа *Adobe Reader* не ниже X.

