

Национальная оценка учебных достижений

Во всем мире эффективная система образования играет ключевую роль в формировании политики, направленной на оптимизацию развития человеческого капитала. Серия «Национальная оценка учебных достижений» знакомит читателя с основными моментами, касающимися оценки образовательных достижений учащихся на разных уровнях обучения. В пяти книгах рассматривается широкий круг вопросов, начиная с политических аспектов и заканчивая моментами, которые касаются проектирования и проведения оценки, такими, как разработка тестов и дизайн анкет, формирование выборки, организация сбора данных, метод статистического анализа, чистка данных, написание отчета и использование результатов национальной оценки для совершенствования качества образования.

Вторая книга серии, «Разработка тестов и анкет для национальной оценки учебных достижений», посвящена формированию надежного инструментария национального оценивания достижений учащихся. Она вводит читателей в сложную систему мероприятий, связанных с конструированием тестов, и затрагивает такие важные аспекты, как разработка структуры теста, написание заданий с множественным выбором и свободно конструируемым ответом, предварительное тестирование, изготовление тестовых буклетов и оценка результатов тестирования вручную.

Во второй части рассматривается процесс разработки анкет: написание вопросов, кодирование ответов и соответствие данных анкет и тестов. В заключительной части освящаются вопросы, связанные с руководством для администраторов по предъявлению тестов, отбором администраторов и информированием школ о национальном оценивании.

Книга будет полезной для лиц, чья деятельность связана с проведением оценочных мероприятий, в особенности специалистам-предметникам, самостоятельно разрабатывающим и/или использующим тесты и другие технологии педагогических измерений.



CICED



Пру Андерсен, Джордж Морган

КНИГА 2

Разработка тестов и анкет для национальной оценки учебных достижений



Национальная оценка учебных достижений

45411

Пру Андерсен, Джордж Морган

Разработка тестов
и анкет для
национальной
оценки учебных
достижений

КНИГА 2



THE WORLD BANK

CICED

**Разработка
тестов и анкет
для национальной
оценки
учебных
достижений**

**Developing Tests
and Questionnaires
for a National
Assessment of
Educational
Achievement**

National Assessments of Educational Achievement

VOLUME 2

Developing Tests and Questionnaires for a National Assessment of Educational Achievement

Prue Anderson and George Morgan

**Vincent Greaney and
Thomas Kellaghan, Series Editors**



THE WORLD BANK

Национальная оценка учебных достижений

Пру Андерсен, Джордж Морган

Разработка тестов и анкет для национальной оценки учебных достижений

КНИГА 2

**Редакторы серии –
Винсент Грини, Томас Келлаган**



**Москва
Логос
2011**

УДК 378
ББК 74.58



*Издание подготовлено при поддержке
Международного банка реконструкции и развития*

A65

CICED

*Перевод подготовлен по инициативе
Центра международного сотрудничества
по развитию образования Российской
академии народного хозяйства
и государственной службы
при Президенте Российской Федерации*

Андерсен, Пру

A65 Разработка тестов и анкет для национальной оценки учебных достижений / Пру Андерсен, Джордж Морган; пер. с англ. В.Н. Симкина; науч. ред. В.И. Звонников. – М.: Логос, 2011. – 196 с.: ил. – (Национальная оценка учебных достижений).

ISBN 978-5-98704-546-6

Представлена технология разработки надежного инструментария для систем оценки учебных достижений. Освещено конструирование тестов учебных достижений, включая разработку структуры заданий, их предварительную проверку, отбор и создание окончательного теста. Раскрыты основные этапы конструирования анкет: их проектирование, написание вопросов, кодирование ответов. Рассмотрен вопрос соответствия данных анкет и тестов. Показано, как спроектировать руководство для администратора тестирования и гарантировать, что учащиеся выполняют тест в равных условиях. Характеризуются обязанности администратора по тестированию. Изложен подход к реформированию школ в национальном оценивании.

Для специалистов по разработке специализированных тестов и специалистов-предметников, а также для методистов, участвующих в организации тестирования и других оценочных мероприятий. Представляет интерес для широкого круга специалистов-предметников – учителей и преподавателей различных дисциплин, самостоятельно разрабатывающих и/или использующих тесты и другие технологии педагогических измерений, которые позволяют оценивать учебные достижения обучаемых и сопутствующие им факторы. Может использоваться в учебном процессе по педагогическим направлениям (специальностям), а также при повышении квалификации педагогических кадров и их подготовке в магистратуре в области тестологии и педагогических измерений, при изучении тематики, связанной с контролем и оценкой знаний обучаемых. Представляет интерес для широкого круга учителей, преподавателей и методистов учебных заведений всех видов и уровней.

УДК 378
ББК 74.58

ISBN 978-5-98704-546-6

© Международный банк
реконструкции и развития/
Всемирный банк, 2008, 2011
© Логос, 2011

1818 H Street NW
Washington, DC 20433
Телефон: 202-473-1000
Internet: www.worldbank.org
E-mail: feedback@worldbank.org

Все права защищены
1 2 3 4 10 09 08 07

Эта книга подготовлена под эгидой Международного банка реконструкции и развития. Представленные в ней материалы не всегда отражают точку зрения исполнительных директоров Всемирного банка или правительства государств, которые они представляют.

Всемирный банк не гарантирует точности приведенных данных. Границы, цвета, географические наименования и иная информация, приведенная на картах, не отражает позицию Всемирного банка относительно правового статуса территорий.

Авторское право

Произведение защищается авторским правом. Копирование и/или передача произведения или его частей без разрешения правообладателя преследуется по закону.

Международный банк реконструкции и развития поощряет работы по распространению книги и гарантирует предоставление разрешения на воспроизведение ее частей.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; telephone: 978-750-8400; fax: 978-750-4470; Internet: www.copyright.com.

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2422; e-mail: pubrights@worldbank.org.

Cover design: Naylor Design, Washington, DC

ISBN-13: 978-0-8213-7497-9

eISBN: 978-0-8213-7498-6

DOI: 10.1596/978-0-8213-7497-9

Library of Congress Cataloging-in-Publication Data

Anderson, Prue, 1954-

Developing tests and questionnaires for a national assessment of educational achievement / Prue Anderson, George Morgan. p. cm.—(National assessment of educational achievement ; volume 2)

Includes bibliographical references and index.

ISBN 978-0-8213-7497-9 (alk. paper)—ISBN 978-0-8213-7498-6

1. Educational tests and measurements—United States. 2. Educational evaluation—

United States. I. Morgan, George, 1945– II. Title.

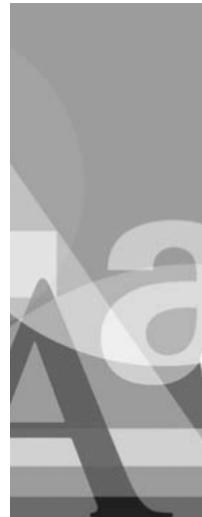
LB3051.A715 2008

371.26”1—dc22

2008002684

Выражение признательности

Научные консультанты, переводчики, редакторы Издательской группы «Логос» выражают признательность Московскому представительству Международного банка реконструкции и развития, особенно Исааку Фрумину и Тиграну Шмису, за помощь и поддержку в подготовке и выпуске этой книги и всей серии книг «Национальная оценка учебных достижений». Мы благодарим В.А. Болотова, вице-президента Российской академии образования, который направлял наши усилия на достижение целей, имеющих исключительное значение для реформирования российской системы образования. Высокой оценки заслуживает вклад Центра международного сотрудничества по развитию образования, который выступил инициатором перевода книг по оценке учебных достижений на русский язык.



СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	13
ОБ АВТОРАХ И РЕДАКТОРАХ	15
БЛАГОДАРНОСТИ	17
СПИСОК СОКРАЩЕНИЙ	19
ЧАСТЬ 1 КОНСТРУИРОВАНИЕ ТЕСТОВ УЧЕБНЫХ ДОСТИЖЕНИЙ	
ГЛАВА 1 ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ	21
ГЛАВА 2 РАЗРАБОТКА СТРУКТУРЫ	29
Проект теста, или Таблица спецификации	31
Валидность	37
Язык теста	37
Форма задания	38
Генеральная совокупность учащихся для проведения оценки	46
Сообщение результатов	46
Контекст	48
ГЛАВА 3 НАПИСАНИЕ ЗАДАНИЙ	51
Трудность задания	53
Систематическая ошибка задания	55

Материал стимула	56
Форма задания	58
Тренировочные задания	72
Расположение и дизайн заданий	73
Команда, пищащая задания	79
Группы экспертов по заданиям	83
Другие эксперты	86
Отслеживание заданий	87
ГЛАВА 4 ПРЕДВАРИТЕЛЬНАЯ ПРОВЕРКА ЗАДАНИЙ 91	
Дизайн предтестовой формы	95
Вычитка и тиражирование предварительного теста ..	100
Проведение предварительного тестирования	103
Оценка предварительного теста	104
Надежность	111
ГЛАВА 5 ОТБОР ТЕСТОВЫХ ЗАДАНИЙ 113	
ГЛАВА 6 СОЗДАНИЕ ОКОНЧАТЕЛЬНОГО ТЕСТА 121	
Проектирование окончательного теста	121
Вычитка и тиражирование	125
ГЛАВА 7 ОЦЕНКА ТЕСТОВЫХ ЗАДАНИЙ ВРУЧНЮЮ 129	
ЧАСТЬ 2 КОНСТРУИРОВАНИЕ АНКЕТ	
ГЛАВА 8 ПРОЕКТИРОВАНИЕ АНКЕТ 137	
Содержание анкет	139
Проект анкеты	144
Задания анкеты	146
Форма заданий	148
Язык анкеты	149
Респонденты	149
Предъявление анкет	150
План анализа данных	151
ГЛАВА 9 НАПИСАНИЕ ЗАДАНИЙ ДЛЯ АНКЕТ 153	
Вопросы	154
Утверждения	154
Варианты ответов	155
Управление проблемами чувствительности	157
Макет анкеты	158
Экспертиза анкет	158

ГЛАВА 10 КОДИРОВАНИЕ ОТВЕТОВ НА АНКЕТЫ	163
Подготовка анкет для ввода данных	165
Кодирование пропущенных или неоднозначных ответов	165
ГЛАВА 11 СООТВЕТСТВИЕ ДАННЫХ АНКЕТ И ТЕСТОВ	167
Анкеты для учащихся	168
Анкеты для родителей	168
Анкеты для опроса преподавателей и директоров школ	169
ЧАСТЬ 3 ДИЗАЙН РУКОВОДСТВА ПО ПРЕДЪЯВЛЕНИЮ ТЕСТА	
ГЛАВА 12 РУКОВОДСТВО ПО ПРЕДЪЯВЛЕНИЮ ТЕСТА	173
Содержание руководства	174
Использование руководства	176
Особенности руководства	176
Сколько деталей необходимо?	178
Тренировочные вопросы	179
Апробация	179
Обзор	182
ГЛАВА 13 АДМИНИСТРАТОР ПО ТЕСТИРОВАНИЮ	183
Выбор администратора	183
Следование инструкциям	184
Проверка качества	186
Контрольный список обязанностей администратора	187
ГЛАВА 14 ИНФОРМИРОВАНИЕ ШКОЛ О НАЦИОНАЛЬНОЙ ОЦЕНКЕ	191
ЛИТЕРАТУРА ДЛЯ ДАЛЬНЕЙШЕГО ЧТЕНИЯ	193
ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ	
2.1. Учебный план по математике в Папуа – Новая Гвинея	32
2.2. Проект Новой Зеландии по английскому языку ..	32
2.3. Примеры заданий с множественным выбором....	39
2.4. Пример задания с кратким конструируемым ответом	39
2.5. Пример открытого задания с кратким ответом...	40
2.6. Пример краткого эссе	41

3.1. Пример иррелевантного материала стимула	58
3.2. Пример задания с неточной или вводящей в заблуждение информацией	59
3.3. Пример задания с множественным выбором	60
3.4. Расстановка знаков препинания в случае, если варианты ответов представляют собой законченные предложения	61
3.5. Расстановка знаков препинания, если варианты ответов даны в форме перечня	61
3.6. Сокращение текста задания	62
3.7. Задание с отрицанием в основной части	62
3.8. Некорректный подбор слов в дистракторах	63
3.9. Корректное сочетание основной части с дистракторами	63
3.10. Пример неграмотно составленного задания с кратким ответом	67
3.11. Пример грамотно составленного закрытого задания с конструируемым ответом	68
3.12. Задание с частичным кредитованием	68
3.13. Пример открытого задания с кратким ответом с руководством по оценке	70
3.14. Пример закрытого задания с конструируемым ответом и руководство по оценке	71
3.15. Использование рисунков для сокращения формулировок	75
3.16. Принцип простоты изображения	76
3.17. Четкая маркировка изобразительного материала	76
3.18. Ясные обозначения на картах	77
3.19. Роль свободного пространства для понимания материала стимула	78
3.20. Пример таблицы стилей для авторов заданий ...	82
4.1. Пример задания в форме с множественным выбором и в открытой форме	105
4.2. Пример формата ввода данных для предтеста	107
6.1. Пример обложки теста	122
8.1. Проект анкеты по отношениям и ценностям	145
9.1. Пример неграмотного расположения квадратиков для ответов	159
9.2. Пример грамотного расположения квадратиков для ответов	159

10.1. Пример кодирования ответов без выделения	164
10.2. Пример рассмотрения каждого варианта ответа в отдельности при вводе данных	164
12.1. Инструкции руководства по применению теста	177
12.2. Информация для преподавателей и руководителей	178
12.3. Предъявление тренировочных заданий	180
13.1. Список обязанностей администратора: пример из Филиппин	188

РИСУНКИ

1.1. Организационная структура национальной оценки	24
1.2. Этапы национальной оценки	25
4.1. Пример круговой связи заданий	97
4.2. Модель для вертикально связующих заданий	98

ТАБЛИЦЫ

1.1. Этапы разработки тестов и анкет	26
2.1. Проект теста по основам математики для средних классов начальной школы	33
2.2. Спецификация теста по математике в исследова- нии TIMSS, 3-й и 4-й классы	34
2.3. Проект содержания теста по математике в Папу – Новой Гвинеи	35
2.4. Преимущества и ограничения форм тестовых заданий	44
2.5. Формы тестовых заданий по математике в Па- пу – Новой Гвинеи	45
4.1. Связующие задания в двух юнитах по чтению	99
4.2. Часть крупноформатной таблицы для отслежива- ния заданий по формам	100
5.1. Пример результатов анализа задания с множе- ственным выбором	115
5.2. Пример результатов анализа открытого дифферен- цированного задания	117
8.1. Компоненты процесса разработки анкет	140
8.2. Функции чтения в международном исследовании	147



ПРЕДИСЛОВИЕ

Качество любого оценочного процесса в образовании зависит от качества используемого инструментария. Как правило, если инструментарий разработан плохо, оценка превращается в пустую трату времени и денег.

«Разработка тестов и анкет для национальной оценки учебных достижений» является второй из пяти книг в серии «Национальная оценка учебных достижений». В ней описана технология разработки надежного инструментария для национального оценивания. При этом особое внимание уделяно развивающимся странам.

В первой книге серии освещаются ключевые цели и особенности системы национальной оценки. Она, в основном, адресована тактикам и руководящим лицам, принимающим решения в сфере образования. В остальных книгах детально рассматриваются частные вопросы – проектирование, применение, анализ и сообщение информации в системе национальной оценки. Они предназначены, главным образом, для команд по национальному оцениванию.

Книга, которую вы держите в руках, посвящена разработке и дизайну двух видов инструментария для сбора данных: тестам учебных достижений и анкетам. Первая часть охватывает разработку структуры оценивания и проекта теста, написание

заданий, проведение предварительного и окончательного тестирования. Во второй части описаны основные этапы конструирования анкет по сбору контекстной информации. Анкеты необходимы для сбора информации об учащихся и их родителях, преподавателях или директорах школ по переменным, которые могли бы помочь объяснить различия в результатах выполнения теста учащимися. В третьей части рассматривается, как спроектировать руководство для администраторов тестирования и гарантировать, что учащиеся выполняют тест в стандартизованных условиях.

В третьей книге серии основное внимание уделяется проблемам, связанным с проведением крупномасштабных программ оценивания, включая логистику, формирование выборки, чистку данных и управление. В четвертой книге рассматривается, как получать данные по заданиям и тестовые оценки, а также связь полученных результатов с другими образовательными факторами. Наконец, пятая книга охватывает вопросы написания отчетов, которые основаны на результатах национального оценивания, и их использования для повышения качества образовательной политики.

После прочтения настоящей книги станет ясно, что разработка оценочного инструментария – это сложная и затратная по времени задача, для решения которой требуются знания, умения и ресурсы. В то же время, как показывает опыт, польза от грамотно спроектированного инструментария может быть существенной с точки зрения качества информации об уровне учебных достижений учащихся, о школьных и внешкольных факторах, которые на них влияют. Высококачественный инструментарий повысит кредит доверия к решениям в образовательной политике. Он также может увеличить вероятность того, что тактики будут использовать результаты национальной оценки при разработке планов и программ, направленных на повышение качества образования. Если тесты и анкеты приведут к достижению этих результатов, то время и усилия, затраченные на их разработку, будут более чем оправданы.

Маргарит Кларк,
главный специалист по образованию,
Международный банк реконструкции и развития



ОБ АВТОРАХ И РЕДАКТОРАХ

АВТОРЫ

Пру Андерсен – старший научный сотрудник в Австралийском совете по исследованиям в области образования. Она разработала оценочные материалы по чтению для программ системы тестирования учащихся начальных и средних классов школы. Участвовала в программах мониторинга образования в Австралии, Бруней-Даруссаламе, Папуа – Новой Гвинее и на Филиппинах. В настоящее время является менеджером проекта программы «Международная оценка школ». Сфера профессиональных интересов: сопоставление данных оценивания с государственными результатами учебных планов, измерения социальных результатов обучения в школе.

Джордж Морган – консультант по образованию. Работал старшим научным сотрудником в отделе по измерениям и руководителем группы разработки тестов по математике и естественным наукам в Австралийском совете исследований в образовании в течение тридцати лет. Разрабатывал учебные планы по математике и естественным наукам и оценочные материалы для всех уровней образования. Работал в программах тестирования на крупномасштабной шкале. Участвовал в проектах по оценке

в Камбодже, Восточном Тиморе, Лаосской Народно-Демократической Республике, Папуа – Новой Гвинеи и Самоа.

РЕДАКТОРЫ СЕРИИ

Винсент Грини – консультант по образованию. Был ведущим специалистом в области образования Международного банка реконструкции и развития, работал в странах Африки, Азии и Ближнего Востока. Бывший преподаватель, научный сотрудник Центра исследований в образовании в колледже Святого Патрика в Дублине; приглашен по программе Фулбрайта, профессор Западно-Мичиганского университета в Каламазу. Член читального зала славы Международной ассоциации по чтению. Сфера профессиональных интересов: оценка, педагогическое образование, чтение, пропаганда социальной сплоченности посредством реформирования учебников.

Томас Келлаган – директор Центра исследований в образовании в колледже Святого Патрика в Дублине, член Международной академии образования. Работал в университете города Ибадан в Нигерии и в Королевском университете в Бельфасте. Сфера профессиональных интересов: оценка и экзамены, недостатки в образовании, проблема связи образования и домашней обстановки. Занимал пост президента Международной ассоциации по оценке образования с 1997 по 2001 год. Участвовал в проектах по оценке в Африке, Азии, Латинской Америке и на Ближнем Востоке.



БЛАГОДАРНОСТИ

Команда редакторов во главе с Винсентом Грини (консультант, центр «Развитие человеческого потенциала», группа по образованию, Международный банк реконструкции и развития) и Томасом Келлаганом (Центр исследований в образовании, колледж Святого Патрика, Дублин), подготовили серию под названием «Национальная оценка учебных достижений», в том числе данную книгу.

В работе над книгой также принимали участие Сильвия Акан (Уганда, Совет по национальным экзаменам), Пру Андерсен (Австралийский совет по исследованиям в образовании), Фернандо Картрайт (Канадский совет по обучению), Жан Дюмей (Группа статистиков, Канада), Крис Фримен (Австралийский совет по исследованиям в образовании), Хью Гоф (Группа статистиков, Канада), Сара Хаун (Университет Претории), Джордж Морган (Австралийский совет по исследованиям в образовании), Т. Скотт Мюррей (Институт статистики, ЮНЕСКО), Джерри Шил (Центр исследований в образовании, Колледж Святого Патрика, Дублин).

Общее руководство осуществляли Руфь Кагиа, директор образовательных программ, и Робин Хорн, руководитель центра «Развитие человеческого потенциала». Инициатором про-

екта был Роберт Праути, он же руководил им до августа 2007 года. На более поздних стадиях руководство осуществляла Маргарит Кларк (подготовка к публикации).

Мы выражаем благодарность за большой вклад в работу группе экспертов: Олу Битону (Бостонский колледж), Ирвину Киршу (Служба тестирования в образовании) и Бенуа Миллоту (Международный банк реконструкции и развития). Также были учтены замечания Хелен Абадзи, Регины Бендокат, Маргарит Кларк, Робина Хорна, Элизабет Кинг, Морина Льюиса, Гарри Патриоса, Карлоса Рохаса, Джи-Пэн Тан, Эдуардо Велеса и Раисы Веналайен.

Мы получали ценные советы от Карли Чиверса, Дэвида Хардинга, Эйдан Малкин, Александры Савицка, Тхи Чана, Хилари Уолш и Ганса Уоджемайкера.

Особую благодарность мы выражаем учреждениям, санкционировавшим выпуск серии: Австралийскому совету по исследованиям в образовании; Центру исследований в области образования, Дублин; Международной ассоциации по эвалюации образовательных достижений; Массачусетскому департаменту образования; Национальному центру статистики в сфере образования США; Организации экономического сотрудничества и развития; Департаменту образования Папуа – Новой Гвинеи.

Редакторские правки, оформление книги и печать были согласованы с Мэри Фиски и Паолой Скалабрин – сотрудниками издательства Международного банка реконструкции и развития.

Также серьезную поддержку оказали Ирландский образовательный доверительный фонд, Банк партнерских программ Нидерландов и Австралийский совет по исследованиям в образовании.



СПИСОК СОКРАЩЕНИЙ

МБРР	Международный банк реконструкции и развития (Всемирный банк)
ООН	Организация Объединенных Наций
ОЭСР	Организация экономического сотрудничества и развития
ЮНЕСКО	Организация Объединенных Наций по вопросам образования, науки и культуры
CTT	Classical Test Theory: классическая теория тестов
IRT	Item Response Theory: современная теория тестов
PIRLS	Progress in International Reading Literacy Study: Международное исследование прогресса в гра- мотности чтения
PISA	Program for International Student Assessment: Про- грамма международного оценивания учащихся
TIMSS	Trends in Mathematics and Science Study: Между- народное исследование тенденций в математи- ческом и естественнонаучном образовании



КОНСТРУИРОВАНИЕ ТЕСТОВ УЧЕБНЫХ ДОСТИЖЕНИЙ

ГЛАВА

1

ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Национальная оценка учебных достижений включает целый комплекс мероприятий – от осознания ее необходимости до ознакомления с полученными результатами.

Книги, вошедшие в серию «Национальная оценка учебных достижений», освещают этот процесс с разных сторон. Особое внимание уделяется развивающимся странам. В них еще не сложились устоявшиеся традиции эмпирических исследований в образовании и потому едва ли отработана технология проведения национальной оценки. Настоящая серия является попыткой обстоятельного анализа мероприятий, проведение которых необходимо для адекватной оценки достижений в сфере образования. Она предназначена для читателей, которые, как мы предполагаем, будут нести ответственность за некоторые аспекты оценивания, чтобы помочь им понять содержание и суть деятельности по оцениванию.

Как правило, общее руководство системой национального оценивания осуществляет министерство образования или назначенный им национальный руководящий комитет, которые несут полную ответственность и оказывают поддержку. Под их непосредственным руководством основную часть работ выполняет исполнительное агентство. Оно, в свою очередь, контро-

лирует менеджера по разработке тестов, а также специалистов-предметников и специалистов по статистическому анализу и отвечает за логистическое сопровождение процесса.

Эта книга, «Разработка тестов и анкет для национальной оценки учебных достижений», посвящена главным образом деятельности менеджера по разработке тестов и специалистов-предметников, а также мероприятиям предварительного тестирования (рис. 1.1). Другие направления деятельности, представленные на рисунке, такие как формирование выборки, логистика (контакты со школами, ввод и чистка данных), рассмотрены в третьей книге серии – «Проведение национальной оценки учебных достижений». Четвертая книга, «Анализ данных национальной оценки учебных достижений», посвящена статистическому анализу.

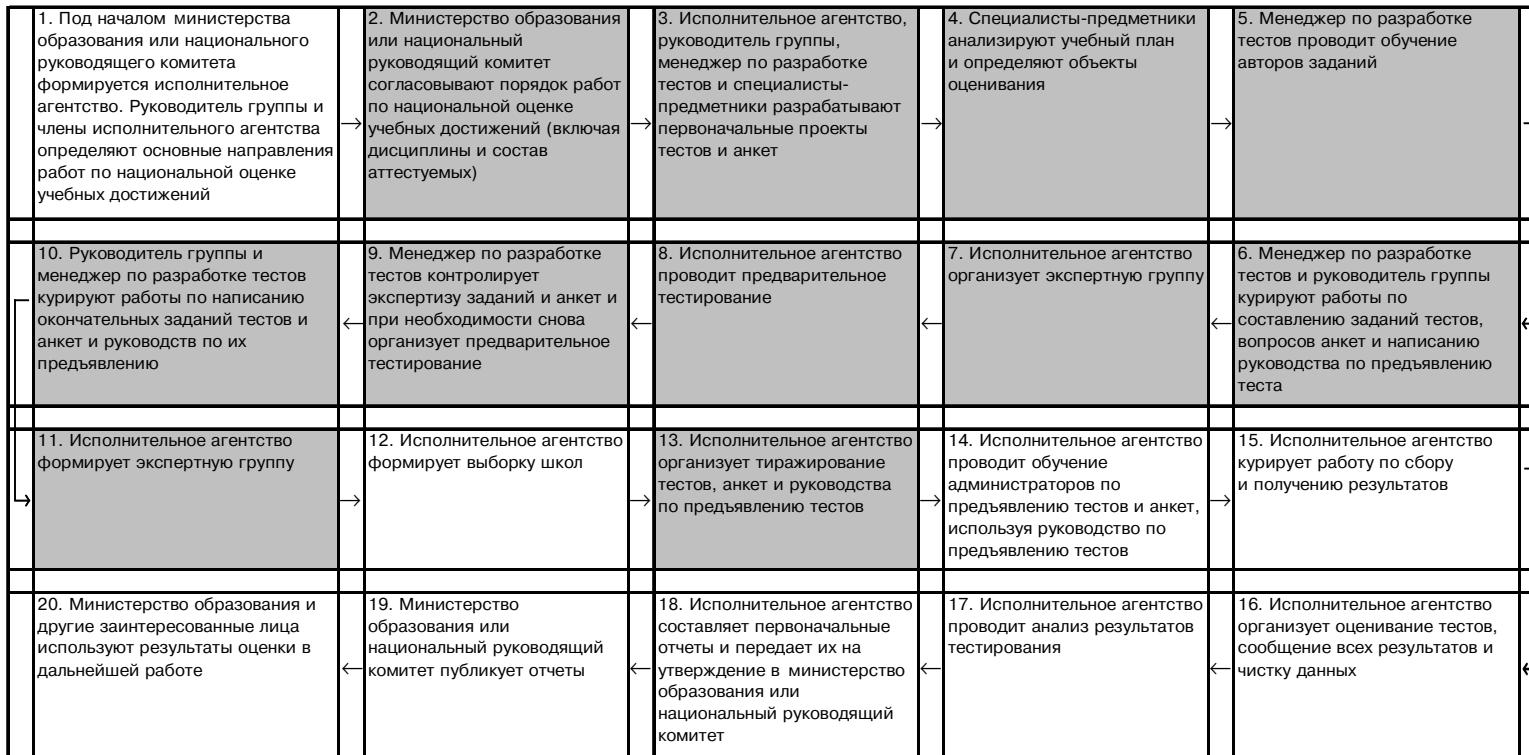
Схема, представленная на рис. 1.2, систематизирует аспекты национальной оценки. Большинство из них весьма подроб-

Рис. 1.1

Организационная структура национальной оценки



Источник: разработки авторов.

Рис. 1.2**Этапы национальной оценки**

Источник: разработки авторов.

ТАБЛИЦА 1.1**Этапы разработки тестов и анкет**

Этап	Деятельность	Срок	Ответственные исполнители
1	Определение цели национальной оценки, разработка структуры, тестов и анкет; формирование выборки	От 4 до 6 недель	Министерство образования или национальный комитет, исполнительное агентство, руководитель группы, специалист по разработке тестов, заинтересованные лица и представители правительственный структур
	Разработка проекта или спецификации, консультации с широкой общественностью	От 4 до 6 недель	Министерство образования или национальный комитет, исполнительное агентство, специалист по разработке тестов, группы экспертов, опытные преподаватели, специалисты-предметники, специалисты по системному анализу, опытные авторы заданий, заинтересованные лица и представители правительенных структур
2	Написание заданий тестов и анкет	От 12 до 14 недель (от 20 до 30 заданий на автора в неделю)*	Менеджер по разработке тестов, специалисты-предметники, авторы заданий и заинтересованные лица
	Разработка предварительных вариантов тестов и анкет	4 недели	Руководитель группы, специалист по разработке тестов, авторы заданий, дизайнеры, верстальщики, корректоры
	Тиражирование предварительных вариантов тестов и анкет	2 недели	Исполнительное агентство и администраторы по предъявлению тестов
	Распространение предварительных вариантов тестов и анкет	От 2 до 3 недель	Исполнительное агентство, руководитель группы, менеджер по разработке тестов и авторы заданий
3	Апробация предварительных вариантов тестов и анкет в школах	От 2 до 3 недель	Исполнительное агентство и администраторы по предъявлению тестов
	Оценка заданий вручную (если необходимо)	2 недели	Руководитель группы, специалист по разработке тестов, специалисты-предметники и авторы заданий
	Введение данных предварительного тестирования	1 неделя	Исполнительное агентство, специалисты по вводу и анализу данных

4	Анализ данных предварительного тестирования и анкетирования	2 недели	Исполнительное агентство, специалисты по вводу и анализу данных, авторы заданий и менеджер по разработке тестов
	Отбор заданий для тестов и анкет	2 недели	Менеджер по разработке тестов, специалисты по системному анализу, авторы заданий и заинтересованные лица
5	Производство окончательных тестов, анкет и руководств для администраторов	2 недели	Исполнительное агентство, менеджер по разработке тестов, дизайнеры, верстальщики, корректоры и авторы заданий
	Тиражирование тестов и анкет	4 недели	Исполнительное агентство, руководитель группы и менеджер по разработке тестов
	Пакетирование тестов и анкет	От 2 до 3 недель (в зависимости от расстояния)	Исполнительное агентство и менеджер по разработке тестов
6	Предъявление тестов и анкет в школах	От 3 до 4 недель	Исполнительное агентство, менеджер по разработке тестов и администраторы
	Оценка заданий вручную (если необходимо)	От 3 до 4 недель	Менеджер по разработке тестов и авторы заданий
7	Ввод и чистка данных	От 4 до 6 недель	Специалисты по вводу и анализу данных
	Анализ данных	От 2 до 3 недель	Специалисты по системному анализу, авторы заданий и менеджер по разработке тестов
8	Подготовка окончательных отчетов	От 4 до 5 недель	Специалисты по системному анализу, авторы заданий и менеджер по разработке тестов

Источник: разработки авторов.

*Если задания необходимо перевести на другие языки, может потребоваться дополнительное время.

но рассматриваются в этой книге. Аспекты, которым авторы уделяют наибольшее внимание, заключены в заштрихованные ячейки. В книге имеются повторы. Их наличие связано с необходимостью помочь читателю, который интересуется только одним аспектом оценки, получить полное и достаточное представление именно о нем, избегая при этом ненужной информации.

Дополнительная информация о разработке тестов и анкет приводится в табл. 1.1. Процесс конструирования тестов учебных достижений и анкет представлен в виде восьми этапов с указанием ответственных исполнителей на каждом из них.

ГЛАВА



РАЗРАБОТКА СТРУКТУРЫ

Структура, которая представляет собой план для руководства процессом разработки тестов, анкет и процедур, крайне важна для понимания сути этого процесса (Линн (Linn) и Данбар (Dunbar), 1992; Муллис (Mullis) и др., 2006). Она помогает со всех сторон рассмотреть конструкт (например, достижения в чтении или математике) и процессы, которые с ним связаны. Этот план должен включать: четкое определение того, что оценивается, идентификацию характеристик задач, которые должны быть использованы при разработке теста, и исходные предпосылки для дальнейшей интерпретации результатов (Кирш (Kirsch), 2001; Мессик (Messick), 1987).

Структура разъясняет цель оценки, выделяет ключевые понятия и может облегчить обсуждение решения, принимаемого руководящими лицами в образовании, с помощью разъяснения ключевых понятий перед выполнением оценивания. Структура также определяет ключевые переменные, которые вероятнее всего связаны с тестовыми оценками, и гарантирует, что эти переменные будут включены в проект национальной оценки.

На первоначальном этапе члены руководящего комитета должны прийти к единому взгляду на предмет измерения. Зачастую определения основных предметных областей содержатся в

национальных учебных планах. Они могут существенно отличаться друг от друга. Так, определение чтения менялось со временем и оно существенно отличается в разных системах образования. В одних системах образования чтение приравнивалось к способности произносить слова. В других оно рассматривалось как способность идентифицировать отдельные слова и объяснять их значение. Также чтение определялось как способность понимать значения слов или выводить их смысл из контекста.

Более свежие определения не ограничивают понимание чтения как совокупности простейших умений. Чтение понимается как способность использовать информацию из текстов, интерпретировать ее. Подчеркивается также, что учащиеся и взрослые читают для достижения самых разнообразных целей – от развлекательных до информационных. Эти новые представления нашли отражение в тестах. В них все чаще включаются текстовые элементы – короткие рассказы, выдержки из газет, рекламные объявления, знаки и схемы.

В процессе разработки теста должна быть четко сформулирована его цель. Для этого необходимо как можно раньше сделать первый шаг – провести консультацию с ключевыми заинтересованными лицами в сфере образования и представителями экспертных групп для выявления предмета оценивания, состава выборки учащихся, чьи достижения нуждаются в оценке, а также времени тестирования и языка, на котором дается тест. К этим решениям должны быть привлечены разработчики учебных планов, а также тактики и менеджеры, которые будут иметь возможность использовать результаты оценивания для выработки образовательной политики, распределения ресурсов и проведения реформ.

Национальная оценка может стать мощным инструментом для эvaluation эффективности разных аспектов учебного плана. Если процесс национальной оценки построен грамотно, то ее результаты могут способствовать улучшению учебного плана посредством включения в него тех знаний и умений, которые учащиеся должны продемонстрировать в ходе проверки. Умения учащихся, наряду с контекстом оценивания, должны способствовать реализации глобальных целей образовательной политики в ключевых сферах учебной деятельности. Примеры контекстов в сфере образования, которые приведены в дополнительных материалах 2.1 и 2.2, отражают некоторые приоритеты в образовании.

ПРОЕКТ ТЕСТА, ИЛИ ТАБЛИЦА СПЕЦИФИКАЦИИ

Проект теста, или таблица спецификации, – важный документ, который является основой для разработки теста, анализа его данных и подготовки письменного отчета. Он включает данные, которые должны быть получены, определяет длину теста и количество заданий по каждому разделу учебного плана. Хороший проект должен включать:

- доли заданий по каждой предметной области учебного плана (например, математика, язык, естественные науки) в окончательной форме теста;
- доли заданий внутри предметной области учебного плана для оценки различных умений (например, в математике – числа, меры, пространство и схемы; в письме – идеи, знание содержания, структура, стиль, словарный запас, правописание и грамматика);
- долю заданий, которые обращаются к различным познавательным процессам проявления навыков (таким, как знание или отклик, интерпретация или рефлексия);
- долю заданий с множественным выбором и открытых заданий;
- долю заданий, связанных со стимулами в различных тестах для чтения (рассказ, описание, процедуры или проблемы для аргументации) или по математике (таблицы, схемы и диаграммы).

Проект теста, представленный в табл. 2.1, основан на учебном плане по математике для средних классов начальной школы. Отдельные субтесты предназначены для измерения способностей учащихся к выполнению простых вычислений, пониманию основных математических понятий и решению задач. Например, ячейка, находящаяся на пересечении содержательной области «Дроби» и интеллектуальной характеристики «Способность решать типовые проблемы», представляет «Способность к решению типичных проблем, включающих действия с дробями». Группа специалистов-предметников, в которую входили преподаватели, решила посвятить пять заданий проверке этой способности. Ячейка, которая содержит задания, проверяющие способность к выполнению операций с целыми числами, получила самый высокий вес (25 заданий). Многие ячейки не содержали ни одного задания. Разработку предварительных, а затем окончательных вариантов тестов определял показатель «Вес относительной важности», присвоенный каждому объекту.

Дополнительный материал 2.1**Учебный план по математике в Папуа – Новой Гвинее**

В Папуа – Новой Гвинее учебный план по дисциплине «Элементарная математика» 2003 года имеет следующее обоснование:

Все граждане должны участвовать в будущем развитии Папуа – Новой Гвинеи. Поэтому учащимся следует учиться устному счету, получать соответствующие знания и умения... На начальном уровне учащиеся должны научиться связывать новые математические понятия содержательных разделов с имеющимся знанием математической культуры, чтобы использовать математику в повседневной жизни. Элементарная математическая культура дает много возможностей для релевантного и осмысленного обучения, которое основано на принципах домашней жизни (Папуа – Новая Гвинея, Департамент образования, 2003: 2).

Данный план (наряду с политическими документами министерства и существенным реструктурированием начального и дошкольного образования) предусматривает преподавание элементарного курса математики в деревенских школах. Реформы делают акцент на приоритете интеграции курса элементарной математики с местной культурой и использовании математического аппарата в повседневной жизни. Недавно проведенная национальная оценка для мониторинга достижений учащихся обозначила необходимость реалистического контекста для вопросов и оценки умений и понимания того, что имеет практические приложения.

Дополнительный материал 2.2**Проект Новой Зеландии по английскому языку**

Общие цели новозеландского государственного учебного плана по английскому языку:

Учащиеся должны обладать следующими умениями: использовать язык и получать удовольствие от общения, понимать и эффективно применять устный, письменный и визуальный язык в разных ситуациях (Новая Зеландия, Министерство образования, 2002: 99).

Эти цели выдвигают на первый план в чтении и изучении текстов проблему интереса и удовольствия. При рассмотрении уровневой системы оценки английского языка привлекательные тексты и задачи для учащихся являются ключевыми. Акцент на языке во всех его вариантах отражает серьезное обязательство по признанию и оценке устной языковой культуры учащихся маори, так же как и письменных форм английского языка. Эти цели отражены в различных национальных системах оценки.

ТАБЛИЦА 2.1

Проект теста по основам математики для средних классов начальной школы

Содержательные области	Интеллектуальные характеристики												Всего заданий	
	Вычисления			Понятия					Решение проблем					
	Знание терминов и фактов (A1)	Способность выполнять операции (A2)	Общее число заданий	Понимание математических понятий (B1)	Понимание математических принципов (B2)	Понимание математической структуры (B3)	Способность к преобразованию элементов из одной формы в другую (B4)	Способность к чтению и интерпретации графиков и диаграмм (B5)	Общее число заданий	Способность решать типовые проблемы (C1)	Способность анализировать и проводить сравнения (C2)	Способность решать нестандартные проблемы (C3)		
Целые числа	1	25	26	1	4	7	2	4	18	14	2	2	18	62
Дроби		4	4	4	1	2			7	5			5	16
Десятичные дроби		8	8			5		1	6	5			5	19
Измерение	2		2				3	2	5	3			3	10
Геометрия			0	2	2				4				0	4
Диаграммы и графики			0						0		4		4	4
Полное общее количество	3	37	40	7	7	14	5	7	40	27	6	2	35	115

Источник: Центр исследований в образовании, 1978: 44.

ТАБЛИЦА 2.2**Спецификация теста по математике в исследовании TIMSS,
3-й и 4-й классы**

Ожидаемые виды деятельности	Общее число заданий*	Задания с множественным выбором	Задания с краткими ответами	Задания с развернутыми ответами
Знания	42	35	7	0
Выполнение стандартных процедур	16	13	3	0
Использование сложных процедур	24	21	2	1
Решение проблем	20	10	3	7

Источник: Международная ассоциация по эвалюации образовательных достижений, <http://timss.bc.edu/timss1995i/TIMSSPDF/AMitems.pdf>.

* Число отражает полную совокупность заданий, которые использовались для формирования 26 тестовых кластеров в восьми различных тестовых буклетах. Ни от одного учащегося выполнять полный тест не требовалось.

Проект теста по математике в третьем Международном исследовании тенденций в математическом и естественнонаучном образовании (TIMSS)¹, представленный в табл. 2.2, определяет формы заданий и познавательные процессы, предназначенные для учащихся 3-х и 4-х классов, несколько иным способом.

Ясно, что проекты могут изменяться в зависимости от измеряемого конструкта и от цели оценки. Каждый, кто принимает участие в разработке теста, должен понимать и одобрять значение проекта с точки зрения того, что должно быть протестировано и что нет.

Из-за того, что время и ресурсы ограничены, проверить знания по каждой области учебного плана невозможно. Тестовые задания должны быть направлены на проверку основных знаний и умений. Чтобы их выявить, нужно проконсультироваться с экспертами по предметам или обратиться к учебному плану.

Важность, которая придается в проекте национальной оценки отдельным разделам учебного плана, зависит, в том числе, от способа, посредством которого сообщаются данные тестирования (с помощью общей оценки или по разделам учебно-

ТАБЛИЦА 2.3**Проект содержания теста по математике в Папуа – Новой Гвинее**

Класс	Числа и их применение	Пространство и форма	Измерение	Вероятность	Схемы и задания по алгебре*	Общее число заданий
3-й	10	7	4	2	2	25
5-й	10	10	7	4	4	35
8-й	10	10	8	6	6	40

Источник: неопубликованный материал Департамента образования Папуа – Новой Гвинеи.

*Задания по алгебре – только в 8-м классе

го плана). Необходимо проконсультироваться с тактиками в образовании, чтобы узнать, в каком виде они хотели бы сообщить пользователям данные тестирования.

Если для каждого учащегося необходимо привести единственную общую оценку, например, в такой дисциплине, как математика, то потребуются, по крайней мере, 25 или 30 заданий. Данные, собранные на основе меньшего числа заданий, могут потребоваться для сообщения результатов тестирования по отдельным разделам проекта, таким как понимание учащимися раздела «пространство» или решение проблем в teste по математике.

Табл. 2.3 содержит пример таблицы спецификации теста по математике для учащихся 3-го, 5-го и 8-го классов в Папуа – Новой Гвинее. Отметим, что в teste для 3-го класса больше 80% заданий охватывают следующие разделы: числа и их использование, пространство и форма, измерения. Только четыре задания относятся к теории вероятности и схемам. Задания для 8-го класса по каждому разделу проекта распределены более равномерно.

Проекты обычно составляют на основе предписанных учебных планов. Однако если не принимается во внимание рассогласование между «реализованными» (которые дает учитель) и «освоенными» учащимися учебными планами, то тест может оказаться слишком трудным или слишком легким при выполнении. Следовательно, он не обеспечит правдоподобную картину учебных достижений учащихся в генеральной целевой совокупности. Если большая часть испытуемых не в состоянии дать правильный ответ на задания теста, это вовсе не означает, что при выполне-

ний чуть менее или гораздо менее трудных заданий они продемонстрируют такие же неважные результаты. Аналогично, если большинство учащихся отвечает на все задания правильно, это не гарантирует того, что на чуть или намного более трудные задания они ответят так же. Результаты, полученные по слишком трудным или слишком легким тестам, используются тактиками, школьами или преподавателями весьма ограниченно.

Трудность теста определяется целью тестирования. Если цель – мониторинг работы всех испытуемых, входящих в целевую генеральную совокупность, то распределение оценок параметра трудности заданий должно соответствовать распределению оценок учебных достижений генеральной совокупности учащихся. Как правило, разработчики тестов исходят из того, что на две трети заданий две трети учащихся дадут правильный ответ с вероятностью от 30 до 70%. (В среднем, вероятность правильного выполнения заданий должна составлять 50%, чтобы получить максимальную дисперсию тестовых оценок учащихся.) Оставшаяся треть должна быть равномерно распределена между заданиями, на которые правильно ответят больше чем 70% испытуемых, и заданиями, которые смогут выполнить менее 30% учащихся.

Хотя подобный подход к оценке уровня учебных достижений группы учащихся очень важен, он не должен вести к автоматическому исключению из теста ключевых разделов учебного плана просто по той причине, что испытуемые выполняют задания по ним очень плохо или очень хорошо. Пригодность заданий по уровню трудности устанавливается с помощью предварительного тестирования, в ходе которого испытуемым дают задания с характеристиками, подобными тем, которые предназначаются для национального оценивания на генеральной совокупности учащихся.

В некоторых случаях в ходе национальной оценки устанавливают уровни учебных достижений или выполнения теста, которые основаны на заданном стандарте (под стандартом понимается пороговый балл, устанавливаемый по тесту. – *Примеч. ред.*) и выявляют испытуемых, которые ему соответствуют. Если стандарт будет очень высок, то тест идентифицирует небольшое количество учащихся, которые продемонстрируют этот уровень достижений владения умением,

при этом будет потеряна информация об уровне учебных достижений остальной части обучаемых (т.е. генеральной совокупности), результаты обучения которых ниже стандарта. Если же требования стандарта будут невысоки, то тест выявит большое число учащихся, которые продемонстрируют этот уровень умений, при этом информации о более высоких уровнях владения умениями, которые, возможно, приобрели испытуемые, будет меньше.

ВАЛИДНОСТЬ

Валидность – весьма многоаспектное понятие, которое подразумевает адекватность интерпретации и использования тестовых оценок и тестовой информации (Мессик (Messick), 1989). Один из аспектов валидности подразумевает выявление степени, в которой содержание теста репрезентативно отражает учебный план или конструкт, который измеряется. Менеджер по разработке тестов несет ответственность за координацию с референтной группой специалистов-предметников, специалистов по учебному плану, чтобы гарантировать, что задания адекватно представляют выборку из учебного плана или конструкт.

В экспертную группу не должны входить авторы заданий. В этом случае проблема валидности решается на основе экспертных суждений, а не путем статистических исследований. Экспертная группа должна определить, дает ли тест адекватное освещение определенного предмета (например, математики для 4-го класса), и рассмотреть, обеспечивают ли результаты выполнения теста адекватное свидетельство наличия учебных достижений учащихся в предметной области.

ЯЗЫК ТЕСТА

Структура теста должна разъяснить и обосновать использование в тестах одного или нескольких языков. Язык теста, как правило, бывает тем, который используется и для обучения. При использовании нескольких языков для обучения перевод заданий бывает дорогостоящим и трудоемким. Переведенные

версии должны быть максимально эквивалентными между собой в тех случаях, когда данные тестирования используются для сравнения. Ниже перечислены некоторые проблемы, которые необходимо рассматривать в процессе принятия решения об использовании одного или более языков для проверки знаний по определенной области учебного плана.

- Если ресурсы ограничены, то предпочтительна оценка учащихся старших классов на языке, на котором велось обучение.
- Сокращение количества слов, используемых в тестовых заданиях, поможет уменьшить затраты на перевод до минимума, но это может нарушить контекст задания, делая его менее аутентичным.
- Желательно исключение из генеральной совокупности некоторых учащихся, чтобы повысить адекватность отражения всех лингвистических групп.
- Иногда официальный язык обучения не является фактическим языком, используемым в обучении. В таких случаях задания в teste для национальной оценки могут быть написаны на фактическом языке обучения.
- В случае тестирования учащихся младших классов, особенно если язык инструкции не является основным, тест может быть проведен в устной форме. Администратор по тестированию громко зачитывает каждый вопрос или воспроизводит его на магнитофоне и дает учащимся достаточно времени для ответа. Эта форма является особенно подходящей для тестов по математике и естественным наукам в младших классах. Учащиеся младших классов смогут продемонстрировать большую степень владения предметом по устному тесту, нежели письменному. Тесты, разработанные для оценки навыков чтения, предъявляются в устной форме, конечно, не могут.

ФОРМА ЗАДАНИЯ

При бланковом тестировании учащиеся отвечают на ряд вопросов или других посылок. Их письменные или представленные в виде символа ответы используются как свидетельство их

Дополнительный материал 2.3**Примеры заданий с множественным выбором**

1. Что из приведенного ниже скорее всего может быть измерено в миллилитрах?

- A. Количество жидкости в чайной ложке
- B. Вес (масса) булавки
- C. Количество бензина в резервуаре
- D. Толщина 10 листов бумаги

Источник: Международная ассоциация по эвалюации образовательных достижений, 1998, пример задания.

Замечание: A – правильный ответ.

2. Бутылка яблочного сока стоит 1,95 \$. Хлеб стоит 2,75 \$. Какое наименьшее количество денег Вам потребуется, если Вам нужно купить яблочный сок и хлеб?

- \$3,75 \$4,00 \$4,80 \$5,00

Источник: Австралийский совет образовательных исследований, пример задания.

Замечание: \$4,80 – правильный ответ.

Дополнительный материал 2.4**Пример задания с закрытым конструируемым ответом**

Дано численное выражение

$$2000 + \square + 30 + 9 = 2739$$

Какое число нужно поставить на место \square , чтобы это выражение стало верным?

Ответ _____

Источник: Международная ассоциация по эвалюации образовательных достижений, 1998, пример задания.

уровня знаний, компетенций и понимания. Наиболее распространенными являются четыре формы заданий:

- задания с множественным выбором;
- задания с закрытым конструируемым ответом;
- задания с кратким открытым ответом;
- эссе или задания с развернутым ответом.

Задания с множественным выбором (см. дополнительный материал 2.3) предполагают, что учащиеся должны выбрать правильный ответ из нескольких, как правило четырех, готовых. Варианты могут быть обозначены буквами (цифрами) или пустыми кружками (квадратами) – для проставления верного ответа. Они могут быть приведены в виде горизонтального ряда, или в од-

Дополнительный материал 2.5**Пример открытого задания с кратким ответом**

А. Как можно установить возраст дерева после того, как оно было срублено?

В. Приведите ниже один пример того, как машины помогают людям выполнять свою работу.

Источник: Международная ассоциация по эвалюации образовательных достижений, 1998, пример задания.

ной или двух колонках. Тестируемые выбирают ответ, заштриховывая соответствующий кружок, обводя букву алфавита или цифру, или отмечают указанным символом выбор части текста или графика. Как правило, задание содержит только один правильный ответ и несколько правдоподобных, но неправильных.

Закрытые задания с конструируемым ответом (см. дополнительный материал 2.4) предполагают, что тестируемый должен сам получить правильный ответ. Способы могут быть различными. В закрытых заданиях с конструируемым ответом требуется вставить одно или два пропущенных слова, подчеркнуть слово или число в тексте или таблице, начертить линию на графике или заштриховать область диаграммы. В закрытых заданиях с конструируемым ответом тестируемым предлагают выбрать несколько вариантов ответа, которые соответствуют определенным критериям, или установить соответствие наборов пар предложений или диаграмм.

Открытые задания с кратким ответом (см. дополнительный материал 2.5) подразумевают, что учащиеся должны дать ответ самостоятельно, который может существовать в различных, но правильных вариантах. Как правило, ответ предполагает объяснение, описание процесса или подробный рисунок (больше, чем одна или две линии). Может потребоваться, чтобы испытуемый написал одно или два предложения, завершил ряд последовательных действий или решение уравнения, до-строил график, диаграмму или схему.

В ходе выполнения задания с развернутым ответом (см. дополнительный материал 2.6) испытуемые должны дать подроб-

Дополнительный материал 2.6**Пример краткого эссе**

Краткое эссе, или задание с развернутым ответом, может быть представлено в виде текста или иллюстрации. Важно, чтобы учащиеся понимали, какой ответ они должны дать, – доказать свою точку зрения, высказать мнение по какому-либо вопросу, написать текст-рассуждение или рассказ. Соответствующая информация об этом обеспечивается в инструкции по предъявлению теста. Рисунок, представленный ниже, использовался в Папуа – Новой Гвинеи, чтобы в процессе написания рассказа оценить языковые навыки учащихся. Администратор по тестированию предупредил учащихся, что они могут взять за основу как изображенный на рисунке сюжет, так и написать историю об охоте самостоятельно.



Напишите историю об охотниках, изображенных на рисунке.

Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

ный, иногда сложный ответ на поставленный проблемный вопрос или написать мини-сочинение на заданную тему. Ответ может занять одну или несколько страниц, иногда он может включать схемы и диаграммы. Существуют определенные «правильные» способы выполнения таких заданий.

Первые три формы заданий по сравнению с четвертой являются более предпочтительными и чаще используются в национальной оценке во многом из-за высокой стоимости надежного оценивания вручную большого числа эссе. В структуре теста должно быть заложено процентное соотношение различ-

ных видов заданий в окончательном тестировании. Она должна также включать образцы всех видов заданий, чтобы помочь членам руководящего комитета и другим заинтересованным лицам ознакомиться с подходом к оценке.

Выбор формы задания и способа оценивания заданий существенно влияет на конечную стоимость теста. Проверка заданий, которые требуют ручного оценивания (имеется в виду экспертная проверка. – *Примеч. ред.*), стоит дороже и занимает больше времени, таким образом задерживая публикацию отчета о результатах исследования. Должны быть разработаны руководства по оцениванию вручную, также следует пригласить и обучить экспертов. Чем сложнее руководства по оценке, тем выше затраты. Поэтому задания с развернутым ответом ведут к увеличению стоимости исследования. Оценка заданий с множественным выбором стоит гораздо дешевле, однако их разработка является более дорогостоящей. Некоторые проблемы финансовых затрат, которые следует учитывать при выборе формы заданий, в общих чертах описаны ниже.

Процедура проверки ответов испытуемых на задания с множественным выбором, как правило, автоматизирована. Сами задания оценивать вручную перед анализом ответов не требуется. Необходимо лишь ввести ответы учащихся в компьютер с помощью сканера или вручную. Сканирование является наиболее экономичным для крупномасштабного тестирования. Оно требует специального оборудования и некоторой технической резервной поддержки. Обычно оно ведет к ограничению форм представления ответа (например, только штриховка кружков). Широкое разнообразие форм (например, обведение слова или буквы кружком, проставление символов в пустых квадратиках или подчеркивание) для обозначения правильного варианта возможно в том случае, если ввод данных осуществляется вручную.

Задания с множественным выбором нельзя оценивать только как правильные или неправильные до начала анализа данных. Они могут стать источником ценной диагностической информации о выполнении теста учащимися путем анализа каждого выбранного варианта ответа. Перед сканированием нужно обязательно убедиться, что все ответы учащихся были введены. Зат-

раты на ручной ввод данных для заданий с множественным выбором уменьшаются, если они расположены по согласованной схеме.

Закрытые задания с конструируемым ответом следует оценивать вручную, потому что могут быть одинаково приемлемы несколько вариантов ответов. Руководства по оценке должны определить диапазон приемлемых и допустимых ответов. Обычно диапазон верных вариантов для закрытых заданий с конструируемым ответом ограничен. Руководства по оцениванию являются однозначными, и время для обучения эксперта может быть предельно кратким. При этом должен осуществляться и поддерживаться постоянный контроль качества проверки ответов.

Открытые задания с кратким ответом также оцениваются вручную. Число и диапазон допустимых и недопустимых вариантов ответов может быть довольно широким. Следовательно, руководства по оценке, как правило, весьма сложные и требуется тщательная разносторонняя подготовка экспертов. Для контроля качества результатов их работы во время экспертной оценки чрезвычайно важна непрерывная перепроверка.

Задания с развернутым ответом имеют сложные руководства, поэтому эксперты должны быть настоящими профессионалами. Существенной для контроля качества является непрерывная двойная экспертная оценка некоторых или всех ответов учащихся. Сами руководства по оценке требуют подробной разработки в учебных целях. Они должны включать примеры ответов, соответствующих каждому из уровней оценивания.

В одном и том же тесте могут быть представлены разные формы заданий. Например, в тест может входить задания с множественным выбором, несколько закрытых заданий с кратким ответом и несколько открытых заданий с конструируемым ответом.

Авторы заданий должны стремиться к тому, чтобы форма каждого задания была выбрана правильно, чтобы оно эффективно оценивало усвоение учащимся планируемого результата обучения. Также следует свести до минимума количество не важных для оценки текстов, описаний или вычислений, требующихся для ответа на отдельное задание.

ТАБЛИЦА 2.4**Преимущества и ограничения форм тестовых заданий**

Задания	Преимущества	Ограничения
С множественным выбором	<ul style="list-style-type: none"> • Большое число заданий в одном teste может охватывать широкий диапазон результатов обучения. • Задания обеспечивают высокую дифференциацию учащихся по знаниям и пониманию учебного материала. • Оценивание вручную не требуется, поэтому проверка относительно недорога 	<ul style="list-style-type: none"> • Для создания высококачественных заданий требуется экспертиза. • Не выявляют понимание учащимися учебного материала. • Имеется вероятность угадывания
Закрытые с конструируемым ответом (одно или два слова в ответе)	<ul style="list-style-type: none"> • Учащиеся определяют местонахождение информации или вспоминают ее самостоятельно. • Оценивание вручную является относительно стандартизованным 	<ul style="list-style-type: none"> • Задания обычно охватывают ограниченный диапазон результатов обучения (главным образом, поиск информации и проверка знания фактов)
Открытые с кратким ответом (одно или два предложения в ответе)	<ul style="list-style-type: none"> • От учащихся может потребоваться проявление высоких уровней знания и понимания. • Задания могут быть направлены на проверку широкого диапазона результатов обучения. • Иногда может быть измерено частичное понимание 	<ul style="list-style-type: none"> • Для написания четко сфокусированных заданий необходима экспертиза. • Требуется привлекать подготовленных экспертов и совершение мер по контролю качества, что способствует повышению затрат. • Задания требуют значительного времени на выполнение, следовательно, уменьшается диапазон результатов обучения, на которые эти задания нацелены
Эссе, или развернутый ответ	<ul style="list-style-type: none"> • Учащиеся могут продемонстрировать глубину понимания учебного материала. • Может быть измерен диапазон частичного понимания материала 	<ul style="list-style-type: none"> • Ограниченный диапазон результатов обучения, к которым задания могут быть обращены. • Требуются подготовленные эксперты и меры по контролю качества, что способствует повышению затрат

Решение о том, какие формы заданий использовать в тесте и в каком соотношении, должно основываться как на уместности выбираемой формы для измерения конструкта в соответствующей области знаний или умений, так и на практических ограничениях, например, со стороны процедуры экспертизы, необходимой для разработки различных форм заданий, и на затратах при проведении оценки вручную. Преимущества и недостатки различных форм заданий представлены в табл. 2.4.

В тестах по математике в Папуа – Новой Гвинее соотношение заданий с множественным выбором и заданий с кратким открытым ответом зависит от класса, в котором проводится проверка (см. табл. 2.5). В тесте для учащихся 8-го класса заданий больше, чем в тесте для учащихся 3-го класса. В основном увеличение длины происходит за счет включения в него большого числа открытых заданий с кратким ответом.

ТАБЛИЦА 2.5

Формы тестовых заданий по математике в Папуа – Новой Гвинее

Класс	Задания с множественным выбором	Открытые задания с кратким ответом
3-й	20	5
5-й	25	10
8-й	25	15

Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

Как правило, учащимся дают достаточно времени на выполнение большей части заданий. Очень важно собрать информацию о времени, которое необходимо учащимся для выполнения предварительного теста при апробации тестовых заданий. Тесты могут различаться по длине, но учащиеся должны быть в состоянии выполнить основную часть заданий приблизительно за 40 минут. Тесты, состоящие главным образом из заданий с множественным выбором, могут включать больше заданий, чем тесты, состоящие преимущественно из заданий с кратким ответом. Так же тесты для учащихся старших классов могут включать больше заданий, а испытуемым – предоставляться больше времени для их выполнения. Если тестируемые незнакомы с использованными

ми в данном тесте формами заданий, то вероятнее всего им потребуется больше времени для ответа на них.

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ УЧАЩИХСЯ ДЛЯ ПРОВЕДЕНИЯ ОЦЕНКИ

Структура теста должна четко определять генеральную совокупность испытуемых для проведения оценки (например, учащиеся 4-го класса) и указывать, почему выбраны именно эти учащиеся. Например, в отдельно взятой стране документ по структуре мог бы, например, обосновывать отбор 3-х классов на том основании, что после 3-го класса значительное число учащихся покинут школу. Выбор для тестирования 4-х классов можно обосновать необходимостью проверки навыков чтения, которые должны быть приобретены по завершении данного этапа обучения. Также можно подчеркнуть, что 4-й класс – это заключительный класс начальной школы, и, соответственно, важно оценить учебные достижения учащихся именно на этом переходном этапе в системе образования.

В структуре теста могут быть определены испытуемые, которые не входят в национальную выборку, например дети с особыми образовательными потребностями или учащиеся маленьких школ, находящихся в весьма отдаленных районах.

СООБЩЕНИЕ РЕЗУЛЬТАТОВ

До проведения национальной оценки необходимо условиться с членами руководящего комитета о том, какие результаты должны быть приведены в отчете. Например, в Ирландии по итогам национальной оценки чтения на английском языке приводятся отдельные результаты, связанные с видом текста и познавательными процессами. Структура теста предполагает построение для учащихся 1-го класса двух шкал по видам текстов (литературный и информационный) и двух шкал по видам процессов (поиск информации и интерпретация выводов). Для учащихся 5-го класса предлагается построить три вида субшкал по тексту (литературный, информационно-не-

прерывный и информационно-дискретный) и три вида шкал для процессов (поиск, вывод и оценочное суждение в форме интерпретации) (Эивес (Eivers) и др., 2005). В Международном обзоре исследования грамотности взрослых использовались дискретные тексты, чтобы оценить выполнение теста по шкале оценок для отдельных напечатанных документов. Структура теста для Международного исследования прогресса в грамотности чтения (PIRLS) определила, что шкала тестовых заданий должна быть построена в двух целях: чтение для получения литературного опыта и чтение для получения информации и ее дальнейшего использования. Структура также предлагает объединить обе шкалы и благодаря этому дать полную оценку грамотности чтения (Кэмпбелл (Campbell) и др., 2001).

Руководящий комитет должен быть проинформирован о том, что отчет о результатах в субшкалах зависит от результатов анализа заданий. Во многих международных и национальных системах оценки адекватность тестовых заданий субшкалам помогает определить моделирование ответов. На данном этапе руководящему комитету можно предоставить концепцию отчета о результатах оценки, подготовленного с использованием уровней учебных достижений, обычно называемых «уровнями мастерства», и выяснить взгляды его членов относительно предпочтительной формы представления результатов тестирования.

Можно привести процент правильно выполненных заданий. Можно обозначить достигнутый уровень усвоения – продвинутый, профессиональный (отвечает ожидаемому стандарту выполнения теста), базовый (не отвечает ожидаемому стандарту) или ниже базового (результаты тестирования ниже базового уровня). В результате анализа данных, полученных в ходе предварительного и окончательного тестирования, число уровней мастерства, возможно, придется пересмотреть. Четвертая книга настоящей серии, «Анализ данных национальной оценки учебных достижений», охватывает как современную теорию тестов (Item Response Theory – IRT), так и уровни мастерства.

В структуре теста следует также обозначить виды отчетов, которые будут опубликованы по результатам национальной оценки. Эти отчеты могли бы включать методический раздел и разделы для заинтересованных читателей, таких как работающие в школах преподаватели, составители учебного плана и тактики в образовании. Также они могут использоваться в пресс-релизах при проведении брифингов.

КОНТЕКСТ

Многие тактики в системе образования используют национальную оценку для сбора дополнительной контекстной информации (синоним «фоновая информация». – *Примеч. ред.*) о факторах, непосредственно влияющих на качество обучения по отдельным предметам учебного плана. Руководящий комитет должен предоставить общее руководство по отбору контекстных переменных, которые должны быть оценены. Эта информация, в свою очередь, может использоваться исполнительным агентством для управления процессом разработки анкет. Контекстную информацию используют тактики для выявления причин появления различий в уровнях достижений учащихся.

Как правило, на учебные достижения влияют две группы контекстных факторов – «домашние» и «школьные». К «домашним» относят: социально-экономический статус, иногда измеряемый имеющейся в семье собственностью, уровень образования родителей, язык общения, отношения в семье, количество членов семьи, внимание, уделяемое в семье процессу обучения, и то, как поставлен этот процесс (например, чтение детям или иные действия, поощряющие детей к учению), контакты с преподавателями². Под «школьными» факторами подразумевают имеющиеся в школе ресурсы, состояние кабинетов, организацию школьного управления, используемые стратегии обучения, квалификацию педагогов, особый подход преподавателя к ученику. В ходе некоторых оценочных мероприятий собирают материал об отношении учеников к школе и отдельным предметам, интересах учеников и их поведении

(например количество времени, затраченное на помощь по дому, работу или чтение для удовольствия).

ПРИМЕЧАНИЯ

- ¹ После того, как исследование было проведено в третий раз, оно получило название «Международное исследование тенденций в математическом и естественнонаучном образовании», и аббревиатура TIMSS была сохранена.
- ² Отметим, что в некоторых странах существует отрицательное отношение к сбору контекстных социально-экономических данных.

ГЛАВА

3

НАПИСАНИЕ ЗАДАНИЙ

В этой главе раскрыты основные характеристики грамотно составленных тестовых заданий, принципы их написания, структурирование и компоновка заданий в teste и их оценка¹. Рассмотрены задачи сотрудников, участвующих в разработке теста. Это команда по написанию заданий и привлекаемые эксперты. Общее руководство осуществляют менеджер по разработке тестов.

Необходимо учитывать, что качество теста в значительной степени определяется тем, в какой степени он отвечает цели создания и насколько предложенные испытуемым задания соответствуют разработанному проекту теста. Хорошее задание является однозначным, релевантным учебному плану и направленным на проверку одной единицы учебного материала. Такие задания обеспечивают включение в тест аутентичных задач, благодаря чему учащиеся, говорящие на разных языках и принадлежащие к семьям с различным культурным фоном, оказываются в равных условиях.

Грамотно написанное задание обладает следующими характеристиками:

- обращается к ключевой области изучения;
- является конструктивным и значимым;

- соотносится с основными характеристиками, заявленными в структуре или проекте теста;
- справедливо;
- направлено на решение основных проблем в учебных стимулах, при этом второстепенные или незначительные детали во внимание не принимаются;
- ясно обозначает задачу, которую испытуемые обязаны выполнить;
- оно не является следствием других заданий и его выполнение не зависит от понимания учебного материала, которое сформировалось на основе выполнения предыдущих заданий;
- трактовка терминов должна рассматриваться в контексте данного задания, а не из общих представлений;
- предпочтительно выражено в форме утверждения, форма отрицания нередко приводит к неоднозначному пониманию проблемы.

Авторам тестов могут помочь исследования моделей высококачественных заданий. Многие организации размещают типовые задания в Интернете. Источником для получения информации по разработке тестовых заданий могут также служить международные тесты, такие как Международное исследование тенденций в математическом и естественнонаучном образовании (TIMSS), Программа международного оценивания учащихся (PISA), Международное исследование прогресса в грамотности чтения (PIRLS), и национальные оценки в других странах, например американская Национальная оценка прогресса в образовании (NAEP).

Опубликованные тестовые задания могут использоваться в других тестах, если их содержание и формулировка соответствуют целям исследования. Это недорогой и полезный источник. Многие задания вместе с ключами доступны в Интернете. Их использование нежелательно, так как испытуемые также могли иметь к ним доступ до выполнения теста. Можно также получить разрешение на использование готовых, но ранее не

рассекреченных заданий. Это намного дешевле, чем разработка новых вариантов. Тем не менее эксперты по учебным планам должны проанализировать имеющиеся задания на соответствие их целям тестирования.

Проведение экспертизы при написании заданий по каждому разделу проекта теста занимает значительное время. Авторы должны придерживаться общей терминологии и иметь общее понимание конструкта, предназначенного для изменения. Для этого им необходимо классифицировать каждое разрабатываемое ими задание проекта по следующим критериям:

- форма задания (задания с множественным выбором, с закрытым конструируемым ответом, с открытым ответом);
- вид текста (для теста по чтению), например рассказ, описание, рассуждение;
- целевая группа тестируемых (например, учащиеся 5-го класса);
- проверяемые результаты учебной деятельности (умение складывать двузначные целые числа или выявлять основную мысль в рассказе);
- познавательный процесс (например, знание, рефлексия, интерпретация или синтез).

Уровень познавательного процесса, требуемого для выполнения определенного задания, невозможно определить точно. Так, если испытуемые не обладают умением подводить итог по пройденному материалу, то может понадобиться задействовать более высокий уровень познавательного процесса.

ТРУДНОСТЬ ЗАДАНИЯ

Определение требуемого уровня трудности заданий является основной задачей для их авторов. Во многих странах предписанный учебный план является для большинства учащихся слишком сложным. Как следствие, авторам заданий приходится обращаться к проверке более простых умений, ниже тре-

бований учебного плана. Например, в тесты учебных достижений по математике, разработанные для 5-го класса, зачастую включают задания, соответствующие уровню 3-го и 4-го классов.

Опытные преподаватели-практики, по сравнению с административными работниками и учеными, должны иметь более правдоподобное представление о необходимом уровне трудности заданий. Однако мнения преподавателей хотя и значимы, но не всегда адекватно отражают требуемый уровень трудности. Предварительное апробирование тестов на выборках испытуемых, подобных тем, что входят в целевую группу учащихся, учебные достижения которых подлежат оценке, является существенным фактором для получения начальных приближений к объективным данным об уровнях трудности заданий. Этот процесс поможет избежать распространенной ошибки – разработки тестов с заданиями, которые являются слишком трудными.

Иногда бывает так, что задания, которые кажутся авторам легкими, на деле оказываются весьма трудными. Или, напротив, задания, которые изначально разрабатываются как трудные, на деле могут быть сформулированы так, что испытуемым не составит труда их выполнить. В случае разработки заданий с множественным выбором авторам необходимо избегать следующего.

- Наличие в задании грамматических или логических подсказок в основной части задания, которые могут косвенно указать на правильный ответ. Например, если задание предполагает ответ – существительное в единственном числе, а предлагаемые варианты ответов, за исключением одного (верного), содержат существительные во множественном числе.
- Использование «абсолютных» терминов, таких как «всегда» или «никогда», которые автоматически исключают некоторые варианты ответов или прямо указывают на правильный ответ.
- Большая длина или детализированность правильного ответа по сравнению с другими вариантами.

- Включение ключевого слова или фразы из основной части задания только в правильный ответ, но не в остальные варианты.
- Расположение вариантов в нелогичном порядке или путаница в самих ответах.
- Написание вариантов ответов и ключевой смысловой основы таким образом, что выбор «наилучшего» ответа в значительной степени будет зависеть от языковых навыков учащихся, а не от знания того, что проверяется.
- Включение вопросов, на которые можно ответить независимо от постановки проблемы в задании (имеется в виду выбор ответа на основе здравого смысла и вербальных ассоциаций между основной частью и правильным ответом. – *Примеч. ред.*).

В случае форм заданий с конструируемыми ответами важно разработать ясные критерии оценки, которые выявляют тот конструкт, который измеряется в задании.

СИСТЕМАТИЧЕСКАЯ ОШИБКА ЗАДАНИЯ

На выполнение теста учащимися немаловажное влияние оказывает своеобразие культурного опыта, которым они обладают. Они не должны оказаться в заведомо лучших или худших условиях из-за контекста, который не имеет прямого отношения к проверяемым знаниям, умениям и пониманию учебного материала, проверяемого в teste. Например, задания о мужских видах спорта могли бы поставить женщин в невыгодное положение при выполнении теста.

Также смещающее влияние на оценки оказывают задания, затрагивающие субъективные ощущения некоторых учащихся. Материал основной части задания не должен затрагивать этические, культурные или иные чувства учащихся. Необходимо полностью исключить вероятность того, что некоторые испытуемые будут оскорблены, напуганы или расстроены из-за материала, представленного в teste. Специалист по разработке

тестов должен научить авторов предвидеть возможные случаи появления систематической ошибки. Экспертам следует уделять внимание поиску «обидных» для отдельных категорий испытуемых заданий тестов или анкет, которые могут оказывать смещающее влияние на оценки учащихся. В четвертой книге настоящей серии, «Анализ данных национальной оценки результатов обучения» рассмотрены статистические методы, которые можно использовать для идентификации подобных заданий как на предварительной, так и на заключительных стадиях тестирования.

МАТЕРИАЛ СТИМУЛА

Материал стимула (под стимулом принято понимать общую информационную часть содержания для нескольких, обычно не более 3–5, тестовых заданий, обеспечивает контекст задания. Это может быть часть текста, диаграмма, график, таблица, карта, схема или их сочетание.

Как правило, разработка теста начинается с подготовки соответствующего материала стимульной части заданий. Так, тесты по чтению обычно основаны на больших текстах, относящимся к ряду заданий или юниту, которые адресованы диапазону релевантных умений. В тестах по математике и естественным наукам материал стимула может быть кратким, например несколько чисел для сложения или уравнение, которое требуется решить, а может быть и весьма сложным – график, схема, таблица или диаграмма – и связанные с ними задания для определенной области умений.

Материал стимула должен содержать основные характеристики, подлежащие оцениванию. Лишние, повторяющиеся или ненужные детали необходимо исключить. Хороший стимул обладает следующими характеристиками:

- обоснован и заслуживает внимательного рассмотрения;
- представляет интерес для потенциальных испытуемых;

- грамотно запланирован и написан;
- оптимален для мотивации испытуемых: не слишком трудный и не слишком легкий;
- не ставит несуществующих проблем;
- содержит достоверные факты;
- дает возможность поиска ответов на вопросы;
- является самодостаточным.

Важно задать некоторый контекст и стимул там, где это возможно, например с помощью заголовка или краткого введения. Так, выдержка из научно-фантастического романа может быть введена следующим образом: «Данный отрывок взят из романа, действие которого происходит в будущем».

Иллюстрации должны быть естественной частью стимула, предоставляя испытуемым дополнительную информацию. Если они включены в тест просто с целью художественного оформления, то они не должны помогать тестируемым.

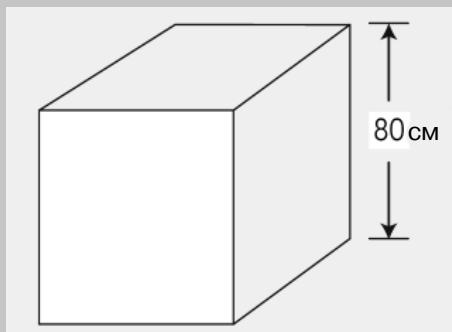
Иногда материал стимула создает искусственный и ненужный для задания контекст. Например, дополнительный материал 3.1 содержит иррелевантную информацию.

Это задание направлено на проверку умения учащихся вычислять площадь поверхности куба. В реальной жизни упомянутая в задании Ирина вряд ли будет обеспокоена проблемой определения наименьшего количества бумаги, необходимой для обертывания куба. Если подумать, то ей необходимо дополнительное количество бумаги для того, чтобы завернуть ее по краям. И вот как раз самые способные учащиеся, догадавшиеся о необходимости соответствующей поправки, могут ответить на вопрос неверно. Поэтому задание лучше сформулировать проще: «Длина одного ребра куба составляет 80 см. Какова площадь поверхности куба?»

Стимул, представляющий реальную жизненную ситуацию, должен точно соответствовать действительности. Информация, приведенная в дополнительном материале 3.2, этим свойством не обладает. Люди обычно не фиксируют изменение своего роста таким образом, как это приведено в примере. Склонность к быстрому росту чаще всего проявляется в раннем возрасте. Если необходимо привести примеры разных гра-

Дополнительный материал 3.1**Пример иррелевантного материала стимула**

Ирина оборачивает этот куб бумагой. Какое наименьшее количество упаковочной бумаги ей потребуется?



фиксов роста в материале стимула, то в целях сравнения более уместно использовать, скажем, заводы, но не людей.

Задания по возможности должны быть написаны на простом и ясном языке. Их формулировка должна быть понятной, такой, чтобы у испытуемых не возникло дополнительных сложностей в восприятии. Поэтому необходимо избегать:

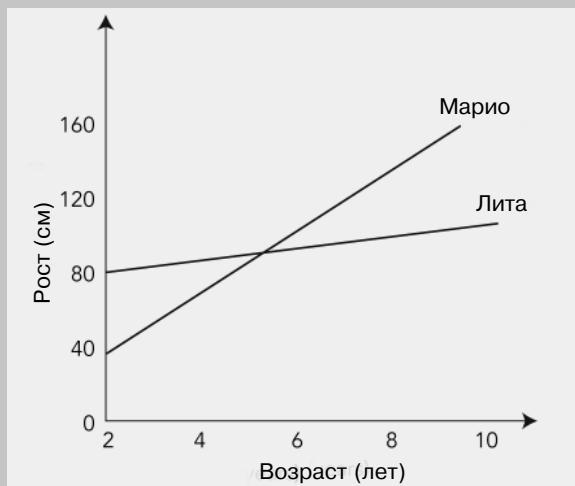
- трудных слов;
- длинных и сложных предложений;
- усложненной логики;
- неоднозначных или неопределенных вопросов;
- двойных отрицаний;
- несогласованности (например, использования различных единиц измерения при характеристике одной и той же величины);
- неопределенных формулировок или незнакомых слов, которым не дается соответствующее определение.

ФОРМА ЗАДАНИЯ

В настоящем параграфе описываются две основные формы заданий: с множественным выбором и с кратким ответом (см. главу 2)². При использовании заданий с множественным выбором важно:

Дополнительный материал 3.2**Пример задания с неточной или вводящей в заблуждение информацией**

График отражает изменения в росте Марио и Литы по мере их взросления



- ограничить число опций ответов;
- привести краткие ответы;
- избегать необходимости переписывания учащимися больших разделов текста из основной части задания (избегать повторов в ответах и т.д.);
- охватить широкий спектр тем.

При подготовке заданий с кратким ответом (закрытым конструируемым или открытым) необходимо:

- уточнить смысл, который тестируемые, вероятнее всего, будут вкладывать в задание;
- учесть различные уровни понимания пути выполнения заданий посредством использования частичного кредитования при его оценивании, присваивая полную оценку ответу, показывающему понимание сложной идеи, и часть оценки для ответа, иллюстрирующего понимание учащимся простых идей;
- обозначить ограниченную и ясно определенную область возможных правильных ответов;

- искать ответ в ситуации, где правильный вариант четко отличается в формате с множественным выбором от вероятных неправильных вариантов.

Испытуемые должны обладать адекватным лексическим запасом, чтобы ответить на такие задания. Если учащиеся, по вашему мнению, скопируют при формулировании краткого ответа большую часть текста стимула, не используйте подобные задания.

Написание заданий с множественным выбором

Задание с множественным выбором состоит из основной части и предложенных готовых ответов. Иногда, когда требуется ответ «истина–ложь», используются только два ответа. Однако такие задания считаются малоэффективными. Чаще используются четыре или пять вариантов ответов. Правильный ответ называется ключом, а неправильные – дистракторами (дополнительный материал 3.3).

Основная часть задания может быть представлена в нескольких формах:

- незаконченное предложение;
- четко сформулированный вопрос;
- предложение с недостающей информацией
(Карлос имеет_____братьев).

Дополнительный материал 3.3

Пример задания с множественным выбором

У девочки 3 цветка. Отец дал ей еще 2 цветка.
Сколько цветков у девочки всего?

← основная часть

- A** 2
- B** 3
- C** 4
- D** 5

дистракторы

ключ (правильный ответ)

Если основная часть имеет форму неполного предложения, она должна быть достаточно информативной, чтобы увидеть ключ к заданию. У испытуемых не должно возникнуть потребности в чтении дистракторов. Варианты ответов на неполное предложение должны:

- быть грамматически согласованными с основной частью задания;
- быть написанными в одинаковом стиле;
- быть правильно акцентированными;
- начинаться со строчной буквы и заканчивать предложение.

При написании заданий с множественным выбором выполняйте следующие правила.

- *Сосредоточьте внимание на правильном завершении предложения.* В дополнительном материале 3.4 варианты ответов являются законченными предложениями (пунктуация соответствующая).
- *Обращайте внимание на перечни.* В дополнительном материале 3.5 варианты ответов даны в форме перечня. Знаки препинания в данном случае не требуются.

Дополнительный материал 3.4

Расстановка знаков препинания в случае, если варианты ответов представляют собой законченные предложения

Что подумал Михо о рынке?

- A** Он был переполнен, и продукты были хорошими.
B Он был дешевым, и продукты были восхитительными.
C Продукты были хорошими, но там никого не было.
D Он был дешевым, но продукты были не очень хорошими.

Дополнительный материал 3.5

Расстановка знаков препинания, если варианты ответов даны в форме перечня

Как долго Джо оставался в доме своего дедушки?

- A** одну неделю
B две недели
C один месяц
D два месяца

- *Сократите текст задания до минимума.* Для этого вынесите повторяющиеся слова в основную часть, и благодаря этому сделайте варианты ответов максимально краткими (см. дополнительный материал 3.6).
- *Избегайте отрицания в основной части.* Его использование может повлечь неправильное понимание задания. Если отрицания избежать нельзя, то частицу «не» вынесите в начало, выделив ее полужирным или курсивным начертанием (см. дополнительный материал 3.7). При этом варианты ответов ни в коем случае не должны включать отрицание.
- *Чередуйте пары дистракторов.* Для подбора дистракторов необходимо использовать разнообразные методы. Единобразие может косвенно помочь учащемуся при выборе правильного ответа. Так, нежелательно объединение ключа (В) с противоположным ответом (А) (дополнительный материал 3.8).

Дополнительный материал 3.6

Сокращение текста задания

Неправильно

Карл поехал

A к реке со своей семьей.

B на море со своей семьей.

C в деревню со своей семьей.

D в горы со своей семьей.

Правильно

Карл и его семья поехали

A к реке.

B на море.

C в деревню.

D в горы.

Дополнительный материал 3.7

Задание с отрицанием в основной части

Что, по словам родителей Марио, их сын **не** должен держать в доме?

A его любимую собаку

B его вонючие ботинки

C попону

D корзину фруктов

Дополнительный материал 3.8**Некорректный подбор слов в дистракторах**

Том не любил пальто, потому что оно было

- A** слишком большое.
- B** слишком маленькое.
- C** неподходящего цвета.
- D** недостаточно теплое.

Дополнительный материал 3.9**Корректное сочетание основной части с дистракторами**

Том не любил пальто, потому что оно было слишком

- A** большое.
- B** легкое.
- C** маленькое.
- D** тяжелое.

Если варианты ответов будут приводиться по определенному принципу (как в дополнительном материале 3.8), то этот принцип вскоре станет очевидным для некоторых учащихся, обладающих значительным опытом тестирования. В результате они будут рассматривать только такие пары ответов, которые объединены противоположным смыслом (как А и В). Выход состоит в том, чтобы написать несколько заданий, в которых ключ не будет соединяться с противоположным по смыслу дистрактором. Другой возможный вариант решения проблемы – включение в задание двух пар противоположных по смыслу ответов (имеется в виду двойная альтернатива: большое – маленькое, легкое – тяжелое. – Примеч. ред.) (см. дополнительный материал 3.9).

- *Избегайте категоричных утверждений в дистракторах.* Не следует использовать дистракторы, которые содержат такие слова, как «всегда», «никогда», «ни один из вышеупомянутых» и «все вышеупомянутые», потому что такие варианты обычно легко исключаются учащимися из рассмотрения.

- *Используйте подходящее число дистракторов.* Если это возможно, разработайте задания с одним ключом и четырьмя правдоподобными ответами (т.е. всего пять вариантов ответа) и предварительно проверьте все дистракторы. Используйте дистракторы, которые имеют лучшие статистические свойства (подробнее см. четвертую книгу из этой серии).
- *Изменяйте месторасположение ключа.* Место для ключа от задания к заданию должно изменяться. Оно не должно быть постоянным. Варианты ответов могут быть ранжированы в зависимости от размера (от самого короткого до самого длинного или наоборот) или располагаться в случайном порядке. Убедитесь в том, что верный ответ не является всегда самым длинным.

Хорошие варианты ответов обладают следующими характеристиками.

- Они имеют сходную длину и написаны в том же стиле, что и ключ. Последний не должен выделяться длиной, формулировкой или иными внешними признаками.
- Стилистика вариантов меняется от задания к заданию. Они не должны повторяться.
- Предлагаемые варианты ответов в задании не являются косвенной подсказкой для выполнения других заданий.
- Недопустимы частично правильные дистракторы – одновременно содержащие правильный и неправильный ответы либо часть правильного ответа.
- Формулировка ответов не вводит учащихся в заблуждение из-за недостаточной ясности или двусмысленности.
- Ответы обладают различной семантикой. Дистракторы не должны быть синонимами. Значение одного дистрактора не может быть одним из значений другого.
- Они включают ключ, который является бесспорно правильным или обоснованно точным ответом на вопрос, а не просто лучшим из представленных вариантов ответов.
- Они имеют дистракторы, которые являются бесспорно неправильными ответами, но при этом они разумные и правдоподобные. Любой дистрактор, который является абсурд-

ным, сокращает количество реальных опций для выбора, доступных для учащегося, и ничего не вносит в задание.

Написание заданий с кратким ответом

Задания с краткими ответами должны быть направлены на оценку именно тех умений, для проверки которых они планируются. Хорошие задания отличаются ясностью и точностью. Одновременно с самими заданиями должно разрабатываться руководство по оценке.

Задания с краткими ответами классифицируются как открытые, когда для ответа требуются одно или два предложения или несколько дополнений к схеме. У них обычно существует определенный диапазон возможных правильных ответов. Если для ответа достаточно одного или двух слов или одной линии на схеме, то задания классифицируются как закрытые. Обычно у таких заданий есть весьма ограниченный набор правильных ответов.

Открытые задания предназначены для проверки наиболее существенных навыков и умений по основным разделам учебного плана, поскольку это оправдывает большие временные затраты на их выполнение. Учащиеся должны уметь давать краткие формулировки ответов на открытые задания. Ведь большая часть времени, которое дается на выполнение задания, должна быть посвящена его решению, а не записи ответа.

Разрабатывая задания с кратким ответом, важно понимать, какой ответ можно считать неправильным. Если все возможные варианты близки к верному, вклад такого задания в оценку отдельного умения будет весьма незначительным. Задание должно быть сконструировано таким образом, чтобы гарантировать наличие правдоподобных, но неправильных ответов.

Следует убедиться, что задания с кратким ответом имеют более чем два возможных варианта ответа. Если существуют только два варианта, такие как, например, «открытый» или «закрытый», то учащихся появляется 50-процентный шанс на «угадывание». Область ответов такого задания можно расширить, попросив учащихся указать причины для того или иного

ответа. Тогда задание можно будет оценить на основе правильного выбора, например «открытый» – «закрытый», и его объяснения. Если учащийся приводит правильный вариант, но не дает объяснение, он получает нулевую оценку.

Задания не должны содержать развернутые объяснения для понимания смысла стимулов. Так, задание не должно суммировать основные идеи или делать вывод для объяснения стимула. Вместо интерпретации смысла стимула предпочтительно его цитировать.

Опасность при разработке открытых заданий кроется в том, что учащиеся могут отвечать на них поверхностно. К примеру, ответ «потому что это важно» формально может быть правильным, но недостаточным. Иногда такой поверхностный ответ следует включить в сам вопрос, тем самым изначально устраивая его из диапазона возможных правильных ответов.

Например, задание может быть сформулировано следующим образом: «Почему несчастный случай с Джемимай важен в этой истории?» При такой формулировке учащиеся не могут ответить на это задание «потому что несчастный случай важен». Чтобы избежать сжатого ответа типа «да» или «нет» инструкции: «Поясните ваш ответ» или «Приведите аргументы» просто необходимы.

Эффективное задание с кратким ответом должно ставить ясную, определенную задачу, которая предполагает определенный ответ. Такое задание позволяет учащимся продемонстрировать свое владение необходимым умением достаточно быстро и полно. Пример в дополнительном материале 3.10 не в состоянии реализовать ни одну из этих целей.

Учащихся не предупреждают, что они должны сделать коробку с максимально возможными размерами. Однако этот ответ является условием для получения максимального балла – 3. Низкое качество задания обусловлено также излишней трудоемкостью. Количество времени, которое учащиеся должны потратить на поиск вариантов сложения коробки, чтобы добиться правильного ответа, не входит в характеристики проверяемого умения. Необходимо упростить задачу таким обра-

Дополнительный материал 3.10**Пример неграмотно составленного задания с кратким ответом**

Открытая коробка должна быть сделана из прямоугольных листов картона со сторонами 150 см и 100 см, исключая квадраты равного размера на каждом углу и с помощью клейкой ленты для соединения граней. Какого размера квадрат Вы вырежете на углу? Приведите аргументы для вычисления этого размера.

Критерий оценки: число баллов – в диапазоне от 0 до 3.

3 балла: учащийся описывает квадрат со сторонами 20 см и коробку с размерами 110 см на 60 см на 20 см; также объясняет, что коробка этого размера имеет наибольшую вместимость.

2 балла: описывает квадрат со сторонами 20 см, который вырезают на каждом углу, но не дает объяснений.

1 балл: описывает любые возможные размеры квадрата со стороной меньше чем 50 см.

0 баллов: предлагает квадраты со сторонами большими, чем 50 см (неприемлемый размер).

9: пропускает задание.

зом, чтобы учащиеся могли продемонстрировать соответствующие умения максимально эффективно.

Если испытуемый по какой-либо причине пропускает задание (например, из-за трудности или из-за отсутствия в выданном тестовом буклете), то заданию присваивается код 9, который не является оценкой, а просто свидетельствует о наличии недостающих данных по тесту. Пропущенные данные подробно рассматриваются в третьей книге настоящей серии – «Проведение национальной оценки учебных достижений».

Задания с кратким ответом должны быть сформулированы ясно и просто (см. дополнительный материал 3.11).

Разработка руководств по оценке заданий с частичным кредитованием

На некоторые задания теста с краткими ответами учащиеся приводят частично правильные ответы, которые можно отне-

Дополнительный материал 3.11**Пример грамотно составленного закрытого задания с конструируемым ответом**

Каждый человек роет яму в саду с одинаковой скоростью.

Один человек может закончить рытье ямы в саду через 12 часов.

Два человека могут закончить рытье ту же самую яму в саду через 6 часов.

Какое время потребуется на рытье ямы четырем работникам? _____

Какое время потребуется на эту работу, если ее выполняют

X человек? _____

сти к двум или более категориям. Такие задания получили название «задания с частичным кредитованием». Руководство по оценке таких заданий должно дифференцировать полные, точные или сложные ответы от частично правильных. Лучшие ответы заслуживают более высокой оценки. Пример в дополнительном материале 3.12 служит руководством по оценке задания с частичным кредитированием (за выполнение которого можно получить максимальную оценку 3 балла).

Дополнительный материал 3.12**Задание с частичным кредитированием**

Длина стороны квадрата – 10 см.

Нарисуйте этот квадрат. Используйте собственную линейку.

Рекомендации по оценке

3 балла: учащийся рисует квадрат с четырьмя сторонами длиной по 10 см и четырьмя прямыми углами.

2 балла: рисует прямоугольник с двумя длинами сторон 10 см и четырьмя прямыми углами.

1 балл: рисует четырехугольник с двумя длинами сторон по 10 см, но не с прямыми углами.

9: пропускает задание.

Следующие виды заданий могут оцениваться как задания с частичным кредитованием.

- Учащихся просят назвать две причины характерного поведения. Те, кто сможет привести обе правильные причины, получат 2 балла, а те, кто вспомнит только одну, – 1 балл.
- Учащиеся получают более высокую оценку за более глубокое понимание; например, в тесте по чтению оценка в 2 балла отражает умение испытуемого видеть в приведенном отрывке иронию, тогда как оценка в 1 балл дается за буквальное понимание текста.
- Оценка 2 балла ставится за верную идентификацию и причин, и следствий, тогда как 1 балл – за правильную идентификацию только одной из вышеназванных категорий.
- В математике 3 балла даются за правильное решение задачи и объяснение способа решения, 2 балла – за правильное решение без объяснения, а 1 балл ставится, когда при верно выбранном способе решения получен неправильный результат вычислений.

Необходимо установить четкое различие между полной оценкой и частичным кредитованием. Также нужно убедиться в том, что ответы, оцениваемые в 1 балл, приведенные в руководстве по оцениванию, не являются плохо сформулированными или сокращенными версиями тех ответов, которые оцениваются в 2 или 3 балла. Также важно установить различия между ответами, за которыедается 1 балл, и полностью неверными ответами. Зачастую сделать это бывает весьма трудно.

Следующие примеры показывают, что открытые задания с краткими ответами не всегда позволяют ввести частичное кредитование при оценивании. Например, в задании дополнительного материала 3.13 учащиеся могут дать множество ответов и получить оценки 1 балл или 0 баллов.

Необходимо провести предварительную проверку заданий с частичным кредитованием, чтобы убедиться в том, что предлагаемые категории оценок по каждому заданию являются статистически устойчивыми (см. главу 5). Пример задания с кратким ответом с нечетким руководством по выставлению балла

Дополнительный материал 3.13**Пример открытого задания с кратким ответом
с руководством по оценке**

Джон и Майкл находят дерево с 400 плодами манго.

Джон говорит, что у Майкла есть шанс 160% на попадание в плод манго.

Вы согласны или не согласны с Джоном?

Докажите свою точку зрения.

Руководство по оценке

1 балл: учащийся не соглашается и объясняет свое несогласие предельной величиной процентов;
не соглашается, потому что невозможно получить 160%;
не соглашается, потому что это невозможно;
не соглашается, потому что 100% – максимально возможная величина.

0 баллов: соглашается (с объяснениями или без них);
не соглашается, но не связывает свой ответ с предельной величиной процентов;
не соглашается, потому что на дереве есть больше чем 160 манго.

9: пропускает задание.

Источник: Филиппинский отдел образования, 2004.

представлен в дополнительном материале 3.14. В предварительном тестировании учащиеся, ответившие «80%» или «80 процентов», получили 2 балла, а те, кто ответил просто «80», – 1 балл. Статистика показала, что у учащихся, которые получили за это задание 2 балла, средняя оценка по математическому тесту была намного выше, чем у учащихся, получивших 1 балл. Последние получили примерно те же средние оценки, как и те, кто выполнил задание неверно и заработал 0 баллов. Благодаря данной информации, полученной в ходе предварительного тестирования, были внесены изменения в руководство по оценке: 1 балл выставлялся учащимся, которые ответили «80%» или «80 процентов», а 0 баллов – тем, кто ответил «80» или дал другой недопустимый ответ.

Дополнительный материал 3.14**Пример закрытого задания с конструируемым ответом и руководство по оценке**

Дерево Манго

Майкл пробует сбить плоды манго на своей ферме с помощью рогатки.

Если на дереве есть 50 плодов манго, то его шанс на попадание составляет 20%.

Если число плодов манго удваивается, то его шанс на попадание также удваивается.

Оцените шанс Майкла на попадание в плод манго, если на дереве имеется 200 плодов манго.

Руководство по оценке

1 балл: 80% или 80 процентов

0 баллов: любой другой ответ, включая «80» без указания «%» или «процентов»

9: пропускает задание.

Источник: Филиппинский отдел образования, 2004.

Написание заданий для юнитов

Юнитами называют группы заданий с общим стимулом. Юнит может содержать краткий рассказ или график, сопровождаемый рядом вопросов. Основные принципы написания заданий с множественным выбором или с кратким ответом применимы и к заданиям юнитов.

При проектировании заданий, основанных на юнитах, необходимо учесть следующие моменты.

- Задания должны быть независимыми друг от друга. Правильный ответ на одно из них не должен зависеть от того, правильно ли выполнены другие задания. Задания не должны подменять друг друга. Каждое задание должно оценивать свой четко дифференцируемый аспект стимула.
- Задания должны оценивать определенный диапазон конкретных умений. Например, задания не должны неоднократно оценивать поиск непосредственно устанавливаемой информации или главной идеи в каждом разделе стимула.

- Юнит должен включать задания разных уровней трудности, начиная, как правило, с самого легкого.
- Информация, приводимая в основной части задания или вариантах ответов, не должна помогать испытуемому при выполнении другого задания.
- Задания должны оценивать значимые аспекты стимула и исключать те, что являются тривиальными.
- Задания должны быть расположены на той же самой странице, что и юнит, или на развороте (в случае длинного стимула).

Юниты с восьмью или большим числом связанных с ними заданий имеют тенденцию к некоторому дублированию, подмене друг друга или тривиальности. Некоторые задания могут быть удалены во время предварительной экспертизы. Или команда по разработке теста может разработать два варианта предварительного теста – благодаря использованию половины заданий юнита в одном варианте, а оставшейся части – во втором.

ТРЕНИРОВОЧНЫЕ ЗАДАНИЯ

Тренировочные задания помогают получить гарантии того, что учащиеся не получили заниженные баллы из-за недостаточного знакомства с формой заданий и способами ответа на них. Обычно администратор по тестированию предъявляет тренировочные задания в соответствии с определенными инструкциями из руководства по применению теста.

Если в teste есть задания с кратким ответом, то особенно важно, чтобы учащиеся понимали, какого рода ответов от них ждут. Учащимся необходимо объяснить, скажем, куда вписать недостающее число, чтобы завершить последовательность чисел, или какой объем должен иметь текст, который требуется для объяснения ответа к заданию. Обязательно нужно подчеркнуть, что грамматические или орфографические ошибки не снижают конечный результат, если правильность написания текста не является предметом измерения. Инструкция должна быть написана таким образом, чтобы, прочитав ее, учащиеся постарались бы ответить на все задания теста.

Тренировочные вопросы должны разрабатываться для всех форм заданий. По этой причине тренировочные вопросы обычно составляются на финальном этапе разработки теста, когда все виды вопросов, которые будут в тесте, уже известны. Необходимо сделать практические задания максимально легкими; например, учащихся можно попросить указать ответ на задание $2 + 2$ на линии, изображенной рядом с суммой:

$$2 + 2 = \underline{\hspace{2cm}}.$$

Акцент делается именно на том, как учащиеся должны показать тренировочные задания или привести свои ответы.

В ходе экспертизы заданий нужно рассмотреть все тренировочные задания, которые также подлежат апробации. Если используются несколько вариантов предварительного теста, те же самые тренировочные задания должны использоваться с каждой формой теста.

РАСПОЛОЖЕНИЕ И ДИЗАЙН ЗАДАНИЙ

Важную роль для ясности теста играют грамотное расположение и дизайн заданий. Учащиеся, скорее всего, будут выполнять те задания, которые хорошо читаются и потому легки для восприятия. Разработчики тестов должны выбрать согласованный, «приятный» стиль расположения заданий. Для этого необходима помочь профессиональных дизайнеров и верстальщиков.

Основные принципы

Ниже приведены несколько важных принципов для успешного расположения заданий и дизайна теста.

- Для нумерации заданий используйте крупный шрифт (например, 36). Это облегчит испытуемым ориентацию в тесте.
- Оставляйте для ответа достаточно свободного места. (Это особенно важно для учащихся начальных классов школы, у которых может быть крупный почерк.)

- Оставляйте между заданиями свободное пространство, чтобы учащиеся видели, где заканчивается одно и начинается другое.
- Покажите учащемуся примерную длину ожидаемого от него ответа с помощью отведенной для него линии. Если она короткая, то ответить на вопрос можно одним словом. Две-три строки, оставленные для ответа, говорят о том, что учащемуся необходимо написать одно или несколько предложений.
- Присвойте каждому заданию уникальную метку и разместите ее слева от номера задания. Желательно использовать мелкий шрифт. Такая метка гарантирует точность информации при наблюдении за заданием. Ведь номер задания может меняться, особенно если оно появляется в тестовых буклетеах многократно.
- Будьте последовательны в использовании кавычек, курсива, подчеркивания, прописных букв и выделений полужирным начертанием (соблюдайте унификацию).
- Позаботьтесь о том, чтобы особенности дизайна и расположения *предтестовых* заданий были как можно ближе к представлению заданий в окончательном варианте. (Предтестовыми называют авторские задания, не прошедшие эмпирическую проверку и не обладающие устойчивыми статистическими характеристиками (см.: Челышкова М.Б. Конструирование и применение педагогических тестов. М.: Логос, 2001). – *Примеч. ред.*) Изменения в дизайне и расположении заданий могут повлиять на их трудность.

Качество изображений

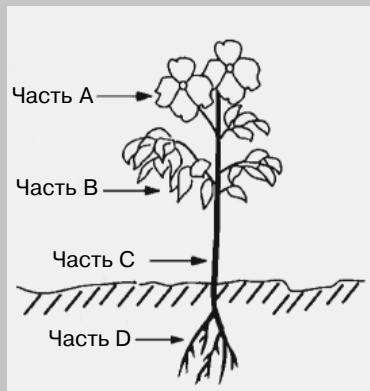
Иллюстрации, используемые в teste, должны быть ясными, четкими, а заголовки и подрисуночные подписи – простыми и понятными. Желательно, чтобы изображения были подготовлены профессионалом. Сканированные или взятые из Интернета рисунки имеют низкое качество. Фотографии также не подходят, потому что их использование увеличивает стоимость печати. Вообще, графический дизайнер должен перерисовывать фотографии. Чтобы достичь единообразного и гармонич-

Дополнительный материал 3.15**Использование рисунков для сокращения формулировок**

Следующий рисунок описывает научный эксперимент намного лучше, чем это мог бы сделать многословный текст.

Какая часть растения получает больше всего воды?

- (A) Часть А
- (B) Часть В
- (C) Часть С
- (D) Часть D



Источник: Международная ассоциация по эвалюации образовательных достижений, 2007, пример задания.

ного оформления, целесообразно привлекать одного и того же специалиста.

Иллюстрации следует использовать настолько часто, насколько это возможно. Они помогают сделать задание более наглядным, кратким и понятным (см. дополнительный материал 3.15). Чем рисунок проще, тем выше его эффективность. Изображение должно отражать суть задания (см. дополнительный материал 3.16). Оно не должно быть слишком «натуральным». Предпочтение желательно отдавать сплошным линиям, а частого использования пунктира следует избегать.

Графики и карты должны быть изображены четко и последовательно (см. дополнительные материалы 3.17 и 3.18).

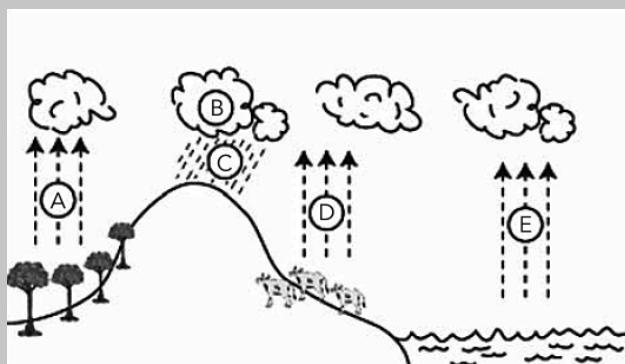
Свободное место на рисунке облегчает прочтение материала стимула (см. дополнительный материал 3.19).

Принимая решение относительно размера шрифта, дизайнерам теста следует придерживаться следующих рекомендаций.

- Используйте кегль 14 в заданиях для учащихся 3-х и 4-х классов и 12 – в заданиях для учащихся старших классов.

Дополнительный материал 3.16**Принцип простоты изображения**

Этот рисунок представляет сложную систему простым способом



Источник: Филиппинский отдел образования, 2004.

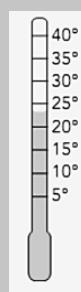
Дополнительный материал 3.17**Четкая маркировка изобразительного материала**

Ясные и последовательные обозначения, подобные тем, которые представлены на следующей иллюстрации, помогут учащимся быстро понять сложные вопросы:

Эта таблица показывает температуру в различное время на протяжении четырех дней.

В какой день и какое время температура, показанная в таблице, такая же, как на термометре?

Температура					
	6.00	9.00	Полдень	15.00	18.00
Понедельник	15°	17°	24°	21°	16°
Вторник	20°	16°	15°	10°	9°
Среда	8°	14°	16°	19°	15°
Четверг	8°	11°	19°	26°	20°



Термометр

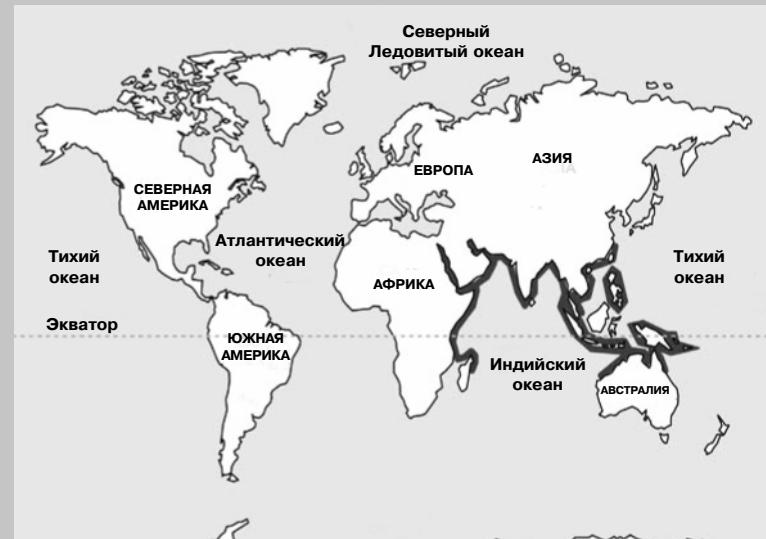
- A. Понедельник, полдень
- B. Вторник, 6.00
- C. Среда, 3.00
- D. Четверг, 15.00

Источник: Муллис (Mullis) и др., 2000.

- Уменьшайте ширину текстов стимулов приблизительно до 10–14 слов в строке.
- Убедитесь, что употребление пунктирных линий не вносит неясности. Не позволяйте, чтобы одно слово оказывалось на одной строке.

Дополнительный материал 3.18**Ясные обозначения на картах**

Необходимо позаботиться о ясных обозначениях названий на картах. На следующей карте названия континентов приведены прописными буквами, в то время как для обозначения океанов используют строчные буквы:



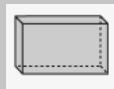
Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

Варианты ответов для выбора должны быть приведены последовательно. Каждый вариант следует определить. Рекомендуется следующие варианты расположения:

- вертикальное: варианты приведены сверху вниз:
A
B
C
D
- горизонтальное: варианты следуют слева направо:
A B C D
- Так могут выглядеть два столбца, в которых варианты ответов расположены сверху вниз:
A C
B D

Дополнительный материал 3.19**Роль свободного пространства
для понимания материала стимула**

Расположение элементов на этом рисунке обеспечивает удобочитаемость материала стимула.

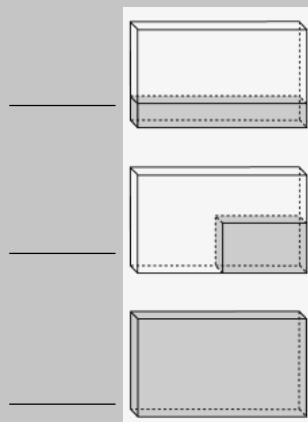


Фигура выше изображает коробку, которая содержит твердое тело, жидкость или газ. Затем содержимое поместили в коробку, большую по размеру в 4 раза.



A. Посмотрите на рисунки ниже. Они показывают, как различные материалы будут выглядеть, если их переместят в большую коробку.

Установите, на каком рисунке изображено твердое тело, на каком – жидкость и на каком – газ. (Напишите слова *твердое тело*, *жидкость* или *газ* на линии слева и ниже рисунка. Используйте каждое слово только один раз.)



B. Свой ответ объясните

Источник: Международная ассоциация по эвалюации образовательных достижений, 2007, пример задания.

КОМАНДА, ПИШУЩАЯ ЗАДАНИЯ

Команду авторов заданий возглавляет менеджер по разработке тестов. Он или она контролирует весь процесс с начала работы, предварительного тестирования и отбора заданий для окончательного тестирования. Менеджер по разработке тестов должен обладать отличными коммуникативными навыками и организационными способностями.

В его обязанности входит:

- подбор команды авторов заданий;
- обеспечение принятия проекта теста авторами;
- чистка проекта;
- разработка правил или протоколов представления, классификации и сохранения заданий;
- предоставление авторам информации о количестве места на странице, которое задания могут занять;
- мониторинг процесса экспертизы заданий;
- мониторинг разработки новых заданий и их усовершенствования по сравнению с проектом теста;
- рассмотрение заданий с экспертными группами или ключевыми заинтересованными лицами;
- мониторинг качества заданий;
- отслеживание выполнения запланированного графика разработки заданий;
- регистрация деталей разработки заданий, включая историю предварительного тестирования и изменения, произведенные во время анализа заданий;
- обеспечение полного отражения проекта в окончательной версии теста.

Написание заданий требует внимания к деталям, творческого потенциала, интеллектуальной строгости, глубокого знания содержания и хорошего понимания уровня развития учащихся. В идеале авторы заданий должны обладать следующими качествами.

- Они проявляют инициативу и готовность проводить масштабный поиск интересных материалов для стимулов и в состоя-

нии разработать высококачественный материал самостоятельно.

- Они могут установить обратную связь и комментировать работу других авторов заданий с достаточной степенью критики.
- Они стремятся к превосходному выполнению работы и, разрабатывая и совершенствуя задания, уделяют внимание деталям.

Для отбора авторов необходимо провести собеседование. Обычно претендентам дается 30 минут для написания заданий, которые основаны на материалах стимулов. Часто эта проверка сопровождается интервью, во время которого у потенциальных авторов спрашивают, какими критериями при написании заданий они руководствовались. Групповое интервью предназначается для оценки умения претендентов принимать критические замечания по своей работе.

Идеальный вариант – когда опытные авторы заданий обучаются новичков. Если таковых в данной стране нет, то целесообразно привлекать в качестве консультантов специалистов из-за рубежа. Они могли бы проводить учебные сессии, делать обзор процесса разработки заданий или выполнять то и другое одновременно. Только долгая упорная работа на протяжении нескольких месяцев позволит новым авторам достичь того уровня мастерства, когда они сами начнут производить задания приемлемого качества.

В ходе обучения должны быть освещены следующие вопросы.

- Какова общая цель теста?
- Каковы соответствующие контексты стимула для заданий?
- К каким разделам учебного плана обращен тест?
- В какой пропорции должны быть распределены задания по разным разделам учебного плана?
- Какой язык (или языки) будет использоваться?
- Каков подходящий уровень сложности лексики и грамматики?

- Какие формы заданий будут использоваться и в каком соотношении?
- Каковы характеристики публикуемых материалов (число страниц в тестовом буклете, формат, число заданий на странице)?
- Сколько заданий будет предложено для окончательной версии теста?
- Сколько заданий должно быть разработано?
- Как будет анализироваться проект тестовых заданий?
- Какой срок отводится на разработку, предварительное тестирования и отбор окончательных вариантов теста?
- Следует ли при подборе материала стимула и разработке заданий рассматривать культурные аспекты или ограничения, связанные с особенностями отдельных групп испытуемых?

Авторы заданий должны выработать единую точку зрения по всем перечисленным вопросам. Необходим последовательный контроль за своей собственной работой и работой своих коллег. У всех авторов должны быть копии окончательного варианта теста и общее представление о его содержании.

Детальное внимание к стилю заданий в самом начале работы позволит сэкономить много времени в дальнейшем. Менеджер по разработке тестов должен предложить авторам специальную таблицу стилей, которая поможет точно определить, как следует представить задания и руководства по оценке. Эта таблица должна охватывать все детали, в том числе такие, как кегль и гарнитура, использование выделений, оформление заголовков и задания, которые должны быть включены в него, как показано в дополнительном материале 3.20.

Этот пример показывает, что авторам следует набрать заголовок своего юнита шрифтом *Arial*, кеглем 16 и выделить его прописными буквами. Остальную часть текста задания следует набрать шрифтом *Times New Roman*, в основном кеглем 12. Вопрос должен быть пронумерован, как в случае «Вопрос 1». Название юнита должно идти после номера. Само задание и место для ответа приводятся под вопросом. Автор заполняет таблицу,

в которой отражены: вид текста, форма задания и процесс, протекающий при его выполнении. Руководство по оценке имеет вид как в приведенном ниже примере. Критерии для выставления баллов выделены курсивом, при этом шрифтом кегля 10 приведены ответы учащихся. Приведенные рекомендации позволяют авторам приобрести уверенность в том, что задания предварительного и окончательного теста согласованы, эффективны и не содержат двусмысленностей.

Дополнительный материал 3.20

Пример таблицы стилей для авторов заданий

ПАСХАЛЬНЫЕ ТРАДИЦИИ (заголовок, шрифт Arial, кегль 16)

Вопрос 1: Пасхальные традиции (подзаголовок, шрифт Times New Roman, кегль 12, начертание полужирное)

Что люди делают на Пасху? (вопрос, шрифт Times New Roman, кегль 12, начертание полужирное)

<вставьте половину линии> (инструкция для печати – набрана курсивом в угловых скобках)

Вид текста	Форма задания	Метод решения
Информационный	Закрытая, конструируемый ответ	Поисковый

(таблица с тремя столбцами и двумя строками, головка: шрифт Times New Roman, кегль 12, начертание полужирное, прографка: шрифт Times New Roman, кегль 12, без выделения)

Руководство по оценке (подзаголовки, шрифт Times New Roman, кегль 12, начертание полужирное)

1 балл: *ответ содержит слово «яйца»* (шрифт Times New Roman, кегль 12, начертание курсивное)

- дают друг другу яйца (буллит, шрифт Times New Roman, кегль 10);
- красят яйца.

0 баллов: *в ответе названы блины, другие подарки или не указано ничего*

- пекут блины;
- дарят друг другу сувениры.

Авторам заданий необходима регулярная обратная связь. Им желательно обмениваться опытом по разработке заданий и способам, с помощью которых они могут установить соответствие готовых заданий проекту. Это позволит им учиться на своих и чужих ошибках, развивать умения и совершенствовать задания. Авторам следует собираться вместе как можно чаще и проводить экспертизу заданий, чтобы критически оценивать проделанную работу. Менеджер по разработке теста должен быть подготовлен к замене тех авторов, которые даже после тщательной подготовки не в состоянии разрабатывать высококачественные задания.

ГРУППЫ ЭКСПЕРТОВ ПО РАЗРАБОТКЕ ЗАДАНИЙ

Группа экспертов по разработке заданий состоит из небольшого числа (обычно от трех до шести) авторов, которые совместно рассматривают материал, разработанный одним или несколькими из них. Их цель – принять, изменить или отклонить представленный тестовый материал. Такой командный подход, являющийся частью контроля качества тестов, помогает рассмотреть процесс развития каждого конкретного задания с разных сторон. Если авторы не обладают высокой квалификацией, то после групповой экспертизы задания обычно подвергаются существенному пересмотру.

Членам группы еще до собрания следует подготовить предварительные критические замечания. У них должно быть достаточно времени для анализа заданий и разработки предложений по усовершенствованию проекта.

Группа должна детально рассмотреть контекст материала стимулов, содержание, формулировки, язык, расположение заданий и их элементов, и иллюстрации, чтобы гарантировать, что каждый аспект стимула релевантен проекту, достоин включения в тест, ясен и краток. Затем необходимо полностью проанализировать каждое задание, чтобы гарантировать, что его формулировка однозначна и форма выбрана верно и что задание напрямую связано с проверяемыми умениями и содержательной областью, определенной в проекте теста. Совокуп-

ность заданий в целом также исследуется, чтобы посмотреть, насколько хорошо окончательные пропорции заданий отражают проект теста. Члены группы должны проанализировать возможности для улучшения имеющихся стимулов заданий, а где необходимо – внести предложения по разработке новых.

В процессе рассмотрения материалов группа экспертов из числа авторов делает готовый отчет о проделанной работе и выслушивает конструктивную критику со стороны коллег. Руководитель группы должен обеспечить согласие всех членов группы относительно изменений, которые необходимо будет внести в отдельные задания. Авторам следует зафиксировать предложенные изменения и впоследствии пересмотреть задания.

В группах, где авторы пишут тесты не на родном языке, может понадобиться эксперт по языку. Причем его языковые навыки должны соответствовать цели тестирования и конкретной генеральной совокупности учащихся.

В группу также может быть включен предметный эксперт, особенно в тех случаях, когда содержательная область теста сложна. Привлечение предметного эксперта помогает разъяснить содержательные проблемы, но это сотрудничество не должно быть постоянным. Предметный эксперт вряд ли будет заниматься тонкостями проблем написания заданий.

В группы экспертов по разработке не желательно привлекать тактиков и заинтересованных лиц. Детали группового обсуждения экспертизы заданий их касаться не должны.

Членам группы необходимо рассмотреть задания со всех сторон.

- Правильное ли оценивается содержание?
- Адекватна ли форма заданий возможностям тех учащихся, для которых они предназначены?
- Действительно ли задания имеют значимое содержание или они чересчур тривиальны?
- Ясны ли формулировки и не являются ли они двусмысленными?
- Есть ли ошибки или пропущенные слова?

- Если задание имеет форму с множественным выбором, то являются ли все предложенные варианты ответов подобными по значимости и конструкции?
- Если задание имеет форму с множественным выбором, то будет ли ответ следовать из имеющейся информации (основной части, материала стимула или из того и другого)?
- Можно ли считать оформление заданий привлекательным? Не слишком ли много текстовой информации на страницах теста?
- Имеет ли большинство заданий такую трудность, которая позволит приблизительно 40 – 80% учащихся дать правильный ответ?
- Если задание требует оценки с частичным кредитованием, то будет ли каждая оценка выбираться, по крайней мере, для 10% испытуемых?
- Отсутствует ли систематическая ошибка задания, т.е. будут ли задания справедливыми для большинства подгрупп в генеральной совокупности учащихся, предназначеннной для тестирования?

Члены группы должны также разработать рекомендации по совершенствованию заданий, чтобы:

- укоротить их;
- добавить больше информации;
- изменить формулировки;
- добавить диаграммы или рисунки;
- представить их в различных формах.

Авторы должны получать регулярную обратную связь с момента, когда они приступают к разработке заданий. Если это возможно, члены экспертной группы по разработке заданий должны встречаться не реже одного раза в неделю.

Пересмотр материалов может стать весьма сложной задачей. В экспертной группе по разработке заданий необходим руководитель, который будет координировать работу группы. Он гарантирует, что ее члены придут к общему мнению относительно рекомендаций и изменений, которые необходимо

сделать. Сам автор не может принять самостоятельное решение относительно того, какие предложения следует принять, а какие проигнорировать. Руководитель группы должен четко выражать свое мнение, чтобы автор задания точно понял, какие изменения следует внести.

Коррекция чужих заданий время от времени может вызывать горячие споры. Руководитель группы должен обеспечить конструктивную работу по продвижению к этой цели.

Пересматривать свою работу приходится всем членам группы, в том числе опытным авторам. Вполне закономерными являются значительные переделки в заданиях, разработанных новичками. Критика заданий ни в коем случае не должна переходить на личность автора. Тех авторов, которые не могут участвовать в здравом обсуждении, не воспринимают критику и отказываются изменить свои задания, необходимо исключить.

ДРУГИЕ ЭКСПЕРТЫ

Члены экспертных групп или ключевые заинтересованные лица должны иметь возможность неоднократно обращаться к разрабатываемым заданиям. Это поможет обеспечить высокое качество заданий и их соответствие проекту. Менеджер по разработке тестов обычно представляет все задания или их часть выбранной с этой целью референтной группе.

Чтобы гарантировать, что авторы заданий работают в правильном направлении, первый обзор следует подготовить как можно раньше. Эксперты могут предложить внести изменения в некоторые части проекта теста, особенно если авторам трудно отобразить некоторые требования в таблице спецификации. Также авторы заданий могут нуждаться в более конкретном руководстве для различия приемлемого и недопустимого материала.

Экспертиза обычно проводится после того, как все задания были запланированы. Это дает гарантию, что ключевые заинтересованные лица в образовании одобрят задания до начала предварительного тестирования. Заключительный обзор позволяет им составить суждение относительно отбора заданий для окончательной формы теста.

ОТСЛЕЖИВАНИЕ ЗАДАНИЙ

Отслеживание заданий необходимо. (Имеется в виду отслеживание истории развития заданий, процесса их коррекции и мониторинг изменений. – *Примеч. ред.*) У каждого задания должна быть уникальная метка, чтобы можно было проследить все стадии его развития от предварительного тестирования до анализа.

Необходимо разработать приблизительно в два с половиной или в три раза больше заданий, чем требуется для окончательных форм теста. Обычно необходимо бывает сделать много-кратные буклеты для каждого класса с целью предварительного тестирования. Некоторые совокупности одних и тех же заданий должны присутствовать в различных буклетах. Это позволит связать все предтестовые задания в одной и той же шкале для сравнения по трудности.

Каждое задание должно иметь свою метку, независимую от порядка расположения заданий в буклете. Это позволит найти одинаковые задания, продублированные в различных буклетах.

Отслеживание заданий на стадии анализа может быть достаточно сложным. Если аналитик не в состоянии присвоить заданию уникальную метку, то его, скорее всего, будет очень трудно отыскать. Еще одной проблемой может стать программное обеспечение. Обычно с его помощью номера присваиваются автоматически. Если задания будут удалены из теста на основе анализа результатов предварительного тестирования, то программное обеспечение перенумерует задания таким образом, что число заданий в процессе анализа больше не будет соответствовать первоначальному числу заданий или их числу в тестовом буклете. Сохранение у каждого задания одной и той же уникальной метки в тестовом буклете и на каждом этапе анализа поможет предотвратить эту проблему.

Метки должны быть настолько значимыми, насколько это возможно. Менеджер по разработке тестов должен согласовать свои действия с аналитиком, чтобы установить, сколько знаков можно использовать для метки. В статистических программных пакетах для анализа данных существуют различные

ограничения. Ниже приведено соглашение о маркировке, которое используется одним агентством по тестированию для тестов по чтению, письму и математике уже на протяжении нескольких лет.

- Первый знак – Ч, М или П – соответственно для тестов по чтению, математике или письму.
- Второй и третий знаки указывают на год, когда задание предъявляли испытуемым (например, 07 для 2007 года).
- Четвертый, пятый и шестой знаки соответствуют номеру задания (начиная с 001).

Например, метка М06003 показывает, что задание относится к тесту по математике, его использовали в 2006 году и оно является третьим по счету.

Метки должны быть присвоены заданиям еще в процессе их разработки и помещены в буклеты для предварительных и окончательных форм теста. Метки могут быть напечатаны мелким шрифтом без абзацного отступа напротив номера задания, как показано ниже:

М06003 **5** Вычислите сумму
 6 + 7 = _____.

Важно сохранять историю развития каждого задания. Анализу необходим набор ключей для заданий с множественным выбором. Также ему нужно знать местоположение заданий в брошюрах с предварительными вариантами тестов. Для подготовки отчетов требуется полная информация о форме заданий и процессах, на оценку которых направлено каждое задание. Менеджер по разработке тестов должен сформировать и поддерживать в активном состоянии крупноформатную таблицу, которая поможет вести учет каждого задания, его классификаций и статуса, а также любых изменений, которые, возможно, были произведены в задании.

Ниже приведен пример из отчета некоторых заголовков крупноформатной таблицы для теста по чтению:

• название юнита	название, данное юниту (например, «Гоночные автомобили»);
• метка задания	метка с шестью цифрами (например, R06003);
• содержание задания	формулировка вопроса теста;
• текущий статус	показывает, в какой форме доступно задание для использования (например, разработано как тренировочное задание; задание, отклоненное клиентом, задание; не разрешенное к копированию);
• ключ	правильный ответ для задания с множественным выбором;
• максимальный балл	максимальная оценка в баллах для задания;
• тип текста	текстовый жанр (например, рассказ, информация);
• форма задания	форма тестового задания (например, с выбором ответа);
• процесс	познавательный процесс (например, воспроизведение);
• аналитические заметки	изменения, сделанные в задании после предварительного тестирования.

Задания нужно хранить в безопасном месте. Материал, связанный с разработкой юнита или задания, должен храниться с тем же заданием. Необходимо сохранить даже тот материал, который не использовался в ходе предварительного тестирования, потому что он может понадобиться позже в тот же самый или на следующий год обучения. Источник документов или иллюстраций должен быть зарегистрирован и помечен так, чтобы разрешение на его копирование можно было бы легко отыскать в случае необходимости. Копии нужно хранить так, чтобы могли быть выявлены любые последующие модификации.

Большинство заданий можно сохранить на электронных носителях информации. Необходимо предусмотреть наличие резервных файлов с заданиями на отдельном компьютере или диске. Правильная маркировка и полная и точная классификация помогут гарантировать, что задания сохранены в соответствующих файлах и могут быть легко восстановлены. Задания подвергаются постоянному пересмотру, включая изменения руководства по оценке и иллюстраций и незначительное усовершенствование формулировок. Последняя версия задания должна быть готова к использованию отдельно от информации в файле.

ПРИМЕЧАНИЯ

- ¹ Для дополнительной информации по разработке тестовых заданий см. *Chatterji (2003), Haladyna (1999), Kubiszyn and Borich (2000), Linn and Miller (2004)*.
- ² Американская национальная оценка образовательного прогресса включает компонент письма (Национальное управление оценки n.d.).

ГЛАВА

4

ПРЕДВАРИТЕЛЬНАЯ ПРОВЕРКА ЗАДАНИЙ

Для разработки тестов национальной оценки результатов обучения, как правило, используется технология, которая была выработана в ходе проектирования тестов, предназначенных для оценки и анализа учебных достижений отдельных учащихся. Поскольку последние необходимы для дифференциации учащихся по результатам тестирования, им выдается по существу один и тот же тест. Цель национальной оценки совершенно иная. Национальная оценка не направлена на дифференциацию учащихся. Ее задача – показать, в какой степени учащиеся в целом в образовательной системе (или в четко определенной ее части) приобрели знания и умения, предписанные учебным планом. Для ее выполнения необходимо обеспечить адекватное отражение в тестах учебного плана. Поэтому они должны намного больше охватывать разделы плана, чем тесты, разработанные для оценки отдельных учащихся. Потребность в широком охвате учебного плана становится еще очевидней в тех случаях, когда оценивание предназначено для идентификации тех его областей, в которых учащиеся показывают высокие или, наоборот, низкие результаты.

Учитывая все вышесказанное, в тесты многих национальных и международных систем оценки включается намно-

го больше заданий, чем в тесты для оценивания отдельных учащихся. Однако чтобы избежать чрезмерной нагрузки, каждый ученик выполняет только часть заданий. Следовательно, в буклете должны быть приведены несколько альтернативных комплектов заданий для их возможной ротации. Точное число заданий в разных системах национальной оценки изменяется.

Данный подход, хотя и необходимый по ряду причин, приносит в процесс предъявления теста много трудностей. Во-первых, довольно сложным становится дизайн теста, потому что нужно обеспечить «наложение» заданий и соответствие между их подвыборками. Во-вторых, необходимо гарантировать правильную выдачу буклетов учащимся и соответствие инструкций для всех буклетов. Наконец, объединение данных, полученных по множеству наборов заданий, требует относительно сложных статистических процедур. По этим причинам многие развивающиеся страны не используют большое число тестовых буклетов.

Большинство комментариев на следующих страницах, так же как и в главе 5, применимы к двум ситуациям, в одной из которых команда по национальной оценке использует множество тестовых буклетов, а в другой – для измерения учебных достижений в предметной области предназначается единственный тестовый буклет. Оба подхода требуют корректного проведения предварительного тестирования.

Предварительное, или пилотное, тестирование является существенным этапом процесса разработки теста. Учащиеся, которые выполняют предварительный тест, должны обладать теми же характеристиками, что и те, которые будут выполнять окончательный тест. Желательно, чтобы в предварительном тестировании участвовали разные по размеру школы из разных районов города с учащимися из семей с варьирующимся социально-экономическим фоном. Идеальный вариант, когда предварительное тестирование проводится в то же самое время, на которое запланировано окончательное тестирование, но годом ранее. Например, предварительный тест для учеников 5-го класса проводится в октябре 2005 года, а окончательный – в октябре 2006 года. Однако на практике этот сценарий

может оказаться невыполнимым, и предварительный тест может быть выдан испытуемым, которые проучились на несколько месяцев больше или меньше, чем целевая генеральная совокупность учащихся. Например, учащиеся 6-го класса могут быть предварительно протестираны в начале учебного года, чтобы обеспечить данные для теста, который будут использовать для оценивания учащихся 5-го класса в конце учебного года.

Учащиеся, которые участвуют в предварительном тестировании, не должны иметь доступа к окончательному варианту теста. Если окончательный тест нужно предъявлять статистически отобранный случайной выборке, то она должна быть построена до того, как будут определены школы для предварительного тестирования.

Считается полезным использовать в предварительном тестировании в два или три раза больше заданий, чем в окончательном. Длина каждой предтестовой формы должна быть подобна той, какая будет в окончательном тесте. Полезно использовать маркировку форм теста в алфавитном порядке согласно году обучения, например, пять форм для 3-го класса тогда были бы обозначены символами 3A, 3B, 3C, 3D и 3E, а для 8-го – 8A, 8B, 8C, 8D и 8E.

Для каждого класса требуется несколько предтестовых форм. Идеально, если формы для каждого года обучения будут случайным образом распределены в каждом классе. Если используются три формы в 5-м классе (5A, 5B и 5C), то каждая школа должны получить «смесь» из заданий всех трех форм. Если эта процедура невозможна, то важно гарантировать, что каждая форма распределена по всем составляющим выборки для предварительного тестирования. Например, форма 5A не должна быть выдана только городским учащимся, форма 5B – только сельским учащимся на севере района и форма 5C – сельским учащимся в южной его части. Формы должны быть настолько эквивалентными, насколько это возможно с точки зрения соответствия проекту теста.

Существенно обеспечить такую связь предтестовых форм, чтобы задания могли быть объединены для сравнения. Связь

означает, что несколько одних и тех же заданий появляются в различных формах теста. Некоторые предтестовые формы неизбежно будут более трудными, чем другие. С помощью связи общая трудность заданий может быть определена независимо от формы, в которой эти задания встречаются. Горизонтальная связь требуется в тех случаях, когда тестируются учащиеся одного года обучения, вертикальная – когда в тестировании принимают участие учащиеся разных лет обучения.

Минимум 200 испытуемых должны попытаться выполнить каждое предтестовое задание для каждого года обучения. Если три предтестовые формы будут разработаны для 5-го класса, то предварительный тест должны выполнять, по крайней мере, 600 учащихся. От этих 200 испытуемых для каждого предтестового задания требуется получить, самое малое, 150 ответов. Неизбежно, что во время предварительного тестирования произойдет некоторая потеря данных. В лучшем случае школа может выйти из программы предварительного тестирования в последнюю минуту или предъявить тест меньшему числу учащихся, чем предполагалось. Также маловероятно, что каждый учащийся попытается выполнить каждое задание в предварительном teste.

Поскольку предварительное тестирование должно проводиться в тех же самых условиях, что и окончательное, промежуток времени для выполнения теста должен быть одинаков. Число заданий, которые учащиеся могут выполнить в отпущенное для тестирования время, возможно, не будет известно. Если это действительно так, то сформируйте пример предтестовой формы и испытайте ее в нескольких классах перед тем, как собирать предварительный тест, чтобы гарантировать «реалистичность» числа заданий в каждой предтестовой форме. В национальном оценивании не применяются тесты на скорость. У большинства учащихся должно быть достаточно времени для выполнения большей части заданий теста.

Желательно завершить предварительное тестирование за две–три недели.

Предварительное тестирование обеспечивает возможность оценить пригодность и качество заданий, а также точно настроить многие аспекты предъявления теста. Администраторы должны предоставить следующую информацию.

- Было ли у учащихся достаточно тренировочных вопросов по применению теста, были ли инструкции и объяснения достаточно ясны?
- Имел ли тест подходящую или чрезмерную длину, и приблизительно сколько учащихся закончили тест раньше, чем за 10 минут до окончания времени, отведенного на тестирование?
- Были ли испытуемые поглощены выполнением теста?
- Было ли достаточно сопроводительных материалов, таких как карандаши и ластики?
- Действительно ли условия в школе были подходящими для проведения тестирования?
- Понимали ли преподаватели и испытуемые цель тестирования?

ДИЗАЙН ПРЕДТЕСТОВОЙ ФОРМЫ

Анализ данных предварительного тестирования дает основания для отбора заданий в окончательный тест. В большинстве случаев в национальной оценке готовят разные формы связанных тестовых буклетов для каждого года обучения. Этот подход помогает шире охватить учебный план по сравнению с единственным тестом и в то же время позволяет гарантировать, что учащимся не предъявлены необоснованно длинные тесты. Дизайн теста и связь данных по формам должны гарантировать, что данные тестирования могут быть представлены в единой шкале. К разработке дизайна предварительного теста необходимо привлечь аналитика, статистику и ведущих специалистов по компьютерному дизайну. Это позволит гарантировать, что при проектировании теста были отражены требования к данным.

Связанные формы имеют общие задания. Обычно требуется от 8 до 10 общих заданий. Есть несколько способов для построения дизайна связанных форм.

Первый способ – единственный единый набор общих заданий для связи, когда одни и те же 8–10 заданий повторяются в каждой форме. Необходимо отметить, что если связующие задания ведут себя плохо (имеют плохие статистические характеристики) при анализе, то связь между формами будет слабой, что негативно повлияет на качество общего анализа результатов тестирования.

Второй способ – круговая связь. Различные наборы заданий используются для связи между парами форм. Например, форма 3A может быть связана с формой 3B с помощью набора из X заданий, форма 3B – с формой 3C с помощью набора из Y заданий, а форма 3C, в свою очередь, связана с формой 3A набором из Z заданий.

Каждая форма также содержит уникальные задания, которые не появляются ни в одной из других форм (рис. 4.1).

Третий способ – линейная связь, которая следует круговой модели связи, но исключает использование для связи набора заданий Z . Таким образом, форма 3A будет связана с формой 3B, а 3B – с 3C, но между формами 3A и 3C не будет никакой связи.

Если неясно, как будут работать задания, лучше использовать круговую связь и большее число заданий для связи, чем это необходимо в окончательном варианте форм. При круговой связи даже в том случае, если один набор связующих заданий окажется неудачным, связи между формами будут сохранены.

Связи могут быть и горизонтальными (для одинаковых классов), и вертикальными (для различных классов, например, связи между результатами тестирования 3-х и 5-х классов). Если конечные формы теста не собираются связывать вертикально, тогда связь при предварительном тестировании должна иметь явно выраженный акцент на горизонтальном способе. Минимальная вертикальная связь может использоваться для сравнения данных предварительного тестирования между различными годами обучения. С точки зрения технологии для связи результатов по разным годам обучения необходимо только 8 или 10 общих заданий. Оценочные процедуры для установ-

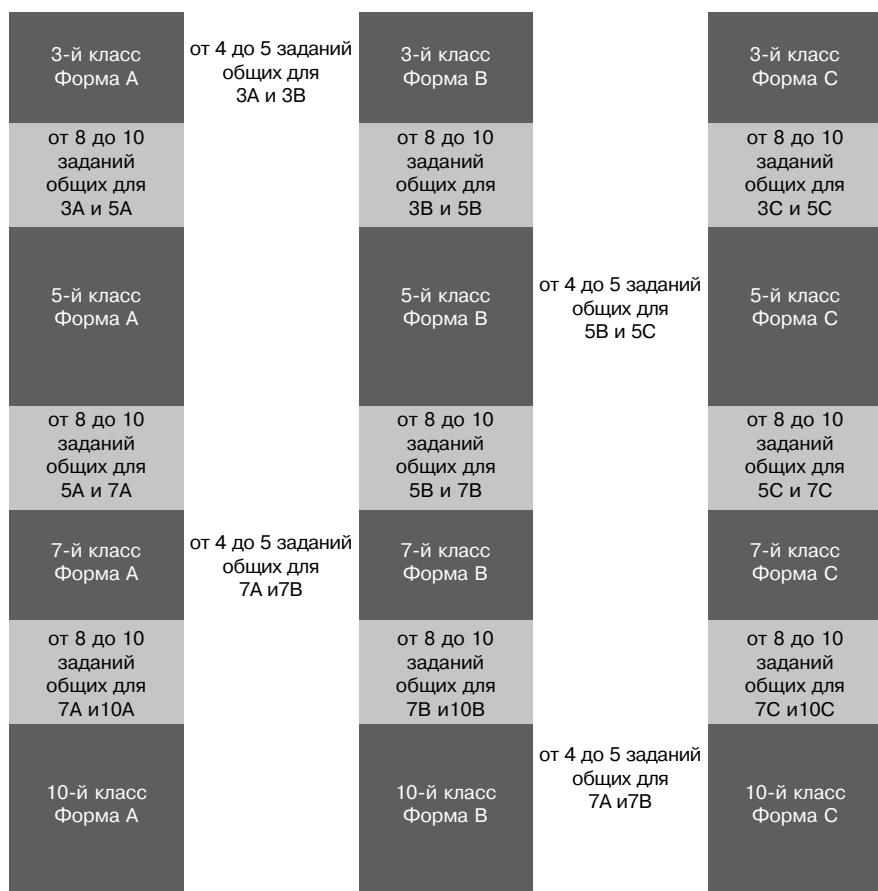
Рис. 4.1**Пример круговой связи заданий**

Источник: разработки авторов.

ления хорошей вертикальной связи между заданиями являются более трудными, чем процедуры для обеспечения хорошей горизонтальной связи. Поскольку в предварительном тестировании качество связующих заданий обычно неизвестно, то желательно иметь, по крайней мере, 16 заданий для установления вертикальной связи, распространенной по каждой из тестовых форм.

Если конечные формы теста должны быть связаны вертикально, то важно иметь большее число запасных заданий для установления вертикальной связи в тестировании по сравнению с тем числом, которое потребуется для окончательного тестирования, чтобы для окончательных форм теста могли быть выбраны лучшие задания. Альтернативная модель предварительного тестирования при вертикальной связи окончательных форм теста показана на рис. 4.2. Модель основана на случайном распределении форм А, В и С в пределах каждого класса.

В этом довольно сложном примере каждая из форм 3А, 5А, 7А и 10А вертикально связана линейным способом с помощью 8–10 заданий. Задания связаны подобным способом и в формах В и С. Есть в целом 8–10 заданий для горизонтальных связей между формами А и В для 3-го и 7-го классов и в целом 8–10 заданий для горизонтальных связей между формами В и С для 5-го и 10-го классов. Это число заданий для горизонтальных связей является вполне приемлемым. Если тестовые формы не собираются распределять в случайном порядке внутри каждого

Рис. 4.2**Модель для вертикально связующих заданий**

Источник: разработки авторов.

го класса или если авторы не уверены в качестве заданий для горизонтальных связей, то должно быть включено больше связующих заданий в тесты для каждого класса.

Связующие задания должны быть помещены в начале или в середине тестовых форм, но не в конце, чтобы избежать ситуаций, когда испытуемые вовсе не приступят к ответу на них. Их необходимо приводить в одном порядке в каждом тестовом буклете и на одинаковых местах так, чтобы любые различия в результатах не могли быть приписаны порядку или месту расположения заданий. Небольшие различия в местоположении неизбежны. Важно избегать значительных отличий.

Связующие задания должны относиться к диапазону средней трудности. В целом, учащиеся со средними способностями в генеральной совокупности должны правильно ответить на эти задания с вероятностью от 40 до 60%. Поскольку в предварительном тесте трудность заданий для выборки испытуемых не будет известна, авторы должны сделать свою приблизительную наилучшую оценку трудности заданий. Если они сомневаются в точности своих приближенных оценок, то желательно увеличить число заданий для установления связи между формами теста.

Если задания разработаны в соответствии с юнитами, то лучше связывать формы с помощью заданий, взятых из двух или больше юнитов, поскольку связь не работает достаточно хорошо, если задания взяты из одного юнита. Нет необходимости использовать для связи все задания из юнита. Некоторые могут быть общими, а другие – уникальными, как показано в табл. 4.1. Оба юнита «Поиск домашнего животного» и «Гора Аварапа» включены в формы 3А и 3В для 3-го класса. Три задания являются общими для каждого юнита и каждой формы. У предварительного теста есть дополнительные уникальные задания.

Если предтест включает комбинацию форм заданий, то связующие задания должны также отражать это сочетание форм.

Задания должны иметь уникальные метки, напечатанные рядом с ними в каждой тестовой форме, в которой они встречаются, и определенным образом выделенные. Задания с одной и той же меткой должны быть идентичными во всех отношениях, за исключением порядка появления в тестовой форме.

ТАБЛИЦА 4.1

Связующие задания в двух юнитах по чтению

Юниты по чтению	Общие задания для форм 3А и 3В	Уникальные задания для формы 3А	Уникальные задания для формы 3В
«Поиск домашнего животного»	3, 4, 6	2, 7	1, 5
«Гора Аварапа»	1, 2, 5	3	4

Источник: разработки авторов.

Задания, имеющие небольшие изменения в их формулировках, должны иметь разные метки.

Автор заданий должен создать крупноформатную таблицу со списком всех заданий; заголовки таблицы должны отражать, какие задания появляются, в какой форме и в каком порядке. Табл. 4.2 показывает часть такой типовой крупноформатной таблицы, которая охватывает три юнита («Собаки», «Ellie» и «Удар») теста по чтению для 5-го класса.

ТАБЛИЦА 4.2

Часть крупноформатной таблицы для отслеживания заданий по формам

Название юнита	Метка задания	Форма 5A	Форма 5B	Форма 5C	Форма 5D
«Собаки»	R070101	1	1	4	4
«Собаки»	R070102	3	3	5	5
«Собаки»	R070103		2		
«Собаки»	R070104	2			
«Ellie»	R070201			1	
«Ellie»	R070202			2	1
«Ellie»	R070203			3	2
«Ellie»	R070204				3
«Удар»	R070301	4			6
«Удар»	R070302	5			7

Источник: разработки авторов.

Задания для трех юнитов показаны слева. Числа в колонках отражают порядок появления каждого из этих заданий в конкретной предтестовой форме. Первые два задания в юните «Собаки» являются общими для всех четырех форм.

ВЫЧИТКА И ТИРАЖИРОВАНИЕ ПРЕДВАРИТЕЛЬНОГО ТЕСТА

Каждое задание предтеста должно появиться в том же виде, в каком оно будет включено в окончательный тест. Это также касается и материалов стимула, графиков и иллюстраций. Идеально, когда и связующие задания приводятся в том же порядке, но на практике это не всегда возможно.

Материал стимула при прочтении задания должен располагаться или на той же самой странице, что и задание, или он может находиться на левой странице разворота, а задания на правой, что позволит испытуемым легко переключать внимание от текста стимулов к заданиям.

На первой странице предтестового буклета не должно быть всех деталей, которые необходимы в конечной форме теста. В последней должны запрашиваться сведения о школе испытуемого, где обучения, а также пол, возраст и родной язык. Поскольку результаты предварительного тестирования учащимся не сообщают вовсе, то бланки подписывать не требуется. Некоторые второстепенные детали, которые необходимы в окончательном тесте, в предварительном тестировании не нужны. Расположение заданий должно быть согласованным по всем тестовым формам.

В процессе подготовки или рассмотрения вопросов тиражирования предтестовых материалов нельзя забывать о следующем контрольном списке:

- заголовки (большие и ясные);
- отступы от краев – вверху, внизу, слева и справа (согласованные);
- номера страниц (последовательные);
- номера заданий (большие и ясные);
- метки заданий (применимые);
- линии для записи ответов испытуемыми (ясные и соответствующей длины);
- тексты заданий (кегль шрифта 12 или 14);
- число слов в строке (10–12);
- материал стимула (ясный, предпочтительно со шрифтом, отличающимся от шрифта текста заданий);
- материал стимула и связанные с ним задания (на одной странице или на страницах разворота);
- заголовки и подстрочные сноски (согласованные и полезные);
- грамотное правописание (завершенная проверка).

Некоторые тесты включают оценочные опции в дихотомической шкале. Например, 0 или 1 можно было бы присвоить

заданию, чтобы оценить неправильный или правильный ответ соответственно. Оценочные опции для заданий, которые не пытались выполнять, обычно принимаются равными 9, как было рекомендовано в предыдущей главе. Эксперты могут просто обвести соответствующую оценку кружком. Вставка оценок в буклет напоминает экспертам диапазон оценочных опций.

Все тестовые буклеты и руководства по предъявлению тестов должны быть тщательно вычитаны. Корректору следует выполнить тестовые задания так, как если бы он был на месте учащихся. Вычитка должна гарантировать, что материалы в тесте соответствуют следующим критериям.

- Входные инструкции и тренировочные задания ясны и недвусмысленны.
- Задания ясны и недвусмысленны.
- Материал стимула ясен и легок для чтения.
- Опции в заданиях с множественным выбором включают один правильный вариант ответа, а другие – все явно неправильные.
- Каждый предложенный вариант ответа для выбора имеет смысл.
- При необходимости обеспечено соответствующее место для записи ответов испытуемых.
- Материал стимула для чтения находится на той же самой странице, что и задание, или он помещен на левую страницу, а задания на противоположной странице разворота.
- Задания в юните независимы, т.е. ответ на одно задание не содержится в основной части или в вариантах ответов на другие задания.
- Связующие задания идентичны.
- Не встречается ни одной ошибки в правописании или пунктуации.
- Расположение различных тестовых форм является согласованным.

Очень важна корректура заданий. Серьезные ошибки могут быть допущены фактически на всех стадиях подготовки. Предварительное тестирование станет пустой тратой времени, сил

и средств, если тест будет содержать опечатки и несогласованности. В этом случае уменьшится полноценность данных, потому что неправильные предтестовые задания не могут использоваться в окончательном тесте. Поэтому необходимо привлекать квалифицированных корректоров и отводить достаточно времени для вычитки и внесения правки.

Предтестовые буклеты должны быть проверены после распечатки на принтере. Выборочная проверка должна быть сделана из каждой связки или коробки с буклетами, чтобы гарантировать, что:

- все страницы были напечатаны четко;
- страницы расположены в правильном порядке;
- страницы не были дублированы;
- материал стимула для чтения для каждого юнита находится на правильной странице;
- иллюстрации четкие.

Предтестовые буклеты должны быть напечатаны и проверены задолго до рассылки по школам. Такой порядок работы позволит найти время для того, чтобы в случае необходимости переиздать буклеты. Поскольку тиражи предтестов обычно маленькие, то переиздание, если оно требуется, стоит относительно недорого.

ПРОВЕДЕНИЕ ПРЕДВАРИТЕЛЬНОГО ТЕСТИРОВАНИЯ

На стадии выполнения предварительного или окончательного теста у учащихся не должно возникнуть никакого сомнения относительно того, как представить свои ответы. Тесты разрабатываются с таким расчетом, чтобы проверить знание важных разделов учебного плана, а не способности испытуемых к угадыванию формы представления ответов. Как в начале предварительного теста, так и в начале каждого его раздела необходимо предоставить учащимся возможность выполнения тренировочных заданий. Особенно важно предоставить испытуемым адекватное число тренировочных заданий (например, три или четыре) в тех системах образования, в которых задания с множественным выбором традиционно не используются.

Заданий в предтестовых формах может быть столько же, сколько и в окончательном варианте, а может быть и немного меньше. Важно, чтобы испытуемые попытались выполнить все предложенные им задания. Если предварительный тест будет слишком длинным или содержать слишком много трудных заданий в конце, то немногие из этих последних заданий будут выполнены учащимися.

Тест лучше начать с нескольких легких заданий. Их выполнение будет стимулом для более слабых учащихся. Желательно чередовать трудность идущих друг за другом заданий, чтобы сохранить у испытуемых определенную мотивацию при выполнении всего теста. Тогда они не бросят тест, когда дойдут до трудных заданий. Попытайтесь сделать общую трудность каждой предтестовой формы примерно одинаковой. Избегайте тестов, наполненных исключительно тяжелыми заданиями, потому что учащиеся могут преждевременно сдаться и бросить тест. Если это случится, то для заданий в конце формы не будет достаточно данных для целей анализа предварительного теста.

Предварительное тестирование предоставляет возможность проверить альтернативные версии заданий в различных формах. Например, задание может быть предварительно проверено в форме с множественным выбором ответа и открытой форме (дополнительный материал 4.1).

В предварительный тест могут быть также включены различные формулировки открытых заданий. Обратите внимание, что альтернативные версии заданий не должны использоваться в качестве связующих, поскольку задания для установки связи между формами должны быть идентичными.

ОЦЕНКА ПРЕДВАРИТЕЛЬНОГО ТЕСТА

Цель сбора данных предварительного тестирования, или экспериментальных данных, состоит в том, чтобы получить информацию, которая поможет отобрать задания хорошего качества для окончательного теста. Вообще, оценки и имена учащихся нет необходимости связывать между собой. Основные проблемы для оценивания в предварительном тестирова-

Дополнительный материал 4.1**Пример задания в форме с множественным выбором и в открытой форме**1) с множественным выбором $13 + 17 + 8 =$

- (A) 28 (B) 30 (C) 38 (D) 110

2) с открытым ответом

$13 + 17 + 8 = \underline{\hspace{2cm}}$

нии связаны с контролем качества и согласованностью в обработке ответов учащихся.

Все оценивание требует процедур контроля качества. Выполнение оценивания по предварительному тесту и ввод данных вручную обычно более экономны, потому что число заданий находится под контролем. Эксперты и персонал по вводу данных непременно должны пройти обучение. Менеджер по разработке тестов обязан проследить, чтобы качество их работы проверялось дважды в день, для получения гарантий согласованности и надежности результатов. На ранних стадиях проверки осуществляются часто, а на более поздних, если работа экспертов оказалась удовлетворительной, их можно проводить реже.

Процент учащихся, которые не попытались выполнить предтестовые задания, помогает сделать важный вывод о том, как структурировать окончательную форму теста. Ниже приводятся общие принципы, регламентирующие такую работу.

- Если не менее 15% учащихся не попытались выполнить несколько заданий в конце теста, то предварительный тест, возможно, был слишком длинным. Рассмотрите возможность создания более легкого или более короткого окончательного теста или измените его трудность, и длину.
- Если не менее 15% учащихся не попытались выполнить задание, расположенное не в конце теста, то, возможно, ошибка допущена в способе представления этого задания либо оно слишком трудное. Возможно, испытуемые про-

пустили задание, не зная, как показать свой ответ, или не поняли его формулировку. Задание необходимо пересмотреть и провести предварительную проверку нового задания.

- Если определенная группа испытуемых (например, не менее 15% девочек) не попытались выполнить задание, но большинство других испытуемых его сделали верно, то, возможно, задание имеет систематическую ошибку. Проанализируйте задания, не включая их в окончательный тест.
- Если не менее 15% испытуемых согласованно не попытались выполнить задания в определенной форме (например, открытые задания), то, возможно, они просто не поняли, как представить свои ответы, или нуждались в подробной инструкции, как отвечать на такие задания. Рассмотрите возможность включения дополнительных тренировочных заданий или примеров их выполнения, использующих эту форму, и снова проведите предварительное тестирование с этими заданиями.

Вообще, пропущенные задания или задания, до которых учащиеся не дошли, кодируются числом 9. Убедитесь в том, что никаким заданиям не присвоен балл 9 за правильный ответ. Если это произошло, то можно использовать X (или другую букву алфавита), для кодирования оценок пропущенных заданий.

Экспертом необходимо разъяснить правила отнесения заданий к категории пропущенных. Задание считается пропущенным, если учащийся не сделал никаких пометок даже карандашом. Любую попытку ответить на задание, даже если она неразборчива, обычно рассматривают как неправильный ответ, а не как пропущенное задание.

Для заданий с множественным выбором должно быть разработано руководство по оценке, что позволит разработчику теста или рецензенту получить максимально возможное количество полезных данных по каждому заданию.

Задание с четырьмя вариантами ответов, например, можно закодировать как 1, 2, 3, 4, 8 или 9. Можно использовать больше цифр, чтобы отразить больше вариантов. Цифры 1, 2,

3 и 4 указывают вариант, который выбрал испытуемый. Кодовое обозначение 7 можно использовать, чтобы показать, что учащийся выбрал два или больше вариантов ответов и, возможно, не понимает, как ответить на задание. В четвертой книге серии с помощью кода 8 мы указываем, что испытуемый не пытался выполнить задание, и 9 – что это отдельное задание не предъявляли учащимся (поскольку оно было в другой тестовой форме), и поэтому его нельзя оценивать как неправильно выполненное.

Данные по заданиям с множественным выбором ни в коем случае не должны вводиться только как результат правильно-го или неправильного выполнения. Для расположения вариантов должен использоваться неявный образец нумерации от 1 до 4 или до 5, в зависимости от их числа.

Эксперт или сотрудник, вводящий данные, просто записывает номер выбранного учащимся варианта для каждого задания. Эксперту или сотруднику, вводящему данные, все не нужно знать, является ли правильным ответ учащегося. Пример формата ввода данных см. в дополнительном материале 4.2.

Дополнительный материал 4.2

Пример формата ввода данных для предтеста

Учащийся	Q1	Q2	Q3	Q4	← Порядок предъявления вопросов в тестовой форме
Ахмед Буга	2	3	2	1	
Мириам Висим	4	3	2	4	
Альмет Дурас	2	3	1	4	} ← Опции, выбранные каждым учащимся для каждого задания

Источник: разработки авторов.

Автор задания должен предоставить список правильных вариантов, или ключей, для каждого задания, которые аналитик данных вводит в компьютерную программу. Программное обеспечение для анализа данных оценивает ответ каждого испытуемого как правильный или неправильный на основе сопоставления с предложенным списком ключей.

Информация об опциях неправильных ответов, выбранных испытуемыми, позволит авторам заданий судить об их качестве и полноценности и принять решение относительно использования в окончательной форме теста. Например, если почти никто не выбрал ни один из двух вариантов неправильных ответов, то они, очевидно, не функционируют как эффективные дистракторы.

Заданиям с закрытым конструируемым ответом и открытым заданиям обычно присваивают 0 (неправильный ответ), 1 (правильный ответ) или 9 (пропущенные задания). Задания с частичным кредитованием могут быть оценены баллами 0, 1, 2 или 9.

Экспертная оценка предтестовых заданий, проводимая вручную, требует специального обучения экспертов. Необходим контроль качества их работы, такой же, как и при экспертной оценке результатов окончательного теста. Опытный автор задания должен провести обучение экспертов и наблюдать за процедурой оценки предтестовых заданий.

Авторам заданий следует использовать ответы на предтесты, чтобы пересмотреть и усовершенствовать свои руководства по оцениванию и оценочные категории до начала оценки результатов предварительного теста. Перед началом оценки вручную авторы должны взять пример законченных предтестовых буклетов и сравнить фактические ответы учащихся с ожидаемыми на задания с кратким ответом. Авторы могут использовать примеры буклетов для включения образцов ответов испытуемых в свои руководства. Оценочные руководства должны включать правила как для верных, так и неверных ответов. В дополнительном материале 3.12 приведены примеры ответов реальных учащихся, которые соответствуют каждой из категорий рекомендаций, включая нулевые оценки.

Чтобы принять во внимание весь диапазон ответов, которые фактически дают учащиеся, авторам необходимо усовершенствовать или расширить свои оценочные руководства. Иногда переделки могут быть весьма значительными. Испытуемые склонны к изобретению непредвиденных, но правильных ответов или необычных, но точных способов выражения своих идей. Такие примеры неожиданных ответов должны быть

добавлены в руководства по оценке, если они разумно обобщены. Если многие учащиеся дают ответы, которые трудно классифицировать как правильные или неправильные, то авторы заданий должны уточнить свои руководства, чтобы прояснить различия между ними. Команда по разработке заданий или менеджер по разработке тестов должны рассмотреть переделанные руководства прежде, чем будут оцениваться буклеты с предварительными тестами.

В ходе оценивания вручную авторы заданий должны получать информацию от экспертов о возможных дальнейших направлениях переработки руководств по оцениванию. Если для некоторого задания должен быть сделан существенный пересмотр руководства по оценке, то само задание также придется повторно оценить, чтобы гарантировать согласованность оценок.

Важно пересмотреть руководства по оцениванию таким образом, чтобы критерии для оценок и образцы ответов соответствовали диапазону фактических ответов учащихся. Если пересмотра не будет, то «потеря» некоторых заданий неизбежна, потому что ни один из ответов не будет соответствовать чрезмерно строгим требованиям руководства. Оценка других заданий может быть ненадежной, потому что эксперты, не зная, как оценить задания в том случае, если ответы на них не соответствуют критериям руководств, будут выносить свои собственные суждения.

Обычно более высокие оценки для заданий с частичным кредитованием обозначают более сложный или развернутый ответ. Оценка 2 балла предполагает «лучший» ответ, чем тот, который получает оценку 1 балл. Данные предварительного тестирования по заданиям с частичным кредитированием могут использоваться для сбора информации о категориях ответов учащихся. Эта информация может помочь скорректировать предтестовые задания или руководства по оценке. Ответы на предтестовое задание с частичным кредитированием могут быть закодированы цифрами 0, 1, 2 или 3, хотя эти оценки, возможно, не являются иерархическими. В таких неиерархических оценочных системах оценка 3 не считается более высокой, чем

оценка 2 или 1. Каждая из оценок 1, 2 и 3 обозначает правильный ответ, различающийся по виду или способу получения. Например, решить задачу по математике можно тремя различными способами. Чтобы учесть три возможных ответа, требуются весьма сложные руководства. Если все учащиеся на предварительном тестировании выбирают один и тот же метод, автор задания может пересмотреть руководство по оцениванию, чтобы сфокусировать его на более популярном методе решения с краткой ссылкой на другие возможные способы. В последнем случае необходимо также скорректировать рекомендации для окончательного теста, чтобы показать, что для правильного ответа должна быть выбрана оценка 1, независимо от использованного способа решения проблемы.

Авторы должны сообщить специалисту по анализу данных о случаях использования заданий с частичным кредитованием, чтобы обозначить категории, а не иерархии, которые позволяют аналитику дифференцировать ответы. Аналитик может присвоить оценку 1 любой категории правильных ответов. Таким образом, очень важно, чтобы эксперты понимали, когда они оценивают задания с частичным кредитованием с иерархическими категориями ответом и когда иерархии нет. Все категории ответов в заданиях с частичным кредитованием в окончательном тестировании нужно рассматривать как иерархические.

В четвертой книге настоящей серии есть специальный раздел по анализу пилотных данных, или данных предварительного тестирования. Этот раздел охватывает подходы к анализу и в классической теории тестов (СТТ), и в современной теории тестов (ИРТ). Современная теория тестов часто используется при анализе тестовых заданий для установления связи тестовых форм и в разработке шкал для сообщения результатов национального оценивания (Beaton and Johnson, 1989). Она обладает рядом преимуществ в ситуациях применения для шкалирования результатов оценивания. Современная теория тестов позволяет получить характеристики задания независимо от выборки испытуемых, которые на него отвечают, а также оценить характеристики испытуемого независимо от выборки заданий, предъявляемых ему. Современная теория тестов

наиболее полезна в тех случаях, когда учащимся предъявляют наборы заданий с множественным выбором. Однако она имеет некоторые недостатки. Один из них – сложность процедур применения, которые требуют от пользователей значительных умений и опыта. В тех случаях, когда необходимые умения и опыт не доступны в какой-либо стране, вполне приемлемым может быть использование классической теории тестов.

НАДЕЖНОСТЬ

В предварительном и заключительном тестировании необходимо привести свидетельства надежности теста. Мера надежности является показателем согласованности результатов тестирования. Надежность зависит от качества тестовых заданий, самого теста, способа, которым тесты предъявлялись, характеристик группы испытуемых (таких, как усердие, с которым они выполняют предварительный тест или тесты для национального оценивания) и качества оценивания тестовых заданий. Надежность рассматривается в четвертой книге серии.

Показатели надежности тестов изменяются в диапазоне от 0 до 1, где 0 соответствует тесту, в котором ответы учащихся полностью несостоительны (например, тест, в котором все испытуемые угадывают ответы на все задания), а 1 – тесту, который измеряет содержательную область с наивысшей согласованностью.

Агентство, занимающееся применением тестов, должно получить свидетельства надежности о степени, с которой отдельные предтестовые задания коррелируют друг с другом.

Эта информация обеспечивает меру внутренней согласованности тестовых заданий. Этот подход предполагает, что отобранные задания измеряют единственный конструкт, или характеристику, например такую, как способность к языку или математике. Обычно национальные и международные команды по оценке склонны удалять задания, которые не являются относительно гомогенными, т.е. не измеряющими единственный конструкт или характеристику. Гомогенность заданий может быть оценена с помощью такой формулы, как альфа Крон-

баха (Cronbach), коэффициент Кьюодера – Ричардсона (Kuder – Richardson) KR20 или KR21, или коэффициента надежности, полученного методом деления теста на две половины. Все эти методы имеются в программном обеспечении – статистическом пакете SPSS®.

Если тесты для оценки включают открытые задания или задания со свободно конструируемым ответом, то агентство, занимающееся применением тестов, должно установить, что метод оценивания надежен. Агентство должно также гарантировать, что каждый эксперт или корректор открытых заданий обучен выносить суждения относительно того, являются ли ответы учащихся приемлемыми. Такое обучение будет предполагать, что эксперты работают с командой по разработке теста, чтобы зарегистрировать список приемлемых и недопустимых ответов для каждого открытого задания. После обучения пары экспертов, работающих независимо, должны оценить каждое открытое задание по крайней мере из 60 претестовых буклетов, выбранных случайным образом, и вычислить процент полной согласованности оценок для всего набора заданий. Агентство должно получить разъяснения от разработчиков тестов, в которых указываются ответы к заданиям, вызвавшие среди экспертов разногласия относительно своей приемлемости. Тщательная работа по оценке открытых претестовых заданий должна обеспечить гарантии того, что остались лишь небольшие основания для разногласий по поводу приемлемых и недопустимых ответов, когда они рассматриваются в национальном оценивании.

ГЛАВА

5

ОТБОР ТЕСТОВЫХ ЗАДАНИЙ

Отбор заданий для окончательного теста детально рассматривается в четвертой книге серии – «Анализ данных национальной оценки учебных достижений». Он в наиболее значительной степени зависит, во-первых, от структуры теста, представленной в спецификации, а во-вторых, от измерительных свойств заданий.

Как правило, используют следующие критерии отбора для каждого задания.

- Задание соответствует проекту теста.
- Процент учащихся, давших на задание правильный ответ, находится в диапазоне от 40 до 80%.
- Для задания характерна низкая доля пропусков.
- Показатель дискриминативной способности задания (корреляция между оценкой по заданию и полной оценкой по тесту) должен быть больше, чем 0,2.
- Надежность теста повышается при включении в него данного задания.
- Систематическая ошибка задания находится в пределах, приемлемых для большинства групп учащихся.

Следующие положения относятся только к заданиям с множественным выбором.

- Коэффициент точечной бисериальной корреляции для ключа положителен и больше, чем 0,2.
- Все дистракторы правдоподобны (т.е. их выбрали по крайней мере 5% учащихся) и имеют нулевой или отрицательный коэффициент точечной бисериальной корреляции.

Табл. 5.1 показывает типичные результаты анализа задания с множественным выбором. С точки зрения статистики задание работает хорошо.

Заголовки столбцов отражают число вариантов ответов в задании (A, B, C, D). Вариант D является ключом, или правильным вариантом ответа, он приведен с оценкой 1 в скобках. Варианты A, B и C представлены с оценкой 0 в скобках. Первая строка таблицы показывает число учащихся, выбирающих тот или иной вариант. В этом задании 254 ученика выбрали правильный ответ. Вторая строка отражает проценты для вариантов ответов (число испытуемых для каждого варианта, выраженное в процентах от их общего числа). В задании 67% учащихся выбрали правильный ответ. Этот результат свидетельствует о том, что задание находится в пределах приемлемого диапазона трудности. Задание довольно легкое. То, что только 3,7% испытуемых выбрали вариант B, дает основания считать, что этот вариант мало привлекателен или неправдоподобен. Его изменение в сторону большей правдоподобности поможет усовершенствовать задание, которое будет нуждаться в повторной предварительной проверке на учащихся.

Следующая строка таблицы содержит значения точечного бисериального коэффициента корреляции для каждого варианта ответа. Для правильного ответа он равен 0,39. Для неправильных ответов все значения являются отрицательными. В заданиях с множественным выбором ответа точечный бисериальный коэффициент корреляции ключа имеет то же самое значение, что и показатель дискриминативности для задания. Последняя строчка показывает среднее значение оценок способности учащихся. У тех испытуемых, которые отве-

ТАБЛИЦА 5.1**Пример результатов анализа задания с множественным выбором**

Критерий	Опции ответов			
	A [0]	B [0]	C [0]	D [1]
Число учащихся	90	14	21	254
Процент	23,7	3,7	5,5	67,0
Точечный бисериальный коэффициент	-0,26	-0,21	-0,16	0,39
Средняя способность	-0,02	-0,48	-0,14	0,54

Источник: разработки авторов.

тили верно, средняя способность значительно выше, чем у выбравших неправильные варианты ответов. Этот результат также говорит о том, что задание работает хорошо.

Для открытых заданий можно привести следующие рекомендации.

- Если задание оценивается дихотомически, то его показатель дискриминативности (корреляция между оценкой по заданию и полной оценкой по тесту) должен быть больше, чем 0,2.
- Если задание предусматривает возможность частичного кредитования при оценивании, то его показатель дискриминативности должен быть положительным и выше, чем 0,3.
- Если каждое задание оценивается дихотомически, то большинство заданий должны иметь трудность в диапазоне от 40 до 80%.
- Если задание предусматривает возможность частичного кредитования при оценивании, то каждой оценочной категории должны соответствовать по крайней мере 5% ответов учащихся.
- Если задание предусматривает возможность частичного кредитования при оценивании, то общий процент учащихся, выполнивших задание, полученный путем сложения процентов тех, кто дал правильные и частично правильные ответы, находится в диапазоне от 40 до 80%.

- Если задание предусматривает возможность частичного кредитования при оценивании, то средняя способность испытуемых должна снижаться по мере перехода от самой высокой до самой низкой категории ответов.

Результаты анализа открытого задания с частичным кредитованием с хорошей статистикой приведены в табл. 5.2. Заголовки столбцов отражают категории ответов учащихся. Испытуемые получают 0, если отвечают на вопрос неправильно. Частично правильному ответу присваивается оценка 1 балл, а полностью правильному ответу – оценку 2 балла. Пропущенные ответы кодируются цифрой 9 и также оцениваются 0. Полный показатель дискриминативности, равный 0,47, является достаточно высоким. Обратите внимание на то, что для заданий с частичным кредитованием показатель дискриминативности не будет таким же, как значение точечного бисериального коэффициента корреляции для наивысшей оценочной категории.

Число и процент учащихся по каждой категории ответов отражены в первых двух строках таблицы. Почти половина испытуемых, пытавшихся выполнить это задание, ответили на него полностью неправильно. Больше чем 5% учащихся приходится на каждую категорию ответов, что позволяет использовать эти категории в дальнейшем. Точечная бисериальная корреляция увеличивается от 0 до 0,2, показывая, что категории оценки работают, как и предполагалось. Средняя способность учащихся, оценки которых соответствуют категории 2, равна -0,9. Учащиеся, оценки которых соответствуют категории 1, имеют среднюю способность – 1,53. Разность между ними больше, чем 0,5, что поддерживает обоснование этих двух категорий в руководстве по оцениванию, потому что они дифференцируют испытуемых по весьма различающимся группам способности.

Процент учащихся, которые не попытались выполнить это задание, весьма высок и составляет 27,3%. Этот результат необходимо рассмотреть в контексте профилей пропущенных заданий по тесту в целом.

ТАБЛИЦА 5.2**Пример результатов анализа открытого задания с частичным кредитованием**

Критерии	Категории ответов учащихся				Показатель дискриминативности = 0,47
	0 [0] 0	1 [1]	2 [2]	9 [0]	
Число учащихся	1,466	425	268	809	
Процент	49,4	14,3	9	27,3	
Точечный бисериальный коэффициент	0,09	0,11	0,45	-0,48	
Средняя способность	-1,66	-1,53	-0,90	-1,90	

Источник: разработки авторов.

В данном тесте большинство открытых заданий имело более 20% пропусков. Нежелание учащихся их выполнять является отдельной проблемой, а не связано со специфическими недостатками этих заданий.

Общий уровень трудности окончательного теста должен быть адекватен его цели. Окончательный тест, который предназначается для мониторинга результатов учебной деятельности всех учащихся в генеральной совокупности, должен иметь диапазон трудности, который соответствует диапазону способности целевой группы. Оценка, предназначенная для различных целей, таких как выявление учащихся, подготовка которых отвечает определенной точке на шкале, может подразумевать большое число легких или, наоборот, трудных заданий в зависимости от того, где на шкале установлена эта точка отсчета.

Опыт, накопленный в разработке предтестов для национальной оценки, свидетельствует о том, что авторы склонны к разработке заданий, которые в итоге оказываются слишком трудными. Частично эта тенденция объясняется тем, что ранее те же авторы разрабатывали задания для массовых экзаменов, в которых принято использовать вопросы достаточно высокого уровня трудности. Кроме того, многие авторы предтестовых заданий живут в городах и не имеют представления о низких результатах обучения, которые могут встречаться в отдаленных сельских районах.

Если большинство предтестовых заданий являются слишком трудными и если предварительный тест не содержит достаточного количества легких заданий, отвечающих критериям проекта теста, то потребуется еще один раунд предварительного тестирования с новым набором более легких заданий. Аналогично, если в тесте недостаточно трудных заданий, то необходимо будет провести еще одно предварительное тестирование с более трудными заданиями.

Специалист, который будет анализировать данные окончательного тестирования, должен рассматривать и данные предварительного тестирования. Совместно с авторами он может решить любые проблемы, связанные с обеспечением информации о классификации заданий, ключах оценки и способах оценивания заданий.

К интерпретации данных анализа предварительного тестирования и отбору заданий для окончательного теста следует привлечь команду авторов заданий. Решения о включении или исключении тех или иных заданий необходимо принимать исходя из проекта теста и предметной области учебного плана, на оценку которого направлен данный тест.

Обычно на каждом уровне обучения используется только одна окончательная форма теста. Она может состоять из двух или трех отдельных тестов, таких как тест по математике, чтению и письму. Эти тесты могут быть напечатаны в отдельных буклетеах или объединены в один.

Проблемы, касающиеся связи заданий, их отбора и размещения, которые возникают в ходе предварительного тестирования, могут также проявиться и в окончательном teste, если он будет иметь несколько форм.

Некоторые задания имеют превосходный концептуальный замысел, но плохую статистику. Эта ситуация может объясняться выбором неправильного способа представления задания. Учащиеся могут быть не знакомы с используемой в задании терминологией, не понять, каким образом они должны привести ответ, или материал стимула может запутать учащихся. Идеально, когда задания с плохой статистикой не просто корректиру-

ются, но и повторно проверяются в пробном тестировании. Однако если задания с плохой статистикой отвечают важным требованиям проекта, а другие задания недоступны, то приходится включать их в окончательный тест.

В принципе, в промежуток времени между созданием предварительного теста и заключительных форм задания не должны изменяться, потому что такие изменения могут повлиять на статистику. На практике тестовые агентства делают только незначительные изменения в нескольких заданиях, часто не больше, чем 4 или 5 в тесте из 30 заданий. Такие незначительные переделки включают:

- замену одного или двух слов, чтобы сделать формулировку задания более ясной или снизить трудность используемой лексики;
- устранение самого «слабого» варианта ответов в задании с пятью вариантами ответов;
- исправление грамматических ошибок или упрощение грамматических конструкций;
- улучшение расположения элементов рисунков в заданиях, таких как символы на диаграммах, или обеспечение согласованности заголовков.

ПРИМЕЧАНИЕ

Примеры, приведенные в этой главе, основаны на анализе заданий с использованием классической теории тестов. В четвертой книге настоящей серии данный подход рассматривается детально. В книге также представлен другой метод анализа, основанный на современной теории тестов, которая использует иной статистический подход и отличающуюся терминологию.

ГЛАВА

6

СОЗДАНИЕ ОКОНЧАТЕЛЬНОГО ТЕСТА

ПРОЕКТИРОВАНИЕ ОКОНЧАТЕЛЬНОГО ТЕСТА

К разработке окончательной формы теста необходимо привлекать специалиста по анализу данных тестирования, или статистика. Он или она должны удостовериться, что дизайн теста отвечает следующим требованиям:

- адекватность формы представления входной информации задачам анализа;
- адекватность метода регистрации ответов на задание задачам анализа;
- адекватность характера и возможностей руководств по оценке.
- горизонтальные связи с данными тестирования предыдущих лет или вертикальные связи, установление которых может потребоваться, являются статистически состоятельными.

Входная информация для учащихся, которую необходимо разместить на обложке тестовых буклетов, должна отражать цель теста и способ сообщения результатов. Учащиеся не должны испытывать затруднений в восприятии входных данных.

Обычно к входной информации относят:

- название школы;
- полное имя испытуемого;
- пол испытуемого;
- возраст испытуемого или дату его рождения;
- уровень обучения или класс испытуемого;
- основной язык испытуемого.

В некоторых странах каждому учащемуся присваивают уникальный национальный идентификационный номер. Использование таких номеров весьма желательно.

Также полезно оставлять на обложке специальную область для администратора тестирования, чтобы он мог сделать отметку о возможном полном или частичном отсутствии учащегося в связи с болезнью или о необходимости оказания спе-

Дополнительный материал 6.1

Пример обложки теста

ОЦЕНИВАНИЕ В ЦЕЛЯХ МОНИТОРИНГА ОБРАЗОВАНИЯ: 4-Й КЛАСС

Раздел для заполнения учащимся:

Школа _____

Месторасположение_____

Класс _____

Имя _____

Фамилия _____

Мальчик Девочка

Возраст лет и месяцев

Язык, на котором я в основном разговариваю дома, является английским

Да Нет

Раздел для заполнения администратором тестирования:

Этот учащийся отсутствовал на следующих тестах:

Чтение

Математика

Этот учащийся получил специальную помощь для выполнения тестов по

Чтению Математике

Опишите специальную помощь, предоставленную учащемуся: _____

Источник: Отдел образования Папуа – Новой Гвинеи, 2004.

циальной помощи испытуемому с ограниченными возможностями (см. дополнительный материал 6.1).

Вообще обеспечить правильную регистрацию идентификационных данных учащегося и связь с соответствующими результатами тестирования легче, если тесты по всем предметам приведены в одном буклете. Однако если информация на обложке каждого отдельного тестового буклета точно и четко введена перед первой сессией тестирования, то потенциальных источников для возникновения проблем с идентификацией не будет. При использовании одного тестового буклета администратор должен проследить за тем, чтобы буклеты вручались соответствующим испытуемым перед каждой последующей сессией тестирования.

Если применяются разные тестовые буклеты, необходимо предусмотреть эффективные процедуры для обеспечения соответствия между испытуемыми и буклетами при очередном использовании. С их использованием связаны следующие проблемы.

- Учащийся может записать свое имя по-разному в различных буклетах.
- Учащийся может использовать разные формы своего имени, например, в одном буклете – указать сокращенную форму, а в других – полную, или в одном буклете привести религиозное или артистическое имя, а в других – фамилию, или в одном буклете привести первое имя, а в других – второе имя или отчество.
- Учащийся может написать свое имя или его часть неразборчиво, по крайней мере, в одном из буклетов.
- Учащийся может быть не в состоянии вписать свое имя в один буклет или более.

Формат расположения и руководства по применению тестов должны быть четкими и последовательными. По возможности расположение компонентов теста в предварительной и окончательной версиях должно быть одинаковым.

Тест необходимо начинать с нескольких легких заданий, чтобы мотивировать слабых учащихся. Для заданий, охваты-

вающих некоторый диапазон уровней трудности, не стоит выбирать такой порядок следования, чтобы учащиеся бросали работу над тестом, не попробовав выполнить ряд трудных заданий. Также важно, чтобы некоторым наиболее медленно работающим учащимся был дан шанс на попытку выполнения трудных заданий, поэтому некоторые из них следует разместить ближе к началу теста. Самые трудные задания следует поместить в конце теста, потому что маловероятно, что учащиеся с более низкими способностями дойдут до конца теста. Задания, относящиеся к общему стимулу (например, параграфу или карте) должны быть представлены вместе, независимо от соображений, связанных с их уровнем трудности.

Основные принципы размещения связующих заданий для заключительной и предварительной форм теста являются одинаковыми (см. главу 4). При горизонтальной связи необходимо соотнести задания с предыдущим тестом, если результаты обучения сравниваются на протяжении долгого времени. Вертикальные связи требуются для сравнения результатов обучения между различными классами. Связующие задания должны быть помещены в начале или ближе к середине буклета и на одной и той же позиции в каждом тестовом буклете.

Метки заданий необходимо выделять в тестовых буклетах, чтобы задания можно было отследить (см. главу 3).

Менеджер по разработке тестов должен предоставить аналитикам данных тестирования крупноформатную таблицу, в которой отражено местоположение заданий в каждом буклете, включая связующие задания.

Решение о том, как учащиеся будут записывать ответы, необходимо принять на стадии разработки проекта теста (см. главу 2). Часто испытуемые отмечают ответы прямо в тестовых буклетах. Нужно предусмотреть достаточно свободного места для записи ответа учащихся, и для экспертов – чтобы они могли проставить свою оценку.

В других случаях, особенно в старших классах, могут использоваться отдельные ответные бланки. Причем учащимся должно быть предельно ясно, как связать ответ на задание или вариант, напечатанный в тестовом буклете, с соответствующей

позицией на бланке ответа. Если задания теста сгруппированы по юнитам, то будет целесообразна такая же организация расположения на бланках ответов.

ВЫЧИТКА И ТИРАЖИРОВАНИЕ

Длину тестового буклета определяют следующие факторы:

- число охваченных в тесте предметных областей;
- широта охвата в пределах каждой предметной области;
- форма заданий;
- использование иллюстраций;
- длина материалов стимулов;
- размер шрифта;
- финансирование, выделенное на тиражирование тестовых буклетов.

Авторы заданий с самого начала должны знать предполагаемое число страниц в окончательной форме теста. Если станет известно, что длина теста будет ограничена, то материал стимула, а также число диаграмм и иных иллюстраций необходимо сократить. Независимо от размеров свободного пространства для ответов макет должен быть ясным и не загроможденным.

Буклеты объемом менее 20 страниц (10 листов) обычно печатаются на листе формата А4 (210×297 мм) и сшиваются с одной стороны. Большие буклеты, как правило, печатают на листе формата А3 (420×297 мм) и сшивают посередине.

Буклеты с большим количеством страниц позволяют использовать более разнообразный и интересный материал стимула и более образные задания. К недостаткам можно отнести высокие расходы на печать и брошюровку. Кроме того, для них требуется больше подсобных помещений для пакетирования и оценивания, что также может существенно увеличить общие затраты на тестирование.

Тесты обычно печатают с двух сторон листа. Качество бумаги должно быть хотя бы таким, чтобы задания, напечатанные на одной стороне листа, не просвечивали и не накладывались на текст, напечатанный на другой стороне того же листа.

Для высококачественной печати фотографий требуется высококачественная бумага. То же самое касается сложных схем с большим количеством деталей.

Если учащиеся должны отвечать прямо в тестовых буклетах, следует использовать достаточно плотную бумагу, чтобы они случайно не разорвали ее и чтобы ответы не просвечивали сквозь страницу.

Бумага для обложки иногда имеет более высокое качество, чем бумага для текста, но это, как правило, ведет к росту затрат и не является обязательным.

Буклеты для разных классов должны различаться по цветовой гамме. Это даст гарантию, что учащимся будут выданы правильные буклеты. Цвет текста должен обеспечить легкость чтения.

Окончательный вариант теста обязательно должен проверить профессиональный корректор. Также стоит попросить компетентных коллег, которые не привлекались к разработке, прочитать окончательную форму теста с точки зрения пользователя, чтобы проверить, что она имеет смысл.

Для корректировки окончательных тестовых форм есть два наиболее подходящих периода. После того, как были собраны окончательные формы, их должны прочитать менеджер по разработке тестов, авторы заданий и, наконец, профессиональный корректор. Авторы заданий должны проверить исправления, сделанные корректором. Второй раз окончательные формы теста должны быть прочитаны, когда принтер выдает «распечатку» для проверки. Распечатка – это изображение страниц теста в том виде, в каком принтер их воспроизводит. Принтер будет обычно выдавать распечатки в течение нескольких дней после получения теста. Специалист по разработке тестов может принять ответственность на себя, а может нанять профессионального корректора. На вычитку и внесение правки нужно отвести по крайней мере две недели. Может потребоваться и больше времени – это зависит от наличия сотрудников для выполнения исправлений в буклете.

Корректоры часто находят сотни маленьких ошибок – в несогласованном использовании заголовков, пунктуации, форматировании, расположении заданий и правописании. Если корректура предварительных тестов была полной и обширной, и в заданиях было сделано минимальное количество существенных изменений, то теоретически корректура окончательных форм должна выявить немного ошибок, если они вообще имеются. На практике это случается редко.

Ошибки могут появиться в окончательных формах теста в тех местах, где раньше их не было. Корректорам обычно нужно несколько дней, чтобы полностью вычитать окончательные тестовые формы и руководства по предъявлению теста даже в тех случаях, когда предтесты прошли корректуру.

Заранее установите порядок работы для больших принтеров на несколько недель или даже месяцев. Принтер имеет характеристики скорости печати. Команда по национальной оценке может финансово поощрить быструю доставку распечатанных тестов или назначить штраф за их доставку с опозданием.

При печати тестовых буклете возможны ошибки. Самая распространенная из них – это пропуск некоторых страниц. Для исключения любых ошибок менеджер по разработке теста должен выборочно проверить распечатанные буклеты.

ГЛАВА



ОЦЕНКА ТЕСТОВЫХ ЗАДАНИЙ ВРУЧНУЮ

Члены команды, осуществляющей национальное оценивание результатов обучения, должны точно знать, что эксперты, оценивающие задания окончательного теста вручную, прошли специальную подготовку. К этому моменту руководства по оценке и задания, которые по итогам предварительного тестирования нуждались в пересмотре, должны быть изменены. До начала заключительной процедуры оценивания вручную авторам заданий следует отобрать образцы окончательных тестовых форм, проверить ясность и эффективность руководств по оценке и при необходимости внести незначительные поправки.

Подготовка специального оценочного центра для проверки заданий вручную и организация эффективной экспертизы результатов тестирования должны быть запланированы заранее. До начала оценивания заданий вручную членам команды следует ответить на следующие вопросы.

- Где будут храниться материалы тестов?
- Каким образом они будут доставлены в центр экспертного оценивания?

- Каковы гарантии безопасности материалов? (Тесты и руководства по оценке не должны покидать комнату, где ведется экспертиза.)
- Каково расписание для работ по оцениванию вручную? Будут ли они проводиться ежедневно в разное время (включая вечерние часы)? Будут ли эксперты работать по выходным?
- Как будут представлены данные в отчете экспертов?
- Какие вспомогательные материалы могут потребоваться? Зачастую эксперты предпочитают красную или зеленую ручку, потому что эти цвета более яркие и заметные. Обозначить задания, в которых у экспертов возникают вопросы, можно с помощью стикеров и флаеров.
- Какова форма оплаты – сдельная (по числу проверенных тестов) или времененная (по времени, затраченному на оценку)? Обе формы имеют преимущества и недостатки. В первом случае эксперты могут торопиться и, как следствие, выполнить работу небрежно, поскольку они будут стараться проверить как можно больше тестов. Поэтому велика вероятность того, что эксперты не будут выкладываться полностью, и результаты такой проверки будут некачественными. Разумным выходом могла бы стать оплата по затраченному времени, но с четким указанием минимального количества работ, которое должно быть проверено каждый день.

В оценочном центре обязательно должны быть: руководитель группы, ведущий специалист и сами эксперты. Лица, ответственные за отбор персонала, должны провести собеседование с кандидатами и проверить их рекомендации.

Руководитель группы экспертов несет ответственность за все ее действия. Он или она гарантирует, что работа идет по расписанию, решает текущие проблемы, наблюдает за процессом контроля качества работы и обеспечивает безопасность тестов, предохраняя их от рассекречивания. Руководитель должен быть опытным экспертом с хорошими управленческими и организаторскими навыками, способный на решительные действия, такие как увольнение сотрудников, не справляющихся со своими обязанностями.

Ведущий эксперт ответственен за мониторинг оценки в конкретной предметной области и осуществление процедур контроля качества. По каждому проверяемому предмету должен быть, по крайней мере, один ведущий специалист (например по математике, чтению и т.д.). Необходимо, чтобы ведущий эксперт имел опыт работы в соответствующей предметной области и в качестве эксперта и внушал уважение подчиненным.

Эксперты оценивают ответы учащихся. Обычно хорошими экспертами являются преподаватели. Они должны быть прилежными, надежными, последовательными в рассуждениях и отлично знать свою предметную область.

Для обучения экспертов менеджер по разработке теста назначает старшего среди авторов заданий по соответствующим предметным областям. Желательно, чтобы автор, обучающий экспертов, был вовлечен в разработку заданий и оценочных руководств, а также имел опыт в данном предмете. Менеджер по разработке теста мог бы взять на себя обучение экспертов в своей предметной области. Идеальный вариант – когда человек, обучавший экспертов на стадии предварительного тестирования, проводит обучение и для окончательного тестирования.

Каждой группе экспертов необходимо предоставить достаточно количество времени на обучение, которое обычно проходит в несколько этапов. В ходе обучения нужно учитывать, что экспертам предстоит оценить целый спектр возможных правильных ответов. Некоторые из них, возможно, не такие, которые выбрал бы сам эксперт, некоторые могут иметь небольшое сходство с ответами в учебнике. Кроме того, учащиеся могут не совсем грамотно представить ответ или использовать нестандартную лексику.

При обучении экспертов необходимо учитывать следующие аспекты.

- Эксперты не обладают свободой в определении степени пригодности ответов оценочных руководств или же эта свобода весьма ограничена. Они не имеют возможности высказывать личное мнение или предпочтения.

- Нельзя снимать баллы за орфографические или грамматические ошибки в тестах по чтению, математике или естественным наукам, если работа учащегося поддается распознаванию.
- Если эксперты не уверены, как оценить отдельный ответ, им следует обращаться за советом к ведущему эксперту.
- Эксперты должны использовать одну и ту же оценку (обычно 0) для всех неправильных и неразборчивых ответов, включая даже одну нечетко написанную букву или небрежно нарисованную линию.
- Эксперты должны использовать один и тот же оценочный код (обычно 9) для всех заданий, на которые учащийся не попытался ответить, т.е. на месте, предусмотренном для ответа, нет никакой отметки карандашом.
- Для облегчения ввода данных эксперт должен использовать для проставления оценки только специально отведенное для этого место в тестовом буклете.
- Эксперты не должны нести ответственность за суммирование баллов при выставлении полной оценки по тесту.

В обучении необходимо сделать акцент на понимании экспертами того, что влечет за собой оценка, и на достижении согласованности в оценках. Методы обучения могут изменяться. Ниже мы представляем один из многих возможных вариантов.

- Преподаватель просит экспертов самостоятельно ответить на каждое задание, оцениваемое вручную. Таким образом, эксперты познакомятся с заданиями, должным образом проучут и поймут их.
- Преподаватель выдает каждому эксперту четыре или пять выполненных тестовых буклетов. Он обсуждает первое задание и руководство по оценке, и затем эксперты оценивают это задание в своих тестовых буклетах. Преподаватель мотивирует группу к анализу любых расхождений или сомнений в том, как оценить ответ. Экспертов просят выделить те части ответов, которые могут отличаться от примеров в оценочном руководстве. После того как было все-

сторонне рассмотрено и проанализировано первое задание, преподаватель переходит к следующему. Обычно на это уходит несколько часов.

- Эксперты проходят повторное обучение, во время которого они работают парами. Они оценивают несколько заданий самостоятельно, а затем проверяют работу друг друга и обсуждают возможные расхождения. Если они не могут прийти к согласию, им следует обратиться за консультацией к преподавателю. После того как обучение закончено, ответственность по управлению группой экспертов принимает на себя ведущий эксперт. Он должен сообщать преподавателю обо всех возникающих проблемах.
- В процессе реальной проверки ведущий эксперт каждый день выбирает несколько проблемных заданий и проводит краткие обсуждения по ним, чтобы сосредоточить на них внимание и добиться согласованности оценок экспертов.

Качественная проверка заданий с кратким ответом для национальных программ оценивания может включать первичную перепроверку почти 100% всех тестовых буклетов. Как правило, ее проводят ведущие эксперты. Постепенно число перепроверяемых тестовых буклетов может быть сокращено до 10–20%, поскольку эксперты постепенно становятся более последовательными и надежными в своих оценках.

Если к проверке привлекается большое число сотрудников, то необходимо назначить несколько ведущих экспертов, чтобы гарантировать ее качество и обеспечить быструю обратную связь по поводу любых возможных ошибок. Ведущий эксперт должен требовать повторной оценки заданий в ранее проверенных тестовых буклетах, где были допущены ошибки при выставлении оценок. Процедуры для устранения несоответствий в оценках также должны быть разъяснены. Обычно учитывается именно та оценка, которую выставляет ведущий эксперт.

Тест по языку может содержать одно или несколько заданий, которые требуют написания эссе. Эссе часто оценивают два эксперта. При этом второй эксперт не знает оценки, присвоенной первым экспертом. Затем оценки сравниваются.

Разница в один балл обычно учитывается, и эти две оценки усредняются. Если разница больше, то это значит, что по крайней мере одна из оценок по результатам дискуссии должна быть изменена. Если эксперты не могут прийти к согласию, то решение должен принять ведущий эксперт.

Оценка вручную требует предельной концентрации. Эксперты не должны работать без перерывов и выходных. Длительность рабочего дня не должна превышать шести – шести с половиной часов. Рабочий день может включать трехчасовую утреннюю сессию с коротким перерывом и трехчасовую дневную сессию также с небольшим перерывом. Экспертам, работающим медленнее своих коллег, могут потребоваться дополнительные полчаса, чтобы выполнить дневную норму. Также эксперты должны заполнять ежедневные бланки посещаемости.

ЧАСТЬ

2

КОНСТРУИРОВАНИЕ
АНКЕТ

ГЛАВА

8



ПРОЕКТИРОВАНИЕ АНКЕТ

Анкета представляет собой набор заданий, предназначенных для получения информации об испытуемых. Запрашиваемые сведения могут быть самыми различными: личные характеристики учащихся, имеющиеся квалификации и опыт, условия работы, краткая автобиография, вероисповедание, политические убеждения, мнение по определенным вопросам.

Цель национальной оценки – получить надежные приближенные оценки достижений учащихся (измеряются с помощью специально разработанного теста), а также ключевых переменных (измеряются с помощью анкеты), связанных с различиями в учебных достижениях. Данные тестирования позволяют оценить результаты обучения, а проводящееся в сочетании с тестированием анкетирование помогает получить данные о переменных, которые связаны или объясняют различия в результатах выполнения тестов. Так, анализируя данные анкет, можно сделать вывод, что плохие результаты испытуемых объясняются отсутствием в их школе библиотеки, а высокие результаты показывают учащиеся тех школ, где преподаватели регулярно участвуют в программах профессионального развития. Данные анкетирования

помогают найти ресурсы для улучшения качества результатов обучения.

Хорошая анкета собирает данные о переменных, по которым тактики хотели бы иметь точную информацию, на которые они могут и хотят повлиять, а также о тех переменных, которые, согласно результатам исследований, могут влиять на учебные достижения.

Типичная ошибка многих анкет состоит в том, что они предназначаются для сбора излишне большого количества информации. Тактиков, как правило, интересует информация только о нескольких ключевых переменных. Кроме того, даже в тех случаях, когда для сбора данных существуют весомые научные причины, возможные политические и социальные последствия могут указать на нежелательность национальной оценки, поскольку она является не самым подходящим механизмом.

Помимо анкет в странах, которые ведут точный и надежный учет характеристик школ, преподавателей и учащихся, могут быть и другие источники информации – правительственные отчеты. Пренебрегать ими не следует. Обращение к таким отчетам может быть значительно дешевле и легче, чем анкетирование.

Дизайн анкеты должен описывать, какие виды данных будут собраны, как они будут анализироваться и сообщаться и какой вклад в совершенствование системы образования они внесут.

Разработка дизайна анкет предусматривает следующие этапы.

- Выберите цель анкетирования и способы использования данных.
- Разработайте проект, который определит реципиентов, целевые области, виды заданий, способы кодирования или оценки и протокол предъявления анкеты (для интервьюера или самостоятельного выполнения).
- Составьте задания. В качестве разработчиков или экспертов привлеките специалистов. Они помогут проанализировать задания, при необходимости скорректировать их

и сделать дизайн анкеты ясным и понятным, что облегчит задачу респондентам и сделает обработку данных более эффективной.

- Составьте план анализа данных для обработки собранной информации, выбора переменных измерения и показателей для последующего статистического анализа.
- Проведите предварительную проверку, или апробацию, анкет, чтобы установить пригодность заданий и предполагаемых категорий ответов.
- Проанализируйте предварительные данные проверки, скорректируйте анкеты и сформируйте окончательные анкеты для предъявления респондентам.

Процесс разработки анкет и перечень привлекаемых специалистов представлен в табл. 8.1.

Анкеты и инструкции по их предъявлению должны быть подготовлены и предварительно апробированы в реальных условиях, как и тесты. Значит, проекты анкет должны разрабатываться в то же самое время, что и проекты тестов. Время написания анкет и окончательных тестовых заданий также должно совпадать. Экспертизу также необходимо проводить одновременно.

СОДЕРЖАНИЕ АНКЕТ

Анкета должна собирать информацию по ключевым переменным, которые могли бы объяснить различия в результатах выполнения учащимися тестов учебных достижений. Таких взаимодействующих переменных множество. Однако анкета может быть сфокусирована только на некоторых из них.

Тактики, как правило, хотят получить информацию о переменных, которые связаны с важными образовательными проблемами в их стране, такими как изучение языка, различия в распределении образовательных ресурсов или отношение к женскому образованию. Возможно, они не уверены в том, какие именно параметры выбрать для исследования. Они могут дать длинный список переменных, которые выбраны на ос-

ТАБЛИЦА 8.1**Компоненты процесса разработки анкет**

Компонент	Описание	Привлекаемые специалисты
Цель	Разъясните цель анкетирования и варианты использования данных анкет	Тактики, заинтересованные лица и менеджер по разработке тестов
Проект	Разработайте дизайн проекта для определения респондентов, целевых областей, видов заданий, кодов и протоколов предъявления анкет	Менеджер по разработке тестов, эксперты по предметам, специалисты по анализу данных тестирования, авторы заданий, опытные преподаватели, тактики и ключевые заинтересованные лица
Задания	Составьте задания анкет	Менеджер по разработке тестов, авторы заданий
	Проведите чистку заданий для ясности и полноценности работы экспертов	Менеджер по разработке тестов, авторы заданий
	Повторно проанализируйте анкеты	Менеджер по разработке тестов, представители правительенных кругов и иные заинтересованные лица
План анализа данных	Определите план обработки информации для выбора переменных измерения и показателей для различных видов анализа	Специалист по анализу данных тестирования и менеджер по разработке тестов
Предварительная проверка анкет	Разработайте дизайн и откорректируйте анкеты для предварительной проверки	Менеджер по разработке тестов, авторы заданий, верстальщики, дизайнеры, корректоры
	Разработайте инструкции по проведению анкетирования для предварительной проверки анкет и обучения организаторов апробации	Менеджер по разработке тестов и авторы заданий
	Проведите предварительную проверку анкет одновременно с апробацией тестов	Менеджер по разработке тестов, менеджер по логистике и организаторы тестирования
Окончательное анкетирование	Проанализируйте данные предварительного анкетирования	Специалист по анализу данных и менеджер по разработке тестов
	Проведите чистку анкет и инструкций по их предъявлению на основе данных апробации и обратной связи по результатам предъявления анкет	Менеджер по разработке тестов, авторы заданий и специалист по анализу данных
	Разработайте окончательную форму анкет	Менеджер по разработке тестов, авторы заданий, верстальщики, дизайнеры, корректоры

Источник: разработки авторов.

новании их личного опыта и наблюдений и которые, как они думают, «должны» быть отражены в анкете. Такой список необходимо сократить до перечня из нескольких требований, которые, вероятно, будут полезны в формировании содержания анкет.

Тактики, возможно, не знают, что процесс анализа и сообщения данных анкетирования дорог и требует методической экспертизы. Поскольку ресурсы всегда бывают ограничены, анкеты должны быть краткими и релевантными. Собранные данные должны также быть методически приемлемыми, если их предполагается использовать для объяснения результатов выполнения тестов учащимися. Модели, используемые в других странах при национальном оценивании, могут обеспечить только приблизительные руководящие указания. Ведь у каждой страны есть свои собственные потребности, которые, тем не менее, должны определить адекватный характер анкет.

Менеджер по разработке теста или сотрудник, ответственный за разработку анкеты, должен подсказать тактикам, какие именно ключевые переменные, вероятно, обеспечат полезную информацию. Для этого потребуется привести релевантные примеры, чтобы тактики могли рассмотреть возможные варианты применения собранной информации. Эта информация поможет в дальнейшем скорректировать список переменных, к которым будут обращаться в процессе анализа.

Поскольку анкеты разрабатываются для разных категорий – учащихся, их родителей, преподавателей и руководителей учебных заведений, то и вопросы в них будут разные. Ниже приведены примерные варианты анкет для каждой из этих групп.

Анкета для учащихся

В анкетах для испытуемых можно запрашивать следующие сведения:

- пол, возраст, основной язык (все данные, которые обычно собирают на первой странице тестового буклета);

- уровень образования: количество лет, проведенных в школе и вне школы;
- возможности для посещения школы;
- стремление к успеху, личное или семейное отношение к ценности школьного образования;
- восприятие характеристик окружающей среды: чувство безопасности в классе, дружелюбное отношение других учащихся или поддержка со стороны преподавателей.

Анкета для родителей

Анкеты для родителей могут запрашивать следующие данные:

- национальность, пол и основной язык;
- характеристики домашней среды: доступность книг, наличие письменных столов и освещения;
- основные характеристики членов семьи: образование родителей и язык домашнего общения;
- характеристики отношения родителей к образованию: считают ли они обязательным посещение школы детьми, понимают ли ценность образования и как воспринимают качество образования;
- характеристики отношения к домашней работе учащихся и условия для выполнения работы дома;
- допустимость и доступность образования для детей;
- ожидания относительно образовательных достижений детей;
- характеристики участия в жизни школы, в работе школьных комитетов;
- характеристика табелей успеваемости учащихся и ценностное отношение к ним;
- финансовая помощь школе в форме платы за учебники и образование.

Анкета для преподавателя

В анкетах для преподавателей может запрашиваться следующая информация:

- пол и возраст;
- родной язык;
- условия работы преподавателя, такие как размеры класса, доступ к учебным ресурсам, процент учащихся, имеющих учебники, возможность замены заболевшего преподавателя, помочь в мотивации учеников;
- квалификация преподавателя, стаж работы с указанием количества лет, проработанных в этой школе;
- развитие в профессиональной области: стремление к повышению квалификации и к профессиональному росту, интерес к обучению и время, затрачиваемое на подготовку к урокам;
- наличие поддержки в учебной работе: визиты руководителей, школьных инспекторов или сторонних наблюдателей;
- характеристики владения учебными методами и средствами, такими как язык обучения, контрольно-оценочная система, стиль обучения;
- удовлетворенность условиями работы, такими как время пребывания на работе, заработка плата и уровень руководства;
- отношения со школьным сообществом – с родителями других учащихся, участие в работе школьных комитетов и событиях школьной жизни;
- расстояние от места проживания преподавателя до школы.

Анкета для директора школы

В анкетах для директоров могут запрашиваться следующие сведения:

- пол и возраст;
- опыт в образовательной и управлеченческой деятельности, квалификация;
- характеристики среды в школе с точки зрения материально-технического оснащения: качество зданий и вспомогательных средств учебного процесса, пригодность ресурсного обеспечения;

- данные школьных отчетов об изменении числа учащихся, пропусках учащимися или преподавателями занятий, частоте смены школ учащимися;
- профессиональные обязательства по развитию руководящих навыков: доступность профессионального роста и интерес к нему, интерес к дополнительному образованию;
- стиль лидерства и использование рабочего времени;
- удовлетворенность условиями работы: временем пребывания на работе, заработной платой, частотой проверок со стороны руководства;
- отношения со школьным сообществом: взаимодействие с родительским комитетом, участие в культурных мероприятиях.

ПРОЕКТ АНКЕТЫ

Проект помогает осуществлять руководство процессом разработки анкеты. В нем должны быть указаны ключевые проблемы образовательной политики, которые определяют основное содержание анкеты, респондентов, перечень ключевых переменных, которые анкета должна будет охватить, форму заданий, виды категорий ответа и протоколов предъявления анкет.

В дополнительном материале 8.1 представлен пример проекта анкеты, используемой для сбора информации о системе ценностей учащихся и их отношении к школе и школьному коллективу в государстве Папуа – Новая Гвинея. Там был проведен ряд реформ в образовании и обновлены учебные планы, которые подчеркивали необходимость акцента на формировании ценностного отношения учащихся к коллективу и приобретению тех навыков, которые помогут им по окончании школы оказывать конструктивную помощь окружающим в налаживании сельской жизни. Цель анкетирования – собрать информацию об ожиданиях учащихся, их нынешнем восприятии школы и школьного сообщества. Анкета была выдана всем учащимся, которые выполняли тесты в процессе национальной оценки.

Дополнительный материал 8.1**Проект анкеты по отношениям и ценностям****Часть I**

Целевая область	Отношение к школе	Представления о жизни в Папуа – Новой Гвинее	Восприятие школьного коллектива
Число вопросов	10	15	15
Респонденты	3-й класс 5-й класс 8-й класс	5-й класс 8-й класс	5-й класс 8-й класс
Категории ответов	Да или нет	Да или нет	Да или нет

Часть II

Отношение к школе	Представления о жизни в Папуа – Новой Гвинее	Восприятие школьного коллектива
Представления о личных успехах, предполагаемая длительность обучения в школе и планы на будущее	Отношения к образованию: обучение на родном языке, обязательное образование, роль школы, женское образование и роль женщин в обществе	Уровень взаимодействия в коллективе: поддержка школы, участие в коллективных мероприятиях и разделение средств
Помощь и восприятие со стороны преподавателей, дружелюбие одноклассников, элементы запугивания, если они есть, готовность приобрести школьных друзей из другой деревни	Отношение к месту проживания: есть ли намерение остаться в нем, если нет, аргументы для переезда	Усвоенное отношение школьного коллектива к женщинам

Часть III

Ценности	Представления о жизни в Папуа – Новой Гвинее	Восприятие школьного коллектива
Ценностное восприятие конфликтных решений	Отношение к конфликтным решениям и борьбе	Уровень конструктивного участия в жизни коллектива и использование компромиссов для решения проблем
Ценностное восприятие личной гигиены и полезных привычек	Отношение к алкоголю и наркотикам	Восприятие коллективом проблем, вызванных наркотиками и алкоголем

Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

ЗАДАНИЯ АНКЕТЫ

Когда принимаются решения относительно заданий анкет – их числе, времени, отводимом на ответы, о средствах, необходимых для анализа результатов, и уровне сложности анализа – необходимо принять во внимание несколько соображений. Краткая анкета, содержащая небольшое число конкретных заданий, должным образом проанализированная, поможет выявить необходимую информацию быстрее, чем длинная, подробная анкета, данные которой полностью никогда не обрабатываются.

Число заданий, необходимых для измерения определенной переменной, зависит от ее характера. Некоторые переменные, такие как пол или возраст, могут быть измерены непосредственно с помощью одного вопроса. А для получения полной информации о социально-экономическом статусе, как правило, необходимо задать несколько вопросов – об уровне образования родителей, их рабочем статусе, месте проживания и наличии собственности. *Сырая переменная* является результатом, полученным на основе прямой меры. *Агрегированная переменная* объединяет данные, полученные на основе двух или более заданий для представления конструкта. Тактиki в образовании вообще считают, что результаты анализа, основанного на сырьих переменных, легче интерпретировать, чем результаты, основанные на агрегированных переменных.

Выбор вида переменной для обоснования конструкта должен быть основан на обширной исследовательской практике и существующих в стране условиях. И в национальных, и международных обзорах используются агрегированные переменные. Одно из международных исследований оценивания конструктов «чтение для утилитарных целей» и «чтение для удовольствия» было основано на агрегированных переменных (см. табл. 8.2).

Принимая решение относительно числа заданий для измерения той или иной переменной, необходимо учитывать ситуацию в данной стране. Там, где условия получения педагоги-

ТАБЛИЦА 8.2

Функции чтения в международном исследовании (весовые коэффициенты использованы для создания двух новых переменных: «чтение для утилитарных целей» и «чтение для удовольствия»)

Задание (сокращенно)	Весовые коэффициенты	
	Утилитарная цель	Удовольствие
Помогает мне в школе	0,75	
Помогает сдавать экзамены	0,74	
Помогает изучать другие школьные предметы	0,73	
Помогает мне лучше работать	0,65	
Я могу поступить в колледж	0,65	
Помогает получить хорошую работу	0,63	
Мои родители думают, что чтение важно	0,58	
Я получаю удовольствие		0,76
Оно является захватывающим		0,72
Оно интересно		0,71
Чтение – вход в другой мир		0,68
Приятно вспоминать о прочитанном		0,54
Приятно воображать себя историческим персонажем в данной истории		0,53
Приятно заниматься этим одному		0,53
Помогает расслабиться		0,50

Источник: данные получены из: Грини (Greaney) и Ньюман (Neuman), 1990 (табл. 8, коэффициенты менее 0,20 исключены).

ческого образования на всей территории примерно одинаковы и преподаватели проходят по крайней мере двух- или трехлетнюю подготовку в профильных учебных заведениях, для измерения педагогического образования может быть вполне достаточно одной сырой переменной. Там же, где педагогическое образование неоднородно, учебные заведения различаются по качественному уровню и преподаватели зачастую вынуждены проходить обучение по месту работы, для адекватного отражения ситуации в педагогическом образовании необходимо агрегирование множества сырых переменных для представления конструкта. Точно так же в богатой стране условия для

домашнего обучения могут быть измерены с помощью единственной сырой переменной – доступности Интернета, а в бедной стране для их характеристики потребуется совокупность показателей, таких как доступность стола, стула, лампы, карандашей, бумаги и учебников.

Решение о том, какую переменную использовать (сырую или агрегированную), чтобы получить адекватную меру, также зависит от значимости возможных сырых переменных. Например, если специалисты, проводящие анкетирование, твердо уверены в том, что на качество преподавания в большой степени влияет место предыдущей работы, необходимо включить в анкеты вопросы о том, где именно работал преподаватель и в течение какого времени он проходил обучение. Если они убеждены, что на результаты выполнения теста учащимися может повлиять стаж преподавателя, независимо от того, где именно он его получал, тогда будет вполне достаточно единственной сырой переменной.

ФОРМА ЗАДАНИЙ

Задания с регламентированным выбором ответов намного легче, быстрее и дешевле обработать, чем открытые задания. Поскольку число готовых ответов ограничено, обработка данных осуществляется после простого ввода в компьютер ответа респондента. Ответы на открытые задания, напротив, необходимо сначала обработать вручную прежде чем они будут введены в компьютер.

Данные анкет обычно суммируются для отчета. Например, ответы на вопрос о времени, ежедневно затрачиваемом учащимися на дорогу до школы, можно разбить на несколько укрупненных категорий: «меньше часа», «от одного до двух часов» и «больше двух часов». В открытой версии этого задания некоторые учащиеся укажут время в минутах, другие – в часах, некоторые могут написать «длительное время», кто-то и вовсе даст неразборчивый ответ. Диапазон ответов будет большим, а категоризация ответов обязательно внесет элемент субъективизма, включая принятие экспертных решений о том,

как классифицировать некоторые ответы, например «длительное время».

Задания с множественным выбором предпочтительны в том случае, если могут быть сделаны правдоподобные предположения о вероятном диапазоне ответов и предлагаемых различиях в категориях ответов для большинства респондентов. Однако если есть некоторая неопределенность, то могут быть использованы более дифференцированные категории ответов. После того как данные введены, необходимо принять решения о том, какие категории ответов несут незначительную информационную нагрузку, какие могут быть объединены, а какие исключены из задания (например, если их никто не выбрал).

Использование открытых заданий может быть вполне уместно, если анкеты предъявляют маленькой выборке и средства для классификации ответов вручную являются доступными. Иногда полезна предварительная проверка, или апробация, заданий в открытой форме, которая предоставляет информацию по разработке категорий ответов в заданиях с множественным выбором.

ЯЗЫК АНКЕТЫ

Анкета должна быть написана на языке, на котором в состоянии читать и связно писать респонденты. Однако выбор языка должен быть согласован с возможностями оценочной шкалы. А вообще, как правило, в анкетах используется тот же самый язык, что и в тестах.

РЕСПОНДЕНТЫ

С учетом того, что учащиеся всегда оставляют некоторые общие сведения о себе на первой странице тестовых буклетов, выбор респондентов для анкетирования зависит от того, что именно хотят знать тактики и от возможности получения этой информации с высокой надежностью и эффективностью.

С респондентами могут быть связаны следующие проблемы.

- Учащиеся могут быть слишком маленькими по возрасту, чтобы заполнить анкеты надежно и точно.
- Недостаток средств может ограничить предъявление анкет небольшим по численности группам респондентов – преподавателям или директорам школ скорее, нежели тысячам учащихся.
- Некоторые родители могут быть неграмотными или не всегда сдавать анкеты.
- Преподаватели и директора школ не захотят отвечать на вопросы, если анкеты покажутся им чересчур длинными или если они почувствуют, что честные ответы могут угрожать их дальнейшей работе.

Независимо от решения, принятого относительно респондентов, образцы заданий, отобранных для анкет, должны быть представлены респондентам. Если учащимся предъявляют анкету, то в анкетировании должна принимать участие та же выборка, что и в выполнении теста. Относительно размеров выборки для предъявления анкет преподавателям, руководству школ и родителям необходимо проконсультироваться с экспертами.

ПРЕДЪЯВЛЕНИЕ АНКЕТ

На вопросы анкеты обычно отвечают в письменной или устной форме (интервью). Во втором случае требуется, чтобы задавал вопросы и записывал ответы специально обученный интервьюер (возможно использование кодов ответов, представленных в самой анкете).

При крупномасштабной оценке большая часть анкет разрабатывается и предъявляется в группах респондентов, чтобы снизить общую стоимость процедуры. Анкеты должны содержать инструкции о том, как нужно отвечать на вопросы. Инструкции могут включать объяснения причин сбора информации.

Сбор данных анкетирования может проводиться под наблюдением специалиста по апробации, который собирает анкеты после того, как респонденты закончили их заполнение.

ПЛАН АНАЛИЗА ДАННЫХ

План анализа данных анкетирования определяет, на получение каких сведений направлено каждое задание в анкете и как они будут использоваться в дальнейшем. Привлечение помощи со стороны опытных специалистов на этапе планирования позволяет гарантировать, что данные анкетирования будут тщательно проверены и что полученные результаты будут корректны и состоятельны (см. четвертую книгу настоящей серии).

План должен содержать следующее.

- Характеристики переменных измерения. Способ анализа данных анкетирования зависит от характеристик измеряемых переменных. Числа, выбранные для номинальных, или категориальных, переменных (например, пол), в действительности являются всего лишь метками, которые могут использоваться только для различия групп респондентов. Числа, выраженные в порядковых шкалах (например, для ответов, указывающих на степень согласия с предложенным утверждением), содержат информацию относительно расположения величин на шкале. Однако часто их рассматривают в статистическом анализе так, будто они обладают свойствами интервальной шкалы (например, шкалы температуры) или шкалы отношений (например, стаж преподавания).
- Каким образом данные по совокупности переменных будут интегрированы для образования новой переменной и как она будет использоваться. Например, переменная «материальное положение семьи учащегося» может быть получена из таких переменных, как «домашний доход», «местожительство», «число комнат в доме», «домашнее имущество», «число детей в семье» и «уровень образования родителей». То, каким образом эти переменные будут агрегированы для представления показателя «материальное положение семьи учащегося», нужно рассматривать на этапе планирования анкеты.

ГЛАВА

9



НАПИСАНИЕ ЗАДАНИЙ ДЛЯ АНКЕТ

Выбор способа формулирования задания должен исходить из того, на получение каких сведений оно направлено. Кроме того, респонденты должны быть в состоянии дать ответ на данное задание. Например, не стоит просить учащихся младших классов вспомнить, сколько дней они отсутствовали в школе в течение учебного года, потому что вряд ли они смогут точно ответить. В лучшем случае они вспомнят, сколько дней пропустили на предыдущей неделе.

Формулировки заданий должны быть настолько простыми и ясными, насколько это возможно. Необходимо использовать знакомую учащимся лексику и короткие предложения. Если анкеты не предъявляют устно, то все респонденты должны уметь прочитать их. Также важно, чтобы все респонденты понимали задания анкеты одинаково. В противном случае достоверная интерпретация результатов будет затруднительной.

В первой части анкеты задание может быть представлено в виде вопроса, неполного предложения или утверждения, которое респонденты должны оценить.

Стиль, в котором обращаются к респондентам, должен быть последовательным. Можно использовать один из следующих вариантов:

- во втором лице: например: «Сколько вам лет?»;
- в первом лице: например: «Я еду в школу на ...».

ВОПРОСЫ

Вопросы должны быть ясными и однозначными. Приведенный ниже вопрос таковым не является.

Как долго Вы были преподавателем?

В данной формулировке непонятно, что имеется в виду: время, которое прошло с момента окончания учебного заведения, или фактический стаж работы по специальности. Любой преподаватель, на время прерывавший преподавательскую деятельность, а затем вернувшийся к ней, например женщина, потратившая часть времени на воспитание детей в семье, будет испытывать затруднения при ответе на этот вопрос. Его точный смысл остается неясным. Для полноценной замены нужны по крайней мере два вопроса:

Когда Вы закончили обучение по специальности?

Какой стаж работы в качестве преподавателя Вы имеете?

Однако при ответе на второй вопрос также возникают проблемы, потому что неясно, как измерять стаж. Например, если респондент работал по специальности на протяжении пяти лет с графиком работы один раз в неделю, то может ли стаж считаться за пять лет или он эквивалентен одному году? Если почти все должности преподавателей в данной стране предполагают полный рабочий день, то двусмысленность маловероятна, однако в ситуациях, когда возможна работа на полставки, неоднозначность неизбежна. Вопрос можно было задать так:

Какой стаж преподавания при полной загруженности рабочего дня (или эквивалентной полному рабочему дню) Вы имеете?

Стоит определить, что имеется в виду под словом «эквивалентный».

УТВЕРЖДЕНИЯ

Задание, начинающееся с утверждения, предполагает, что респонденты должны выразить свое отношение к нему. Оно может быть выражено словами: «полностью согласен», «согласен», «не согласен» или «полностью не согласен».

Старайтесь избегать отрицательных утверждений, таких как, например, «Я не люблю школу», потому что это может привести к путанице. Учащиеся, которые действительно любят школу, отвечая на данный вопрос, должны выбрать категорию «не согласен». Двойные отрицания часто оказываются трудными, особенно для учащихся младшего возраста.

По возможности используйте нейтральные формулировки. Например, утверждение «мне нравится школа», лучше чем «я люблю школу». Более теплое отношение к школе учащиеся могут выразить с помощью ответа «полностью согласен».

Утверждение должно быть сосредоточено только на одной проблеме. Такие утверждения, как «Я упорно тружусь и хорошо учусь», употреблять не следует. Учащимся, которые получают хорошие и отличные оценки, не прикладывая к этому больших усилий, будет трудно определиться с ответом. Те же учащиеся, которые серьезно и ответственно относятся к учебе и тратят много времени и сил на подготовку к занятиям, могут согласиться с данным утверждением, хотя, возможно, они не преуспевают в оценках. Поэтому приведенное высказывание лучше разбить на два: «Я вкладываю в учебу много времени и сил» и «Я хорошо учусь».

ВАРИАНТЫ ОТВЕТОВ

Хорошие варианты ответов не допускают многозначного толкования. Следующие ответы учащиеся с большой долей вероятности поймут по-разному.

Сколько книг находится в классной библиотеке?

- A. ни одной
- B. немногого
- C. некоторое число
- D. много

Ответы должны быть точно определены количественно, так, чтобы значение каждого было однозначно ясно:

Сколько книг находится в классной библиотеке?

- A. библиотеки в классе нет

B. от 1 до 10

C. от 11 до 20

D. более 20

Иногда ответы могут допускать различные толкования для разных респондентов, но эти различия является частью искомой информации, как в следующем задании.

Оцените вашу школьную библиотеку:

A. библиотеки в школе нет

B. бедная

C. адекватная потребностям учащихся

D. хорошая

E. превосходная

Если задание направлено на выявление уровня удовлетворенности респондента школьной библиотекой, независимо от объективной меры ее качества, то его можно считать хорошим. Если это задание объединить с другими заданиями о количестве книг, такими как, например, сколько приблизительно книжных полок или компьютерных средств обслуживания имеется в библиотеке, то это позволит сравнить субъективное восприятие респондента с более объективными характеристиками качества библиотеки.

Предлагаемые варианты должны учитывать степень точности, с которой респонденты смогут ответить на поставленный вопрос. Вряд ли учащиеся знают точное количество книг в библиотеке, если только она не является очень маленькой.

Необходимо привести все возможные ответы на вопрос. Если существует ряд второстепенных вариантов, то после перечисления основных ответов, необходимо включить вариант «и другие». Выявить основные варианты поможет предварительная проверка заданий.

Категории ответов, между которыми необходимо сделать выбор, не должны «накладываться» друг на друга или, напротив, оставлять «пропуски». В следующем вопросе допущены обе ошибки.

Как долго Вы преподавали в этой школе?

- A. менее 5 лет
- B. менее 10 лет
- C. более 10 лет

Преподаватели с опытом работы меньше 5 лет будут сомневаться при выборе между первым и вторым вариантами. А у преподавателей с опытом работы ровно 10 лет и вовсе не будет возможности дать верный ответ.

Задания, которые начинаются с утверждения, предназначеннного для оценивания, могут иметь много различных вариантов ответа. Важно, чтобы они не пересекались. Ниже приведены некоторые варианты ответов, которые использовались в анкетах для преподавателей в исследовании TIMSS.

- Да, нет.
- Полностью согласен, согласен, не согласен, полностью не согласен.
- Почти каждый день, раз или два раза в неделю, несколько раз месяц, никогда или почти никогда.
- Совсем ничего, немного, довольно много, много.
- Не важно, довольно важно, очень важно.

УПРАВЛЕНИЕ ПРОБЛЕМАМИ ЧУВСТВИТЕЛЬНОСТИ

Некоторые вопросы анкет могут быть довольно деликатными, например вопрос о том, берут ли преподаватели дополнительную работу, чтобы повысить свои доходы. Если существует вероятность, что на какой-либо вопрос большинство респондентов вряд ли ответят честно, то его следует исключить. Тактики могут быть крайне заинтересованы в подобных сведениях, но от сбора ненадежных данных будет мало толку. Иногда взамен может запрашиваться смежная, не столь деликатная, информация, которая не будет столь чувствительна к контексту вопросов.

Респонденты, столкнувшись с подобными деликатными вопросами, могут быть до такой степени оскорблены, что вовсе

не станут отвечать на оставшиеся вопросы или откажутся сдать анкету. Поэтому, если существуют хотя бы малейшие опасения в этом, вопросы, вызывающие сомнения, предпочтительнее удалить.

МАКЕТ АНКЕТЫ

При разработке макета и дизайна анкеты необходимо исходить из двух соображений: легкость использования для респондента и легкость обработки данных. Удобные для использования анкеты имеют следующие характеристики:

- простой, последовательный способ ответов на вопросы;
- лаконичность;
- легко опознаваемые отдельные вопросы;
- варианты ответов, которые четко связаны с каждым вопросом;
- согласованность заголовков, шрифтов и иных элементов оформления;
- варианты ответов, закодированные для ввода данных.

Варианты ответов могут быть представлены разными способами – вертикально в виде столбца или горизонтально в строчку. Респондентам предлагается обвести кружком букву алфавита или номер варианта ответа или сделать специальную отметку в квадратике. В стиле ответов важна последовательность.

В дополнительном материале 9.1 представлен пример задания, в котором квадратики для ответов нечетко соотнесены с предлагаемыми ответами. В дополнительном материале 9.2 представлен лучший вариант.

ЭКСПЕРТИЗА АНКЕТ

Составить анкету гораздо труднее, чем это может показаться на первый взгляд. Все задания должны быть тщательно отобраны и рассмотрены, чтобы гарантировать четкость и однозначность формулировок. Для экспертизы анкет мы настоятельно рекомендуем привлечь экспертную группу. В нее должны вой-

Дополнительный материал 9.1**Пример неграмотного расположения квадратиков для ответов**

Ниже приведен пример неграмотного расположения квадратиков и вариантов ответов.

Сколько времени у Вас, как правило, занимает дорога до школы?

- | | | |
|---|-----------------------------------|---------------------------------------|
| менее 15 минут <input type="checkbox"/> | 15 минут <input type="checkbox"/> | 30 минут <input type="checkbox"/> |
| 45 минут <input type="checkbox"/> | 1 час <input type="checkbox"/> | более 1 часа <input type="checkbox"/> |

Квадратики помещены между вариантами, вместо того чтобы четко располагаться только около одного из них.

Источник: разработки авторов.

Дополнительный материал 9.2**Пример грамотного расположения квадратиков для ответов**

Ниже приведен пример грамотного расположения квадратиков и вариантов ответов.

Как часто члены Вашей семьи помогают Вам делать домашнюю работу?

Член семьи	Никогда или почти никогда	Несколько раз в год	Раз в месяц	Несколько раз в месяц	Несколько раз в неделю
Мать	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Отец	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Братья и сестры	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Бабушка или дедушка	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Источник: разработки авторов.

ти авторы заданий, а также специалисты, хорошо знающие генеральную совокупность респондентов и способные подтвердить соответствие заданий культурному фону страны. Иногда в процессе национальной и международной оценки непосредственно перед анкетированием проверяют, как интерпретируют задания сами респонденты.

Будет полезно, если члены экспертной группы сами попытаются заполнить анкету так, как если бы они были респондентами. Этот процесс поможет выявить те варианты ответов, которые могут показаться респондентам неясными, частично подменять друг друга либо вовсе не должны быть представлены в качестве вариантов ответов.

Члены экспертной группы должны критически пересмотреть задания, чтобы добиться ясности формулировок и пригодности вариантов ответов. Они должны гарантировать, что использованные формулировки настолько просты, насколько это возможно, что задания написаны в одном стиле и представлены в логическом порядке с соответствующими инструкциями.

Членам экспертной группы также следует проверить, что задания соответствуют проекту анкеты и гарантировать, что каждое из них обеспечивает выявление необходимой информации. Они также должны удостовериться, что число используемых заданий является достаточным для измерения каждой переменной с приемлемой точностью.

После того как анкеты прошли процесс чистки, нужно, чтобы их просмотрели и одобрили тактики. В особой степени это касается заданий, связанных с политическими проблемами. Также тактики должны подтвердить, что задания обеспечивают полезную информацию.

После того как анкеты были скорректированы, наряду с тестовыми материалами они должны быть апробированы в «полевых условиях». Предварительная проверка позволяет повысить качество заданий, а также сократить время и стоимость обработки данных окончательных анкет. Вопросы, которые не «работают» (в частности те, что вводят респондентов в заблуждение) могут быть удалены, а варианты ответов – расширены или оставлены в прежнем виде.

После предъявления анкеты администратор должен получить необходимую информацию от респондентов (учащихся или преподавателей) о заданиях, которые являются неясными или не содержат нужные сведения (механизм обратной свя-

зи). Администраторы должны убедиться, что ни одно из заданий не воспринимается респондентами как раздражающее, потому что оно затрагивает деликатные вопросы.

Более формальный статистический анализ ответов может указать на те варианты, которые нуждаются в более четкой дифференциации. Например, в том случае, если большинство учащихся выбирает какой-либо один вариант, для получения более точной информации его следует разбить на несколько.

Если диапазон возможных ответов на задание является очень большим и трудным, то в ходе предварительной проверки анкеты задание нужно предъявить в открытой форме. После этого ответы респондентов могут быть классифицированы и использованы для разработки вариантов для заданий с множественным выбором в окончательной анкете.

ГЛАВА



КОДИРОВАНИЕ ОТВЕТОВ НА АНКЕТЫ

Для последующего ввода данных варианты ответов должны быть закодированы. Кодирование бывает буквенным или цифровым.

Буквенные коды подразумевают, что респонденты должны обвести букву, которой обозначен выбранный ими вариант. Для учащихся младших классов этот метод является не совсем подходящим. Люди с ограниченной грамотностью могут отвечать, заштриховывая соответствующий кружок или помечая нужный квадратик «галкой». В последнем случае предпочтительнее цифровое кодирование.

Если используется цифровое кодирование, то первому варианту ответа обычно присваивается цифра 1, второму – 2 и т.д. Ввод данных более эффективен, если коды напечатаны на анкете. Может использоваться шрифт без каких-либо выделений, как показано в дополнительном материале 10.1. В приведенном примере номера вариантов ответов обозначены под квадратиками: пешком – ответ 1, на общественном транспорте – 2 и т.д. Учащийся ставит метку в выбранном ими квадратике. Специалист по вводу данных вводит номер квадрата, выбранного учащимся.

Дополнительный материал 10.1

Пример кодирования ответов без выделения

Сегодня я добирался до школы

- пешком на общественном транспорте на личном транспорте
1 2 3
на каком-либо животном
4

Источник: разработки авторов.

Дополнительный материал 10.2

Пример рассмотрения каждого варианта ответа в отдельности при вводе данных

Если Вы отсутствовали на прошлой неделе, укажите одну или несколько причин.

- Я болел.
1
 Я должен был помочь родителям.
1
 Была плохая погода.
1
 У меня не было еды.
1
 У меня были проблемы в семье.
1
 У меня не было чистой или подходящей по погоде одежды.
1
 Другое _____.

Источник: разработки авторов.

Если на какое-либо задание можно дать несколько ответов, то при вводе данных и их обработке каждый из них нужно рассматривать в отдельности. Это позволяет проследить, из каких вариантов каждый респондент выбирал подходящие для себя. В дополнительном материале 10.2 приведен вопрос и ряд воз-

можных ответов. При вводе данных он рассматривается как восемь отдельных вопросов. Первый вариант («Я болел») кодируется или цифрой 1, или как пропущенный (если он не выбран), подобным образом кодируются второй («Я должен был помочь родителям»), третий («Была плохая погода») и так далее для каждого из восьми вариантов.

ПОДГОТОВКА АНКЕТ ДЛЯ ВВОДА ДАННЫХ

Данные анкет могут быть обработаны с помощью специального оборудования или введены вручную. Если будет использоваться сканер, макет анкеты необходимо разрабатывать исходя из задачи сканирования.

Если ввод данных осуществляется вручную, в том случае, когда варианты ответов были закодированы прямо в анкете, информация может быть введена непосредственно из нее. Специалисты по вводу данных могут столкнуться с определенными трудностями, поскольку им необходимо постоянно поддерживать высокий уровень точности ввода, особенно если они незнакомы с этим видом работы. На точность ввода также может негативно повлиять частое изменение расположения заданий или большое количество возможных вариантов ответов.

Ввод данных можно облегчить, если эксперты обозначат код для выбранного варианта слева от номера задания. Тогда ввод данных просто сводится к занесению в компьютер соответствующих кодов. Если для кодов добавить специальные чуть заштрихованные квадратики, то процесс ввода данных станет еще более эффективным.

КОДИРОВАНИЕ ПРОПУЩЕННЫХ ИЛИ НЕОДНОЗНАЧНЫХ ОТВЕТОВ

Иногда респонденты не отвечают на задания или отвечают на них двусмысленно, например выбирают больше одного варианта ответа в тех случаях, когда предложенные ответы взаимно исключают друг друга.

Информация о пропущенных ответах позволяет определить, были ли респонденты единодушны в пропуске определенных заданий, поскольку не могли ответить на них, и выявить определенные тенденции, если они имеются. Например, если никто не ответил на задания в конце анкеты, то это указывает на ее чрезмерную длину. Или если в нескольких анкетах пропущен ответ на определенный вопрос, то это может означать, что он расположен слишком близко к другим заданиям и поэтому его легко пропустить. Сведения о неоднозначных ответах позволяют судить о ясности и точности формулировки.

Специалист по вводу данных должен знать, как закодировать пропущенные или неоднозначные ответы. Коды, используемые для пропущенных или неоднозначных ответов, должны отличаться от кодов, используемых для категорий ответов, чтобы не возникло путаницы.

Для обозначения пропущенных заданий или заданий, которые никто не пытался выполнить, можно использовать букву алфавита, например *X*. Неоднозначные ответы можно обозначить другой буквой, например *Y*. Для пропущенных заданий с множественным выбором в тестах используют код 9 и код 8 – для заданий, в которых учащиеся выбрали два или больше вариантов ответа. Эти коды обычно не используются для анкет, потому что некоторые вопросы в них вполне вероятно будут иметь 8 или 9 вариантов ответов.

ГЛАВА



СООТВЕТСТВИЕ ДАННЫХ АНКЕТ И ТЕСТОВ

При установлении соответствия между результатами тестов и данными анкет необходимо руководствоваться потребностями анализа и сообщения испытуемым результатов. Все связующие элементы перед сбором данных должны быть ясно и однозначно установлены. Любую ошибку в соответствии, обнаруженную после сбора данных, выявить, скорее всего, будет невозможно или очень трудно. Такие ошибки могут привести к отказу от некоторых запланированных исследований.

АНКЕТЫ ДЛЯ УЧАЩИХСЯ

Самый легкий способ установления соответствия между данными анкетирования и тестирования – напечатать тесты и анкеты в одном буклете. Учащийся записывает в нем свое имя, а администратор по тестированию гарантирует, что испытуемый работает в его или ее собственном буклете на каждой сессии тестирования.

Если тесты и анкеты являются отдельными документами, то один из методов установления соответствия данных – напечатать или маркировать и тесты и анкеты именами учащихся.

Имена берутся из школьной ведомости и должны быть идентичными. Ясность и четкость написания имен должен обеспечить тот же администратор по тестированию.

Если отдельные буклеты и анкеты не могут быть предварительно помечены именами, то анкеты учащихся должны включать достаточно информации для идентификации соответствия, чтобы обеспечить согласованность данных тестирования и анкетирования. Желательно поместить на обложке идентификационный номер (ID), удостоверяющий личность учащегося, и гарантировать, что учащиеся используют один и тот же номер на тестовом буклете и анкете. За этим должен следить администратор по тестированию. Для специалиста по анализу данных также необходим список имен учащихся и их идентификационных номеров, потому что имена могут быть «запасным вариантом» для установления соответствия в тех случаях, когда встречаются ошибки в номерах ID.

Использование только одних имен учащихся для установления соответствия форм не желательно. Если на каждой паре теста и анкеты имена абсолютно идентичны и вводятся тождественно специалистом по обработке данных без проверки ошибок, то компьютер не сможет установить соответствие. Соответствие придется устанавливать вручную, что отнимет много времени и будет дорого по осуществлению. Некоторые учащиеся дополнительно усложняют установление соответствия имен за счет использования различных их вариантов (например, сокращенные формы имени или фамилии, религиозные имена) в тесте и в анкете, либо напишут неразборчиво на одной или более формах, или пропустят свое имя на одной или более формах.

АНКЕТЫ ДЛЯ РОДИТЕЛЕЙ

Данные анкет для родителей обычно соотносят с данными учащихся. Соответствие будет, скорее всего, устанавливаться с помощью имен учащихся. Здесь будут возникать те же самые проблемы, которые были отмечены для анкет учащихся. Про-

цедуры анализа данных должны проводиться таким образом, чтобы гарантировать их согласованность.

АНКЕТЫ ДЛЯ ОПРОСА ПРЕПОДАВАТЕЛЕЙ И ДИРЕКТОРОВ ШКОЛ

Данные анкет для преподавателей и директоров школ обычно соотносят только с данными по классу и школе. Если класс учащегося известен, то информация о преподавателе может использоваться для анализа данных испытуемого. Администратор по тестированию должен проверить, чтобы преподаватели и директора школ обеспечили эту информацию по анкете.

Когда анкеты возвращают из школ, данные по анкетам по каждой школе должны быть сохранены в отдельной матрице. Таким образом, даже если информация по школе не была собрана по некоторым анкетам, можно получить релевантную информацию из данных по другим анкетам в той же самой матрице.

ЧАСТЬ

3

ДИЗАЙН
РУКОВОДСТВА
ПО ПРЕДЪЯВЛЕНИЮ
ТЕСТА

ГЛАВА

12

РУКОВОДСТВО ПО ПРЕДЬЯВЛЕНИЮ ТЕСТА

Руководства по предъявлению тестов вводят в процесс национальной оценки элемент стандартизации, который обеспечивает равные условия для всех испытуемых. Главная цель руководства – определить основные условия тестирования, включая требования к подготовке и процедуре, которые помогут предотвратить расскречивание данных и будут способствовать сохранению информационной безопасности. Учащиеся должны выполнить одни и те же тренировочные задания и иметь одинаковые инструкции по внесению ответов. Всем учащимся необходимо предоставить одно и то же время для выполнения теста и в равной степени следить за тем, как они соблюдают дисциплину.

Результаты теста в национальной оценке должны отражать меру способность учащихся выполнять предложенные им задания без посторонней помощи. Испытуемые должны ясно понимать, что нужно сделать и как представить ответы, при этом никакая иная помощь им предоставляться не может. Также следует закрыть доступ к любым вспомогательным средствам, которые не являются частью теста. После ознакомления с процедурами, установленными в руководстве по предъявлению теста, необходимо гарантировать их выполнение.

СОДЕРЖАНИЕ РУКОВОДСТВА

В руководстве по предъявлению тестов должна быть представлена следующая информация.

- *Для чего создан тест?*
 - Краткое обоснование цели тестирования, описание возможных путей использования результатов.
- *Какие тесты выдаются учащимся, какие учащиеся тестируются и когда?*
 - Какие тесты предъявляются в школе?
 - Какие учащиеся должны выполнить определенный тест?
 - Даты и время предъявления тестов.
 - Порядок предъявления тестов.
 - Длительность предъявления каждого теста.
 - Длительность перерывов между предъявлениями тестов.
 - Варианты для обеспечения гибкости графика предъявления тестов.
- *Какие материалы необходимы для тестиования?*
 - Список всех материалов для тестиирования.
 - Необходимое количество материалов на одного учащегося или преподавателя.
 - Список материалов, которые должна обеспечить школа, например карандаши и ластики.
- *Какое помещение должно быть предоставлено для тестиования?*
 - Материально-технические ресурсы школы, например парты и стулья.
 - Средства, которые могут косвенно помочь учащимся при выполнении теста и которые должны быть удалены из помещения или спрятаны, например схемы, таблица умножения или плакаты с грамматическими правилами.
- *Какая предварительная подготовка должна быть проведена?*
 - Как классный руководитель или директор школы могут привлечь внимание преподавателей и учащихся, чтобы те поддержали тестиирование до его фактического начала?

- Какая информация может потребоваться администратору по предъявлению теста, например списки учащихся классов со сведениями о них?
- Как нужно рассортировать тестовые буклеты – по номенклатуре или по именам учащихся?
- По каким признакам должны быть сформированы группы учащихся для тестирования?
- *Как следует проводить тестирование?*
 - Каким образом учащиеся должны приводить свои имена на буклетах и основные сведения на обложке?
 - Когда и как администратор должен проверить, правильно ли учащиеся внесли информацию о себе на обложке и в тестовом буклете?
 - Как следует предъявлять и объяснять тренировочные вопросы?
 - Какие инструкции должны получить учащиеся о teste?
 - Какую помощь администратор может оказать испытуемым в процессе тестирования?
 - За какое время учащиеся должны выполнить тест?
 - Какие условия должен поддерживать администратор во время тестирования?
 - Кому разрешено заходить в помещение во время тестирования?
- *Как должны храниться тестовые материалы?*
 - Процедуры для гарантии информационной безопасности тестовых материалов до, во время и после тестирования.
- *К кому можно обратиться за помощью?*
 - Контакты людей, которые помогут решить возникающие проблемы или предоставить дополнительную информацию.

Дополнительная информация может быть включена в руководство, чтобы упростить распространение тестовых буклетов по школам и передачу их из школ после тестирования. Эти сведения, вероятно, будут изменяться в зависимости от того, кто проводит тестирование – школьные преподаватели или независимое агентство.

ИСПОЛЬЗОВАНИЕ РУКОВОДСТВА

Руководством должны пользоваться директор школы (или завуч) и администратор по предъявлению теста. Иногда специальные руководства составляют для руководителей или тех, кто несет полную ответственность за национальную оценку внутри отдельных школ.

Директору или завучу руководство поможет должным образом подготовить школу к проведению тестирования. Он должен как можно больше знать о предстоящем тесте, чтобы направлять работу преподавателей и учащихся и мотивировать последних выполнить тест наилучшим образом. Директор школы или школьный координатор по национальной оценке (если таковой назначен) должен обладать информацией, достаточной для того, чтобы организовать работу коллектива. Он должен быть уверенными, что в тестировании примут участие отобранные учащиеся, что оно пройдет в нужное время, что испытуемые будут обеспечены всеми необходимыми материалами и правильно рассажены, а сами тесты будут храниться в атмосфере полной секретности.

Администраторам руководство поможет узнать, что следует делать при предъявлении теста, когда и по каким правилам. Они должны убедиться, что доступно достаточное количество тестовых материалов и что были отобраны подходящие (с точки зрения представительности выборки. – *Примеч. ред.*) учащиеся. Также им следует знать, какая информация о тесте может быть предоставлена учащимся, как ответить на практические вопросы и за какое время испытуемые должны выполнить тест. Помимо этого, администраторам необходима информация о процедурах для обеспечения секретности при хранении тестовых материалов.

ОСОБЕННОСТИ РУКОВОДСТВА

Хорошее руководство является максимально полным и удобным в использовании. Информация в нем логически упорядочена, инструкции ясны и однозначны, а язык – простой и пра-

вильный. Выделение ключевых позиций, использование рамок и таблиц облегчают работу с ним. В хорошем руководстве должно быть оглавление с ясными обозначенными заголовками (см. дополнительный материал 12.1).

Дополнительный материал 12.1

Инструкции руководства по применению теста

В национальной оценке следующая информация печатается крупным шрифтом (Arial, кегль 14) и занимает всю вводную страницу руководства по предъявлению тестов.

Пожалуйста, перед тем как ваши учащиеся будут выполнять тест, прочтите эти рекомендации.

Учащиеся должны выполнять этот тест не менее ДВУХ ДНЕЙ.

- Тест разделен на четыре раздела. Учащиеся должны выполнять каждый день по два раздела.
- Перед выполнением очередного раздела учащимся следует предоставить перерыв.
- Не позволяйте учащимся проходить весь тест сразу.

Правила предъявления

- Преподаватели должны контролировать все сессии тестирования на всем их протяжении.
- Учащиеся НЕ должны выносить тестовые буклеты из класса или работать с ними после окончания тестирования.
- Учащиеся должны пользоваться карандашами с ластиками, которые были им предоставлены.
- Учащиеся не должны пользоваться никакими дополнительными материалами, которые есть в классе, такими как учебные пособия, словари или калькуляторы.
- Учащимся ни в коем случае нельзя помогать. Если испытуемый не понимает, что делать, объясните ему на примере тренировочных заданий еще раз и посоветуйте ему / ей стараться изо всех сил, но не подсказывайте.

Информационная безопасность теста

- Тестовые материалы должны ОБЯЗАТЕЛЬНО ХРАНИТЬСЯ В УСЛОВИЯХ СЕКРЕТНОСТИ.
- Тестовые буклеты учащихся НЕ должны копироваться ни для каких целей.
- Учащиеся НЕ должны брать тестовые буклеты домой.

Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

СКОЛЬКО ДЕТАЛЕЙ НЕОБХОДИМО?

Информация об условиях предъявления теста и подготовке тестовых материалов должна быть исчерпывающей, но в то же время краткой (см. дополнительный материал 12.2).

Дополнительный материал 12.2

Информация для преподавателей и руководителей

Информация о тестовых материалах должна быть предельно краткой. Ее следует приводить в виде списка, который легко проверить. Приведенный ниже фрагмент программы крупномасштабного оценивания содержит информацию для директора или завуча о том, какие материалы были отправлены в школу и как узнать, какие классы будут участвовать в тестировании.

Тестовые материалы

Старший инспектор сообщит Вам, какие классы отобраны для участия в тестировании.

Вы должны были получить следующие материалы:

- письмо с указаниями для завуча;
- тестовый буклет для каждого учащегося, принимающего участие в тестировании;
- руководство по предъявлению теста для каждого преподавателя, участвующего в тестировании;
- анкеты для сбора базовой информации о каждом преподавателе, участвующем в тестировании;
- карандаш с ластиком на конце для каждого учащегося, участвующего в тестировании.

Если какие-нибудь материалы отсутствуют или Вы не имеете их в достаточном количестве, пожалуйста, свяжитесь со старшим инспектором по начальной школе.

Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

Инструкции по выполнению теста должны быть полностью изложены и заключены в рамки. Все, что администратор должен сказать учащимся о самом тесте, тренировочных заданиях или условиях предъявления теста, обязательно необходимо напечатать. Администратор по тестированию должен зачитывать выдержки из руководства, не делая никаких изменений в фор-

мулировках. Это гарантирует, что учащиеся, выполняющие тест, точно получат одни и те же инструкции.

ТРЕНИРОВОЧНЫЕ ВОПРОСЫ

В дополнительном материале 12.3 представлены основные методические рекомендации и описана цель включения в тест тренировочных вопросов. Инструкции, которые администратор дает учащимся, приведены в выделенной области в полном объеме. Администратор должен зачитать их вслух. Изображения монет, которые являются частью тренировочных вопросов, также приведены в руководстве для администратора. Благодаря этому ему не нужно одновременно читать руководство и просматривать тестовый буклет, чтобы узнать, на чем сосредоточено внимание учащихся.

АПРОБАЦИЯ

Руководство должно быть полностью готово к аprobации в ходе предварительного тестирования, или испытания в полевых условиях. Предварительная проверка помогает выявить возможные недоразумения или двусмысленные моменты, которые необходимо устраниить в заключительной версии. Поскольку условия предварительного тестирования максимально приближены к окончательным, ко времени аprobации руководство должно быть готово в той мере, насколько это возможно.

Как правило, инструкции по предъявлению тестов могут быть написаны в любое время после того, как разработан их проект. В проекте должны быть определены требования относительно числа тестов, их длины и о том, какие учащиеся должны выполнять тест.

В течение предварительного тестирования администратор должен собрать такую информацию, как приведенная ниже, чтобы помочь менеджеру по разработке теста подготовить окончательный вариант.

Дополнительный материал 12.3

Предъявление тренировочных заданий

В следующем фрагменте приведены несколько инструкций для предъявления некоторых тренировочных вопросов.

Первый день: сессия 1

ТРЕНИРОВОЧНЫЕ ВОПРОСЫ ПО МАТЕМАТИКЕ ДЛЯ 3-ГО КЛАССА (приблизительно 10 минут)

Удостоверьтесь, что у каждого учащегося есть тестовый буклет, на обложке которого он указал свое имя. Тренировочные вопросы позволяют увидеть разные варианты представления ответов в teste.

Попросите, чтобы учащиеся открыли свои буклеты на странице с тренировочными вопросами по математике (первая страница).

СКАЖИТЕ

Сегодня мы выполним часть теста по математике и благодаря этому узнаем о ваших достижениях. Сначала мы ответим на некоторые тренировочные вопросы, чтобы Вы узнали, какие задания Вам предстоит выполнить и каким образом нужно показать свои ответы.

Откройте тестовый буклет и покажите учащимся тренировочные вопросы. Проверьте, что каждый из них правильно нашел вопросы.

СКАЖИТЕ

Давайте рассмотрим одно тренировочное задание. Я зачитаю его вслух. На рисунке изображены монеты.



Сколько всего денег изображено на рисунке?

2 тойя, 7 тойя, 25 тойя или 205 тойя?

Пожалуйста, закрасьте маленький кружок около правильного, по Вашему мнению, ответа.

Закрасить можно только один кружок.

Подождите, пока все учащиеся выполнят задания, и затем проверьте ответы.

СКАЖИТЕ

Правильный ответ – 25 тойя. Вы должны были закрасить кружок около 25 тойя. Если Вы ошиблись, сотрите ластиком неправильный вариант и закрасьте кружок для верного ответа.

Дождитесь, пока все учащиеся исправят свои ответы, если это необходимо.

Источник: Департамент образования Папуа – Новой Гвинеи, 2004.

- Все ли тренировочные задания необходимы, достаточно ли их и ясны ли приведенные объяснения?
- Можно ли считать длину теста оптимальной или в нем слишком много заданий? Сколько учащихся закончили его выполнение раньше, чем за 10 минут до конца отведенного времени? (Если в одном и том же классе используются разные формы теста, то администратор может сравнить промежутки времени, требующиеся учащимся для выполнения каждой из них.)
- Казались ли испытуемые поглощенными работой над тестом?
- Были ли у учащихся необходимые средства и в достаточном ли количестве, например карандаши или ластики?

- Были ли условия в школе подходящими для проведения тестирования?

ОБЗОР

Менеджер по разработке теста и авторы заданий несут ответственность за составление обзора и тренировочных вопросов для инструкций. Тренировочные вопросы и инструкции по предъявлению тестов должны быть предъявлены той же самой группе экспертов, которая будет анализировать задания.

Менеджер по разработке и специалист, ответственный за логистику и распространение тестов, должны рассмотреть и усовершенствовать процедуры передачи тестовых буклетов в школы и обратно.

Несколько человек с подготовкой, подобной той, которую имеет администратор по тестированию, также должны просмотреть все руководство полностью, чтобы проверить инструкции и удостовериться, что они достаточно ясны, и разъяснить возможные неточности.

Как и все тестовые материалы, руководство по тестированию должно быть вычитано опытными специалистами-корректорами. Инструкции для тренировочных вопросов могут корректироваться, только если у корректора есть копии соответствующих тестовых буклетов для учащихся.

ГЛАВА

13

АДМИНИСТРАТОР ПО ТЕСТИРОВАНИЮ

ВЫБОР АДМИНИСТРАТОРА

Участвующие в процессе национальной оценки лица должны быть твердо уверены в том, что условия предъявления теста стандартизированы и равны для всех. Поэтому администраторы должны пользоваться большим доверием и заслуженным авторитетом.

Выбор администратора по тестированию определяется условиями данной страны. В одних странах идеальными администраторами становятся школьные инспекторы, а в других это будет проблематично. Если инспекторы рассматривают выполнение возложенных на них функций администратора по тестированию как дополнительную работу, которая не входит в их обязанности и недостаточно оплачивается, то эта работа не представляет для них какого-либо интереса, и у них нет мотивации выполнить ее должным образом.

Иногда в процессе национальной оценки приходится привлекать внешних администраторов. Идеально, если это специалисты, которые умеют точно следовать инструкциям, у которых есть время и средства для решения задач и у которых нет никакой специальной заинтересованности в результатах

тестирования, кроме желания провести предъявление теста правильно. Например, в Восточном Тиморе для работы в качестве администратора были специально обучены специалисты по сбору информации. Они занимались предъявлением тестов в школах, и эта работа оплачивалась. Они понимали важность сбора данных систематическим способом и не оказывали учащимся никакой помощи. Внешний контроль их работы позволял гарантировать, что она отвечает соответствующим стандартам.

Иногда роль администраторов выполняют преподаватели, что в некоторых странах серьезно подрывает доверие к полученным результатом, потому что преподаватели могут намеренно или неумышленно оказать помощь учащимся. На это есть ряд причин. Некоторые волнуются из-за того, что по результатам теста будет оцениваться их собственная работа в качестве преподавателя. Они непременно захотят помочь учащимся, чтобы удержаться на рабочем месте или поддержать свое реноме. Еще в большей степени угрозу профессиональному статусу ощущают директора школ.

Некоторые преподаватели считают, что по результатам одного теста судить об их работе будет несправедливо. Они чувствуют, что просто обязаны помочь учащимся, чтобы сделать тестирование по-настоящему «честным». Преподаватели могут иметь искреннее намерение участвовать в тестировании в качестве инструктора, но на деле окажутся неспособными отойти от своей роли учителя. Иногда преподаватели помогают учащимся, даже не осознавая, что они делают, или потому что просто не могут видеть учащихся, попавших в затруднительное положение, и не предложить им помочь.

Подробнее о процессе отбора администраторов тестирования, их достоинствах и недостатках смотрите в третьей книге настоящей серии.

СЛЕДОВАНИЕ ИНСТРУКЦИЯМ

Администратор должен помнить, что одни указания, приводимые в руководстве, необходимо понимать дословно, а другие

являются более общими и дают возможность для интерпретации и адаптации к условиям в классе. Администратор не может игнорировать какие-либо инструкции. Выявить ошибки или неоднозначности в указаниях, если они существуют, поможет предварительная проверка руководства.

Помощь администратора учащимся сводится лишь к ясному и четкому объяснению того, что они должны сделать и в какой форме представить свои ответы. Необходимо, чтобы все учащиеся, выполняющие тест, понимали, что никакая иная помощь им не может быть оказана. Администратор не должен пояснять вопросы, интерпретировать значения слов, подсказывать возможные способы решения. Если учащийся прямо просит о помощи, администратор может только предложить ему сделать все возможное, чтобы ответить на вопрос самостоятельно.

Администратор не должен переводить задания на другой язык для учащихся, если только этого не допускают инструкции.

В некоторых тестах администратор должен зачитывать задания вслух. Читать нужно медленно, отчетливо, весь тест вслух, вопрос за вопросом, или отдельные вопросы по просьбе учащихся на том языке, на котором написан тест. Администратор может повторять отдельные вопросы по просьбе учащихся. Администратору необходимы часы – наручные или настенные – чтобы записать на доске или бумаге точное время начала и конца тестирования. Администратор должен гарантировать, что учащиеся знают о времени, за которые они должны выполнить тест. Для этого он обязательно объявляет в начале тестирования, сколько времени есть у испытуемых и напоминает, когда до конца остается 10, 5 или 2 минуты (в зависимости от длины теста).

Если администратор видит, что учащиеся тратят чересчур много времени на решение одного задания, он может посоветовать пропустить задание и попытаться выполнить остальные. Он может предложить, чтобы учащийся выбрал ответ, который кажется ему наиболее подходящим, попробовал ответить на следующие вопросы, а затем вернулся бы к заданию, вызвавшему трудности.

В помещении, где проводится тестирование, разрешено иметь только те материалы, которые предусмотрены в руководстве. Зачастую учащиеся приносят свои собственные карандаши и ластики. Пеналы и сумки перед началом тестирования необходимо сдать. Все, что может оказать учащимся косвенную помощь, должно быть удалено. Кроме того, испытуемые не должны обращаться к словарям или калькуляторам, если условиями тестирования не предусмотрено иное.

Во время тестирования в классе могут находиться только администратор, учащиеся, выполняющие тест, и, возможно, наблюдатель. Директору школы, классному руководителю или другим преподавателям заходить не разрешается. О любых нарушениях необходимо сразу уведомить ответственного за проведение тестирования.

В процессе тестирования администратор должен фиксировать любые изменения условий, касающиеся отдельных учащихся. Часто на обложке тестового буклета есть специальная область, где администратор может указать, что учащийся отсутствовал в течение определенного времени или на протяжении всего теста. Если учащийся не смог закончить тест из-за болезни и покинул класс раньше, администратор обязательно должен записать эту информацию.

ПРОВЕРКА КАЧЕСТВА

При отборе администраторов нужно руководствоваться соображениями их пригодности для решения поставленной задачи. Они должны хорошо владеть языком, на котором написано руководство для тестирования. Администраторам следует целиком посвятить себя выполнению возложенных на них обязанностей.

Администраторы нуждаются в обучении, независимо от возраста или академической квалификации. Им следует посещать занятия, на которых им объясняют цель тестирования и их функции при его проведении. Администраторы должны понимать, почему так важно следовать указаниям. Желательно предоставить им возможность предварительно по-

практиковаться вместе с коллегами, а также задать вопросы о процедурах, выделенных в руководстве по проведению тестирования.

Если в роли администраторов выступают преподаватели, которые предъявляют тесты своим учащимся, то в ходе обучения необходимо четко дать им понять, что по результатам теста не будут судить об их работе и поэтому помогать учащимся ни в коем случае не следует.

Работу администраторов необходимо контролировать, по крайней мере, на протяжении части времени отведенного на предъявление теста. Проверка работы каждого из них невозможна, но выборочная проверка работы некоторых должна быть выполнена обязательно.

Чтобы администратор был уверен в том, что полностью завершил свою работу, можно попросить его заполнить и подписать контрольный список обязанностей.

КОНТРОЛЬНЫЙ СПИСОК ОБЯЗАННОСТЕЙ АДМИНИСТРАТОРА

Контрольный список обязанностей администратора может изменяться в зависимости от того, кто проводит тестирование, и от того, какие процедуры для отслеживания буклетов и гарантий информационной безопасности были разработаны. В дополнительном материале 13.1 приведен пример контрольного списка обязанностей администратора, используемый на Филиппинах. (Еще один пример вы найдете в третьей книге серии.)

Дополнительный материал 13.1**Список обязанностей администратора:
пример из Филиппин**

Администратор должен проверить каждый пункт списка, чтобы показать, что он или она полностью выполнили все пункты, и в конце этого перечня поставить свою подпись.

Имя		Дата	
Задача	Источник	Время	Выполнено
1. Закончить распределение тестовых буклетов (по форме STBA), вписывая номера тестов в порядке их следования и внося имена учащихся в алфавитном порядке	Форма STBA	10 мин	<input type="checkbox"/>
2. Раздать анкету для преподавателей	Анкета для преподавателей	15 мин	<input type="checkbox"/>
3. Заполнить форму для обратной связи	Форма для обратной связи преподавателей	10 мин	<input type="checkbox"/>
4. Раздать тесты каждому учащемуся и отметить отсутствующих, поставив напротив имени учащегося соответствующую отметку	Форма STBA	10 мин	<input type="checkbox"/>
5. Прочитать введение из руководства	Руководство для администратора, с. 7	5 мин	<input type="checkbox"/>
6. Попросить учащихся закончить заполнение данных о себе на обложке тестового буклета	Руководство для администратора, с. 9	5 мин	<input type="checkbox"/>
7. Проверить, что каждый учащийся закончил заполнение необходимых данных о себе на обложке тестового буклета		10 мин	<input type="checkbox"/>
8. Следовать инструкциям для сессии 1	Руководство для администратора, с. 15–17	60 мин	<input type="checkbox"/>
9. Перед перерывом попросить учащихся выйти из помещения по очереди и оставить свои тесты на столах		15 мин	<input type="checkbox"/>

10. Следовать инструкциям для сессии 2	Руководство для администратора, с. 15–17	60 мин	<input type="checkbox"/>
11. Перед перерывом попросить учащихся выйти из помещения по очереди и оставить свои тесты на столах		15 мин	<input type="checkbox"/>
12. Следовать инструкциям для сессии 3	Руководство для администратора, с. 19–21	70 мин	<input type="checkbox"/>
13. Собрать все тестовые буклеты и отметить их сдачу, используя форму STBA	Форма STBA	10 мин	<input type="checkbox"/>
14. Посчитать все тесты и удостовериться в том, что каждый тест был сдан	Форма STBA	5 мин	<input type="checkbox"/>
15. Сдать класс		2 мин	<input type="checkbox"/>
16. Подписать форму STBA	Форма STBA	2 мин	<input type="checkbox"/>
17. Собрать и упаковать все тестовые материалы в коробку, включая: I. Форму STBA. II. Анкеты для преподавателей. III. Формы обратной связи с преподавателями. IV. Все выполненные тесты. V. Все неиспользованные тесты.		10 мин	<input type="checkbox"/>
18. Надежно спрятать материалы на складе		10 мин	<input type="checkbox"/>
19. Возвратить материалы вашему старшему инспектору по району (SDS) для регионального оценивания по математике, естественным наукам и английскому языку (RAMSE)	Формы распределения SDS RAMSE	Время на поездку	<input type="checkbox"/>
20. Возвратить этот завершенный список вашему инспектору по району	Административный список для RAMSE	2 мин	<input type="checkbox"/>
Подпись администратора _____			

Источник: Департамент образования Филиппин, 2004.

ГЛАВА

14

ИНФОРМИРОВАНИЕ ШКОЛ О НАЦИОНАЛЬНОЙ ОЦЕНКЕ

Чтобы учащиеся в процессе национальной оценки выполнили тест наилучшим образом, у них должен быть стимул. Обычно мотивации учащихся способствуют слова преподавателей о том, что результаты тестирования будут использоваться исключительно для совершенствования процесса обучения, а не для оценивания уровня подготовки самих испытуемых.

Учащиеся должны чувствовать значимость своего участия в тестировании, особенно те из них, чьи успехи в учебе оставляют желать лучшего. Выбор наиболее подходящего времени для сообщения учащимся даты тестирования зависит от ряда обстоятельств. Если учащиеся боятся тестирования и есть вероятность, что они пропустят школу в назначенный день, то дату лучше вовсе не объявлять. Если испытуемые, хотя и волнуются, но, скорее всего, придут в школу, то предпочтительнее предупредить их заранее, чтобы они успели настроиться.

Агентство по проведению тестов должно гарантировать, что цель тестирования была поставлена перед школами заранее, до его проведения. Информацию можно обеспечить с помощью семинаров, информационных писем или телефонных контактов. Желательно честно и ясно сообщить о том, какие

данные собираются по результатам тестирования, в каком виде о них будут сообщать, как они будут использоваться и какова будет обратная связь.

Руководителей и преподавателей школ, принимающих участие в тестировании, необходимо предупредить, какие классы были отобраны, чтобы помочь получить информацию о том, что учащиеся освоили и чего они не знают. Цель национальной оценки – помочь улучшить систему образования, а не оценить отдельные школы или классы. Следует также объявить, что результаты тестирования и ответы на анкеты будут конфиденциальны.

Необходимо гарантировать, что школы согласились участвовать в тестировании до назначения внешних администраторов. Внешние администраторы должны иметь все необходимые документы, например письменное соглашение. Это даст им уверенность в том, что руководство школы окажет им поддержку. В третьей книге серии вы найдете дальнейшие предложения по информированию школ, включая проект такого соглашения.



ЛИТЕРАТУРА ДЛЯ ДАЛЬНЕЙШЕГО ЧТЕНИЯ

Allen, N. L., J. R. Donoghue, and T. L. Schoeps. 2001. *The NAEP 1998 Technical Report*. Washington, DC: National Center for Education Statistics. Australian Council for Educational Research. No date. *Literacy and Numeracy*

National Assessment (LANNA), Sample Questions, Numeracy Year5. http://www.acer.edu.au/documents/LANNA_Y5Numeracy_Questions.pdf.
Baker, F. 2001. *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.

Beaton, A. E., and E. G. Johnson. 1989. "Overview of the Scaling Methodology used in the National Assessment." *Journal of Educational Measurement* 29: 163–75.

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl. 1956. *Taxonomy of Educational Objectives: Handbook 1—Cognitive Domain*.

London: Longmans, Green. Campbell, J. R., D. L. Kelly, I. V. S. Mullis, M. O. Martin, and M. Sainsbury.

2001. *Framework and Specifications for PIRLS Assessment 2001*. Chestnut Hill, MA: International Study Center, Boston College.

Chatterji, M. 2003. *Designing and Using Tools for Educational Assessment*. Boston: Allyn and Bacon.

Educational Research Centre. 1978. *Drumcondra Attainment Tests, Manual, Level II, Form A*. Dublin: Educational Research Centre.

- Eivers, E., G. Shiel, R. Perkins, and J. Cosgrove. 2005. *The 2004 National Assessment of English Reading*. Dublin: Educational Research Centre.
- Forster, M. 2000. *A Policy Maker's Guide to International Achievement Studies*. Camberwell, Australia: Australian Council for Educational Research.
- . 2000. *A Policy Maker's Guide to Systemwide Assessment Programs*. Camberwell, Australia: Australian Council for Educational Research.
- Greaney, V., and S. B. Neuman. 1990. "The Functions of Reading: A Cross-Cultural Perspective." *Reading Research Quarterly* 25 (3):172–95.
- Haladyna, T. M. 1999. *Developing and Validating Multiple-Choice Test Items*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Harlen, W., ed. 2008. *Student Assessment and Testing*. Vols. 1–4. London: Sage IEA (International Association for the Evaluation of Educational Achievement).
1998. *Third International Mathematics and Science Study-TIMSS Sample Items*. Chestnut Hill, MA: International Study Center, Boston College. <http://www.edinformatics.com/timss/pop1/mpop1.htm>, <http://timss.bc.edu/timss1995i/TIMSSPDF/BSItems.pdf/>, and http://www.ed.gov/inits/Math/timss4_8.html. —. 2007. *TIMSS 2003, Science Items, Released Set. Fourth Grade*.
- S011026. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center, Boston College. timss.bc.edu/PDF/T03_RELEASED_S4.pdf.
- Kirsch, I. 2001. *The International Adult Literacy Survey (IALS): Understanding What Was Measured*. Research Report RR-01-25. Princeton, NJ: Educational Testing Service.
- Kubiszyn, T., and G. Borich. 2000. *Educational Testing and Measurement*. New York: Wiley.
- Linn, R. L., and S. B. Dunbar. 1992. "Issues in the Design and Reporting of the National Assessment of Educational Progress." *Journal of Educational Measurement* 29 (2): 177–94.
- Linn, R. L., and M. D. Miller. 2004. *Measurement and Assessment in Teaching: Student Exercise Manual*. Upper Saddle River, NJ: Prentice Hall.
- Messick, S. 1987. "Large-Scale Educational Assessment as Policy Research: Aspirations and Limitations." *European Journal of Psychology and Education* 2 (2): 157–65.
- . 1989. "Validity." In *Educational Measurement*, 3rd ed., ed. R. L. Linn, 13–103. New York: American Council on Education/Macmillan.
- Mullis, I. V. S., A. M. Kennedy, M. O. Martin, and M. Sainsbury. 2006.

Assessment Framework and Specifications: Progress in International Reading Literacy Study. 2nd ed. Chestnut Hill, MA: TIMSS and PIRLS International

Study Center, Boston College. Mullis, I. V. S., M. O. Martin, E. J. Gonzalez, and S. J. Chrostowski. 2004.

TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth

Grades. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

Mullis, I. V. S., M. O. Martin, E. J. Gonzalez, K. D. Gregory, R. A. Garden, K. M. O'Connor, S. J. Chrostowski, and T. A. Smith. 2000. *TIMSS 1999*

International Mathematics Report. Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade. Chestnut

Hill, Mass.: The International Study Center, Boston College. timssandpirls.bc.edu/timss1999i/pdf/T99i_Math_2.pdf.

National Assessment Governing Board. No date. *Writing Framework and Specifications for the 1998 National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education.

New Zealand Ministry of Education. 2002. *English in the New Zealand Curriculum.* Wellington: Learning Media for the New Zealand Ministry of Education.

Nitko, A. J. 2004. *Educational Assessment of Students.* 4th ed. Upper Saddle River, NJ: Pearson, Merrill, Prentice Hall.

Papua New Guinea Department of Education. 2003. *Cultural Mathematics Elementary Syllabus.* Port Moresby: Papua New Guinea Department of Education. —. 2004.

National Curriculum Standards Monitoring Test. Port Moresby: Papua New Guinea Department of Education.

Philippine Department of Education. 2004. *Region-wide Assessment for Mathematics, Science, and English (RAMSE): Basic Education Assistance for Mindanao (BEAM).* Manila: Philippine Department of Education.

PISA (Programme for International Student Assessment). 2004. *Learning for Tomorrow's World: First Results from PISA 2000.* Paris: Organisation for Economic Co-operation and Development.

Научное издание

Национальная оценка учебных достижений

**Андерсен Пру
Морган Джордж**

**РАЗРАБОТКА ТЕСТОВ И АНКЕТ
ДЛЯ НАЦИОНАЛЬНОЙ ОЦЕНКИ
УЧЕБНЫХ ДОСТИЖЕНИЙ**

Ответственная за выпуск, литературный радактор *С.А. Есина*
Корректор *Ю.Г. Слизун*
Компьютерная верстка *И.В. Кравченко*
Оформление *И.В. Кравченко*

Подписано в печать 04.03.2011. Формат 70x100/16
Печать офсетная. Бумага офсетная. 12,25 печ. л.
Тираж 3000 экз. Заказ

Издательская группа «Логос»
111024, Москва, ул. Авиамоторная, д. 55, корп. 31, офис 305
Тел.: (495) 644-38-04, 642-59-89
Электронная почта: universitas@mail.ru
Дополнительная информация на сайте: <http://logosbook.ru>