

## 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 от 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.

Винсент Грини, Томас Келлаган

КНИГА 1

Национальная оценка учебных достижений

В современный период залогом экономического процветания стран и народов во всем мире являются вовсе не сырьевые ресурсы и не человеческий труд, а знания. Доступность образования во многом определяет успехи в развитии страны и ее конкурентоспособность на международном рынке. Все более широкое использование возможностей национальной оценки позволяет сотрудникам министерств образования описать уровни достижений в образовании по ключевым параметрам. Также становится возможным сравнить результаты обучения основных категорий учащихся, различающихся по гендерному признаку, национальности, месту проживания (город или сельская местность), школе, в которой они проходят обучение (государственная или частная). Кроме того, национальная оценка доказывает необходимость разработки специальных образовательных стандартов.

Несмотря на существенную активизацию деятельности по национальной оценке, ей по-прежнему не придается того значения, которое она объективно имеет. В связи с этим необходимо совершенствовать процесс ее проведения. Ведь даже в тех странах, которые уже осуществляют национальную оценку или участвуют в международных оценочных процедурах, информация, полученная в ходе исследований, не используется должным образом. Книга «Оценка образовательных достижений на национальном уровне» описывает основные цели и особенности национальной оценки: проектирование, проведение, анализ результатов и составление отчета. В ней также рассмотрены основные международные, региональные и национальные оценочные системы различных стран.

Издание представляет интерес для сотрудников министерств образования, а также для научных работников и преподавателей высших учебных заведений.

Оценка образовательных достижений  
на национальном уровне

Винсент Грини, Томас Келлаган

# Оценка образовательных достижений на национальном уровне

КНИГА 1



THE WORLD BANK

CICED



ISBN 978-5-98704-547-3



THE WORLD BANK

CICED

Оценка  
образовательных  
достижений  
на национальном  
уровне

Assessing National  
Achievement  
Levels in  
Education

**National Assessments of Educational Achievement**

VOLUME 1

# Assessing National Achievement Levels in Education

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



**THE WORLD BANK**

**Национальная оценка учебных достижений**

**Винсент Грини, Томас Келлаган**

**Оценка  
образовательных  
достижений  
на национальном  
уровне**

**КНИГА 1**

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



**Москва  
Логос  
2011**

УДК 378  
ББК 74.58  
Г85



*Издание подготовлено при поддержке  
Международного банка реконструкции и развития*

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

**Грини, Винсент**

Г85 Оценка образовательных достижений на национальном уровне / Винсент Грини, Томас Келлаган; пер. с англ. Т.Н. Леоновой; науч. ред. М.Б. Чельшкова. – М.: Логос, 2011. – 208 с.: ил. – (Национальная оценка учебных достижений).

ISBN 978-5-98704-547-3

Освещены особенности национальной и международной оценки учебных достижений обучаемых в сфере образования. Дано обоснование национальной оценки достижений, раскрыты ее цели, методы, основные элементы и организация. Рассмотрены решения, которые принимаются в связи с национальной оценкой достижений в образовании. Показано, кто может принимать эти решения и как они должны проводиться в жизнь. Изложены вопросы дизайна, применения, анализа, отчетности и использования национальной системы оценки. Охарактеризованы особенности международного оценивания учебных достижений. Книга содержит приложения, где показаны примеры программ оценивания достижений в разных странах, а также международные и региональные примеры. Приведены аргументы, доказывающие, что перенос центра тяжести в сфере образования с обеспечения всеобщей доступности обучения на оценку его результатов является этапным шагом вперед в деле повышения эффективности образования в экономическом и социальном развитии, особенно в развивающихся странах.

Для лиц, принимающих участие в разработке и реализации образовательной политики, руководителей и специалистов органов управления образованием и учебных заведений. Будет полезной специалистам и экспертам, принимающим участие в программах оценки качества образования. Может использоваться при подготовке и повышении квалификации педагогических кадров по тематике управления качеством образования и менеджмента в образовании. Представляет интерес для широкого круга учителей и преподавателей учебных заведений различных видов и уровней.

УДК 378  
ББК 74.58

ISBN 978-5-98704-547-3

© 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](http://www.worldbank.org)  
E-mail: [feedback@worldbank.org](mailto: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 правительств 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](http://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](mailto: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 Исаку 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

<b>ПРЕДИСЛОВИЕ</b> .....	13
<b>БЛАГОДАРНОСТИ</b> .....	15
<b>СПИСОК СОКРАЩЕНИЙ</b> .....	17
<b>ГЛАВА 1 ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ</b> .....	19
<b>ГЛАВА 2 НАЦИОНАЛЬНАЯ ОЦЕНКА ДОСТИЖЕНИЙ УЧАЩИХСЯ</b> .....	27
Каковы основные элементы системы национальной оценки? .....	33
Каковы различия между национальной оценкой и государственными экзаменами? .....	35
<b>ГЛАВА 3 ПОЧЕМУ ПРОВОДЯТ НАЦИОНАЛЬНУЮ ОЦЕНКУ?</b> ....	39
<b>ГЛАВА 4 РЕШЕНИЯ В НАЦИОНАЛЬНОЙ ОЦЕНКЕ</b> .....	47
Кто должен проводить национальную оценку .....	49
Кто будет предъявлять тесты и анкеты? .....	54
Какая категория учащихся будет оцениваться? .....	55
Будет ли оцениваться генеральная совокупность учащихся или выборка? .....	58
Что будет оцениваться? .....	61
Как будут оцениваться учебные достижения? .....	66



Как часто будет проводиться оценка? .....	71
Как должны сообщаться данные по образовательным достижениям учащихся? .....	72
Какие виды статистического анализа должны быть проведены? .....	75
Как результаты национальной оценки будут передаваться и использоваться? .....	77
Какова стоимость компонентов национальной оценки? .....	78
Итоги решений .....	82
<b>ГЛАВА 5 ПРОБЛЕМЫ ПРОЕКТИРОВАНИЯ, ПРОВЕДЕНИЯ, АНАЛИЗА, СООБЩЕНИЯ РЕЗУЛЬТАТОВ И ИСПОЛЬЗОВАНИЯ НАЦИОНАЛЬНОЙ ОЦЕНКИ .....</b>	<b>83</b>
Проектирование .....	83
Проведение оценки .....	85
Анализ данных .....	88
Написание отчета .....	90
Распространение и использование результатов .....	92
<b>ГЛАВА 6 МЕЖДУНАРОДНАЯ ОЦЕНКА ДОСТИЖЕНИЙ УЧАЩИХСЯ .....</b>	<b>93</b>
Рост активности в международной оценке .....	95
Преимущества международной оценки .....	97
Проблемы в международной оценке .....	104
<b>ГЛАВА 7 ВЫВОДЫ .....</b>	<b>113</b>
<b>ПРИЛОЖЕНИЯ</b>	
<b>А. ОБРАЗОВАТЕЛЬНЫЕ ИССЛЕДОВАНИЯ</b>	
<b>В СТРАНАХ .....</b>	<b>123</b>
А.1. Индия .....	123
А.2. Вьетнам .....	125
А.3. Уругвай .....	129
А.4. Южная Африка .....	132
А.5. Шри-Ланка .....	136
А.6. Непал .....	139
А.7. Чили .....	141
А.8. Соединенные Штаты Америки .....	144
А.9. Уганда .....	147

<b>В. МЕЖДУНАРОДНЫЕ ИССЛЕДОВАНИЯ</b> .....	153
В.1. Международное исследование тенденций в изучении математики и естественных наук ....	153
В.2. Международное исследование прогресса грамотности чтения .....	160
В.3. Программа международного оценивания учащихся .....	164
<b>С. РЕГИОНАЛЬНЫЕ ИССЛЕДОВАНИЯ</b> .....	175
С.1. Южно- и Восточноафриканский консорциум по мониторингу качества образования .....	175
С.2. Программа анализа систем образования .....	185
С.3. Латиноамериканская лаборатория по эвалюации качества образования .....	190
<b>СПИСОК ЛИТЕРАТУРЫ</b> .....	197
<b>ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ</b>	
2.1. Эфиопия: цели национальной оценки .....	31
2.2. Вьетнам: вопросы, поставленные в системе национального оценивания .....	32
2.3. Основные элементы системы национальной оценки .....	33
4.1. Состав комитета по регулированию в Сьерра-Леоне .....	48
4.2. Примеры заданий с множественным выбором .....	68
4.3. Примеры открытых заданий .....	68
6.1. Опыт участия Южной Африки в программах международной оценки .....	110
<b>РИСУНКИ</b>	
3.1. Пробелы в достижениях в США для девятилетних учащихся: достижения по чтению в программе NAEP, 1971–1999 годы .....	42
3.2. Учащиеся 4-х классов, достигшие профессионального уровня мастерства по чтению или превысившие его, NAEP, 1992–2003 годы, % .....	42
4.1. Средний процент оценок правильных ответов по содержательным областям в математике, полученных учащимися Лесото .....	73

А.9.1. Распределение тестовых оценок грамотности в 6-х классах в Уганде .....	151
В.3.1. Пример заданий PISA по математике .....	167
В.3.2. Средние оценки грамотности чтения в программе PISA и подшкалы оценок по чтению .....	170
В.3.3. Квалификационные уровни учащихся для математики в программе PISA .....	172
В.3.4. Учащиеся каждого квалификационного уровня по шкале PISA (математика), % .....	173
В.3.5. Учащиеся каждого квалификационного уровня по шкале PISA (чтение), % .....	174
С.1.1. Учащиеся 6-х классов, достигшие квалификационных уровней в SACMEQ, 1995–1998 годы, % .....	183
С.1.2. Изменения в оценках грамотности в промежуток времени между SACMEQ I и SACMEQ II .....	184
С.2.1. Учащиеся 5-го класса с низкими достижениями, PASEC, 1996–2001, % .....	188
С.3.1. Социо-экономические градиенты для 11 латиноамериканских стран, LLECE .....	195

## ТАБЛИЦЫ

2.1. Различия между национальной оценкой и государственными экзаменами .....	36
4.1. Варианты проведения национальной оценки .....	50
4.2. Преимущества и недостатки оценивания, основанного на генеральной совокупности учащихся для повышения ответственности школ .....	60
4.3. Понимание чтения в исследовании PIRLS .....	63
4.4. Процент учащихся по классам, достигших целевого уровня или уровня мастерства, Коннектикут, 2006 .....	76
4.5. Учреждения, несущие основную ответственность за принятие решений в программах национальной оценки ..	82
6.1. Сравнение TIMSS и PISA .....	98
6.2. Процент учащихся, достигших международных норм в программе TIMSS по математике, 8-й класс: страны с высокими и низкими оценками .....	108
А.2.1. Проценты и стандартные ошибки оценок учащихся для разных уровней умений по чтению .....	128

А.2.2. Связь между выбранными переменными преподавателей и достижениями учащихся по математике .....	128
А.5.1. Контекстные данные и их источники в программе национального оценивания в Шри-Ланке .....	137
А.5.2. Процент учащихся, достигших уровня мастерства по перво- му языку, по провинциям .....	138
А.7.1. Показатель награждения за заслуги для школ в Чили, 1998–1999 годы .....	143
А.9.1. Учащиеся 3-х классов в Уганде, ранжированные по квалификационным уровням по английскому языку, 2005 год, % .....	149
В.1.1. Планируемые доли содержательных разделов и когнитивных процессов в 4-х и 8-х классах в тестах по математике в программе TIMSS 2007 года, % .....	156
В.1.2. TIMSS, распределение достижений по математике, 8-й класс .....	158
В.2.1. Учащиеся 4-го класса, достигшие реперных точек в достижениях по чтению в программе PIRLS, % .....	163
С.3.1. Учащиеся, которые достигли каждого классификационного уровня по языку по видам школ и их расположению, LLECE, 1997 год, % .....	193
С.3.2. Учащиеся, которые достигли каждого классификационного уровня по математике по видам школ и их расположению, LLECE, 1997 год, % .....	193





## ПРЕДИСЛОВИЕ

Председатель Международного банка реконструкции и развития Роберт Зоеллик в ознаменование 100 дней своего пребывания в этой высокой должности выступил со специальным докладом. В нем он перечислил шесть стратегически важных задач, решение которых позволит вести работу Всемирного банка в русле устойчивой глобализации. Он подчеркнул роль Банка как «уникального и особенного института знаний и изучения общества – “мозгового треста” прикладного опыта» и отметил, что исполнение данной роли требует «постоянной сосредоточенности на результатах и оценке эффективности».

В наибольшей степени эти слова относятся к сфере образования. Взаимосвязь образования и экономического благосостояния государств очевидна, что подтверждают результаты многочисленных исследований. Ясно, что высокие показатели приема и выпуска обучаемых являются необходимыми, но недостаточными условиями для сокращения бедности и улучшения качества жизни в отдельно взятой стране.

Ключевое значение для снижения бедности и повышения экономической конкурентоспособности имеют учебные достижения, которые выражаются в форме новых знаний и познавательных навыков. Более того, высокий уровень учебных достижений поможет сохранить имеющуюся на сегодняшний день доступность образования. Иными словами, экономичес-

кий потенциал образовательной сферы может быть реализован в полной мере только при условии высокого качества предлагаемых услуг и достойной подготовки обучаемых.

Судя по имеющимся данным, качество и результаты учебной деятельности в развивающихся странах остаются очень низкими. В то же время немногие из них проводят систематический контроль учебных достижений путем национальной оценки или участия в региональных или международных оценочных процедурах. Недостаток подобной информации на уровне системы образования затрудняет анализ достижений страны в целом. Это приводит к тому, что оценить работу ответственных лиц и отследить происходившие в ней изменения становится невозможно. Кроме того, определить эффективность разработанных правительством мер, направленных на улучшение результатов в той или иной сфере образования, также оказывается проблематично.

Таким образом, основная задача Всемирного банка и его стран-клиентов заключается в том, чтобы перейти от просто доступного образования к образованию качественному. К сожалению, в развивающихся странах в этом отношении наблюдается недостаток инструментов и ресурсов.

Серия книг «Национальная оценка учебных достижений», подготовленная под редакцией Винсента Грини и Томаса Келлагана, поможет преодолеть разрыв между доступностью обучения и его результатами. Именно учебные достижения должны занять центральное место в образовательной повестке дня малообеспеченных стран. Благодаря данной серии представители правительственных кругов научатся измерять национальный образовательный уровень более действенными, жизнеспособными и систематическими способами. Как мы надеемся, это приведет к тому, что политика в сфере образования будет строиться на основе точных фактических данных. Это, в свою очередь, приведет к заметному росту качества обучения, что будет способствовать реальному прогрессу образования в развивающихся странах и их дальнейшему экономическому процветанию.

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



## БЛАГОДАРНОСТИ

Работу над данной серией книг осуществляла команда 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).

В проекте также участвовали Сильвия Акана (Уганда, Национальная экзаменационная комиссия), Пру Андерсен (Австралийский совет по исследованиям в образовании), Фернандо Картрайт (Канадский совет по образованию), Жан Дюмей (Группа статистиков, Канада), Крис Фримен (Австралийский совет по исследованиям в образовании), Хью Гоф (Группа статистиков, Канада), Сара Хауи (Университет Претории), Джордж Морган (Австралийский совет по исследованиям в образовании), Т. Скотт Мюррей (Институт статистики, ЮНЕСКО) и 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 проекта



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 учитывались дополнительные комментарии и экспертные оценки многих сотрудников Международного банка реконструкции и развития – Карлоса Рохаса, Эдуардо Велеса, Элизабет Кинг, Гарри Патриноса, Хелен Абадзи, Джи-Пэн Тан, Маргарит Кларк, Морина Льюиса, Раисы Веналайен, Регины Бендокат, Роберта Праути и Робина Хорна.

Особую благодарность мы хотим выразить Эйдану Малкину и Саре ПлUFF. Также огромную поддержку нам оказали Синтия Гаттмен, Мацеко Рамокоена, Александра Савицка, Пэм Спагноли, Беата Торстенсен, Мириам Вайсер, Питер Уиноград и Ганс Уоджемайкер. Спасибо Патрисии Аррегуи, Харше Атурупан, Луису Бенвенисте, Жан-Марку Бернару, Карли Чиверсу, Зеуду Герберкидану, Венито Каулю, Педро Равеле и Кинг Бинг Ву.

Также мы хотим поблагодарить следующие учреждения за разрешение воспроизводить их материалы: Экзаменационный Совет Лесото, Международную ассоциацию по эвалюации образовательных достижений, Национальный центр статистики образования американского Министерства образования, Организацию экономического сотрудничества и развития и Министерство образования Папуа – Новой Гвинеи.

Хилари Уолш помогла подготовить рукопись к изданию. Редактирование, работы по оформлению книги и производственный процесс были организованы Мэри Фиски и Паолой Скалабрин, Издательский офис Международного банка реконструкции и развития.

Также серьезную поддержку серии оказали Ирландский образовательный доверительный фонд, Банк Партнерских программ Нидерландов, Центр исследований в образовании Дублина 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
PIRLS	Progress in International Reading Literacy Study: Международное исследование прогресса в грамотности чтения
PISA	Program for International Student Assessment: Программа международного оценивания учащихся
TIMSS	Trends in Mathematics and Science Study: Международное исследование в математическом и естественнонаучном образовании



ГЛАВА

# 1



## ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В данной вводной книге мы рассмотрим основные особенности национальной и международной систем оценки учебных достижений, которые вышли на первый план в качестве средств определения качества образования в 1990-х и 2000-х годах.

Рост популярности национальной и международной оценки отражает две тенденции развития общества. Во-первых, он является следствием глобализации и увеличения интереса к масштабным проектам, таким как, например, «Образование для всех» (ЮНЕСКО, 2000). Во-вторых, он демонстрирует радикальное смещение в оценке качества образования от интереса к «входным» данным (оценки учащихся на входе в обучение, материально-техническая база, материалы учебного плана и кадровая подготовка преподавателей) к результатам обучения (знания и умения, которые приобрели учащиеся в процессе обучения в школе) (Kellaghan and Greaney, 2001b). Этот акцент на результатах обучения свидетельствует о важности развития человеческого капитала, о том, что знания заменяют сырье и рабочую силу в качестве экономических ресурсов и что при определении темпа экономического развития страны и ее конкурентоспособности на мировом рынке доступность чело-

веческих знаний и умений является решающей (Kellaghan and Greaney, 2001a).

Таким образом, необходима точная и достоверная информация о результатах функционирования системы образования. Это, в свою очередь, требует перехода от традиционного использования тестов учебных достижений для оценки индивидуальных достижений учащихся к их применению для получения сведений о достижениях системы образования в целом (или одной ее четко определенной составляющей).

Развитие систем национальной оценки позволило министерствам образования, осуществляющим управленческие функции, описать национальные уровни учебных достижений, особенно по основным предметам, сравнить учебные достижения групп учащихся, выделенных по разным признакам – пол, возраст, национальность, место жительства (город или деревня), школа (государственная или частная). Кроме того, благодаря национальной оценке представители министерств образования могут подтвердить или опровергнуть утверждения о том, что требования стандартов со временем ужесточаются или, наоборот, смягчаются.

Несмотря на активизацию деятельности в области национального и международного оценивания, все еще ощущается недостаток информации о значимости результатов его проведения, а также умений, необходимых для грамотной методической поддержки результатов. Даже когда страны проводят национальную оценку или участвуют в международных оценочных процедурах, полученная информация зачастую используется не в полной мере. Существует ряд причин этого: возможно, тактики привлекались к оценке опосредованно и посвятили ей не все свое время; возможно, результаты анализа данных не были представлены в форме, понятной тактикам; возможно, тактики сознательно не стремились использовать результаты оценки для разработки социальной политики в целом или образовательной политики в частности, особенно в той ее части, которая касается составления учебных планов, распределения ресурсов, практики обучения

и профессионального роста преподавателей. (Тактики – правительственные чиновники, которые формируют образовательную политику. – *Примеч. ред.* (см. 2-ю книгу данной серии, приложение А.))

Эта серия помогает решить такие проблемы. Она погружает читателей в сложную технологию, которая была разработана в процессе проведения национальной и международной оценок. В данной вводной книге описаны ключевые понятия и процедуры национальной оценки. Она предназначена, прежде всего, для тактиков и лиц, принимающих решения в системе образования. Цели и основные особенности процедур национальной оценки охарактеризованы во второй главе (см. также приложение А). Причины проведения национальной оценки рассмотрены в третьей главе, а основные вопросы, на которые необходимо ответить в процессе разработки дизайна и планирования оценивания, – в четвертой. Проблемы и самые распространенные ошибки, допускаемые в дизайне заданий, применении, анализе, сообщении и использовании результатов национальной оценки, отражены в пятой главе. В шестой главе описаны программы международной оценки достижений учащихся, приведены процедурные особенности национальной оценки (такие как формирование выборки, предъявление тестов, сбор основных данных и методы анализа – см. приложение В).

Сравнение национальной и международной систем оценки позволяет выявить достоинства и недостатки последней. С одной стороны, международная оценка обеспечивает получение данных из многих стран, благодаря чему тактики могут сравнить результаты своих учащихся с достижениями их сверстников из других стран. С другой стороны, требования к тестовому инструментарию должны быть приемлемы для всех стран, принимающих участие в международной оценке, а это затрудняет точное отражение диапазона учебных достижений учащихся в отдельных странах.

Еще одна особенность заключается в том, что многие участвующие в международной оценке страны проводят собственные исследования, которые основаны на данных, собранных

в стране. Таким образом, данные международного исследования могут использоваться и для национальной оценки. Однако на практике всегда встречаются проблемы, и данные, полученные таким способом, являются менее востребованными тактиками, чем те, которые были собраны в ходе специализированной программы национальной оценки.

Также существует региональный уровень, который занимает промежуточное положение между системами национальной оценки в отдельных странах и крупномасштабными международными исследованиями. В региональной оценке, как правило, участвуют страны из одного региона с одинаковыми социально-экономическими и культурными особенностями, выделяемыми в исследовании (см. приложение С).

Другим вариантом таких процедур является субнациональное оценивание, в ходе которого исследования проводятся в какой-то отдельной области (провинции или штате) внутри страны. Субнациональные оценки проводились в ряде больших стран (Аргентина, Бразилия и Соединенные Штаты Америки) с целью получения информации об образовании в отдельных областях страны. Эти процедуры являются относительно независимыми. Их основное отличие от национальной оценки состоит в том, что в каждой области (провинции, штате) применяются различные инструментарий и методики, поэтому прямое сравнение учебных достижений учащихся невозможно.

В заключительной главе этой книги приводятся некоторые общие выводы и условия развития и институционализации возможностей национальной оценки для оптимального использования ее результатов. В конце рассматриваются основные особенности национальной оценки в девяти странах (приложение А), сопровождаемые описаниями трех международных исследований (приложение В) и трех региональных исследований (приложение С).

Остальные книги серии освещают отдельные аспекты подготовки и проведения национальной оценки. Эти книги адресованы тем, кто принимает непосредственное участие в

конструировании тестов и анкет, сборе, анализе и описании результатов исследований с рассмотрением как ключевых умений, так и методических особенностей решения этих задач.

Вторая книга, «Разработка тестов и анкет для национальной оценки учебных достижений», состоит из трех частей, посвященных разработке тестов учебных достижений, анкет и руководств по предъявлению тестов. В первой части рассматривается дизайн тестов учебных достижений, освещается роль структуры теста и его проекта, или таблицы спецификации, в дизайне тестов. В ней раскрыт процесс написания задания и приведены примеры их различных форм, таких как задания с множественным выбором, краткими и развернутыми открытыми для завершения ответами. Также рассмотрены: экспертиза тестовых заданий, основные приемы достижения высокой содержательной валидности тестов, основные принципы проведения предварительного тестирования, отбора заданий для окончательного тестирования и разработки окончательной версии теста. В завершение вкратце описаны процессы обучения экспертов и процедуры оценивания тестовых заданий вручную.

Вторая часть посвящена конструированию анкет: их проектированию, написанию заданий, оценке и кодированию ответов, установлению связи полученных данных в анкетировании с оценками учебных достижений учащихся. В заключительной части приводятся оформление и содержание руководств для администраторов по проведению тестирования, рассматривается их роль и процедура отбора для участия в тестировании.

Третья книга, «Проведение национальной оценки учебных достижений», также разделена на три части. В первой освещены практические проблемы, которые связаны с проведением крупномасштабных программ национальной оценки. Затронуты вопросы планирования, составления бюджета, кадрового обеспечения, ранжирования средств и оборудования, установления контактов со школами и отбора администраторов тестирования, упаковки и пересылки, а также обеспечения информационной безопасности тестов. Также рассматриваются



вычислительные аспекты оценки с помощью тестов, чистки данных и подготовки отчета. Во второй части представлена пошаговая инструкция, предназначенная для команд по оцениванию, по созданию подходящей национальной выборки. Затронутые вопросы раскрываются в зависимости от того, какие учащиеся выбраны для оценки. Они включают: создание фрейма выборки, вычисление соответствующего размера выборки, формирование выборки вероятностно-пропорциональным методом и проведение многошагового отбора. Чистка данных и управление данными рассматриваются в заключительной части. Описаны процедуры по верификации и валидации данных, включая случайные, внутрифайловые и межфайловые проверки их согласованности.

Первая половина четвертой книги, «Анализ данных национальной оценки учебных достижений», посвящена генерации данных на уровне заданий с использованием как классической, так и современной теории тестов. (Под современной теорией понимается Item Response Theory, см. *Крокер Линда. Введение в классическую и современную теорию тестов / Линда Крокер, Джеймс Алгина, М.: Логос, 2010. – Примеч. ред.*)

В ней рассмотрены такие темы, как: анализ заданий предварительных и окончательных тестов, мониторинг изменения результатов тестирования с учетом фактора времени, построение тестов из ранее созданных заданий, выравнивание и определение пороговых баллов для различных уровней мастерства при выполнении тестов.

Вторая половина книги помогает проводить базовый анализ результатов национальной оценки. В ней рассматриваются меры центральной тенденции и дисперсия, различия средних значений, идентификация высоких и низких уровней достижений учащихся, корреляции, регрессионный анализ и визуальное представление данных.

Заключительная книга серии, «Сообщение и использование результатов национальной оценки учебных достижений», посвящена подготовке отчетов, которые впоследствии оказывают существенное влияние на политику в образовании.

В ней раскрывается методология проектирования, распространения и разработки коммуникационных стратегий для национальной оценки. Также описан процесс подготовки методического отчета, пресс-релизов, брифингов для тактиков и отчетов для преподавателей и других специалистов. Во второй части освещаются пути фактического использования результатов национальной оценки в странах для выработки политики, реформирования учебных планов, распределения ресурсов, обучения преподавателей, выявления и мониторинга изменений учебных достижений и других переменных с течением времени.

Если вы внимательно изучите эти книги, то приобретете основные умения, необходимые для проведения национальной оценки. Однако необходимо принять во внимание три фактора. Во-первых, нельзя расценивать представленный материал как набор простых формул или алгоритмов, которые можно применять механически. Необходимо сформировать свою точку зрения по различным аспектам национальной оценки (отбор содержания теста, построение выборки испытуемых, анализ данных). Причем суждения по этим вопросам должны совершенствоваться по мере роста практического опыта. Во-вторых, вы можете попросить совета более опытных практиков. В-третьих, вы должны быть готовы к тому, что в будущем вам придется подготовиться к адаптации, изменить свои взгляды, так как изменения в используемых технологиях неизбежны.



ГЛАВА

# 2



## НАЦИОНАЛЬНАЯ ОЦЕНКА ДОСТИЖЕНИЙ УЧАЩИХСЯ

Мы начнем главу с определения национальной оценки и перечисления вопросов, на которые она должна ответить. Затем мы приведем основные элементы национальной оценки. В завершение мы рассмотрим отличия между национальной оценкой и государственными экзаменами.

Национальная оценка предназначена для описания достижений учащихся с целью получения приблизительной оценки уровня достижений по учебной программе учащихся определенного возраста или класса. Она обеспечивает данные для аудита системы национального образования, чтобы информировать тактиков о ее ключевых аспектах. Как правило, оценка включает в себя предъявление тестов учебных достижений или выборке испытуемых или генеральной совокупности учащихся, при этом внимание уделяется определенному контингенту (например, пятиклассникам или тринадцатилетним испытуемым). Преподавателей, а иногда и родителей, директоров школ и самих учащихся, могут попросить предоставить «фоновую» информацию (другой термин – контекстная информация. – *Примеч. ред.*), которая связана с достижениями учащихся. Как правило, для этого используются анкеты. Подобная информация помо-

гает понять зависимость результатов обучения от влияния внешних факторов, таких как домашние условия, уровень подготовки преподавателей, их отношение к учебным программам и способность преподавателей представлять учебный материал.

В целом разные системы национальной оценки по всему миру имеют общие черты. Везде оцениваются знание учащимися родного языка, или грамотность, а также математические способности (способности к количественному мышлению). В некоторых системах оцениваются достижения учащихся в естественных науках, искусстве, музыке, владении вторым языком, а также в социальных науках. Практически во всех национальных системах оцениваются достижения учащихся начальных классов. Во многих системах национальная оценка проводится также и в средней школе, обычно в период обязательного обучения.

В каждой стране есть свои национальные особенности. Во-первых, национальные оценки отличаются частотой проведения. В некоторых странах оценка бывает ежегодной, хотя оцениваемые учебные программы могут каждый год меняться. В других системах оценка проводится не так часто. Во-вторых, системы отличаются структурами, ответственными за оценку. Где-то эту функцию выполняет министерство образования, где-то – национальный исследовательский центр либо консорциум образовательных органов, университет или экзаменационная комиссия. Третье отличие заключается в том, что в одних странах участие школ бывает добровольным, а в других – обязательным. В случае добровольного участия отсутствие некоторых школ в программе оценки будет приводить к смещению результатов и неточному отражению уровней учебных достижений в системе образования.

Несмотря на то что большинство индустриально развитых стран уже имеет сложившиеся системы национальной оценки, только с начала 1990-х годов проведение оценки стало возможным в других частях мира. Так, на протяжении 1990-х годов становление систем национального оценивания

наблюдалось в Латинской Америке и странах Карибского бассейна, что объяснялось необходимостью обеспечения данных для проведения реформ в образовании (Rojas and Esquivel, 1998). В ходе развития произошел сдвиг в оценке качества образования: акценты сместились с оценки «входных» данных на оценку результатов образования, что соответствовало духу Жомтьенской декларации (см. Всемирная декларация об образовании для всех, 1990).

В ст. 4 Жомтьенской декларации утверждается, что базовое образование должно быть ориентировано на «действительное получение образования и на результаты больше, чем на выборочное приобщение к материалам, продолжительное участие в организованных программах и заполнение сертификационных требований» (Всемирная декларация об образовании для всех, 1990, 5). Позже Дакарский проект (ЮНЕСКО, 2000), который был создан по истечении десятилетнего периода существования Жомтьенской декларации, снова выдвинул на первый план важность результатов образовательной деятельности. Было достигнуто соглашение по семи целям, которые необходимо воплотить в жизнь к 2015 году для совершенствования «всех аспектов качества образования... чтобы признанные и измеримые результаты образования были достигнуты всеми, особенно результаты по грамотности, способности к количественному мышлению и осуществлению существенных жизненных навыков» (ЮНЕСКО, 2000).

Эти заявления подразумевают, что в странах, принимающих участие в программе «Образование для всех», улучшение качества образования должно сопровождаться процедурами, которые предоставят информацию об обучении. В результате национальное правительство и спонсорские организации значительно увеличили поддержку мониторинга учебных достижений на основе национальной оценки. Нередко предполагалось, что национальная оценка не только обеспечит информацию об уровне образования в государстве, но и само использование подобных сведений приведет к росту достижений учащихся. Произойдет ли этот рост,

нам еще предстоит выяснить. Пока же ожидания того, что программа «Образование для всех» и регулярный мониторинг уровней учебных достижений приведут к улучшению результатов образования, не оправдываются (Postlethwaite, 2004). Это не могло произойти, потому что резкое увеличение числа учащихся, посещающих школу (которому способствовала программа «Образование для всех») не сопровождалось увеличением ресурсов (особенно квалифицированных преподавателей). Кроме того, информация, полученная на основе оценки, зачастую имела низкое качество, и даже в тех случаях, когда это было не так, она не учитывалась в процессе принятия решений в образовании с должной систематичностью.

Все системы национальной оценки предназначены для получения ответов на один или несколько следующих вопросов.

- Насколько хорошо построен процесс обучения (по отношению к общим ожиданиям, целям учебного плана и подготовке учащихся к дальнейшему обучению или к жизни)?
- Демонстрирует ли оценка сильные и слабые аспекты знаний и умений учащихся?
- Действительно ли отдельные подгруппы генеральной совокупности учащихся плохо выполнили тест? Например, существуют ли различия в результатах тестирования между: мальчиками и девочками; учащимися городской и сельской местности; учащимися, принадлежащими к различным языковым или этническим группам; учащимися из разных регионов страны?
- Какие факторы связаны с результатами обучения? До какой степени учебные достижения меняются в зависимости от характеристики образовательной среды (например, ресурсов школы, подготовки преподавателей и их компетентности и вида школы) или от домашних условий и общественных обстоятельств?
- Используются ли государственные стандарты при решении вопросов о наличии необходимых ресурсов (учебников, квалифицированных преподавателей и качества других входных данных)?

- Изменяются ли достижения учащихся с течением времени? Этот вопрос особенно интересен, если в образовании проводятся реформы. Ответ на него требует анализа оценок, которые могут предоставить сравнительные данные по различным аспектам с учетом фактора времени (Kellaghan and Greaney, 2001b, 2004).

Эти вопросы ставились в ходе проектирования и введения национальной оценки в Эфиопии (дополнительный материал 2.1.)

Особенностью подхода к национальной оценке во Вьетнаме, в дополнение к оценке учебных достижений, является повышенное внимание к ключевым входным характеристикам, таким как материально-технические условия обучения в школах, доступность учебных материалов и квалификация преподавателей (дополнительный материал 2.2).

#### Дополнительный материал 2.1

##### Эфиопия: цели национальной оценки

1. Определение уровня развития академических достижений и воспитания в начальном образовании.
2. Анализ различий достижений учащихся в зависимости от региона, пола, расположения школы и языка преподавания.
3. Выявление факторов, влияющих на достижения учащихся в начальном образовании.
4. Мониторинг позитивных изменений учебных достижений учащихся в образовании с начала первых исследований в 1999–2000 годах.
5. Формирование способности системы образования к проведению национального оценивания.
6. Создание устойчивых базовых данных для будущего.
7. Формирование рекомендаций для тактиков с целью повышения качества образования.

*Источник:* Эфиопия, Национальная организация по экзаменам, 2005.



**Дополнительный материал 2.2****Вьетнам: вопросы, поставленные в системе национального оценивания****Вопросы, связанные с входными характеристиками**

- Каковы характеристики учащихся 5-х классов?
- Какие условия обучения в классных комнатах для 5-х классов и для начальных классов?
- Каковы в целом условия в школьных зданиях?

**Вопросы, касающиеся стандартов оборудования для образования**

- Соблюдались ли государственные стандарты по отношению:
  - к размерам классов;
  - мебелировке классных комнат;
  - квалификации персонала?

**Вопросы, связанные с вложениями в обеспечение «входных» условий**

- Существует ли разница в обеспеченности ресурсами школ между провинциями и в каждой провинции:
  - по материальным вложениям на входе в обучение?
  - по человеческим ресурсам на входе в обучение?

**Вопросы, связанные с учебными достижениями**

- Каков процент учащихся, достигших различных уровней умений по чтению и математике?
- Каков был уровень преподавателей 5-х классов по чтению и математике?

**Вопросы, связанные с факторами влияния на учебные достижения**

- Каковы основные факторы, которые необходимо учитывать для анализа отличий в достижениях учащихся по чтению и математике?
- Каковы основные переменные, которые позволяют дифференцировать наиболее и наименее эффективные школы?

*Источник:* Всемирный банк, 2004.

## КАКОВЫ ОСНОВНЫЕ ЭЛЕМЕНТЫ СИСТЕМЫ НАЦИОНАЛЬНОЙ ОЦЕНКИ?

Несмотря на то что системы национальной оценки могут варьироваться в зависимости от того, как они были созданы, они имеют много общих элементов (см. дополнительный материал 2.3 и Kellaghan and Greaney, 2001b, 2004)

### Дополнительный материал 2.3

#### Основные элементы системы национальной оценки

- Министерство образования или формирует внутреннее исполнительное агентство, или выбирает независимую внешнюю организацию (например, подразделение университета или исследовательскую организацию) и обеспечивает ее финансирование.
- Министерство образования определяет политику оценки, иногда путем консультаций с основными субъектами образовательного процесса (преподаватели, специалисты по учебному плану, спонсоры и родители).
- Министерство образования или комитет по регулированию определяет генеральную совокупность учащихся, которая будет оцениваться (например, учащиеся 4-х классов).
- Министерство образования указывает предметную область, в которой будет проводиться оценка (например, чтение или математика).
- Исполнительное агентство определяет область учебных достижений и описывает ее с точки зрения содержания и когнитивных умений.
- Исполнительное агентство готовит тесты учебных достижений, сопровождающие их анкеты и руководства по предъявлению тестов, и принимает меры по обеспечению их валидности.
- Тесты и сопутствующие документы проходят предварительную проверку и последовательно рассматриваются комитетом по регулированию и другими уполномоченными органами для определения соответствия учебным планам и подтверждения того, что все задания учитывают чувствительность конкретной категории учащихся в соответствии с этническими, гендерными и культурными факторами.
- Исполнительное агентство формирует целевые выборки (или всю генеральную совокупность) школ или учащихся, организует тиражирование материалов и устанавливает связи с выбранными школами.
- Исполнительное агентство формирует обучение администраторов по тестированию (преподавателей, школьных инспекторов или выпускников университетов).

(продолжение на обороте)

- Материалы для исследования (тесты и анкеты) предъявляются в школах в определенный день и под общим руководством исполнительного агентства.
- Исполнительное агентство несет ответственность за сбор инструментария исследования, оценку, чистку данных и их подготовку для анализа.
- Исполнительное агентство устанавливает надежность инструментария и оценочных процедур.
- Исполнительное агентство проводит анализ данных.
- Исполнительное агентство готовит предварительные отчеты, которые анализируются национальным комитетом по регулированию.
- Исполнительное агентство готовит заключительные отчеты, а затем уполномоченные на то лица распространяют их.
- Министерство образования и другие релевантные ключевые структуры анализируют результаты с позиций потребности принятия политических решений, носящих адресный характер, и определения соответствующего курса в своей деятельности.

*Источник:* разработки авторов.

Как видно из материала 2.3, перед тем, как учащиеся начнут выполнять оценочные задания, необходима значительная теоретическая и практическая подготовка. Следует назначить орган, ответственный за сбор данных, понять, кому будут адресованы политические решения, разработать и апробировать тесты и анкеты. При подготовке к фактическому тестированию должны быть сформированы выборки (или вся генеральная совокупность) школ и учащихся, со школами необходимо установить связь, необходимо выбрать и обучить администраторов по тестированию. В ряде стран (например, в Индии, Вьетнаме и некоторых африканских странах) преподаватели должны выполнять те же задания, что и учащиеся (см. А.1 и А.2 в приложении А и С.1 в приложении С). После предъявления теста требуются значительные усилия и длительное время для подготовки данных к анализу, проведения самого анализа и написания отчетов.

Страны с неразвитой экономикой, помимо общих проблем, сталкиваются с дополнительными трудностями, связанными с тем, что бюджетное финансирование образования может быть

крайне низким. По данным на 2005 год (Всемирный банк, 2007), некоторые страны (например, Бангладеш, Камерун, Чад, Доминиканская Республика, Гвинея, Казахстан, Лаосская Народно-Демократическая Республика, Мавритания, Пакистан, Перу, Республика Конго, Объединенные Арабские Эмираты и Замбия) выделяют на государственное образование не более 2% от валового внутреннего дохода по сравнению с более чем 5% у большинства стран со средним или высоким уровнем доходов населения.

Потребности образовательного сектора, реализуемые на конкурентной основе, для деятельности по строительству школ, подготовке преподавателей и обеспечению учебных материалов могут снизить до минимума доступность средств для мониторинга образовательных достижений. Кроме того, многие страны с низким или средним уровнем дохода имеют слабый институциональный потенциал для проведения национальной оценки. Они могут также столкнуться с дополнительными административными и коммуникационными проблемами, вызванными недостатком хороших дорог, налаженной почтовой службы и телефонной связи. Наконец, значительные отличия достижений учащихся в некоторых странах с низкими доходами требует большой выборки (см. Национальная экзаменационная комиссия Уганды, 2006; Всемирный банк, 2004).

### **КАКОВЫ РАЗЛИЧИЯ МЕЖДУ НАЦИОНАЛЬНОЙ ОЦЕНКОЙ И ГОСУДАРСТВЕННЫМИ ЭКЗАМЕНАМИ?**

Государственные экзамены играют важную роль во многих системах образования при сертификации достижений учащихся, отборе учащихся для дальнейшего обучения и стандартизации того, что преподается в школах. Существует мнение, что государственные экзамены обеспечивают ту же информацию, что и национальная оценка, благодаря чему потребность в проведении национальной оценки в стране, имеющей систему государственных экзаменов, отпадает. Тем не менее государственные экзамены не могут обеспечить ту информацию, которую предоставляет национальная оценка.

Во-первых, так как государственные экзамены играют важную роль в отборе учащихся (для перехода на следующий уровень обучения, а иногда и для работы), они предназначаются для выявления различий между относительно высокими достижениями обучаемых и поэтому не могут обеспечить адекватный охват учебных планов. Во-вторых, содержание экзаменов, так же, как и характеристики сдающих их учащихся, изменяется из года в год, поэтому получить выводы, основанные на сравнении учебных достижений с течением времени, нельзя. В-третьих, тот факт, что если на результаты экзаменов возлагается «высокая ответственность» (т.е. результаты экзаменов имеют важные последствия для учащихся и, возможно, для их преподавателей), то преподаватели (и учащиеся) могут сосредоточить свои усилия только на тех разделах учебного плана, которые включены в экзамен, в ущерб другим важным аспектам, которые на экзамене не проверяются (например, практические навыки). Поэтому результаты экзаменов не обеспечивают корректного отображения разделов учебного плана. Хотя и существуют некоторые исключения, но, как правило, решения по поводу отдельных учащихся, их преподавателей или школ не принимаются по результатам национального оценивания.

ТАБЛИЦА 2.1

**Различия между национальной оценкой и государственными экзаменами**

Признак	Национальная оценка	Государственные экзамены
Цель	Обеспечить обратную связь для тактиков в образовании	Аттестовать и отобрать учащихся
Частота	Для отдельных предметов предлагается на регулярной основе (например, каждые четыре года)	Регулярно и более часто в тех случаях, когда образовательная система допускает повторное изучение предмета
Продолжительность	Один или два дня	Могут продолжаться несколько недель
Кто проверяется?	Обычно выборка учащихся определенного возраста или класса	Все учащиеся, желающие пройти экзамен на проверяемом уровне образования

Форма заданий	Обычно задания с множественным выбором и краткими ответами	Обычно эссе и задания с множественным выбором
Ответственность: важность для учащихся, преподавателей и всех остальных заинтересованных пользователей	Низкая важность	Высокая важность
Охват учебного плана	В целом охватывает один или два предмета	Охватывает основные разделы предметов
Воздействие на обучение	Очень небольшое прямое воздействие	Значительное воздействие: преподаватель стремится обучать тому, что ожидается на экзамене
Дополнительные льготы для учащихся	Очень редки	Частые
Узнают ли учащиеся свои результаты?	Изредка	Всегда
Собирается ли дополнительная информация об учащихся?	Часто, в анкетах для учащихся	Редко
Оценивание	Обычно включает статистически сложные методики анализа результатов	Обычно простой процесс, основанный на предварительно определенной схеме оценки
Воздействие на последующие достижения учащихся	Воздействие нежелательно	Низкие результаты или перспектива неудачи могут привести к раннему отсеву
Польза для мониторинга трендов уровней учебных достижений с учетом фактора времени	Существует, если тесты разработаны с предусмотренной возможностью мониторинга	Не существует, так как экзаменационные вопросы и испытуемые меняются каждый год

В-четвертых, информация в национальном оценивании, как правило, собирается на более ранних этапах обучения по сравнению с периодом проведения государственных экзаменов. В-пятых, определенная «внешняя» информация (о процессе, ресурсах, учащихся и их семьях), которая используется при интерпретации данных об учебных достижениях, собранных по итогам проведения национальной оценки, не доступна для интерпретации результатов государственных экзаменов (Kellaghan, 2006).





## ПОЧЕМУ ПРОВОДЯТ НАЦИОНАЛЬНУЮ ОЦЕНКУ?

Решение о необходимости проведения национальной оценки может быть принято по целому ряду причин. Зачастую национальная оценка связана с деятельностью правительства по модернизации системы образования путем внедрения в нее методов бизнес-менеджмента (Kellagan, 2003). Данный подход опирается на понятия, используемые в мире бизнеса, такие как стратегическое планирование и усиление внимания к конечным продуктам и результатам, и требует введения системы отчетности по результатам деятельности. С этой точки зрения национальная оценка является инструментом для обеспечения обратной связи по определенному числу измеряемых результатов, которые тактики, политики и широкая педагогическая общественность признают важными.

Основная цель данного подхода – обеспечение информации по функционированию системы образования. Многие правительства испытывают недостаток в основной информации по различным аспектам системы образования, в особенности по уровню достижений учащихся и характеристикам образовательных систем. Национальная оценка может предоставить такую информацию, которая является ключевой предпосылкой для



объявляемых политических решений. Например, национальная оценка во Вьетнаме помогла установить, что многие классные комнаты были плохо оборудованы (Всемирный банк, 2004). Точно так же оценка в Занзибаре выявила, что 45% учащихся не хватало рабочих мест в классах (Nassor and Mohammed, 1998). Национальная оценка в Бутане установила, что некоторым учащимся приходилось тратить по несколько часов в день на дорогу в школу и из школы (Бутан, экзаменационный совет, Министерство образования, 2004). Оценка в Намибии показала, что многие преподаватели недостаточно владеют базовыми умениями в английском языке и математике (Makuwa, 2005).

Необходимость получения информации о том, что учащиеся изучают в школе, резко возросла по мере развития так называемой экономики знаний. Некоторые аналитики утверждают, что учащимся потребуется более высокий уровень знаний и умений, особенно в области математики и естественных наук, чем ранее, если они собираются занять достойное место на будущей работе. Кроме того, поскольку свобода доступа к товарам и услугам по мере глобализации увеличивается, способность стран успешно конкурировать друг с другом в значительной степени зависит от уровня квалификации работников и управления использованием капитала и технологий. Данный фактор указывает на необходимость сравнения учебных достижений учащихся в одной образовательной системе с результатами обучаемых в других системах. При этом существует риск придать слишком большую важность совокупным достижениям учащихся, которые учитываются в показателях экономического роста, предполагая, что принимаются во внимание также и многие другие факторы (Kellaghan and Greaney, 2001a).

Когда национальная оценка проводится в течение определенного периода времени, она может определить, изменились ли стандарты результатов обучения или они остались прежними. Многие развивающиеся страны сталкиваются с проблемой роста расходов, увеличением количества новых школ и масштабов подготовки учителей, одновременно стараясь повысить качество образования, нередко в противовес возможностям снижающихся бюджетных расходов. В этой ситуации прави-

тельства должны вести мониторинг уровня учебных достижений, чтобы определить, как изменения в приеме учащихся и возможностях бюджета влияют на качество обучения. В противном случае существует риск, что рост норм набора учащихся может быть воспринят как доказательство повышения качества образования.

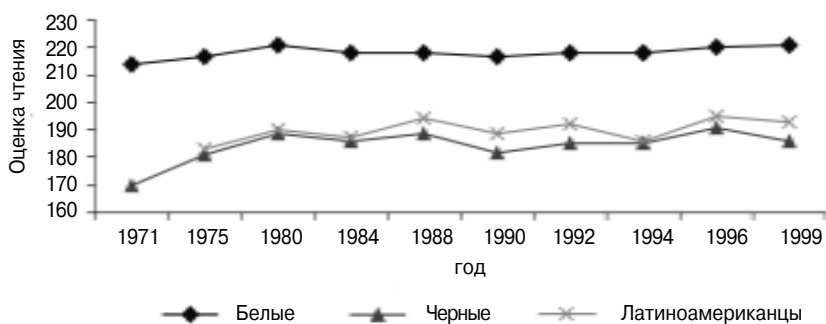
Данные национальной оценки использовались в мониторинге достижений в течение долгого времени. Ряд исследований в Африке между 1995–1996 и 2000–2001 годами показал значительное снижение оценок по грамотности чтения в Малави, Намибии и Замбии (см. рис. С.1.2 в приложении С). Национальная оценка прогресса в образовании (NAEP) в Соединенных Штатах Америки, осуществляющая мониторинг уровней учебных достижений в чтении на протяжении почти трех десятилетий, обнаружила, что, хотя разрыв между учебными достижениями девятилетних афроамериканских и латиноамериканских детей и детей белой расы постоянно сокращался приблизительно до 1980 года, дифференциация оценок по тестам сохранялась (рис. 3.1). Кроме того, в США программа NAEP помогла выявить вариации уровней чтения в различных штатах (рис. 3.2). В Непале результаты национальной оценки использовались для мониторинга: изменений учебных достижений за период с 1997 по 2001 год; эффективности стратегических решений, касающихся бюджета, учебных планов, учебников, обучающих материалов и повышения квалификации преподавателей (см. А.6 в приложении А).

Когда данные национальной оценки используются для мониторинга учебных достижений длительное время и каждый раз применяется один и тот же тест или различные тесты, часть заданий должна быть общей, таким образом результаты по тестам могут быть выровнены или связаны. В любом случае такие общие задания должны быть защищены от рассекречивания, чтобы знание содержания этих заданий учащимися или преподавателями не привело к снижению валидности результатов сравнений.

Другой аспект использования результатов национальной оценки зависит от того, были ли данные собраны по определен-

Рис. 3.1

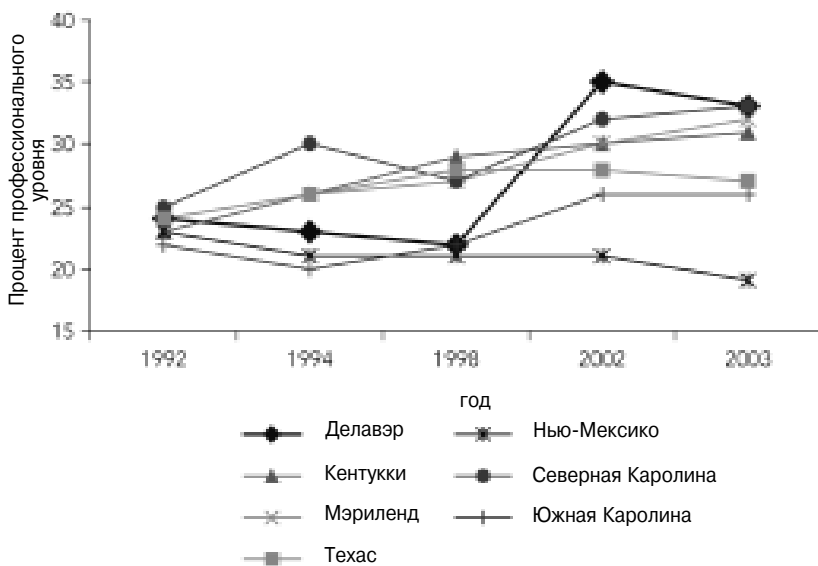
Пробелы в достижениях в США для девятилетних учащихся:  
достижения по чтению в программе NAEP, 1971–1999 годы



Источник: Winograd and Thorstensen, 2004.

Рис. 3.2

Учащиеся 4-х классов, достигшие профессионального уровня  
мастерства по чтению или превысившие его, NAEP, 1992–2003 годы, %



Источник: Winograd and Thorstensen, 2004.

ной выборке школ или по всем школам (или большинству). В обоих случаях результаты могут быть использованы тактиками, заинтересованными в повышении качества образования, для выбора направлений развития образовательной системы. Например, результаты оценки могут помочь правительству в установлении меры связи между качеством обучения учащихся и различными факторами, над которыми правительственные чиновники имеют некоторый контроль (например, качество учебников, нормы наполняемости классов и квалификации преподавателей).

Анализ результатов может привести к решениям, затрагивающим предоставление ресурсов в системе образования в целом (например, для совершенствования учебных планов и учебников или для повышения квалификации преподавателей) или в отдельных категориях школ с определенными характеристиками (например, школ в сельских районах или школ в районах с неблагоприятными социально-экономическими условиями). Можно привести много примеров использования для этих целей результатов национальной и международной систем оценивания. Они применялись в Австралии для информационного обеспечения программ, направленных на рост познавательной активности и результатов девочек при изучении математики и естественных наук (Keeves, 1995). Подобные результаты послужили побудительным мотивом реформ учебных планов в странах с низким и средним уровнем дохода (Elley, 2005), помогли привлечь финансовые ресурсы для наиболее бедных школ в Чили (см. А7 в приложении А) и способствовали росту профессионализма преподавателей в Уругвае (см. А.3 в приложении А).

Результаты национальной оценки также могут быть использованы для совершенствования текущего учебного процесса в классе (Horn, Wolff and Velez, 1992). Получение информации для преподавателей и совершенствование их мастерства для существенного роста достижений учащихся является весьма легкой задачей. В тех случаях, когда требуется максимальное влияние на совершенствование учебного процесса в школах

и классах или когда результаты национальной оценки используются для ответственных административных решений, она должна проводиться не на выборке, а на генеральной совокупности учащихся.

Нельзя предпринимать никаких конкретных действий или применять какие-либо санкции (наказания или поощрения), если они не основаны на опубликованной информации о результатах (например, на ранговых таблицах). Поощрения могут носить форму вознаграждений за улучшение учебных результатов (например, школы, преподаватели или и те и другие получают материальные поощрения, если учащимся удается достичь определенной цели). Чтобы наказать учащегося за плохую работу, можно не перевести его в следующий класс или уволить преподавателей (см. А.7 в приложении А для краткого описания программы наград в образовании Чили).

Когда в национальной оценке получают информацию о достижениях учащихся во всех школах (или их большинстве), некоторые тактики используют эти данные для анализа качества работы школ и преподавателей. Очевидно, что преподаватели и учащиеся должны нести определенную ответственность за обучение, но роль институтов, учреждений и отдельных лиц, которые осуществляют контроль ресурсов и деятельности школ, должна быть отражена в системе отчетности. Очень важное значение имеет правильное распределение обязанностей всех заинтересованных сторон, будь то оценивание выборки или генеральной совокупности учащихся. Национальная оценка в Уругвае является хорошим примером подтверждения ответственности различных ключевых структур (включая государство) за достижения учащихся (см. А.3 в приложении А).

В некоторых случаях национальная оценка может играть символическую роль, которая предназначена для оправдания государственной деятельности на основе применения моделей, принятых в международном сообществе, и придания процессу принятия политических решений вида научной обоснованности (Benveniste, 2000, 2002; Kellaghan, 2003). Если ставится такая цель, то сам процесс оценки приобретает большее значе-

ние, чем его результаты. Если национальная оценка проводится для обеспечения соответствия требованиям спонсорского агентства или выполнения международных обязательств правительства страны по программе «Цели развития тысячелетия», то оценка также будет носить символический характер, и ее данные нельзя будет серьезно применять в управлении системой образования или при разработке тактики.



ГЛАВА

# 4



## РЕШЕНИЯ В НАЦИОНАЛЬНОЙ ОЦЕНКЕ

В этой главе мы рассмотрим 12 решений, которые приходится принимать при планировании национального оценивания (см. Greaney and Kellaghan, 1996; Kellaghan, 1997; Kellaghan and Greaney, 2001b, 2004).

### **КТО ДОЛЖЕН ДАВАТЬ СТРАТЕГИЧЕСКИЕ УКАЗАНИЯ ДЛЯ НАЦИОНАЛЬНОЙ ОЦЕНКИ?**

Министерство образования назначает национальный комитет по регулированию, чтобы обеспечить общее руководство агентством, осуществляющим оценку. Комитет может помочь выполнению гарантий того, что национальная оценка приобретает определенный статус и что будут учтены ключевые политические вопросы, представляющие интерес для министерства образования и других правительственных органов. Он решает серьезные административные и финансовые проблемы, которые могут возникнуть во время проведения национальной оценки. Предоставление национальному комитету права выбора направления и цели национального оценивания также увеличивает вероятность того, что его результаты будут играть должную роль в принятии будущих политических решений.



Состав комитета будет меняться от страны к стране в зависимости от руководящей структуры в образовательной системе. Помимо представителей министерства образования, в комитет могут войти представители основных этнических, религиозных и языковых групп, а также лица, деятельность которых влияет на результаты обучения (педагоги, обучающие преподавателей, сами преподаватели, школьные инспекторы, методисты). В дополнительном материале 4.1 перечислен предполагаемый состав комитета в Сьерра-Леоне, предложенный участниками международного семинара. Обращение к информационным потребностям различных заинтересованных лиц должно гарантировать, что национальная оценка не приведет к критике и не будет проигнорирована из-за того, что «не обращает внимание» на «правильные» вопросы.

Деятельность комитета не должна быть перегружена встречами или выполнением второстепенных рутинных задач, связанных с национальной оценкой. В некоторых случаях комитет может обеспечить направление работ на начальной стадии путем постановки цели на основе выделения разделов учебного плана и классов, которые будут оцениваться, или с помощью назначения агентства или агентств для проведения оценки,

#### Дополнительный материал 4.1

##### Состав комитета по регулированию в Сьерра-Леоне:

- основная образовательная комиссия;
- общественные движения;
- децентрализованный секретариат;
- генеральный директор по образованию (председатель);
- Департамент образовательной политики;
- Межрелигиозный совет;
- Национальный исследовательский центр по разработке учебных планов;
- Союз учителей Сьерра-Леоне;
- Управление статистики образования Сьерра-Леоне;
- педагогические колледжи;
- Западноафриканский экзаменационный совет.

хотя эти задачи могут быть выполнены еще до назначения комитета. Представляется, что комитет должен быть наиболее активен на начальной стадии оценочного процесса, тогда как агентство выполняет основную часть конкретных работ, таких как разработка инструментария, формирование выборок, анализ результатов и подготовка отчетов. Однако исполнительное агентство должно предоставлять комитету копии проектов тестов и анкет с описанием предложенных процедур, чтобы члены комитета могли обеспечить руководство и гарантировать, что первоочередные информационные потребности, вызванные оценкой, нашли адекватного получателя. Члены комитета также должны проанализировать проекты отчетов, подготовленные агентством.

*Ответственность за обеспечение политического руководства лежит на министерстве образования.*

#### **КТО ДОЛЖЕН ПРОВОДИТЬ НАЦИОНАЛЬНУЮ ОЦЕНКУ?**

Национальную оценку должна проводить команда или организация, члены которой пользуются уважением, что увеличивает вероятность одобрения результатов оценивания со стороны широкой общественности. В разных странах ответственность за национальную оценку брали на себя самые разные организации, начиная с команд, что входят в состав Министерства образования, до автономных учреждений (университетов, исследовательских центров) или международных методических команд. Мы предполагаем, что на решение относительно ответственной группы влияет множество факторов, таких как уровень национальных возможностей в методическом обеспечении работ, административные и политические особенности страны. В табл. 4.1 перечислены некоторые потенциальные преимущества и недостатки различных агентств, которые необходимо учитывать, принимая решение о том, кто должен проводить национальную оценку.

В некоторых случаях ограничения на деятельность Министерства образования при выборе исполнительного агентства

**ТАБЛИЦА 4.1**

**Варианты проведения национальной оценки**

Состав агентства	Преимущества	Недостатки
Сотрудники министерства образования	<p>Вероятно, будет пользоваться поддержкой министерства.</p> <p>Гарантирует наличие свободного доступа к ведущим специалистам, материалам и данным (например, данные по выборке школьников).</p> <p>Скорее всего, не нужно будет выделять дополнительные средства на содержание штата сотрудников</p>	<p>В интересах министерства результаты могут сознательно искажаться или даже скрываться.</p> <p>Результаты могут быть скептически приняты другими заинтересованными лицами.</p> <p>От штата сотрудников могут потребовать выполнения многих других задач.</p> <p>Методическая подготовка может быть недостаточной</p>
Сотрудники из числа специалистов по проведению государственных экзаменов	<p>Пользуются доверием.</p> <p>Имеют опыт в обеспечении информационной безопасности оценочных процедур и результатов.</p> <p>Финансовые средства не нужно будет сохранять в течение работы.</p> <p>Некоторые навыки (например, по разработке тестов) могут быть использованы для повышения качества оценочных процедур.</p> <p>Вариант привлечения сотрудников более устойчивый по сравнению с другими моделями</p>	<p>От сотрудников могут потребовать выполнения многих посторонних задач.</p> <p>Методическая подготовка может быть недостаточной.</p> <p>Может быть ограничен свободный доступ к данным.</p> <p>Опыт участия в государственных экзаменах может привести к выбору слишком трудных тестовых заданий</p>
Сотрудники исследовательского / университетского сектора	<p>Результаты могут быть в большей степени научно обоснованными.</p> <p>Больше вероятность некоторой методической компетентности.</p> <p>Данные могут быть использованы для дальнейших исследований в системе образования.</p>	<p>Появляется необходимость в увеличении финансирования для оплаты труда сотрудников.</p> <p>Модель привлечения сотрудников может быть менее устойчивой, чем другие модели.</p> <p>Результаты работы могут вступать в конфликт с интересами министерства образования</p>

Сотрудники, принятые на работу в качестве зарубежной методической помощи	Скорее всего, методически компетентны. Имеются источники финансирования, которые могут помочь гарантировать своевременное завершение работ	Скорее всего, является более дорогостоящим. Могут быть не чувствительными к образовательному контексту данной страны. Трудно гарантировать устойчивость оценки. Возможность небольшого повышения требований к достижениям на уровне национальной оценки
Национальная команда, обладающая международной методической поддержкой	Может повысить методический потенциал национальных команд. Гарантирует своевременное завершение работ. Можно повысить доверие к результатам	Возможно, будет трудно координировать работу членов национальной команды и осуществлять методическую поддержку. Могут возникнуть трудности при передаче национальной команде международного опыта
Министерская команда, имеющая национальную методическую поддержку	Гарантирует поддержку министерства во время получения национальной методической поддержки. Меньше расходов, чем при международной методической поддержке	Национальной методической поддержке недостает необходимого методического обеспечения. Могут проявиться другие потенциальные недостатки, которые перечислены министерством образования

Источник: разработки авторов.

могут накладывать существующие в стране традиции и действующее законодательство. Так, в Аргентине провинции должны утвердить содержание учебных планов для их использования при эвалюации в национальной оценке. Поэтому первоначально к провинциям обращаются с просьбой об участии в разработке тестовых заданий. Однако не везде есть методические возможности для такой работы. В дальнейшем провинциям предоставляют примеры наборов заданий для одобрения. Национальное информационное агентство по эвалюации Министерства образования (DiNIECE) формирует окончательный инструментарий для оценки из совокупности предварительно апробированных тестовых заданий. Позже тестовые задания разрабатываются независимо специалистами университетов и одобряются национальным Федеральным советом. DiNIECE несет ответственность за проекты тестов учебных достижений, анализ результатов и общую координацию ежегодной деятельности по национальной оценке.

Для того чтобы решить, кто отвечает за те или иные задачи, необходимо обсудить широкий спектр умений, требующихся для проведения национальной оценки. Данная проблема более подробно рассмотрена в третьей книге серии. Национальная оценка в основе своей является командной работой. Члены команды должны проявлять гибкость, способность работать в условиях соблюдения дисциплины и сотрудничества, демонстрировать готовность к изучению новых методических и оценочных подходов. Руководитель должен обладать хорошими навыками управления. От него или нее будет требоваться организация сотрудников, координация деятельности, поддержка подготовки сотрудников и мониторинг финансов. Руководитель команды должен обладать определенной проницательностью в образовательной политике, потому что ему необходимо представлять отчеты комитету и осуществлять связь с национальными, региональными, а в некоторых случаях и с районными органами власти и представителями заинтересованных сторон (например с преподавателями и религиозными организациями).

Члены команды должны быть высокодисциплинированными и иметь оперативные исполнительские навыки. Задачи, которые будет решать команда, включают: организацию семинаров для авторов заданий и администраторов по тестированию, организацию тиражирования, распространение тестов, анкет и руководств, контакты со школами, разработку учебных материалов, а также сбор и кодирование данных. Небольшая специальная команда разработчиков тестов будет необходима для анализа учебных планов, разработки спецификаций, или проектов тестов, написания проектов заданий, отбора заданий после предварительного тестирования, а также подготовки рекомендаций по оценке. После предъявления теста должны быть оценены открытые задания и задания с множественным выбором. Команде потребуется поддержка со стороны одного или нескольких человек, компетентных в статистических и аналитических процедурах, для формирования выборки, взвешивания данных, при вводе данных и подготовке файлов, в анализе заданий по результатам тестирования, а также в общем статистическом анализе всех результатов тестирования и в подготовке файлов данных для других специалистов (например для исследователей образования и аспирантов) при проведении вторичного анализа. Многие развивающиеся страны не имеют таких возможностей в этой последней области, что приводит к ситуациям, в которых данные собираются, но никогда должным образом не анализируются и не публикуются.

Команда должна иметь необходимый персонал для первичной обработки и распространения результатов, пресс-релизов, брошюр и целенаправленных информационных бюллетеней. Она также может играть ключевую роль в организации семинаров для преподавателей и других представителей системы образования, чтобы они могли обсудить результаты, понять их важность и осознать их последствия для обучения и преподавания.

Большинство членов команды имеют возможность работать неполный рабочий день по мере необходимости. К этой категории можно причислить авторов заданий – особенно

опытных преподавателей с хорошим знанием учебных планов – и экспертов по формированию выборок и проведению статистического анализа. К работе в команде могут привлекаться специалисты и со стороны, а не только из сферы образования. Например, хорошим источником специалистов по экспертизе выборок учащихся может быть Бюро национальной переписи населения. Системные администраторы и программисты с соответствующим опытом могли бы помочь в чистке данных, а журналисты – в ярком представлении материалов пресс-релизов. Ни в Камбодже, ни в Эфиопии члены команды для национальной оценки никогда не работают полный рабочий день.

*Ответственность за проведение национальной оценки лежит на исполнительном органе (министерство образования, экзаменационная комиссия, исследовательское агентство, университет).*

#### **КТО БУДЕТ ПРЕДЪЯВЛЯТЬ ТЕСТЫ И АНКЕТЫ?**

Национальные традиции административной деятельности и специфика восприятия уровней доверия, так же как и источники финансирования, обычно влияют на отбор персонала, ответственного за предъявление тестов и анкет в национальной оценке. От страны к стране практика оценки изменяется. Так, в некоторых странах к оценке привлекались выпускники учебных заведений, а, например, в Замбии – школьные инспекторы и чиновники Министерства образования. В других странах в оценивании участвовали опытные преподаватели из школ, не занимающиеся оценкой или учебной деятельностью. На Мальдивах администратор по тестированию должен быть сотрудником школы, расположенной на другом острове, а не на том, где находится школа, в которой проводится оценка.

Необходим тщательный отбор администраторов по тестированию. Они должны иметь хорошие организаторские навыки, опыт работы в школах и быть последовательными и точными в выполнении руководств по тестированию и анкетированию. В идеале они должны иметь опыт работы в клас-

се, говорить на том же языке и с тем же акцентом, что и учащиеся, и иметь авторитетный, но не авторитарный стиль общения. В третьей книге серии рассмотрены преимущества и недостатки привлечения в качестве администраторов преподавателей, инспекторов, преподавателей педагогических колледжей, экзаменационного персонала и студентов университетов.

Хотя привлечение в качестве администраторов тестирования преподавателей тех учащихся, которые принимают участие в национальном оценивании, может оказаться удобным с административной точки зрения и очень эффективным по стоимости, есть множество аргументов против. Одни преподаватели решают, что оценивается качество их работы. Другие не смогут удержаться от привычной практики помощи учащимся и приспособиться к формальному подходу в тестировании. Третьи могут сделать копии тестов или тестовых заданий, таким образом исключая возможность использования этих заданий в будущем. Приглашение преподавателей для проведения тестирования у собственных учащихся может снизить общественное доверие к результатам оценивания.

*Ответственность за предъявление тестов и анкет лежит на исполнителем агентстве.*

### **КАКАЯ КАТЕГОРИЯ УЧАЩИХСЯ БУДЕТ ОЦЕНИВАТЬСЯ?**

В обычном понимании национальная оценка представляет собой исследование в системе образования. Однако в данном смысле она трактуется не всегда. Впервые национальная оценка проводилась в США в 1969 году – как среди молодежи, окончившей школу (17–18 лет и 26–35 лет), так среди школьников, продолжающих обучение по гражданскому праву, чтению и естественным наукам. Затем оценка генеральной совокупности выпускников была отменена из-за ее стоимости (Jones, 2003). Последующие исследования грамотности взрослого населения проводились вне системы национального оценивания.



Проблема оценки детей младшего возраста, не посещающих школу, более актуальна для развивающихся стран, чем для Соединенных Штатов Америки, потому что именно в развивающихся странах большое количество детей школьного возраста не ходят в нее. Очевидно, учебные достижения (или их отсутствие) у детей входят в сферу интересов тактиков и политиков и могут иметь особую значимость для сектора неофициального образования. Тем не менее их включение в обычную национальную оценку трудно предусмотреть. Отдельные группы детей, не посещающих школу, могут быть оценены с помощью тестов для национальной оценки по определенным предметам. Однако методы оценки и процедуры формирования выборок будут сильно отличаться, поскольку необходимо принимать во внимание различные условия жизни таких детей (например специальные потребности, социально-экономическое положение семьи или расстояние от места проживания до школы).

Что касается детей, посещающих школу, то тактики нуждаются в информации об уровне их знаний и умений в определенные моменты развития их образовательной карьеры. Должно быть принято решение о том, будет ли выборка испытуемых формироваться на основе выделения возрастной группы, или по признаку принадлежности к определенному классу обучения, или на основе сочетания первого и второго признака. В странах, где учащиеся сильно отличаются по возрасту, в котором они впервые идут в школу, и где не проводится поощряющая учебу политика, учащиеся-ровесники не обязательно попадут в одинаковые классы. В таком случае есть серьезные аргументы в пользу подхода для построения выборки на основе принадлежности учащихся к одинаковым классам, а не по возрасту.

Класс, в котором будет проводиться оценка, обычно выбирается в соответствии с информационными потребностями министерства образования. Если, например, министерство заинтересовано в получении информации об уровне достижений учащихся, заканчивающих начальную школу, то это означает, что национальная оценка должна быть проведена в конце последне-

го года начальной школы (во многих странах это 5-й или 6-й класс). Министерство могло бы также потребовать проведения национальной оценки в 3-м или 4-м классе, если оно нуждается в данных о результатах учебной деятельности учащихся в середине основного образовательного цикла. Эта информация затем может быть использована для введения коррекционных мер (таких как штатные курсы для преподавателей) по отношению к проблемам по специальным разделам учебного плана, идентифицированным в национальной оценке.

Целевые классы для национальной оценки меняются от страны к стране. В США уровни учебных достижений учащихся оцениваются в 4-м, 8-м и 12-м классах; в Колумбии – в 3-м, 5-м, 7-м и 9-м классах; в Уругвае – в дошкольном возрасте и в 1-м, 2-м и 6-м классах, в Шри-Ланке – в 4-м, 8-м и 10-м классах. В англоязычных странах Африки региональный консорциум системы образования, Южно- и Восточноафриканский консорциум по мониторингу качества образования (SACMEQ) оценивали учащихся 6-х классов. Консорциум франкоязычных африканских стран, программа d'Analyse des Systemes Educatifs de la CONFEMEN (Conférence des Ministres de l'Éducation des Pays ayant le Français en Partage) оценивали учащихся 2-х и 5-х классов.

Иногда выбор класса продиктован прагматическими соображениями. Нигерийское федеральное министерство образования решило оценивать учащихся 4-го класса, поскольку тестирование на более низком уровне потребовало бы перевода тестов на многие местные языки. Старшие классы не считаются подходящими, потому что учащиеся и преподаватели сосредоточены на подготовке к вступительным экзаменам в среднюю школу.

В относительно немногих странах проводится широкомасштабная оценка в классах с 1-го по 3-й. Учащиеся могут быть не в состоянии следовать инструкциям, справиться с когнитивными задачами оценки или с проблемами при выполнении заданий с множественным выбором. Исследование на Ямайке показало, что значительное число первоклассников не в состоянии различать буквы алфавита (Lockheed and Harris, 2005). Тем не менее мы должны иметь в виду, что, поскольку инфор-

мация об учащих­ся младшего школьного возраста может иметь решающее значение для реформирования образования, для мониторинга этих результатов должны быть приняты альтернативные процедуры.

*Ответственность за отбор категорий учащихся для оценивания лежит на министерстве образования и национальном комитете.*

### **БУДЕТ ЛИ ОЦЕНИВАТЬСЯ ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ УЧАЩИХСЯ ИЛИ ВЫБОРКА?**

Большинство национальных, а также все региональные и международные исследования при определении уровней национальных достижений используют подходы, основанные на выборках. Некоторые системы национальной оценки используют оба способа – как основанный на генеральной совокупности, так и на выборках учащихся (например Коста-Рика, Куба, Франция, Гондурас, Иордания, Мексика и Уругвай), тогда как большинство локальных программ национального оценивания собирают данные по генеральной совокупности учащихся (например, Минас-Жерайс, Парана и Сан-Пауло, Бразилия, Богота, Колумбия и Агуаскальентес, Мексика) (см. Crespo, Soares and deMello e Souza, 2000). Некоторые факторы способствуют использованию выборок испытуемых, если цель состоит в том, чтобы получить информацию для политических целей о функционировании системы образования в целом. Эти факторы включают: сокращение стоимости предъявления тестов, чистку данных и управление ими; уменьшение времени, требуемого для анализа данных и подготовки отчета; повышение точности, так как появляется возможность обеспечения усиленного контроля выполнения тестов и подготовки данных (Ross, 1987).

Как отмечалось в третьей главе, при определении того, будут ли рассматриваться выборки или целая целевая генеральная совокупность учащихся, основополагающей является цель национальной оценки. С одной стороны, решение о привлечении к оценке всей генеральной совокупности учащихся может отражать намерения повысить ответственность за резуль-

таты тестирования школы, преподавателей или даже обучаемых. Это облегчает применение санкций (поощрений или наказаний), обеспечение обратной связи с отдельными школами по результатам тестирования и публикацию сравнительных таблиц, а также идентификацию школ, которые больше всего нуждаются в помощи (например, как в Чили и Мексике). С другой стороны, подход, основанный на выборках учащихся, позволит обнаружить проблемы только на системном уровне. Он не позволит выделить отдельные школы, нуждающиеся в поддержке, хотя может определить виды или категории школ (например малые сельские школы), которые требуют внимания. Он сможет также определить проблемы, связанные с гендерным или этническим равенством.

Серьезным аргументом против использования подхода, основанного на выборках, является тот факт, что такая оценка не позволяет получить результаты тестирования для принятия особо ответственных решений в образовании, поскольку некоторые учащиеся не будут мотивированы к ответственному выполнению теста. Это не всегда имеет место, однако во многих странах, например в Южно-Африканской Республике, некоторые учащиеся опасались, что их результаты в «Международном исследовании тенденций в математическом и естественнонаучном образовании» (TIMSS) будут учитываться в качестве официальных школьных результатов. Интересно отметить, что случаи обмана во время проведения тестирования встречались, по-видимому, из-за ощущения, что с результатами выполнения тестов будут связаны относительно высокие ставки (смотри А.4 в приложении А).

Преимущества и недостатки использования национальной оценки для повышения ответственности школ, преподавателей и иногда учащихся приведены в табл. 4.2. Перечисленные темы получены по большей части из исследования последствий высоких ставок государственных экзаменов для принятия ответственных административных решений, а не из опыта национальной оценки. Тем не менее они должны соответствовать национальным оценочным системам, основанным на генеральной совокупности, по крайней мере тем, которые действуют

ТАБЛИЦА 4.2

**Преимущества и недостатки оценивания, основанного на генеральной совокупности учащихся для повышения ответственности школ**

Преимущества	Недостатки
Сосредоточено на важных аспектах образования	Имеет тенденцию к игнорированию содержательных областей, которые не отражены в тесте
Выявляет знание важных разделов отдельных предметов	Имеет тенденцию к игнорированию тех разделов предметов, которые не отражены в тесте (такие как свободное владение разговорной речью)
Помогает гарантировать, что учащиеся достигли приемлемых стандартов обучения перед дальнейшим продвижением	Содействует раннему отчислению учащихся и переводу на повторное обучение
Позволяет сравнивать школы напрямую	Приводит к несправедливому ранжированию школ, где учатся представители разных социальных групп и где результаты не имеют значимых различий
Формирует общественное доверие к результатам функционирования системы образования	Приводит к обманам со стороны учащихся во время проведения тестирования и необходимости их последующего исправления
Заставляет учащихся учиться	Делает акцент на запоминании и зубрежке
Приводит к росту уровня выполнения тестов в некоторых школах и у отдельных учащихся	Повышение уровня может быть ограничено конкретными тестами и не будет наблюдаться в других тестах по этой же предметной области
Позволяет родителям судить об эффективности деятельности отдельной школы и преподавателей	Ведет к несправедливой оценке эффективности на основе результатов тестирования, поскольку не принимает во внимание другие установленные факторы, связанные с учебными достижениями
Пользуется популярностью у политиков и СМИ	Редко расценивается политиками в качестве причины неудач в обучении для поддержки выделения образовательных ресурсов

Источник: разработки авторов.

в качестве замены государственных экзаменов (как в Соединенных Штатах Америки и некоторых странах Латинской Америки).

*Ответственность за решение относительно использования выборки или генеральной совокупности испытуемых лежит на министерстве образования.*

### ЧТО БУДЕТ ОЦЕНИВАТЬСЯ?

Все системы национальной оценки измеряют когнитивные результаты обучения или учебные умения в области языка / грамотности и математики / операций с числами, что отражает важность этих результатов для базового образования. В некоторых странах в практику оценивания включаются исследования по естественным и общественным наукам. Независимо от оценочной области важно обеспечение соответствующей структуры содержания, в первую очередь для конструирования инструментария, а во вторую – для интерпретации результатов. Структура может быть доступна в учебном плане, если, например, этот документ включает ожидаемые результаты обучения, которые четко установлены и используются в работе преподавателей. В большинстве случаев, однако, такая структура оказывается недоступной, и ее вынуждены создавать специалисты, которые связаны с проведением национальной оценки. По этой причине требуется тесное сотрудничество между агентствами по оцениванию, специалистами, ответственными за учебные планы, и другими заинтересованными лицами.

Структура предназначается для подробного разъяснения, что оценивается в широкомасштабном исследовании, как это оценивается и почему это оценивается (см. Kirsch, 2001). Цель создания структуры заключается в том, чтобы сделать процесс оценивания и предположения прозрачными не только для разработчиков тестов, но также и для более многочисленной аудитории, включая преподавателей, методистов и тактиков. Структура обычно начинается с общего определения или постановки цели, которая направляет оценку и определяет, что должно быть измерено в терминах знаний, умений и других характеристик. Она идентифицирует и описывает различные результаты деятельности или поведения, устанавливаемые конструктами, с помощью идентификации определенного числа характеристик задач или переменных, которые будут использоваться. Структура указывает, как эти результаты учебной деятельности должны быть использованы при оценке результатов выполнения тестов учащимися (Mullis and others, 2006).

Многие системы национальной оценки были основаны на содержательном анализе того, что учащиеся должны освоить в результате следования предписанному или намеченному учебному плану в определенном классе. Как правило, этот анализ проводится на матрице с характеристиками познавательных процессов по горизонтальной оси и с темами или содержательными разделами по вертикальной оси. Таким образом, пересечение характеристики познавательного процесса и содержательной области будет представлять цель изучения. Ячейкам матрицы можно присвоить весовые коэффициенты в зависимости от их важности.

Недавние национальные (и международные) оценки опирались на исследования, связанные с развитием у учащихся литературных навыков и умений выполнять действия с числами, которые не обязательно могут быть представлены в национальных учебных планах.

Например, в Международной ассоциации по эвалюации образовательных достижений в документе по структуре и спецификациях для Международного исследования прогресса в грамотности чтения (PIRLS, 2006), грамотность чтения определяется как «способность понимать и использовать те формы письменного языка, которые необходимы обществу и/или индивидуальным лицам. Читатели младшего возраста могут выделить смысл из различных текстов. Они читают, чтобы учиться, чтобы принимать участие в сообществах читателей в школе и в повседневной жизни, или для удовольствия» (Mullis and others, 2006, 3). Из этого определения видно, что чтение – это намного больше, чем расшифровка текста или выявление смысла отрывка либо стихотворения. В исследовании PIRLS дополнительно уточняется, что именно они предполагают измерять с помощью указания процессов и задач, которые предназначаются для оценки, и приводятся проценты тестовых заданий, выделяемых для каждой цели измерения (табл. 4.3).

Документ по структуре указывает, что в оценке будут использоваться тестовые буклеты с пятью литературными и пятью информационными отрывками, а каждый отрывок будет сопровождаться двенадцатью вопросами, половина из которых бу-

**ТАБЛИЦА 4.3**  
**Понимание чтения в исследовании PIRLS**

Процесс понимания	Примеры задач	Число заданий, %
Фокусируется на получении явно заданной информации	Рассмотрение для определенных идей. Поиск определений или фраз, идентифицирующих установки истории (например время и место). Поиск смысла темы или главной идеи (явно представленных)	20
Делает прямые выводы	Вывод о том, что одно событие является причиной другого. Определение обобщения в тексте. Описание связей между персонажами. Определение референтного местоимения	30
Интерпретирует и интегрирует идеи и информацию	Определение общей предпосылки или темы. Противопоставление текстовой информации. Выводы о стиле или настроении истории. Интерпретация фактического применения текстовой информации	30
Анализирует и оценивает содержание, язык и текстовые элементы	Оценка вероятности того, что описанные события могли бы произойти. Описание того, как автор разрабатывает неожиданную концовку. Суждения относительно завершенности текста или ясности информации в тексте. Определение перспектив автора	20

Источник: Campbell and others, 2001; Mullis and others, 2006.

дет в форме заданий с множественным выбором, а половина – в форме с конструируемым ответом. В нем также указывается, что, поскольку отношение к чтению имеет важное значение для развития привычки к чтению на всю жизнь, исследование PIRLS включает в анкеты учащихся вопросы для оценки их отношения к чтению. В нем обосновывается отбор учащихся 4-х классов как целевой популяции для оценки на основе того, что 4-й год обучения является переходной стадией от обучения чтению к чтению для обучения.

В структуре оценки PIRLS устанавливает две основные цели чтения:

- чтение для получения опыта грамотности;
- чтение для приобретения и использования информации.

В ней также дано подробное обоснование акцентов в исследовании PIRLS, которое преимущественно нацелено на



получение информации об окружающей среде, где учащиеся учатся читать. Этот акцент привел к включению в анкеты заданий о характеристиках домашней обстановки, которые могут поощрять детей учиться читать: действия родителей, связанные с грамотностью, язык, на котором говорят дома, контакты между домом и школой и внешкольные занятия учащихся по приобретению грамотности. Задания на школьном уровне охватывали школьные ресурсы, которые могут прямо или косвенно влиять на достижения в чтении. Документ по структуре также обосновывал оценку переменных, характеризующих обстановку в классе, таких как подходы к обучению и природа педагогического образования.

Дальнейшей альтернативой для обоснования оценочного инструментария на основе включенных в учебный план планируемых результатов обучения или предписаний, которые применимы для учащихся старшего возраста, является отражение в структуре теста тех знаний и умений, которые, вероятно, будут необходимы учащимся в дальнейшей взрослой жизни. Программа международного оценивания учащихся (PISA) приводит пример этого метода, когда предполагается оценить «математическую грамотность» пятнадцатилетних учащихся, определенную как «способность идентифицировать и понимать роль, которую математика играет в мире, приводить хорошо обоснованные суждения и использовать их вместе с математикой в деятельности, отвечающей потребностям жизни индивидуумов, характерной для заинтересованных и ответственных граждан» (ОЭСР, 2003, 24) (см. В.3 в приложении В).

Этот подход отвечает потребностям международных исследований. Однако, учитывая, что возможность разработки инструментария, который бы в равной степени соответствовал множеству учебных программ, очевидно проблематична, его можно было бы также использовать в национальной оценке.

Ряд национальных оценочных систем объединили информацию об аффективных результатах обучения (например отношение учащегося к школе и его самооценка). В Колумбии, например, оценивается отношение учащихся к миру на земле. Хотя эти результаты очень важны, их измерения, как правило,

менее надежны, чем измерения когнитивных результатов, и данные анализа на их основе было трудно интерпретировать. В Чили методические трудности измерения отношения учащихся к учебе привели к необходимости отказа от этих исследований (см. А.7 в Приложении А).

В одном широкомасштабном исследовании (Мониторинг учебных достижений) оценивались «жизненные навыки», определенные как знание учащимися чего-либо и их отношение к здоровому образу жизни и питанию, окружающей среде, гражданской ответственности, естественным наукам и технологиям (Chinapah, 1997). Хотя в целом принято считать, что жизненные навыки важны и должны преподаваться, существуют значительные разногласия по поводу их точной природы. Их измерение также достаточно затруднительно.

Большинство программ национальной оценки собирает информацию об учащихся, школах и домашних факторах, которые считаются связанными с учебными достижениями учащихся (например, пол учащегося и его «образовательная карьера», включая сведения о том, сколько раз он оставался на второй год; ресурсы школы, в том числе наличие учебников; уровень образования и квалификации преподавателей, а также социально-экономический статус семей учащихся). Информация, как правило, собирается с помощью анкет (или интервью), предъявляемых учащимся, преподавателям, директорам школ, а иногда и родителям в то же время, когда предъявляется оценочный инструментарий.

Идентификация контекстных факторов, связанных с учебными достижениями, может помочь установить управляемые переменные, т.е. факторы, которые могут быть изменены тактиками: изменение времени, предназначенного для того или иного учебного плана, обеспечение учебниками учащихся и размер классов. Контекстные данные, собранные в некоторых национальных и международных исследованиях, однако, не могут играть эту роль, так как они неадекватно измеряют условия, в которых живут учащиеся. Экономический статус, например, может быть основан на заданиях, которые включают: наличие автомобиля, телевизора и водопровода в

стране, где большинство населения живет, по крайней мере, несколько месяцев в году на средства, составляющие менее 1 американского доллара в день. Кроме того, несмотря на актуальность проблемы состояния здоровья и питания, о них нет никакой информации (Naumann, 2005).

В некоторых исследованиях оценивались достижения преподавателей (так же, как и учащихся). Во Вьетнаме (см. А.2 в приложении А) и некоторых африканских странах в исследованиях SACMEQ (см. С.1 в приложении С) преподаватели должны были выполнить те же самые тестовые задания, что и учащиеся. Это позволило получить некоторое представление о квалификации преподавателей. В Уганде были получены сведения о степени знакомства преподавателей с ключевыми официальными документами по учебному плану.

*Ответственность за принятие решения о том, что будет оцениваться, лежит на министерстве образования, национальном комитете и исполнительном агентстве.*

### **КАК БУДУТ ОЦЕНИВАТЬСЯ УЧЕБНЫЕ ДОСТИЖЕНИЯ?**

Инструментарий должен быть разработан таким образом, чтобы обеспечить получение информации, необходимой для национальной оценки. Поскольку цели и предлагаемое использование программ национальной оценки меняются, то же самое будет происходить с инструментарием, используемым при оценке, и способах, которыми результаты сообщаются.

Некоторые национальные оценки представляют результаты в терминах характеристик распределения тестовых оценок, например, средний процент заданий, выполненных учащимися правильно, и характер распределения оценок вокруг среднего. Или результаты могли бы быть шкалированы путем приведения к произвольному среднему (такому как 500) и стандартному отклонению (такому как 100). Хотя эти оценки могут использоваться для сравнения результатов подгрупп в выборке, они имеют ограниченное применение прежде всего потому, что они очень мало говорят нам об уровне знаний учащим-

ся материала или фактических умениях, которые приобрели обучаемые.

Для того чтобы решить эту проблему и сделать результаты оценки более значимыми для заинтересованных лиц, все чаще результаты в отчетах приводятся так, чтобы отобразить, что учащиеся знают и что не знают. Это позволяет определить сильные и слабые стороны их знаний и умений. Данный подход включает соответствующие оценки учащихся с описанием тех заданий, которые они в состоянии выполнить (например «может читать на определенном уровне понимания» или «может выполнять основные математические операции»). Результаты могут быть разбиты на категории различными способами (например «удовлетворительный уровень» или «неудовлетворительный уровень», «базовый уровень», «профессиональный уровень» или «продвинутый уровень»). При таком разбиении для каждого уровня должна быть определена соответствующая ему доля учащихся. Установление соответствия оценок учащихся различным уровням выполнения теста является сложной задачей, сопряженной с привлечением экспертов по учебным программам и специалистов по статистическому анализу.

Способ описания результатов должен быть рассмотрен еще на стадии разработки теста. Таким образом, разработку теста можно начать с определения его структуры, фиксирующей ожидаемые результаты обучения, на основе которой пишутся тестовые задания для оценки степени, в которой подготовка учащихся отвечает этим ожиданиям. Однако если задания не соответствуют определенным критериям при их выполнении, включая степень, в которой они дифференцируют учащихся, то они не могут быть включены в окончательный оценочный инструментарий. Необходимо позаботиться о гарантиях того, что в оценке отражены важные учебные цели, даже если ни один из учащихся при выполнении теста не представляет свидетельств их достижения.

Большинство систем национальной и международной оценки при разработке своего инструментария отдает значительное предпочтение форме заданий с множественным выбором.

Они часто дополняются открытыми заданиями, которые требуют от учащегося написания слова, фразы или предложения. Примеры заданий с множественным выбором и открытых заданий представлены в дополнительных материалах 4.2 и 4.3 соответственно.

#### Дополнительный материал 4.2

##### Примеры заданий с множественным выбором

**Предмет: география**

Река Волга протекает в:

- A. Китае
- B. Германии
- C. России
- D. Швеции.

**Предмет: математика**

Тюлень должен дышать, если он спит. Мартин наблюдал за тюленем на протяжении 1 часа. В начале его наблюдений тюлень опустился на дно моря и начал спать. Через 8 минут он медленно поднялся наверх и вдохнул воздух. Через 3 минуты он уже опять был на дне моря, и весь процесс повторялся с такой же частотой. Через час тюлень был:

- A. на дне
- B. на пути к поверхности
- C. дышал на поверхности
- D. опускался на дно.

*Источник:* примеры по математике: ОЭСР, 2007. Воспроизводится с разрешения правообладателя.

#### Дополнительный материал 4.3

##### Примеры открытых заданий

**Предмет: язык**

Слова ВЫСОКИЙ – это антоним к слову НИЗКИЙ

Подберите антонимы к словам:

Быстрый \_\_\_\_\_ Темный \_\_\_\_\_

Тяжелый \_\_\_\_\_ Старый \_\_\_\_\_

**Предмет: математика**

Используйте свою линейку, чтобы нарисовать прямоугольник с периметром 20 сантиметров. Отметьте ширину и длину.

В некоторых системах национальной (например исследование NAEP в США и Национальное оценивание чтения на английском языке в Ирландии) и международной (например, TIMSS и PISA) оценки каждый учащийся отвечает только на часть общего числа заданий (см. A8 в приложении А, B1 и B3 в приложении В). Данный подход увеличивает общий охват учебного плана в тесте без излишне большой нагрузки на отдельных испытуемых. Он также позволяет использовать расширенные отрывки (например, короткий рассказ или газетную статью) в оценивании понимания прочитанного. В других оценочных системах все учащиеся выполняют один и тот же набор заданий. Хотя есть определенные преимущества в подходе, когда отдельные учащиеся отвечают только на часть заданий, недостатки также существуют, особенно для стран, только начинающих реализовывать программы национальной оценки.

Процесс подготовки и предъявления тестов (например, тиражирование и распределение по пунктам проведения тестирования) становится более сложным, так же как оценка и шкалирование результатов, вместе с усложнением анализа данных отдельных учащихся или школ (Sofroniou and Kellaghan, 2004).

Проблеме оценке по языку, как правило, уделяют меньше внимания, чем она того заслуживает. Это связано с двумя причинами. Во-первых, хотя во многих исследованиях представлены группы учащихся, значительная часть (а иногда и большинство) которых обучаются не на родном языке, они, как правило, оцениваются на языке обучения. В Уганде, например, подавляющее большинство учащихся выполняют тесты на втором языке (см. A.9 в приложении А). Низкие результаты тестирования являются столь же типичными для такой практики, как и в целом отсутствие явного прогресса в достижениях учащихся и раннее отчисление из школы многих из них (Naumann, 2005).

Во-вторых, проблема, связанная с языковым тестированием, возникает в тех случаях, когда оценочный инструментарий должен быть переведен на один или более языков. Если проводится сравнение результатов тестирования на разных языках, то анализ результатов должен принимать во внимание возмож-

ность появления различий, вызванных изменением трудности заданий из-за особенностей перевода. Эта проблема частично связана с неполным соответствием слов при переводе. Например, в международной оценке, проводимой в Южной Африке, такие слова, как газолин «gasoline» (petrol) и «flashlight» (torch) были изменены. Гана заменила слово снег «snow» словом дождь «rain». Если языковые различия меняются в сочетании с культурными и экономическими факторами, то проблема усугубляется, поскольку трудно гарантировать эквивалентность способа формулирования вопросов и адекватность их содержания культурному фону страны во всех языковых версиях теста. Например, материал, который подходит по контексту для учащихся в сельских областях, охватывающий охоту, местные рынки, хозяйство и местные игры, может быть незнакомым для учащихся в городских районах.

Независимо от деталей метода, оценка должна обеспечить валидную и надежную информацию. Валидизация имеет несколько аспектов, включая адекватность оценочного инструментария выборке учащихся и представления конструкта (например, грамотности чтения) или предметной области учебного плана (например, общественные науки), определенного в структуре оценки. При валидизации экспертные суждения специалистов по учебным планам являются очень важными. Кроме того, оценочный инструментарий должен измерять только то, что он был предназначен измерять. Например, тест по математике или естественным наукам должен оценивать знания и умения учащихся в соответствующих областях, а не их компетентность в языке.

Надежность оценочных процедур обычно включает приближенную оценку степени согласованности, в которой отдельные задания в тесте оценивают общий конструкт, для измерения которого предназначался тест, а в случае открытых заданий – степень, в которой два или более экспертов достигают согласия в своих оценках.

*Ответственность за решение о том, как будут оцениваться учебные достижения, лежит на исполнительном агентстве.*

### КАК ЧАСТО БУДЕТ ПРОВОДИТЬСЯ ОЦЕНКА?

Частота, с которой проводится национальная оценка, варьируется от страны к стране и составляет от 1 года до 10 лет. Может появиться искушение оценивать учебные достижения по одним и тем же областям учебного плана на одинаковой выборке ежегодно, но такая частота проверок не является необходимой, поскольку она приводит к значительным финансовым затратам, если целью является мониторинг национальных стандартов. В США чтение и математика оцениваются каждые два года, а остальные предметы гораздо реже. Исследование PIRLS имеет разрыв в пять лет между первым и вторым проведением тестирования (2001–2006). В Японии учебные достижения по основным областям учебных планов оцениваются каждые десять лет, что обеспечивает необходимую информацию для коррекции учебных планов и пересмотра учебников (Ishino, 1995).

Если целью оценки является поддержка школ, преподавателей и учащихся для повышения ответственности в обучении, то она может проводиться каждый год. Более того, так как такая оценка сосредоточена и на результатах индивидуумов, и на уровне функционирования системы в целом, то должны быть оценены все (или большинство) учащиеся в системе образования. Такая система была создана в Чили и Англии.

Если же целью является лишь обеспечение информации о функционировании системы образования в целом, то вполне достаточно будет оценивания выборки учащихся в определенной содержательной области учебных планов каждые три года или пять лет. Так как образовательные системы не изменяются быстро, более частые оценки, вероятно, не будут отражать никаких изменений. Излишняя частота, вероятно, уменьшит воздействие на результаты образования и приведет к лишним затратам.

*Ответственность за решение о частоте оценки лежит на министерстве образования.*



### **КАК ДОЛЖНЫ СООБЩАТЬСЯ ДАННЫЕ ПО ОБРАЗОВАТЕЛЬНЫМ ДОСТИЖЕНИЯМ УЧАЩИХСЯ?**

Хотя тактики отдают предпочтение суммарным статистикам, существующие свидетельства многомерности учебных достижений предполагают, что единственный показатель результатов выполнения теста, такой как итоговая оценка, может скрыть важную информацию. Альтернативным подходом является обеспечение дифференцированной информации, которая отражает сильные и слабые стороны в учебных планах в стране. Информация будет еще более ценной, если она выявляет различия между знаниями основных фактов, а также умениями их использовать на простейшем уровне и их углубленным пониманием на более высоком уровне.

Для описания достижений учащихся в национальной оценке использовалось множество процедур, которые отражали разнообразие данных, обеспеченных оценкой (см. пятую книгу серии). Выбор одной или нескольких процедур должен опираться на информационные потребности министерства образования и остальных ключевых заинтересованных лиц.

#### **Информация на уровне задания**

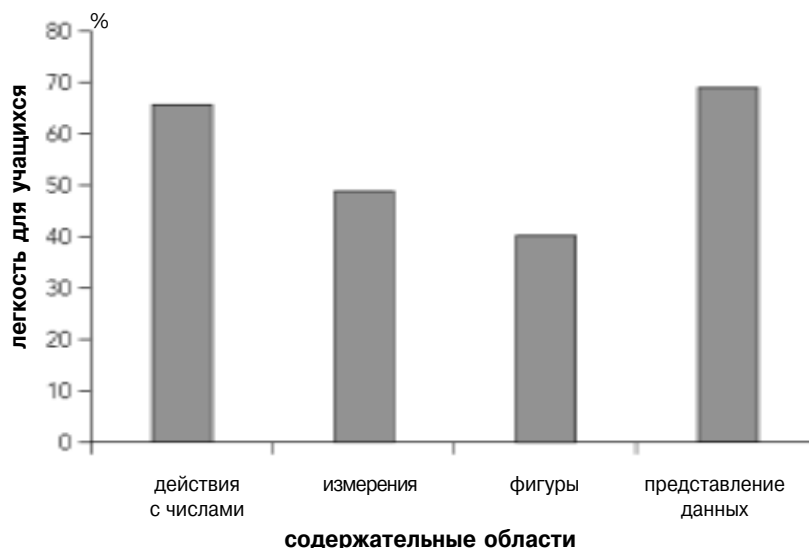
Данная информация немного более обширна, чем простое сообщение процента учащихся, правильно ответивших на отдельные задания. Данные национальной оценки могут показать, что большинство обучаемых, участвующих в ней, справляется плохо с заданиями по математике, касающимися использования показателей, или что фактически все учащиеся были в состоянии найти ассоциации между простыми словами и рисунками. В Гане, например, только 1% учащихся правильно ответили на вопрос о преломлении света в исследовании TIMSS (Гана, Министерство образования, молодежи и спорта, 2004). Этот вид информации, являющийся слишком фрагментарным для тактиков в образовании, вероятно, будет представлять интерес для персонала по разработке учебных планов, преподавателей и, возможно, авторов учебников.

### Результаты выполнения тестов по содержательным разделам учебных планов

Задания могут быть сгруппированы в соответствии с юнитами учебных планов или разделами содержания, а тестовые оценки могут сообщаться в терминах результатов выполнения теста по каждой содержательной области. Задания по чтению, например, были классифицированы в соответствии со способностями выделять информацию из текста, делать выводы из него, интерпретировать и интегрировать информацию, а также проверять и оценивать текстовую информацию (Eivers and others, 2005). Рис. 4.1 показывает, как сообщаются результаты выполнения тестов по содержательным областям математики в Лесото.

Рис. 4.1

Средний процент оценок правильных ответов по содержательным областям в математике, полученных учащимися Лесото



*Примечание:* Под легкостью понимается процент правильного выполнения заданий, называемый трудностью заданий в классической теории тестов.

*Источник:* Лесото, Экзаменационный совет Лесото и Центр развития национального планирования, 2006.

### Стандарты выполнения тестов

Стандарты в национальной и международной оценке показывают, насколько хорошо учащиеся выполняют тесты для достижения «базового», «профессионального» или «продвинутого» уровней мастерства в планируемых областях. Число уровней может варьироваться (см. А2 в приложении А для описания шести уровней профессионализма по чтению, использованных в программе национального оценивания во Вьетнаме, и см. С1 в приложении С для восьми уровней по чтению и восьми уровней математических умений, использованных в SACMEQ). Выбор пороговых баллов для дифференциации уровней требует использования статистических данных и субъективных экспертных суждений.

### Стандарты мастерства

Уровни мастерства могут быть основаны на общей тестовой оценке (например, правильные ответы по определенному проценту тестовых заданий). В Шри-Ланке для установления уровня мастерства для 4-го класса было выбрано 80%. Менее 40% учащихся достигли данного уровня по родному языку или по математике и менее 10% – по английскому языку (Perera and others, 2004). Уровни мастерства также могут быть основаны на достижении определенного результата выполнения теста. В США пять уровней выполнения тестов (ниже базового, базовый, профессиональный, планируемый, продвинутый) использовались в Коннектикуте. Планируемый уровень рассматривается как проблемный, но приемлемый уровень в соответствии с ожиданиями для учащихся и принимается в качестве уровня мастерства. Данные в табл. 4.4 показывают, что более чем половина учащихся 3-х и 4-х классов достигли планируемого уровня мастерства по всем трем предметным областям учебных планов.

*Ответственность за решение о том, в какой форме будут представлены достижения учащихся, лежит на исполнительном агентстве при участии национального комитета.*

### **КАКИЕ ВИДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДОЛЖНЫ БЫТЬ ПРОВЕДЕНЫ?**

Некоторые виды анализа будут продиктованы политическими вопросами, которые включены в программу национальной оценки в первую очередь. Большинство исследований обеспечивает данные по учебным достижениям с учетом ряда показателей. К ним относятся: гендерные признаки, региональные особенности, принадлежность школы к городскому или сельскому району, принадлежность учащегося к определенной языковой или этнической группе, статус посещаемого образовательного учреждения (государственное или частное). Некоторые системы оценивания также предоставляют данные по качеству школьных услуг (например в Кении). Анализ, включающий эти переменные, относительно прост и очень важен для тактиков и политиков. Однако он не отображает в полной мере сложную структуру данных. Часто бывают необходимы развернутые формы анализа, если мы, например, намереемся осветить школьные и базовые факторы, которые вносят свой вклад в формирование учебных достижений. Примеры использования сложных статистических процедур приведены в описании системы национальной оценки во Вьетнаме (см. А.2 в приложении А).

При проведении анализа должны быть учтены его возможные ограничения и рассмотрены вопросы установления причинной взаимосвязи в исследованиях, в которых данные собираются в одно и то же время по учебным достижениям и другим переменным. Хотя довольно трудно, а иногда просто невозможно дифференцировать эффекты воздействия сообщества, семьи и школьных факторов на обучение учащихся, эти трудности не помешали выполнить в некоторых исследованиях причинную интерпретацию данных, собранных в программах по национальному и международному оцениванию.

*Ответственность за выбор методов статистического анализа лежит на исполнительном агентстве.*

**ТАБЛИЦА 4.4****Процент учащихся по классам, достигших целевого уровня или мастерства, Коннектикут, 2006**

Класс	Математика		Чтение		Письмо	
	На целевом уровне или выше, %	На продвинутом уровне или выше, %	На целевом уровне или выше, %	На продвинутом уровне или выше, %	На целевом уровне или выше, %	На продвинутом уровне или выше, %
3-й	56	22	54	17	61	22
4-й	59	22	58	16	63	22

*Источник:* Департамент образования Коннектикута, 2006.

### **КАК РЕЗУЛЬТАТЫ НАЦИОНАЛЬНОЙ ОЦЕНКИ БУДУТ ПЕРЕДАВАТЬСЯ И ИСПОЛЬЗОВАТЬСЯ?**

Если предполагается, что результаты национальной оценки будут влиять на национальную образовательную политику, то они должны быть сообщены как можно скорее после завершения анализа данных. В прошлом методические отчеты, которые содержали значительный объем данных, как правило, были единственной формой отчетности. Однако некоторые группы пользователей (например, преподаватели в Чили, см. А.7 в приложении А) представляли данные отчеты с излишними углубленными методическими акцентами. В результате в настоящее время значительно возросла необходимость разработки других форм отчетов. Такие альтернативные варианты включают краткие сообщения, которые сконцентрированы на основных результатах для привлекаемых тактиков, пресс-релизов, специальных докладов для радио и телевидения, а также отдельные доклады для школ, преподавателей, разработчиков учебных планов и преподавателей педагогических колледжей. В некоторых странах, например в Шри-Ланке, для каждой провинции готовятся отдельные отчеты. Отчет в Эфиопии был переведен на четыре основных языка. Информационные потребности ключевых лиц должны определить содержание дополнительных отчетов.

Министерство образования должно предусмотреть соответствующее бюджетное обеспечение на этапе планирования, подготовки и распространения отчетов. В сотрудничестве с Национальным комитетом управления следует разработать процедуры для сообщения результатов национальной оценки ключевым лицам. Соответствующие стратегии для обобщения результатов должны принимать во внимание тот факт, что пользователи (будь то администраторы или преподаватели) сильно различаются по своей способности понимать и применять статистическую информацию для принятия решений. Очевидно, нет смысла в подготовке докладов, если содержащаяся в них информация не представлена надлежащим образом. Таким образом, стратегии распространения результатов исследований требуются для того, чтобы соответствующая

информация достигала все ключевые заинтересованные стороны. Стратегия должна выявлять потенциальных пользователей (ключевые институты и отдельные лица) и их уровень методической экспертизы.

Результаты национальной оценки были использованы при определении точек отсчета для мониторинга уровней учебных достижений (в Лесото), реформирования учебных планов, предоставления исходных данных о количестве и качестве учебных материалов в школах (во Вьетнаме), определения корреляций учебных достижений и диагностики тех аспектов учебных планов, которые не освоены учащимися. В Уругвае, например, результаты национальной оценки использовались для помощи в подготовке пособий для преподавателей и определения содержания учебных планов и поведенческих областей, которые напрямую помогли большому количеству сильно различающихся по подготовке учителей в реализации учебных программ (см. А.3 в приложении А).

В пятой книге серии имеется обширный материал по написанию отчета и использованию результатов национальной оценки.

*Ответственность за распространение и использование результатов национальной оценки лежит на исполнительном агентстве, министерстве образования, комитете, организаторах переподготовки преподавателей, ответственных за учебные планы и отдельных преподавателях.*

### **КАКОВА СТОИМОСТЬ КОМПОНЕНТОВ НАЦИОНАЛЬНОЙ ОЦЕНКИ?**

Стоимость национальной оценки существенно меняется. Она зависит от страны, в которой проводится оценивание, от уровня зарплаты персонала и стоимости услуг. Внутри страны стоимость будет также варьироваться в зависимости от всех или некоторых из нижеследующих факторов (Пон, 1996).

- *Исполнительное агентство.* Стоимость будет меняться в зависимости от того, есть ли у агентства возможность предоставления необходимых услуг и экспертизы или оно нужда-

ется в модернизации либо найме консультантов, полностью или частично занятых в течение рабочего дня. Стоимость предоставления услуг и оборудования, включая компьютеры и программное обеспечение, также должна быть принята во внимание.

- *Отбор содержания инструментария и его конструирование.* Состав работ по отбору содержания и формы оценивания должен быть рассмотрен в контексте их стоимости, так же как и другие факторы, такие как валидизация результатов оценки и проведение тестирования. Задания с множественным выбором более дорогие по стоимости их конструирования, чем открытые задания, но оценка результатов их выполнения обычно требует меньших затрат. Стоимость перевода тестов, анкет и пособий, а также обучения авторов заданий также должна учитываться.
- *Число участвующих школ и обучаемых.* Оценка, основанная на генеральной совокупности учащихся, очевидно, будет более дорогой, чем в случае, когда привлекаются лишь выборочные совокупности учеников. Затраты увеличиваются в тех случаях, когда требуются надежные данные для компонентов государственной системы (например, штата или провинции). Планирование оценки по возрастному уровню, вероятно, является более дорогим, чем планирование оценки по классам, потому что учащиеся разного возраста могут быть распределены по разным классам, что требует применения дополнительного оценочного материала и сессий тестирования.
- *Предъявление тестов.* Сбор данных, как правило, оказывается наиболее дорогим компонентом национальной оценки. Он включает получение предварительной информации от школ до процедуры оценки; проектирование, тиражирование, упаковку и рассылку тестовых материалов и анкет и установление системы по предъявлению инструментария. Факторы, которые влияют на полную стоимость, включают: число участвующих школ и учащихся; перевозку учащихся; трудности в получении доступа к школам; заключение договоров по привлечению лиц для учета числа учащихся; сбор и возвращение выполненных тестов и анкет.



- *Оценка, управление данными и ввод данных.* Затраты изменяются в соответствии с числом участвующих школ, числом учащихся, учителей и родителей, числом открытых заданий; способом оценки, которая может проводиться вручную или автоматизированно, числом исследований внутрирейтерной надежности (см. *Крокер Линда. Введение в классическую и современную теорию тестов / Линда Крокер, Джеймс Алгина. М.: Логос, 2010. – Примеч. ред.*) и качеством процедур предъявления тестов и оценивания их результатов.
- *Анализ.* Стоимость анализа будет зависеть от вида используемых процедур оценки и доступности технологии для шкалирования результатов тестирования и их анализа. Хотя компьютерная оценка, как обычно считают, более дешевая, чем ручная, это сокращение стоимости, возможно, не будет иметь место в стране, где технологические затраты высоки, а стоимость рабочей силы низка.
- *Сообщение результатов.* При калькуляции работ по подготовке отчетов необходимо принимать во внимание тот факт, что для тактиков, преподавателей и широкой общественности будут необходимы различные версии отчетов. На стоимость также влияет выбор стратегии по характеру и степени распространения отчетов по оцениванию.
- *Последующие действия.* При составлении бюджета, вероятно, придется учитывать обеспечение внутренней системы повышения квалификации штатного педагогического состава, которая основана на результатах национальной оценки, инструктаж специалистов по учебным планам и вторичный анализ данных. Вероятно, потребуется финансовое обеспечение для вложений в ключевые области профессиональной деятельности по оценке (например, статистический анализ). Бюджетные ассигнования должны быть выделены на повышение заработной платы участников работ в течение всего срока действия программы оценивания (обычно 2–3 года), с учетом инфляции, так же как неожиданных событий (условные).

Некоторые программы национальной оценки не достигли своей основной цели, потому что бюджет был недостаточ-

ным. Хотя за общий бюджет несет ответственность министерство образования, при составлении калькуляции для крупномасштабных проектов в бюджетных дискуссиях должны участвовать опытные сотрудники. Представители министерства, которые не знакомы с крупномасштабными проектами по сбору данных, вряд ли правильно оценят размеры бюджетных затрат для таких мероприятий, как апробационное тестирование и чистка данных.

Данные по США в программе NAEP обеспечивают приблизительные ориентиры для калькуляции: сбор данных (30%), разработка инструментария (15%), анализ данных (15%), представление и распространение (15%), формирование выборок (10%), обработка данных (10%) и расходы на руководство (5%) (Пон, 1996). В некоторых странах, где, например, министерство или экзаменационная комиссия проводит процедуры тестирования, рассматривая их как часть своих обычных обязанностей, отдельные бюджетные ассигнования могут не применяться для некоторых видов деятельности. Затраты и заработная плата будут варьироваться в зависимости от национальных экономических условий. В Камбодже (которая не входит в первую сотню стран по валовому национальному доходу), авторы заданий тестов в 2006 году получали зарплату, эквивалентную 5 долларам США в день.

Страны с очень ограниченными ресурсами могут не иметь возможности для увеличения затрат на потребности национальной оценки и не счесть их целесообразными, особенно если их система образования имеет много нерешенных финансовых проблем. Если они действительно хотят принимать участие в национальной оценке, то желательно ограничить число оцениваемых содержательных областей учебного плана (возможно, по одной для каждого класса) и искать методическую помощь и поддержку спонсоров.

При рассмотрении затрат нужно принять во внимание, что стоимость программы оценки в целом и национальной оценки в частности является очень маленькой по сравнению со стоимостью других образовательных программ (Нохбу, 2002). Зат-

раты на анализ того, что работает и что не работает в системе образования по результатам оценивания, вероятно, будут намного больше, чем стоимость самого оценивания. Книга третьей этой серии, «Осуществление национального оценивания образовательных достижений», обсуждает проблемы, касающиеся стоимости национального оценивания.

*Ответственность за оценку составляющих затрат на национальную оценку лежит на министерстве образования.*

## ИТОГИ РЕШЕНИЙ

В табл. 4.5 приводятся учреждения, которые несут основную ответственность за принятие решений по 12 компонентам национальной оценки, рассмотренным в этой главе.

**ТАБЛИЦА 4.5**

**Учреждения, несущие основную ответственность за принятие решений в программах национальной оценки**

Решения	Приоритетная ответственность			
	Министерство образования	Национальный исполнительный комитет	Агентства	Другие структуры
Дать политические указания	•			
Провести национальную оценку			•	
Предъявить тесты и анкеты			•	
Выбрать популяции для оценки	•	•		
Сделать выбор между выборкой или генеральной совокупностью	•			
Решить, что должно быть оценено	•	•	•	
Решить, как оценивать достижения учащихся			•	
Определить частоту проведения оценки	•			
Выбрать методы представления отчетов		•	•	
Определить статистические процедуры			•	
Определить методы связи и использования результатов	•	•	•	•
Оценить стоимость компонентов работ	•			•

Источник: разработки авторов.

ГЛАВА

# 5

## ПРОБЛЕМЫ ПРОЕКТИРОВАНИЯ, ПРОВЕДЕНИЯ, АНАЛИЗА, СООБЩЕНИЯ РЕЗУЛЬТАТОВ И ИСПОЛЬЗОВАНИЯ НАЦИОНАЛЬНОЙ ОЦЕНКИ

В этой главе мы рассмотрим ряд проблем, связанных с формированием доверия, которое ключевые лица могут проявлять по отношению к результатам национальной оценки. Для пяти компонентов деятельности по национальной оценке (проектирование, проведение оценки, анализ данных, написание отчета и распространение или использование результатов) мы предлагаем ряд действий, которые повысят доверие к результатам, что, в свою очередь, будет способствовать их оптимальному использованию. Для каждого компонента мы также определим распространенные ошибки, которые делаются в системах национальной оценки и о которых эвалюаторы должны знать, чтобы избежать их впредь.

### **ПРОЕКТИРОВАНИЕ**

Проектирование оценки устанавливает широкий набор параметров практической деятельности: учебные достижения, которые будут оцениваться, учащиеся какого класса или возраста будут оцениваться, политические вопросы, на решение которых направлена оценка, и будет ли оценка включать всю генеральную совокупность или выборку учащихся.

**Рекомендуемые действия**

- С самого начала привлекайте тактиков высокого уровня, чтобы гарантировать политическую поддержку и помочь в создании фрейма дизайна оценивания.
- Определите информационные потребности тактиков и обратитесь к ним при отборе разделов учебных планов, классов и популяции подгрупп (например, регион страны или пол учащихся) для оценки.
- Получите поддержку преподавателей путем их привлечения к решениям, связанным с оценкой.
- Имейте в виду, что придание слишком большого значения результатам учащихся может привести к тому, что преподаватели «уйдут в оппозицию» и сосредоточат обучение на том, что будет оцениваться, и это будет способствовать снижению эффективности учебного плана.

**Распространенные ошибки**

- Отказ от подготовки соответствующих финансовых условий для ключевых аспектов национальной оценки, включая написание отчета и его распространение.
- Отказ от создания национального руководящего комитета и его использования в качестве источника информации и руководящей структуры в ходе проведения национальной оценки.
- Отказ от получения правительственных гарантий по поддержке процесса национальной оценки, который проявляется в отказе от идентификации ключевых политических проблем, которые должны были бы быть учтены на стадии проектирования; в отсутствии национального руководящего комитета; в дроблении системы национальной оценки, происходящей в то же самое время (часто поддерживаемой внешними спонсорами).
- Отказ от привлечения ключевых заинтересованных лиц (например, преподавателей или специалистов по их подготовке) к планированию национальной оценки.

- Пропуск подгрупп оцениваемой генеральной совокупности учащихся при формировании выборки, что, вероятно, серьезно сместит результаты оценки (например, отсутствие в выборке учащихся частных или небольших школ).
- Постановка нереалистичных целей для повышения оценок по тестам (например, двадцатипроцентный прирост оценок за четырехлетний период).
- Выбор неадекватного времени для разработки теста.

### **ПРОВЕДЕНИЕ ОЦЕНКИ**

Процесс проведения оценки включает обширный диапазон действий: от разработки инструментария, соответствующего задачам оценки, до отбора тех учащихся, которые будут отвечать на задания инструментария, и предъявления инструментария в школах.

### **Рекомендуемые действия**

- Опишите в деталях содержательные и когнитивные умения и базовые переменные, которые будут оцениваться.
- Поручите разработку тестов персоналу, который знаком как со стандартами учебных планов, так и с уровнями подготовки учащихся (специально для преподавателей-практиков).
- Используйте инструментарий, адекватно оценивающий те знания и умения, по которым собирается информация, и обеспечивающий конкретную информацию о компонентах знаний или умений (например, решение проблем), а не только общую оценку.
- Разработайте четкие и недвусмысленные задания тестов и анкет и представьте их в ясной и привлекательной манере.
- Обеспечьте гарантии того, что проведены соответствующие процедуры для оценивания эквивалентности языковых версий при переводе тестов и анкет.
- Разработайте пилотные тесты, анкеты и руководства по проведению тестирования.

- Проведите анализ заданий, чтобы выявить случаи двусмысленности и возможные смещения результатов оценки характеристик учащихся (например по гендерным признакам, местоположению школы или принадлежности к этнической группе) и пересмотрите задания или удалите некоторые из них в случае необходимости.
- Тщательно скорректируйте все материалы.
- Установите процедуры для получения гарантий информационной безопасности всех национальных оценочных материалов (тестов и анкет) во всех оценочных процессах таким образом, чтобы материалы не попали в руки неправомочных людей.
- Обеспечьте информационную безопасность индивидуальных результатов учащихся и единиц учебного материала с помощью выборочной экспертизы.
- Определите целевую генеральную совокупность учащихся (категорию, из которой будут формироваться выборки, – т.е. выборочный фрейм) и исключаемую категорию (например подгруппы генеральной совокупности учащихся, которые тяжело будет привлечь к оцениванию или которые будут не в состоянии выполнять задания инструментария). По исключенным группам должны быть предоставлены точные данные.
- Обеспечьте гарантии того, что предлагаемая выборка является репрезентативной и имеет достаточные размеры для обеспечения информации о генеральной совокупности учащихся с приемлемым уровнем ошибки.
- Отберите членов выборки из выборочного фрейма в соответствии с установленной вероятностью отбора.
- При предъявлении тестов и анкет следуйте стандартной процедуре. Подготовьте руководства для администраторов.
- Обеспечьте гарантии того, что администраторы тестирования полностью знакомы с содержанием тестов, анкет, руководствами и процедурами предъявления инструментария.
- Подготовьте и примените механизм обеспечения качества, охватывающий среди прочего валидность теста, формиро-

вание выборки, тиражирование материалов, предъявление теста и подготовку данных к анализу.

### Распространенные ошибки

- Привлечение для разработки заданий тестов специалистов, которые незнакомы со средним уровнем выполнения заданий учащимися (например, ученых), что приводит к появлению слишком трудных тестов.
- Неадекватное представление учебных планов в тестах, выражающееся, например, в отказе от включения в содержание тестов важных аспектов учебного плана.
- Провал апробации заданий или предварительного тестирования из-за их проведения на непредставительной выборке из генеральной совокупности учащихся.
- Использование недостаточного числа тестовых заданий в окончательной версии теста.
- Провал попытки дать ясное определение оцениваемого конструкта (например, в тестах по чтению).
- Включение недостаточного числа примеров заданий для учащихся, которые незнакомы с формами заданий теста.
- Отсутствие мер поощрения учащихся к получению разъяснений от наблюдателей перед началом работы над тестом.
- Отказ от выдачи соответствующих уведомлений при тиражировании тестов, анкет и руководств.
- Недостаточное внимание к корректуре тестов, анкет и руководств.
- Использование неадекватных или устаревших национальных данных по числу учащихся и школ для формирования выборок.
- Неудачное выполнение надлежащих процедур формирования выборки, включая отбор предварительно определенного процента школ (например, 5%).
- Проведение неадекватного обучения администраторов по тестированию и анкетированию.
- Разрешение внешнего вмешательства (например, директора школы) во время предъявления теста.



- Отсутствие правильной рассадки учащихся, которые могут находиться близко друг к другу во время тестирования (провоцирует списывание).
- Провал попытки введения практики сверхурочной работы в случаях, когда нужно решить ключевые задачи вовремя.

### **АНАЛИЗ ДАННЫХ**

Статистические исследования организуют, суммируют и интерпретируют данные, собранные в школах. Они должны быть соотнесены с политическими проблемами, определенными в дизайне национальной оценки.

### **Рекомендуемые действия**

- Обеспечьте компетентные статистические услуги.
- Подготовьте шифры с определенными указаниями для подготовки данных к анализу.
- Проверьте и обеспечьте чистку данных, чтобы избавиться от ошибок (например, ошибки в числах, оценки вне допустимой шкалы, несоответствия между данными, собранными на разных уровнях).
- Вычислите выборочную ошибку, принимая во внимание сложности в формировании выборки, такие как стратификация и кластеризация.
- Взвесьте данные таким образом, чтобы вклад различных страт в выборке при агрегировании оценки учебных достижений отражал их пропорции в целевой генеральной совокупности.
- Определите процент учащихся, результаты которых отвечают приемлемым уровням стандартов выполнения теста.
- Проанализируйте данные оценивания, чтобы выявить факторы, которые могли бы объяснить вариацию уровней учебных достижений учащихся, для передачи информации тактикам.
- Проанализируйте результаты оценивания путем сопоставления их с содержательными разделами учебного плана.

Предоставьте информацию по подразделам предметных областей учебного плана (например, подразделы учебного плана по чтению, математике).

- Не забывайте, что вариативность измерений, отличия учебных планов и различные социальные факторы должны учитываться при анализе результатов выполнения тестов.

### Распространенные ошибки

- Использование неадекватных методов статистического анализа, включая неудачное взвешивание выборочных данных при их анализе.
- Получение результатов по небольшому числу респондентов (например, группа меньшинства из числа выбранных преподавателей, которые, возможно, могут ответить на специфический вопрос).
- Противопоставление результатов выполнения тестов в различных предметных областях учебного плана и утверждение о том, что учащиеся добиваются больших успехов по одному из предметов на основании анализа разностей средних оценок.
- Неудачная попытка подчеркнуть произвольный характер соглашения по выбору порогового балла для тестовых оценок (например, для противопоставления группы мастерства по сравнению с группой учащихся, не овладевших уровнем мастерства в предметной области, либо зачет – незачет) при дихотомическом делении результатов и неудачная попытка в установлении широкого диапазона тестовых оценок в группе.
- Скрытие информации о стандартных ошибках измерения, связанных с отдельными статистическими показателями.
- Установление рангов и разглашение результатов ранжирования школ на основе данных по тестам учебных достижений без учета ключевых контекстных факторов, вносящих свой вклад в установление рангов. Смещение в рангах проявляются в тех случаях, когда результаты по школам сравниваются путем использования оценок по невыровненным

результатам тестирования вместо оценок, учитывающих контекстные факторы (например, процент учащихся из бедных семей с низким социально-экономическим фоном), и оценок, скорректированных с учетом более ранних учебных достижений.

- Установление причинной зависимости в тех случаях, где она не имеет места (например, приписывание появления различий в результатах обучения одной из переменных, такой как руководство частной школой или размеры класса).
- Сравнение результатов тестирования по двум или более периодам времени в тех случаях, когда использовались неэквивалентные тестовые задания.
- Сравнение результатов тестирования по двум или более периодам времени без учета важных базовых условий (например, учебный план, зачисление, характеристики семьи или уровень гражданской активности), которые могли измениться за время между моментами, выбранными для сравнения. Хотя большинство образовательных взаимосвязанных переменных не склонны к быстрым изменениям за короткий промежуток времени (например, 3–4 года), некоторые страны ввели политику, которая привела к существенным изменениям в зачислении учащихся в учебные заведения. Так, например, после отмены платы за обучение число учащихся, зачисленных в школы, резко увеличилось в Малави и Уганде.
- Ограничение видов основного анализа при составлении данных по средним оценкам для географических или административных регионов.

## НАПИСАНИЕ ОТЧЕТА

Нет никакой необходимости в проведении национальной оценки, если результаты не предоставляют в форме, которая согласована с потребностями различных заинтересованных лиц.

### Рекомендуемые действия

- Подготовьте отчеты своевременно в соответствии с потребностями клиентов и представьте их в форме, понятной за-

интересованным сторонам, особенно тем, кто имеет возможность принимать решения.

- Сообщите результаты с учетом гендерных и региональных факторов, если это позволяет дизайн выборки.
- Обеспечьте в отчете или в методическом руководстве адекватную информацию, позволяющую воспроизводить оценивание.

#### **Распространенные ошибки**

- Написание громоздких методических отчетов.
- Провал попыток анализа ряда основных результатов.
- Создание рекомендаций в соответствии с определенной переменной даже при том, что результаты анализа подвергаются сомнению валидность данных по этой переменной.
- Провал попыток установления связи результатов оценки с учебным планом, учебниками и проблемами подготовки преподавателей.
- Отсутствие подтверждения того, что факторы, находящиеся вне контроля преподавателя и школы, вносят свой вклад в оценку результатов выполнения тестов.
- Неспособность признать, что различия между средними оценками могут не быть статистически значимыми.
- Несвоевременная подготовка отчета, исключая возможность влияния его результатов на соответствующие политические решения в образовании.
- Подготовка излишне большого списка литературы в отчете.
- Провал попытки предать гласности ключевые выводы отчета для отдельных соответствующих групп заинтересованных лиц.

#### **РАСПРОСТРАНЕНИЕ И ИСПОЛЬЗОВАНИЕ РЕЗУЛЬТАТОВ**

Важно, чтобы результаты национальных оценок не остались на полках тактиков, а были переданы на соответствующем языке всем, кто может повлиять на качество обучения учащихся.

**Рекомендуемые действия**

- Предоставьте результаты заинтересованным лицам, особенно ключевым тактикам и менеджерам.
- Используйте результаты в тех местах, где они повлияют на выработку тактики и улучшение обучения и совершенствование учебных планов.

**Распространенные ошибки**

- Игнорирование результатов оценки, когда дело доходит до выработки политики в образовании.
- Неудачные попытки применения результатов оценки ключевыми заинтересованными лицами (например, в системе обучения преподавателей или разработчиками учебных планов).
- Провал попыток использовать извлеченные уроки и принимать их во внимание при последующих процедурах оценки национальной командой.

ГЛАВА

# 6



## МЕЖДУНАРОДНАЯ ОЦЕНКА ДОСТИЖЕНИЙ УЧАЩИХСЯ

В этой главе мы рассмотрим программы международной оценки образовательных достижений учащихся. Они используются во многих странах, чтобы обеспечить данные для систем национальной оценки. Сначала мы в общих чертах обозначим основные особенности международной оценки, приведем сходства и отличия по сравнению с национальной оценкой. Далее мы опишем активизацию деятельности по международной оценке. Затем мы определим преимущества международной оценки и проблемы, связанные с ней.

Международная оценка достижений учащихся во многом подобна национальной оценке. Обе эти системы используют схожие процедуры (по конструированию инструментария, формированию выборок, проставлению баллов и анализу результатов). Также они имеют похожие цели: определить, насколько успешно проходит обучение в системе образования; идентифицировать отдельные сильные и слабые стороны в знаниях и умениях, которые приобрели учащиеся; сравнить достижения подгрупп генеральной совокупности учащихся (например, в терминах гендерных различий или местоположения школ); определить связь между достижениями учащихся

и вариацией характеристик образовательной среды в школе, семье, обществе. Кроме того, обе системы оценивания позволяют установить, изменяются ли достижения учащихся с учетом фактора времени (Kellaghan and Greaney, 2004). Однако на практике не всегда понятно, почему страна принимает решение об участии в международной оценке (Fetter, 2006).

Главное преимущество международной оценки по сравнению с национальной состоит в том, что первая предназначена для обеспечения тактиков, педагогов и широкой общественности информацией о характеристиках системы образования данной страны по отношению к системам образования других стран (Beaton and others, 1999; Husen, 1973; Postlethwaite, 2004). Эта информация может стать источником давления на тактиков и политиков с целью повышения качества образовательных услуг. Кроме того, возможно, эта информация внесет свой вклад в развернутое объяснение факторов, изменяющихся в зависимости от страны, которые обуславливают появление различий в достижениях учащихся.

К предметам учебного плана, которые имели наибольшие нормы участников программ международного оценивания, относятся: чтение, математика и естественные науки. Исследования проводились на уровне начальной и средней школы. Как правило, комбинация класса и возраста используется для определения тех учащихся, которые будут принимать участие в оценке (например, учащиеся двух смежных классов, которые имеют наибольшие доли девятилетних и тринадцатилетних детей, учащиеся классов с наибольшими долями девятилетних и четырнадцатилетних детей, более двух смежных классов с наибольшей долей девятилетних учащихся). В другом международном исследовании были отобраны учащиеся определенного возраста (пятнадцатилетние дети).

Результаты программ международного оценивания, таких как TIMSS и PISA, и программы региональной оценки могли быть и были использованы для подготовки отдельных национальных отчетов о результатах тестирования на уровне страны. Международные базы данных доступны для выполнения таких исследований.

Страны значительно отличаются и по той степени, в которой они полагаются на результаты международной и национальной оценки для выработки политики в образовании. Многие индустриально развитые страны проводят свои собственные исследования и при этом участвуют в международных программах. У Соединенных Штатов есть собственная программа национального оценивания образовательных достижений для 4-х, 8-х и 12-х классов; при этом США также участвуют в международных программах оценки учебных достижений. Некоторые страны с развитой индустрией принимают участие в программах международной оценки, но не проводят национальную оценку (например, Российская Федерация и Германия). Точно так же некоторые развивающиеся страны использовали программу международной оценки для обеспечения собственной единственной формы национальной оценки (Braun and Kanjee, 2007). Многие из самых бедных стран в мире ни разу не принимали участие в международной оценке или не проводили национальную оценку, хотя в последние годы ситуация существенно изменилась.

### **РОСТ АКТИВНОСТИ В МЕЖДУНАРОДНОЙ ОЦЕНКЕ**

Деятельность по международной оценке началась в тот период, когда группа исследователей встретилась в 1958 году для рассмотрения возможности принятия обязательств по изучению измеряемых результатов образования и определяющих их факторов внутри системы и между системами образования (Husen and Postlethwaite, 1996). С тех пор более 60 стран приняли участие в международных исследованиях учебных достижений по одному или нескольким различным предметам учебного плана: чтению, математике, естественным наукам, письму, литературе, иностранным языкам, гражданскому образованию и компьютерной грамотности. Самые известные программы международной оценки – TIMSS (см. В.1 в приложении В), PIRLS Международной ассоциации по эвалюации образовательных достижений (см. В.2 в приложении В) и PISA Организации по экономическому сотрудничеству и развитию (ОЭСР) (см. В.3 в приложении В).



Региональные системы оценки по чтению и математике были созданы на юге и востоке Африки (см. С.1 в приложении С), во франкоязычной Африке (см. С.2 в приложении С) и в Латинской Америке (см. С.3 в приложении С). Ряд отличительных особенностей программ TIMSS и PISA представлен в табл. 6.1 (см. также В.1 и В.3 в приложении В).

Количество стран, принимающих участие в международных исследованиях, за прошедшие годы увеличилось. В то время как до 1980-х годов в международной оценке участвовали не более 20 стран, в 1991 году программа Исследования грамотности чтения привлекла уже 32 страны. В 2003 году 52 страны приняли участие в TIMSS и 41 страна – в PISA (30 государств – членов ОЭСР и 11 стран – партнеров). Кроме того, международные исследования в последние годы имеют явно выраженную ориентацию на мониторинг результатов учащихся с учетом фактора времени. Все три основные текущие программы международной оценки (TIMSS, PIRLS и PISA) проводятся на циклической основе и имеют характер трендовых исследований.

Участие развивающихся стран в международных исследованиях раньше было крайне низким. Однако в соответствии с увеличением общего количества стран, участвующих в международных исследованиях, количество стран со слабо развитой индустрией также возросло. Большая их часть вошла в программу TIMSS в 2003 году (семь стран Африки) и в 2007 году (шесть стран Африки). За время участия в международных исследованиях развивающиеся страны проявили больший интерес к участию в исследованиях по математике и чтению, чем по другим предметам учебного плана.

Рост интереса к международным исследованиям, отмеченный в последнее время, может быть следствием процессов глобализации, стремления к предоставлению качественных услуг в здравоохранении и образовании на уровне международных стандартов и заинтересованности в получении соответствующего признания. Результаты некоторых исследований свидетельствуют о том, что качество образования (в особенности в плане учебных достижений по математике и естественным наукам) играет важную роль в экономическом росте страны, хотя эта тен-

денция не полностью прослеживается по странам на протяжении длительного периода (Coulombe, Tremblay and Marchand, 2004; Hanushek and Kimko, 2000; Hanushek and Wossmann, 2007; Ramirez and others, 2006). Тем не менее, каковы бы ни были причины, образовательная политика во всем мире все сильнее направлена на развитие мониторинга совокупности учебных достижений учащихся в международном контексте.

### **ПРЕИМУЩЕСТВА МЕЖДУНАРОДНОЙ ОЦЕНКИ**

Для привлечения стран к участию в международной оценке было выдвинуто большое количество аргументов. Возможно, наиболее очевидным из них является то, что международные исследования обеспечивают сравнительную структуру, которая позволяет оценить достижения учащихся и обеспечение учебного процесса в стране, а также разработать процедуры для устранения выявленных недостатков (Straus, 2005). С помощью сравнительного анализа международной практики на основе результатов оценки страны могут выявить свои достижения, установить характер их распределения и связи, существующие между средними оценками учебных достижений и их распределением. Например, могут ли высокие средние оценки учебных достижений совмещаться с узким диапазоном распределения результатов выполнения теста? Если верить результатам исследования PISA, то такая ситуация встречается на практике.

Данные по учебным достижениям обеспечивают только ограниченную информацию о качестве образования в отдельно взятой стране. Это позволяет утверждать, что преимущество международных исследований состоит в том, что они могут использовать вариативность результатов, существующую в системах образования разных стран, для расширения диапазона условий, которые возможно изучить, в отличие от тех, которые действуют в отдельной стране (Husen, 1973). Исходя из этих соображений, анализ данных, собранных в этих исследованиях, обычно проводится с учетом взаимосвязи между учебными достижениями и широким диапазоном контекстных переменных. Диапазон рассматриваемых переменных включает

**ТАБЛИЦА 6.1**  
**Сравнение TIMSS и PISA**

Критерий сравнения	TIMSS 2003	PISA 2003
Цели	Обеспечить сравнительные данные о степени, в которо	Обеспечить сравнительные данные о результатах школьной системы в ведущих развитых странах и оценить, могут ли учащиеся применять свои знания и компетенции в чтении, математике и естественных науках в реальных жизненных ситуациях.
	Проводить мониторинг изменений уровней учебных достижений с течением времени.	Проводить мониторинг уровней учебных достижений и вклада учебных результатов с течением времени.
	Проводить мониторинг отношения учащихся к математике и естественным наукам.	Проводить мониторинг подходов учащихся к изучению математики, естественных наук и чтения.
	Анализировать связь между областью обучения, школьными факторами и учебными достижениями.	Обеспечить базу данных для разработки политики в образовании
	(Чтение охвачено отдельно в оценивании программы PIRLS.)	
Структура	Разработана экспертами по содержанию из нескольких стран – участниц программы	Разработана экспертами по содержанию из нескольких стран – участниц программы
Целевая генеральная совокупность	4-е и 8-е классы	Учащиеся 15 лет
Адекватность учебному плану	Спроектирована для оценивания официальных учебных планов, структурированных в соответствии с признанными предметами учебных планов, общих для стран – участниц программы	Спроектирована для охвата знаний, полученных как в школе, так и вне школы, определена в терминах пересечения идей и компетенций, применимых к личностным, образовательным, профессиональным, общественным и научным ситуациям

Содержание различных заданий (математика, 8-й класс)	8-й класс, распределение заданий:	Математика, охваченные разделы:	Распределение заданий:
	• действия с числами – 30%	• количество	• действия с числами – 31,8%
	• алгебра – 25%	• пространство и формы	• геометрия – 21,2%
	• работа с данными – 15%	• изменения и связи	• статистика – 21,2%
	• геометрия – 15%	• изменчивость	• функции – 10,6%
	• измерения – 15%		• дискретная математика – 5,9%
			• вероятность – 5,9%
			• алгебра – 3,5%
Когнитивные процессы	8-й класс	Распределение заданий:	
	• решение типичных проблем – 40%	• установление связей – 47%	
	• использование концепций – 20%	• репродукция – 3%	
	• знание фактов и процедур – 15%	• рефлексия – 22%	
	• рассуждения – 25%		
Виды заданий (математика)	Приблизительно две трети – задания с множественным выбором, оставшаяся часть – задания с конструируемым ответом или с открытым ответом	Приблизительно одна треть – задания с выбором, оставшаяся часть – как правило, закрытые задания (один возможный правильный ответ) или открытые задания (более одного возможного правильного ответа), или задания с конструируемым ответом	
Частота	Каждые 4 года: равные пропорции заданий по математике и естественным наукам в каждом цикле	Каждые 3 года: полный охват содержания одного предмета за каждые 9 лет (чтение в 2000 году, математика в 2003 году, естественные науки в 2006 году), плюс менее полный охват других двух предметов каждые 3 года	
Территориальный охват	48 стран: 20 – с высоким доходом, 26 – со средним доходом, 2 страны с низким доходом	30 стран ОЭСР так же, как и 11 других стран	
Анализ	Четыре уровня с точками отсчета и оценки средних, которые основаны на данных всех участвующих стран	Семь уровней освоения по математике и средние оценки, которые основаны на данных по странам ОЭСР	

Источник: структуры исследований TIMSS и PISA, Национальный центр образовательной статистики, база данных по показателям Всемирного банка.

содержание учебных планов, время, продолжительность учебного процесса, педагогическое образование, наполняемость классов и организацию системы образования. Ясно, что ценность международных исследований возрастает, поскольку они предоставляют исследователям и политикам информацию, выдвигающую гипотезы о причинах отличий учащихся по их достижениям от страны к стране. Исследования также обеспечивают информационную базу для проведения эвалюации образовательной политики и практики.

Потенциально программы международной оценки способны пролить свет на те концептуальные аспекты развития образования, на которые раньше в стране не обращали внимание (например, на определение понятия грамотности или его осмысление в учебных планах в контексте планирования, выполнения и достижения; см., например, Elley, 2005). Иногда программы оценки помогают идентифицировать вопросы и приводить к тем выводам, которые могут считаться само собой разумеющимися (например, ценность всеобщего образования по сравнению с селективным образованием, меньшие размеры класса, связываемые с более высокими достижениями учащихся или польза повторного обучения в том же классе для учащегося).

Международные исследования могут привлечь внимание СМИ и большого числа заинтересованных лиц, таких как политические деятели, ученые, преподаватели и общественность. Различия между странами в уровнях учебных достижений наглядно представлены в виде дескриптивной статистики, содержащейся в исследовательских отчетах. Действительно, эти различия обычно представляются в виде упорядоченных таблиц, в которых страны ранжируются по уровням учебных достижений. Сравнительные данные по результатам этих исследований оказывают большее шокирующее воздействие на общественность страны, чем результаты национальной оценки. Низкие результаты могут способствовать дебатам, которые, в свою очередь, обеспечат политическим деятелям и политикам аргументы для увеличенной бюджетной поддержки сектора образования, особенно если низкие результаты связаны с более низким уровнем расходов на образование.

Важная особенность программ международной оценки состоит в том, что они предоставляют данные, которые отдельные страны могут использовать для выполнения анализа своих достижений и подготовки отчета, который, в действительности, будет отчетом по национальной оценке. Этой практике следует ряд стран, которые участвуют в программе PISA (см. В.3 в приложении В) и SACMEQ (см. С.1 в приложении С). Практическая значимость этих результатов возрастает, если, в дополнение к данным, собранным в международном исследовании, будут получены данные по специальным проблемам, характерным для отдельных стран.

Участие в международной оценке обеспечивает для стран практические преимущества, особенно в тех случаях, когда у них нет возможности развивать при обучении в университетах те виды умений, которые необходимы для национальной оценки. Во-первых, центральное агентство может проводить анализ на национальном уровне, результаты которого используются в отчетах по отдельной стране. Во-вторых, исследования могут способствовать развитию методического потенциала для оценочной деятельности в стране, связанного с освоением методов формирования выборок, определения учебных достижений, разработки тестов, анализа статистики и подготовки отчетов. В-третьих, требования к штату и затраты (например, на разработку инструментария, чистку и анализ данных) могут быть ниже, чем в программах национальной оценки, потому что эти затраты распределяются между странами.

Исследование воздействия программы TIMSS на процесс обучения по математике и естественным наукам и их освоение учащимися в странах – участниках предоставляет свидетельства разнообразных последствий, к которым приводит это международное исследование (Robitaille, Beaton and Plomp, 2000):

- Результаты TIMSS использовались в парламентских дискуссиях по планированию изменений в образовательной политике (Япония).
- Министр образования учредил специальную комиссию по математике и естественным наукам (Новая Зеландия).

- Президент страны принял решение о том, чтобы был осуществлен «комплекс мер по спасению» для улучшения результатов обучения по естественным наукам и математике (с преимущественным акцентом на обучении преподавательского состава школ) (Филиппины).
- Были установлены национальные нормы в области грамотности и действий с числами (Австралия).
- Результаты внесли вклад в разработку новых образовательных стандартов по математике и естественным наукам (Российская Федерация).
- Результаты помогли изменить природу общественных обсуждений в области образования путем перехода от дискуссий, основанных на субъективных суждениях, к дискуссиям, основанным на фактах (Швейцария).
- Результаты привели к разработке учебных материалов, которые были основаны на анализе распространенных заблуждений и ошибок учащихся при выполнении заданий программы TIMSS (Канада).
- Результаты ускорили пересмотр учебных планов для их изменения (Чехия, Сингапур).
- Результаты программы TIMSS были определены как один из многих факторов, повлиявших на изменение образовательной политики в области обучения математике (Англия).
- Были сформированы комитеты для пересмотра учебных планов по математике и естественным наукам (Кувейт).
- В учебный план по математике были внесены новые темы (Румыния).
- В учебные планы по математике и естественным наукам были внесены новые разделы содержания, связанные с жизненными ситуациями (Испания).
- Результаты помогли выявить потребность в совершенствовании баланса между формальными и контекстуальными математическими знаниями (Швеция).
- Результаты программы TIMSS подтвердили предположение о существовании гендерных различий и отрицательном отношении к естественным наукам и математике

и были использованы как основа для проведения реформ учебного плана и профессионального развития преподавателей (Республика Корея).

- Результаты повлияли на итоги дискуссий об улучшении организации и перенос акцентов на подготовку преподавателей (Исландия).
- Результаты программы TIMSS привели к выполнению ряда шагов по усилению мер для развития профессионализма преподавателей по математике и естественным наукам (Норвегия, США).
- Частично в ответ на результаты программы TIMSS была введена централизованная система экзаменов (Латвия).
- Результаты TIMSS внесли существенные изменения в систему обучения, организацию деятельности школ и формирование классов, подготовку преподавателей и постановку целей для школ (Шотландия).
- Результаты TIMSS повлияли на проведение исследований в образовании, разработку стандартов и документов для учебного плана, обучение преподавателей, разработку методик преподавания математики и естественных наук и разработку учебников (Словацкая Республика).

Результаты анализа данных программы PISA привели к следующему.

- Подверглась сомнению целесообразность интенсивного применения компьютеров в классных комнатах для улучшения учебных достижений.
- Было определено, что уровень национальных расходов на образование не связан с уровнем достижений (среди стран – участников).
- Начались дебаты по вопросам общей политики в образовании (Германия).
- Внесен вклад в разработку учебных планов по естественным наукам для средней школы (Ирландия).
- Была подчеркнута сложность связей между социально-экономическим статусом семьи учащегося и его достижениями в чтении по ряду стран.



- Оценена связь между учебными достижениями, видом школы и характером отслеживания учебных планов внутри школы.
- Нашло подтверждение мнение о том, что государственные и частные школы оказывают на одинаковые категории обучающихся равное воздействие по результатам обучения, но при этом частные школы, деятельность которых зависит от правительства, являются более эффективными для учащихся из семей с более низким социально-экономическим статусом.
- Была подчеркнута потребность в интенсивных программах по языку и чтению для иностранных учащихся с целью повышения уровня их учебных достижений (Швейцария).

#### **ПРОБЛЕМЫ В МЕЖДУНАРОДНОЙ ОЦЕНКЕ**

Несмотря на очевидные преимущества, которые имеют программы международной оценки, прежде чем та или иная страна решит принять участие в одной из них, необходимо рассмотреть проблемы, связанные с таким участием (см. Kellaghan, 1996).

Во-первых, трудно спроектировать оценочную процедуру, которая будет адекватно измерять результаты обучения по множеству учебных планов. Хотя учебные планы во всем мире имеют общие элементы, особенно на уровне начальной школы, существенные различия между странами также существуют. Они касаются того, что именно преподается, когда преподается и каковы стандарты ожидаемых достижений.

Анализ заданий TIMSS, проведенный в Южно-Африканской Республике, показал, что только 18% заданий по естественным наукам соответствовали национальному учебному плану для 7-го класса, в то время как 50% соответствовали учебному плану для 8-го класса (Howie and Hughes, 2000). Чем больше различий между учебными планами и уровнями учебных достижений в странах, участвующих в программах международной оценки, тем труднее разработать процедуру, которая будет устраивать все страны. Поэтому большие сомнения вызывает ва-

лидность любых выводов, которые делаются по результатам сравнения учебных достижений.

Мы вправе надеяться, что тест учебных достижений, основанный на содержании национального учебного плана, обеспечит более валидные меры результатов освоения учебных планов, чем тест, который был спроектирован в качестве общего знаменателя учебных планов, предлагающихся в 30–40 странах. Например, национальное руководство по разработке учебных планов и специалисты по проектированию международной оценки могут присваивать умениям совершенно различные весовые коэффициенты в соответствии с их важностью, например, умению сделать вывод из предложенного текста. Национальная оценка, рассматриваемая в контексте противопоставления международной, может проверять разделы учебного плана, которые являются уникальными для отдельных стран.

Разработка единого оценочного инструментария для одних предметов учебного плана (например, для естественных и общественных наук) сопряжена с большими трудностями, чем для других предметов (чтение). Например, в случае оценки достижений учащихся по естественным наукам наборы достижений, как было доказано, являлись более гетерогенными, чем по математике. Кроме того, требовалось учитывать большее число факторов для анализа отличий результатов выполнения тестов учащимися по естественным наукам, чем по математике. Таким образом, тест по естественным наукам, который был бы пригоден для множества систем образования, спроектировать трудно.

Вторая проблема, связанная с международной оценкой, заключается в том, что раньше ставилась амбициозная цель – извлечь позитивные следствия из отличий, существующих в системах образования, чтобы оценить влияние разнообразия школьных ресурсов и процесса обучения на его результаты. Но эта цель оказалась практически недостижимой. Поскольку относительный эффект переменных зависит от контекста, в который они включены, методы обучения, ассоциируемые с высокими достижениями в одной стране, могут не иметь

подобные связи в других странах. Фактически, теснота корреляционных связей между основными факторами и учебными достижениями изменяется в широком диапазоне от страны к стране (см., например, ОЭСР и ЮНЕСКО, Институт статистики, 2003; Wilkins, Zembylas and Travers, 2002). Особые трудности возникают в тех случаях, когда развивающиеся страны принимают участие в исследовании, предназначенном для индустриально развитых стран, потому что они существенно отличаются по социально-экономическим факторам. Так факторы бедности семьи учащегося, плохого питания и медицинского обслуживания, неразвитости образовательной инфраструктуры, низкого уровня ресурсного обеспечения могут оказать существенное влияние на результаты.

В-третьих, генеральные совокупности и выборки учащихся, участвующих в программах международной оценки, не всегда строго сопоставимы. Например, отличия в результатах тестирования могут быть следствием того, что страны отличаются в настоящее время по степени, в которой отдельные категории учащихся удаляются из доминирующих страт в выборке и таким образом исключаются из оценки (например, учащиеся специальных программ или учащиеся в школах, в который язык обучения отличается от языка оценивания).

Наиболее очевидные проблемы возникают, когда в странах существенно отличаются: возраст зачисления в школы; задержки в переводе из класса в класс; проценты отсева. Особенно это касается тех исследований, в которых одновременно участвуют индустриально развитые и развивающиеся страны. В некоторых развивающихся странах большие доли учащихся выбывают из школы задолго до конца периода обязательного обучения. В то время как зачисление в начальную школу в Западной Европе и Северной Америке имеет почти стопроцентный характер, в странах Африки района Сахары этот показатель в среднем меньше 60% (ЮНЕСКО, 2002). Состав групп учащихся, отчисляемых из школы в начальных классах, может отличаться от страны к стране. В латиноамериканских и арабских странах мальчики не заканчивают 5-й класс чаще, чем де-

вочки. Эта тенденция также сохраняется в некоторых африканских странах (например, в Гвинее и Мозамбике). Проблема формирования выборки для программы TIMSS появилась в Республике Йемен, где у нескольких школ не было 4-го класса и где школа для детей из кочевых семей не имела определенного месторасположения.

Подобные проблемы сопоставимости результатов могут возникнуть и в национальной оценке. Например, дифференциация результатов тестирования учащихся в штатах Индии была приписана отличительным показателям выживания (см. 1 в приложение А).

В-четвертых, вариация оценок результатов выполнения тестов является важным фактором, если она используется для: адекватного описания достижений учащихся в системе образования; определения корреляции учебных достижений. Поэтому тщательно спроектированные национальные тесты должны гарантировать относительно широкий диапазон распределения тестовых оценок. Однако многие задания программ международной оценки оказались слишком трудными для учащихся из менее развитых стран, что привело к ограничению величины дисперсии тестовых оценок. Этот результат отражен в табл. 6.2, данные которой основаны на выбранных странах, участвовавших в программе TIMSS с 2003 года.

Данные показывают процент учащихся 8-го класса, которые достигли определенных уровней или норм результатов выполнения тестов по сравнению со всеми учащимися, которые выполняли тест. 7% всех тех учащихся, кто выполнял тест по математике, достигли международной нормы, обозначенной термином «продвинутый уровень», 23% учащихся – нормы, обозначенной термином «высокий уровень», половина учащихся – нормы «промежуточный уровень» и примерно три четверти учащихся выполнили нормы «низкого уровня». В качестве контраста можно привести другие результаты: 2% учащихся из Ганы достигли промежуточного уровня и 9% – низкого уровня. Продвинутого и высокого уровней не достиг ни один учащийся.

ТАБЛИЦА 6.2

Процент учащихся, достигших международных норм  
в программе TIMSS по математике, 8-й класс:  
страны с высокими и низкими оценками

Страны	Продвинутый	Высокий	Средний	Низкий
Сингапур	44	77	93	99
Китай	38	66	85	96
Республика Корея	35	70	90	98
<b>Среднее значение баллов по международному</b>	<b>7</b>	<b>23</b>	<b>49</b>	<b>74</b>
Филиппины	0	3	14	39
Бахрейн	0	2	17	51
ЮАР	0	2	6	10
Тунис	0	1	15	55
Марокко	0	1	10	42
Ботсвана	0	1	7	32
Саудовская Аравия	0	0	3	19
Гана	0	0	2	9

*Примечание:* Определения уровней использованы в соответствии с программой TIMSS 2003 года. *Продвинутый уровень:* учащиеся могут структурировать информацию, делать обобщения, решать творческие задачи, извлекать и обосновывать выводы на основе данной информации. *Высокий уровень:* учащиеся могут применять свои знания для широкой области относительно сложных ситуаций. *Средний уровень:* учащиеся могут применять базовые математические знания для простых решений. *Низкий уровень:* учащиеся имеют некоторые базовые математические знания.

*Источник:* Mullis and others, 2004, 64.

Подобным образом в исследовании PISA 2003 года ограниченное использование результатов для внутренних политических решений было вызвано низкой вариацией тестовых оценок у ряда стран; большинство пятнадцатилетних учащихся в Бразилии, Индонезии и Тунисе получили оценки ниже уровня 1. (Уровень 2 был предложен в качестве минимального требования для учащихся, приступающих к трудовой деятельности или к получению дальнейшего образования.) Ясно, что информация, обеспеченная этими исследованиями для тактиков и политиков, принимающих решения относительно учебных достижений учащихся в образовательных системах этих стран, носила ограниченный характер. Кроме того, из-за ограниченной дисперсии оценок учебных достижений, корреляции

между достижениями и основными или школьными переменными слабо отражали влияние тех факторов, которые способствуют росту учебных достижений.

В-пятых, в исследованиях возникают проблемы в тех случаях, когда основное внимание в отчете о результатах международной оценки уделяется ранжированию стран по средним баллам их учащихся, к которым обычно проявляют наибольший интерес средства массовой информации. Ранжирование само по себе ничего не говорит нам о многих факторах, которые могут лежать в основе появления различий между результатами учащихся разных стран. Кроме того, ранжирование может вводить в заблуждение в тех случаях, когда не принимается во внимание статистическая значимость различий средних оценок достижений. Важно учитывать, что ранг страны может измениться в зависимости от участвующих в исследовании стран и с течением времени в случаях, когда проводятся ранговые сравнения. Таким образом, например, если число стран, имеющих, как правило, высокий уровень достижений в международных исследованиях, будет уменьшаться, а число стран с традиционно низким уровнем учебных достижений – расти, то ранг страны может повыситься без реальных позитивных изменений учебных достижений в системе образования данной страны.

В-шестых, низкие результаты международной оценки (так же, как и в национальной) могут спровоцировать политические риски в решениях ключевых официальных лиц в системе образования, включая министра образования и его подчиненных. Риск, вероятно, возрастает по мере того, как международный ранг страны снижается по сравнению с рангами стран, постоянно участвующих в исследованиях. В некоторых странах, где были собраны данные, официальные лица в системе образования отказывались от включения результатов своей страны в публикуемый список межгосударственных сравнений. (Программа «Образование для всех» больше не допускает участия тех стран, которые уклоняются от сравнений.) Получение сравнительных данных для соседних стран или стран одного региона кажется более уместным, чем получение данных для стран по всему миру и значительно отличающихся по уров-

ню социально-экономического развития. Пример можно найти в Латинской Америке и в странах Карибского бассейна, где 13 стран совместно выполнили программу оценки ключевых компетенций по языку и математике в 1997 году (см. С.3 в приложении С). Программы оценки SACMEQ в Южной и Восточной Африке, которые были выполнены под руководством сети министерств в 1990-х годах, также предусматривали международные сравнения внутри региона (см. С.1 в приложении С).

В-седьмых, требования выполнения работы в срок могут оказаться очень затруднительными в странах, которые испытывают недостаток в административном персонале и имеют неразвитую коммуникационную инфраструктуру (см. дополнительный материал 6.1). Время, позволяющее выполнить различные задачи (например, тиражирование или распределение тестовых буклетов), связанные с международной оценкой и являющиеся традиционными для индустриально развитых стран, может оказаться недостаточным при наличии диапазона основных проблем во многих развивающихся странах, включая неэффективные системы связи.

#### Дополнительный материал 6.1

##### **Опыт участия Южной Африки в программах международной оценки**

Опыт Южной Африки в программе TIMSS ярко иллюстрирует те проблемы, с которыми сталкиваются все участники программ международной оценки. Крайние сроки работ, установленные организаторами, могут быть трудно достижимыми или вообще не реальными в ситуации, когда почтовая служба, телефонная связь или фонды для оплаты поездок в школы являются неадекватными.

Другие проблемы: отсутствие точных данных о генеральной совокупности учащихся в школах; низкий уровень управленческих навыков; недостаточное внимание к деталям, особенно при редактировании тестов, кодировании их результатов и сборе данных; нехватка финансирования для поддержки исполнителей проектов и трудности в обеспечении качественной своевременной печати тестов. Инструкции для администраторов тестирования (например, рекомендуемые администратору подняться наверх в помещении для тестирования или спуститься вниз по проходу) могут оказаться бессмысленными в тех случаях, когда в помещении нет соответствующих условий.

Источник: Howie, 2000.

Наконец, к недостаткам можно отнести также существенные затраты, связанные с участием в международных исследованиях. Страна, принимающая участие в программе TIMSS для 8-х классов, должна была заплатить 40 000 долларов в дополнение к расходам на тиражирование и распределение тестов, проведение тестирования, ввод данных и оценивание результатов. Национальная оценка, конечно, также требует значительных затрат.







## ВЫВОДЫ

Читатели, которые ознакомились с предыдущими главами и дошли до этой, заключительной, главы, получили представления об основных особенностях программ национальной и международной оценки, их сходствах и различиях, о причинах, побуждающих страны к участию в оценке, и о сопутствующих проблемах. Читатели узнали об основных видах деятельности, включая определение ключевых вопросов политики, разработку инструментария, отбор школ и учащихся для репрезентативного представления системы образования, анализ данных для описания достижений учащихся и их корреляций и связь формы представления результатов и запросов потребителей. Для решения всех этих задач требуются специальные знания и умения.

Если читатель является руководителем группы тактиков или сотрудником министерства образования, то он или она вряд ли будут обладать специальными знаниями и умениями, которые необходимы для детального выполнения национальной оценки. Данный недостаток не означает, что эти руководители и специалисты не будут играть важную роль в оценке от момента ее планирования и выбора общего дизайна до со-

проведения процесса ее выполнения, интерпретации и применения результатов. В этой главе мы обращаем особое внимание на роль тактиков или менеджера по развитию в институционализации деятельности по национальному оцениванию и в оптимальном использовании его результатов.

Руководитель группы тактиков или менеджеры, которые принимают решение о проведении национальной оценки (или об участии в международной оценке), должны быть убеждены в том, что информация, которую она предоставит, будет полезна для идентификации проблем системы образования и информирования тех политиков и практиков, которые способны эти проблемы рассмотреть. Их обязанности, вероятно, будут расширены, если оценка обеспечит выполнение пяти условий.

Во-первых, оцениваемые достижения учащихся рассматриваются как важные результаты обучения и, соответственно, адекватно отражают учебный план. Во-вторых, у инструментария, используемого для оценивания, есть потенциал для предоставления диагностической информации (под диагностической информацией понимаются данные о причинах неправильного выполнения отдельных заданий теста, собранные в соответствии с пробелами в профиле ответов на задания теста (см. *Челышкова М.Б. Конструирование и применение педагогических тестов / М.Б. Челышкова. М.: Логос, 2002). – Примеч. ред.*) относительно достижений каждого учащегося, в частности, сильных и слабых сторон в профиле его достижений.

В-третьих, метод формирования выборки (если оценка базируется на выборочных данных) должен гарантировать, что собранные данные адекватно представляют учебные достижения в системе образования в целом (или ясно идентифицируют некоторую часть достижений). В-четвертых, чтобы идентифицировать и описать основные особенности данных, необходимо использовать подходящие методы анализа, включающие связи между значимыми переменными. В-пятых, методические аспекты оценки должны соответствовать существующим профессиональным стандартам в таких областях, как разработка тестов, формирование выборки и статистический анализ.

Вся эта деятельность требует значительных ресурсов и политической поддержки. Например, тактики или менеджеры должны играть важную роль в обеспечении гарантий того, что знания и умения, которые требуются для проектирования, проведения и интерпретации результатов национальной оценки, являются доступными. Во многих странах эти условия будут невыполнимы, поэтому их необходимо создавать специально. Такое создание потребует введения долгосрочных или краткосрочных программ обучения. Согласно этим программам, необходимо обеспечить повышение методической подготовки с помощью программ обучения в стране, организации профессиональных встреч и введения долгосрочных программ обучения всех тех специалистов, которые будут привлечены к администрированию национальной оценки на постоянной основе.

В некоторых странах деятельность по национальной оценке влияет на те области системы образования, которые выпадают из ее государственной структуры и процессов принятия политических решений. В этой ситуации нет никаких гарантий того, что информация, полученная в ходе оценки, будет использована для реализации образовательной политики или что программы национальной оценки будут выполняться для проведения мониторинга изменений учебных достижений с течением времени. Чтобы найти адресное решение этих проблем, деятельность по национальной оценке должна стать регулярной при функционировании системы образования. Эта деятельность потребует активного участия руководителей и тактиков в общем планировании и то ли участия, то ли выдвижения представителей национального руководящего комитета. Она также потребует выделения соответствующего бюджета и принятия решения о том, кто будет осуществлять оценочную деятельность, которая изменяется от страны к стране в зависимости от их особенностей.

Долгосрочные правительственные обязательства очень важны в формировании мощной институциональной базы для проведения национальной оценки на регулярной основе. В ее рамках агентству могут дать разрешение для приема на работу

и обучения специалистов по таким ключевым областям, как разработка тестов, формирование выборок и проведение статистического анализа. Пассивная роль правительства может найти свое отражение в модели «распыления» ответственности при реализации национальной оценки по различным агентствам, в выборе стратегии, ограничивающей или полностью исключающей возможности для проведения необходимой методической экспертизы. Существует практика, когда в ряде стран множество агентств выполняли несвязанные программы национальной оценки, используя диапазон подходов с ограниченными возможностями для проведения образовательной политики.

В некоторых случаях правительственные обязательства могут быть расширены, когда специальное подразделение внутри министерства, деятельность которого поддерживается отдельной строкой в бюджете системы образования, проводит оценку. В Чили, например, правительственные обязательства и интерес общественности к результатам программы Sistema de Medicion de la Calidad de la Educacion (SIMCE) возросли после того, как функции по национальной оценке были переданы от университета министерству образования. Ежегодная оценка, своевременное сообщение результатов и оценка ценности результатов для принятия политических решений помогли усилить легитимность SIMCE, институционализировать ее работу и гарантировать дальнейшие долгосрочные правительственные обязательства и поддержку. Во многих других латиноамериканских странах центры оценки, которые независимы от министерств образования, преуспели в развитии компетентности и автономности, проводя оценивание со значительной гибкостью и согласованностью (Fetter, 2006).

Одной институционализации недостаточно, хотя она и обеспечивает гарантии того, что не возникнет ситуация, когда результаты национальной оценки не достигают ключевых правительственных фигур. Также существует необходимость приложить усилия к введению процедур для передачи результатов ключевым заинтересованным лицам внутри и вне министерства. Кроме официальных правительственных лиц, по-

требность в результатах оценки также ощущается в деятельности разработчиков учебных планов, экзаменационных комиссий, педагогов, обучающихся преподавателей и самих преподавателей в их каждодневной работе в школах. Обращение к информационным потребностям этого широкого спектра заинтересованных лиц требует подготовки ряда отчетов и принятия различных стратегий их распространения. Стратегии должны идентифицировать потенциальных пользователей (базовые структуры и индивидуумы) и уровень их методической подготовки. Требуется методический отчет (который обеспечивает достаточную информацию для представления исследования). Но методические данные также нуждаются в приведении к формам, которые будут доступны для пользователей без методической подготовки и могут быть представлены в итоговом отчете (например, для общественности) или в более подробном отчете для тактиков, где указывается: есть ли в образовательной системе какая-либо особая группа учащихся, заслуживающая особого внимания; существуют ли корректирующие воздействия; могут ли быть выявлены факторы, ассоциирующиеся с очень высокими результатами выполнения тестов.

Во многих странах разработка политических решений подвержена влиянию политических приоритетов и восприятию министров и высокопоставленных официальных лиц. Она часто строится на личном опыте и искаженной информации, так же как зависит от политического давления. Гораздо реже решения строятся на информации по результатам анализа валидных и надежных данных о функционировании системы образования, которые могут быть обеспечены хорошо спроектированной и правильно функционирующей системой национальной оценки.

Тактики должны взять на себя лидерство в обеспечении гарантий того, что объективные надежные характеристики функционирования системы образования, полученные в ходе национальной оценки, используются для улучшения общего качества политических решений в образовании. Они могут сделать это с помощью проверки и отражения релевантности

результатов национальной оценки разработке политических решений в таких областях, как гендерное и региональное равенство, обеспечение учебными материалами школ, повышение квалификации преподавателей и организация штатных курсов для преподавателей. Тактики могут зафиксировать, вводились ли изменения, влияющие на достижения учащихся по результатам проведения предыдущей национальной оценки. Они могут поощрять и поддерживать организаторов (внештатных и штатных) системы повышения квалификации преподавателей, чтобы анализировать результаты и адаптировать существующую практику функционирования этой системы там, где имеются доказательства необходимости ее совершенствования. Тактики могут также консультировать руководителей по разработке учебных планов относительно изменений в их содержании, когда есть явные признаки того, что учебный материал оказывается слишком легким или, что более вероятно, слишком трудным для учащихся.

Тесная причастность тактиков к проектированию системы оценки в самом начале работ и обсуждению релевантности ее результатов после завершения может гарантировать, что они действительно разделяют представления о ценности системы национальной оценки. На протяжении ряда лет эти воззрения могут давать надежды на то, что тактики будут воспринимать национальную оценку как ключевой инструмент выработки политических решений в образовании.

Краткое описание опыта проведения национальной оценки в девяти странах представлено в приложении А. Описание не является исчерпывающим, а сами примеры опыта не презентуются в качестве совершенных моделей практики по оценке. Некоторые из примеров имеют фактические недостатки по ряду методических аспектов. Однако в действительности они иллюстрируют сходство и различия в подходах к оценке, что представляет несомненный интерес. Общие черты, отраженные в этих примерах, состоят в том, что во всех странах оценка учебных достижений проводится по языку / грамотности и математике / работе с числами на одном или нескольких уровнях начальной школы. Во всех странах процессы оценки были

основаны на выборочных совокупностях учащихся. Кроме этого, в Чили и Уругвае оценка достижений проводилась по генеральной совокупности школ страны.

Различия между странами проявляются в частоте проведения оценки, которая изменяется от одного года до четырех лет. Агентства, ответственные за проведение оценки, также отличаются по составу. Как правило, в них входят представители министерства образования, исследовательского института, деятельность которого поддерживается правительством, и национальной экзаменационной комиссии. Существенная вне-национальная поддержка была предоставлена исполнительному агентству по оценке в нескольких странах. По крайней мере, в двух странах (Чили и Южно-Африканская Республика) исполнительное агентство менялось в процессе выполнения программ оценивания.

Способы описания достижений учащихся также были различными. Они менялись от приведения средних оценок и распределения числа заданий, на которые учащиеся ответили правильно, до представления процента учащихся, чьи результаты выполнения теста достигли «ожидаемых» стандартов или определенного процента, оцениваемого на различных уровнях «мастерства». Методы анализа данных также значительно отличались, что отражало уровень методической подготовки команд по национальной оценке. В некоторых странах использовались сложные аналитические подходы (например, Соединенные Штаты Америки и Вьетнам).

Использование результатов оценивания значительно варьируется от страны к стране, хотя это заключение не является бесспорным, потому что доступно не так много информации, как хотелось бы, о той степени, в которой результаты могут быть распространены по системе образования или эффективно использованы для содействия формированию политики. Так же, как в случае описания гендерных различий, некоторые страны используют результаты национальной оценки для того, чтобы:

- обеспечить политические рекомендации для образования (Шри-Ланка, Вьетнам);



- представить документальное подтверждение региональных различий в достижениях учащихся (Непал, Южная Африка, Шри-Ланка);
- спроектировать основную программу повышения квалификации для преподавателей (Уругвай);
- обеспечить финансовые и другие формы поддержки школ с низкими оценками учебных достижений (Чили);
- включить информацию о сильных и слабых аспектах достижений учащихся в отчеты для преподавателей (Уганда);
- описать изменения достижений учащихся в группах меньшинства с течением времени (Соединенные Штаты Америки);
- предложить снижение усиленного внимания к алгебре и геометрии в учебном плане (Бутан).

Участникам проектирования программ национальной оценки, возможно, было бы интересно рассмотреть несколько необычного практического опыта, который связан с особенностями программ, описанных в приложении А:

- проведение рекламной кампании по осведомлению общественности до начала оценки (Чили);
- сбор дополнительных данных вместе с данными по достижениям учащихся с целью мониторинга степени совершенствования образовательных услуг, предоставляемых школами, с течением времени (Вьетнам);
- предъявление тестов учебных достижений преподавателям наравне с учащимися (Индия, Вьетнам);
- работа в тесном сотрудничестве с объединениями преподавателей для проведения оценки (Уругвай).

В приложении В представлено описание основных особенностей трех главных крупномасштабных международных исследований, которые охватывают весь земной шар. Эти исследования сосредоточены на оценке чтения / грамотности, математики / действий с числами и естественных наук, являющимися тремя областями знаний и умений, которые можно, вероятно, рассматривать как «ядро» в образовании школьников во всех странах. Все три исследования также касаются мониторинга достижений учащихся с учетом фактора времени.

Уровень методической компетентности в международных исследованиях очень высок, и учащиеся могут улучшить свои знания и умения, участвуя в них. Много стран, как мы видели, также используют данные, собранные при международной оценке, для выполнения анализа на национальном уровне, как результат использования данных международного исследования для национальной оценки. Эта процедура может быть обогащена, если основная информация на национальном уровне объединится с той, которая требуется для международной оценки.

Проектирование международных исследований подобно планированию программ национальной оценки за исключением того, что должен быть учтен факт проведения оценивания в разных странах. Таким образом, инструментарий, возможно, не будет в равной степени пригоден для проведения оценивания во всех странах, или потому, что он не везде будет адекватно представлять учебные планы школ, которые изменяются при переходе от страны к стране, или потому, что он не будет адекватно отражать возможные диапазоны достижений учащихся, которые могут существенно отличаться в разных странах.

Существует два подхода для адаптации инструментария применительно к вариации учебных планов школ. В Международном исследовании тенденций в математическом и естественнонаучном образовании TIMSS (В.1 в приложении В), как и в более ранних исследованиях, выполненных под эгидой Международной ассоциации по эвалюации достижений в образовании, тесты разрабатываются с учетом соглашений среди стран-участников, на основе которых общие элементы учебных планов включаются в тест. Подход в программе Международного оценивания учащихся PISA (В.3 в приложении В) основан не на использовании инструментария, разработанного на базе анализа учебных планов, а на обращении к «экспертным» оценкам для определения знаний и умений, которые должны были приобрести пятнадцатилетние учащиеся по окончании обязательного образования для того, чтобы быть полноценными членами общества.

Тот факт, что достижения учащихся связаны с экономическим развитием страны, означает, что программы оценивания (такие как TIMSS и PISA), спроектированные для развитых стран, вряд ли обеспечат удовлетворительное описание достижений в развивающихся странах. В связи с этой проблемой были спроектированы региональные исследования для стран, менее развитых в индустриальном отношении, и результаты трех таких исследований – двух в Африке и одного в Латинской Америке – представлены в приложении С. Эти исследования подходят как для программ национальной, так и международной оценки.



## ПРИЛОЖЕНИЕ А

# ОБРАЗОВАТЕЛЬНЫЕ ИССЛЕДОВАНИЯ В СТРАНАХ

### **А.1. Индия**

*Цель.* Оценка была разработана для получения основных данных о качестве образования в каждом штате с целью предоставления информационной помощи правительству Индии. Оценивание было частью программы правительства, названной в честь Сарва Шикша Абхиян (Sarva Shiksha Abhiyan – SSA), которая была направлена на достижение всеобщего начального образования в стране к 2010. Ранее программы крупномасштабной оценки достижений выполнялись в определенных школьных округах как часть правительственного проекта по начальному образованию в отдельных районах страны (Prakash, Gautam and Bansal, 2000). В рамках этих проектов средние оценки по математике и языку сравнивались по районам, предметным областям и классам обучения. Оценка показала, что учащиеся имели лучшие результаты по языку и что средние достижения по выборке учащихся старшего возраста не были столь же внушительными, как в группе учащихся более младшего возраста. Большинство различий в районах между результатами по математике и языку для мальчиков и девочек не были статистически значимыми. В дополнение к этой

оценке на уровне районов были выполнены программы крупномасштабной оценки в 22 штатах страны в начале 1990-х годов (Shukla and others, 1994).

**Частота оценивания.** Каждые 3 года.

**Классы.** Оценка учащихся 5-х классов была проведена в 2001–2002 годах. Также оценивались учащиеся 3-х классов и последних классов начальной школы (число лет обучения в которых изменяется от штата к штату).

**Оцениваемые достижения.** Учебные достижения по языку и математике.

**Кто выполнял оценку?** Национальный совет по исследованиям и обучению (Дели) при поддержке районных институтов образования, которые контролировали сбор данных.

**Выборка или генеральная совокупность.** Выборка.

**Анализ.** Отчет об оценках учащихся 5-х классов каждого штата, представленный в виде процентов правильных ответов учащихся на задания теста.

**Использование результатов.** Результаты 5-х классов показали незначительные отличия учебных достижений групп учащихся, сформированных по гендерному признаку и по принадлежности к городским и сельским школам. Эти данные будут использоваться для мониторинга изменений уровней образовательных достижений и идентификации тех факторов внутри и вне образовательной среды, которые могут помочь принять во внимание различия в достижениях учащихся.

**Темы, представляющие интерес для обсуждения.** При более ранней крупномасштабной оценке один и тот же тест был предъявлен учащимся и преподавателям в 22 штатах страны. В одном штате с очень низкими средними значениями оценок учащихся только 1 из 70 преподавателей выполнил тест, ответив на все 40 заданий по арифметике правильно. Из числа протестированных преподавателей 10% правильно выполнили менее половины заданий теста (Shukla and others, 1994).

Национальная оценка будет использоваться для оказания помощи по проведению мониторинга эффектов, полученных за счет развития инициатив программы SSA. В отличие от большинства других программ, оценки учащихся приводятся в форме общего процента заданий, которые они выполнили верно. Штатам с особенно низкими уровнями достижений предполагается уделить особое внимание. Некоторые штаты с устойчивыми образовательными традициями с точки зрения рейтингов участвующих в исследовании школ (например, Керала и Химачал-Прадеш) получили относительно низкие средние оценки при тестировании 5-х классов, в то время как некоторые из штатов с относительно низким рейтингом участвующих в исследовании школ (например, Бихар, Орисса и Западная Бенгалия) показали более высокие результаты. Эти данные, которые также ранее сообщались в программах оценки учащихся 22 штатов, объясняются тем фактом, что в штатах с более низкими результатами выборки учащихся, выполнявших тесты, стремились остаться в системе образования, а многие из учащихся с меньшими возможностями, учитывая их способности и семью, перестали посещать школу к 5-му классу.

*Источник:* Индия, Национальный Совет по исследованиям в образовании и обучению, Департамент образовательных измерений и эвалюации, 2003.

## **А.2. Вьетнам**

*Цель.* Измерить качество образования с акцентом на достижениях учащихся начальных классов.

*Частота.* Предварительные программы маломасштабной оценки проводились с 1998 по 2000 год в 3-х и 5-х классах, но их результаты оказались неподходящими для обеспечения реперных точек отсчета при проведении мониторинга тенденций в образовании с учетом фактора времени.

*Классы.* Учащиеся 5-го класса.

**Оцениваемые достижения.** Чтение на вьетнамском языке и математика в 2001 году.

**Инструментарий.** Тесты учебных достижений, анкеты для учащихся, преподавателей и школ.

**Кто выполняет?** Министерство образования и обучения при поддержке других национальных агентств и международная команда при поддержке Всемирного банка и Департамента международного развития Объединенного Королевства Великобритании и Северной Ирландии.

**Выборка или генеральная совокупность.** Дизайн выборки был выполнен таким образом, чтобы обеспечить репрезентативность национальной генеральной совокупности и генеральных совокупностей для каждой из 61 провинции.

**Анализ.** Анализ включал перекрестное табулирование данных по учебным достижениям и школьным данным по регионам, корреляционный анализ данных по учебным достижениям, факторный анализ данных, применение современной теории тестов к результатам по тестовым заданиям и иерархическое линейное моделирование для идентификации факторов, связанных с учебными достижениями.

**Использование результатов.** Официальные представители правительства разработали 40 политических рекомендаций в образовании, которые были основаны на итоговых результатах оценки достижений.

**Темы, представляющие интерес для обсуждения.** Тесты включали задания Международной ассоциации по эвалюации образовательных достижений для исследования грамотности чтения 1991 года (Elley, 1992, 1994), которые использовались для сравнения результатов вьетнамских учащихся с результатами учащихся других стран. Преподавателям и обучаемым были предъявлены одни и те же тесты. По итогам тестирования 12% учащихся получили более высокие оценки, чем 30% преподавателей. Менее 3% школ обладали обязательными школьными ресурсами (например, библиотекой, водопроводом).

Более 80% учащихся занимались в классных комнатах с минимальными ресурсами (доска, мел и т.д.), в то время как 10% обучались у преподавателей, которые сами не окончили среднюю школу.

Для результатов выполнения теста по чтению были установлены шесть уровней мастерства освоения грамотности чтения:

- *Уровень 1.* Учащийся устанавливает соответствие слов или соединяет их в предложения при помощи иллюстраций. В своих действиях он ограничен словарным запасом, связанным с иллюстрациями.
- *Уровень 2.* Учащийся определяет местонахождение текста, выраженного в коротких повторных предложениях, и может работать с текстом без помощи иллюстраций. Текст ограничен короткими предложениями и фразами с повторяющимися наборами слов.
- *Уровень 3.* Учащийся читает и понимает более длинные отрывки. Он может определить из текста предыдущую или последующую информацию. Понимает перефразирование. Обладает расширенным словарным запасом и в состоянии понять смысл предложений с достаточно сложной структурой.
- *Уровень 4.* Учащийся в состоянии связать информацию из различных частей текста. Отбирает и соединяет текст для получения вывода и различных возможных трактовок.
- *Уровень 5.* Учащийся устанавливает связи и определяет замысел автора на основе информации, представленной разными способами в различных видах текста и в документах, где информация не приведена напрямую.
- *Уровень 6.* Учащийся объединяет текст с внешней информацией для вывода разнообразных смысловых трактовок, для понимания скрытого смысла. Идентифицирует цели автора, отношения, ценности, верования, побуждения, неустановленные предположения и аргументы.

При выполнении программы наблюдалась значительная вариация уровней выполнения тестов учащимися как по чтению, так и по математике. Например, значительно меньше учащихся достигло двух высших уровней чтения в провинциях Хазянг



и Тьензянг, чем в провинции Дананг (табл. 2.1). Были проанализированы связи между характеристиками преподавателя и оценками учащихся после того, как были приняты во внимание основные характеристики домашнего окружения (табл. 2.2).

*Источник:* Всемирный банк, 2004.

**ТАБЛИЦА А.2.1**

**Проценты и стандартные ошибки оценок учащихся для разных уровней умений по чтению**

Провинция	Названия показателя	Уровень 1	Уровень 2	Уровень 3	Уровень 4	Уровень 5	Уровень 6
Хазянг	Процент	7,5	22,1	27,4	18,7	18,5	5,7
	Стандартная ошибка	1,66	3,23	3,06	2,97	3,07	2,09
Тьензянг	Процент	2,8	13,4	28,8	20,2	22,4	12,5
	Стандартная ошибка	0,7	2	2,49	1,8	2,46	2,78
Дананг	Процент	0,8	5,7	15,4	21,3	32,9	24,1
	Стандартная ошибка	0,34	0,88	1,79	1,89	1,98	3,23
Вьетнам	Процент	4,6	14,4	23,1	20,2	24,5	13,1
	Стандартная ошибка	0,17	0,28	0,34	0,27	0,39	0,41

*Источник:* Всемирный банк, 2004, т. 2, табл. 2.3.

**ТАБЛИЦА А.2.2**

**Связь между wybranными переменными преподавателей и достижениями учащихся по математике**

Переменная преподавателя	Простая корреляция	Частная корреляция после принятия во внимание основных характеристик домашнего окружения учащегося
Пол преподавателя	0,17	0,14
Академическое образование	0,08	0,04
Знания по математике	0,29	0,25
Классифицирован как «превосходный преподаватель»	0,18	0,13
Ресурсное обеспечение класса	0,24	0,15
Число часов на подготовку и оценку	0	0,01
Частота встреч с родителями	0,05	0,04
Число инспекционных проверок	0,13	0,11

*Источник:* Всемирный банк, 2004, т. 2, табл. 4.38.

*Примечание:* корреляции более 0,02 являются статистически значимыми.

### **А.3. Уругвай**

**Цель.** Национальная оценка предназначалась для определения: степени, в которой выпускники начальной школы имели развитое «фундаментальное понимание» по языку и математике; социокультурных факторов, которые могли оказывать воздействие на достижения учащихся. Программа оценки придавала особое значение характеристикам профессионального развития преподавателя, которые включали диагностику учебных проблем, предоставление преподавателю информации о результатах выполнения теста учащимися и помощь преподавателю для совершенствования процессов обучения и эвалюации. Программа также преследовала цель улучшения условий в школах на основе использования результатов тестов и анкет.

**Частота и классы.** В программе участвовали учащиеся 6-х классов (каждые 3 года) в 1996, 1999, 2002 и 2005 годах. Помимо этого в 2001 году проводилось оценивание учащихся 1-х, 2-х и 3-х классов с целью анализа развития квалификации преподавателей. В 1999 году были протестированы учащиеся 9-х классов, а в 2003 году – 12-х классов. С 2003 года проводится оценка пятнадцатилетних учащихся, которая является частью Программы международного оценивания учащихся (PISA).

**Оцениваемые достижения.** Достижения по математике (решение задач) и пониманию прочитанного в 6-х классах, достижения по математике, языку, естественным и социальным наукам в 9-х и 12-х классах.

**Инструментарий.** Тесты учебных достижений, анкеты для родителей, преподавателей и директоров школ.

**Кто выполняет?** Первоначально подразделение «Unidad de Medicion de Resultados Educativos (UMRE)», созданное как часть финансируемого Всемирным банком проекта, несло ответственность за проведение оценки учащихся 6-х классов, в то время как программа «Programa de Modernizacion de la Educacion Secundaria y Formacion Docente (MESyFOD)» – проект, финансируемый Межамериканским банком, – отвечала за национальную оценку учащихся на уровне средней школы.

С 2001 года оценочная деятельность была объединена и институционализирована под эгидой структуры «Gerencia de Investigacion y Evaluacion» (Отдел по исследованиям и оценке), являющейся частью Национального управления общественного образования. Финансирование проектов обеспечивалось международными агентствами.

**Выборка или генеральная совокупность.** В исследованиях участвовали: генеральная совокупность и выборка учащихся 6-х классов, за исключением учащихся очень маленьких сельских школ, генеральная совокупность учащихся 9-х классов, выборка учащихся 1-х, 2-х, 3-х и 12-х классов и выборка учащихся для программы PISA.

**Анализ.** В программе оценивания подразделения UMRE в качестве критерия адекватности результатов тестирования учащихся использовался пороговый балл на уровне 60% правильных ответов по тесту. Индивидуальные школьные оценки сравнивались со средними национальными оценками, средними оценками по департаменту или региону, а также по школам, обслуживающим учащихся со сходными социально-экономическими условиями проживания. Данные по тестам учебных достижений учащихся связывались с основными факторами.

**Использование результатов.** Результаты оценки использовались, главным образом, преподавателями, директорами школ и школьными инспекторами. Правительство страны использовало результаты для определения школ, которые нуждались в специальной поддержке, и для крупномасштабных программ повышения квалификации и подготовки преподавателей. Результаты оценки на национальном уровне были представлены широкой общественности. Через сорок дней после тестирования и перед концом учебного года участвующие в оценке школы получали конфиденциальные сообщения со средними результатами, представляющими задание за заданием. Сообщения не включали в себя индивидуальные результаты учащихся или результаты по отдельным классам. Подразделение UMRE подготовило руководства для преподавателей, чтобы помочь им

обратиться к выявленным пробелам по языку и математике и организовать внутриштатные программы повышения квалификации преподавателей в школах, находящихся в неблагополучных районах проживания, подготовило отчеты для контролирующего персонала и провело семинары для инспекторов, которые рассматривали результаты тестирования. Часть тестов, кроме тех, которые были предъявлены выборкам учащихся, были предоставлены школам. Каждая школа получила отчет о средних национальных результатах по каждой проверенной компетентности. Школам, не участвовавшим в тестировании, были разосланы нормы для целей сравнения. Приблизительно 80% этих школ провели тестирование и сравнили свои результаты с представленными национальными нормами. Инспекторы проводили свои собственные семинары, способствующие развитию углубленного понимания результатов оценки, выявлению эффекта влияния социального неблагополучия семьи учащегося на результаты его учебной деятельности и выдвигению планов действий для повышения качества образования.

*Темы, представляющие интерес для обсуждения.* Первоначально профсоюз учителей начальных классов был настроен решительно против проведения национальной оценки. В частности, он выступал против публикации результатов отдельных школ. В конечном счете, профсоюз добился согласия правительства на то, чтобы отменить публикацию результатов оценки по отдельным школам или преподавателям, но разрешить использовать результаты тестирования в диагностических целях. Публикации подлежали только совокупные данные по школам. Кроме того, правительство пригласило преподавателей участвовать в группах, которые планировали оценку, и в других консультативных группах. Преподаватели также были в значительной степени привлечены к разработке тестов. К этому времени среди общественности наблюдалась лишь незначительная оппозиция формальной оценке начального обучения в рамках обсуждаемого подхода. Значительную поддержку получило решение о том, что на преподавателей или школы не будет наложено никаких

штрафных санкций за низкие результаты тестирования. Профсоюз преподавателей средней школы также не был расположен к оценке и занял выжидательную позицию. Принятие преподавателями инициатив подразделения UMRE и результатов оценки связано с конфиденциальностью результатов тестирования, скоростью подготовки отчетов, рассмотрением тестовых оценок в контексте социокультурного фона и признанием того, что результаты учащихся зависят от комбинации факторов (включая характеристики домашнего уклада, школы, школьного сообщества и преподавателей).

Хотя правительства в некоторых странах ищут пути для поддержки той точки зрения, что школы и преподаватели ответственны за результаты учащихся, Уругвай выбрал другой подход. Государство берет на себя ответственность за обеспечение возможностей окружающей среды для содействия в достижении равенства внутри системы образования.

*Источник:* Beneviste, 2000, Ravela, 2005.

#### **А.4. Южная Африка**

*Цель.* Южная Африка провела серию программ по национальной оценке в 3-х, 6-х и 9-х классах. Она также участвовала в трех международных исследованиях для обеспечения основных данных, по сравнению с которыми можно будет проводить мониторинг будущего прогресса в образовании и для получения сравнительных данных по учебным планам и достижениям учащихся по математике и естественным наукам в Южной Африке и в индустриальных странах. Каждое из международных исследований можно считать национальной оценкой образовательных достижений. Участие в программах международной оценки обеспечило возможности для развития системы образования.

Южная Африка была единственным африканским участником TIMSS в 1995 году, она принимала участие в TIMSS в 1999 году вместе с Марокко и Тунисом, а в том же составе, к которому присоединились Ботсвана, Гана и Египет, вошла в программу TIMSS

в 2003. Южная Африка также участвовала в Южном и Восточном африканском консорциуме для осуществления программы «Мониторинг качества образования» на основе оценки учащихся 6-х классов в 2000 году и в программе «Мониторинг учебных достижений» на основе оценки учащихся 4-х классов, которая началась в 1992 году.

**Частота.** Программа TIMSS 1995, 1999 и 2003 годов.

**Классы.** Учащиеся 8-х классов.

**Инструментарий.** Тесты учебных достижений; анкеты для учащихся, преподавателей и директоров школ.

**Оцениваемые достижения.** Учебные достижения по математике и естественным наукам.

**Кто выполняет?** Научный совет гуманитарных наук в 1995 и 1999 годах и Университет Претории в 2003 году.

**Выборка или генеральная совокупность.** Выборка. Один 8-й класс был выбран целиком в каждой отобранной школе.

**Анализ.** Исследование сравнивало результаты тестирования по математике и естественным наукам с показателями других стран, представленными в виде средних результатов выполнения тестов в каждой стране и результатов в шкале процентилей на 5-м, 25-м, 50-м, 75-м и 95-м процентилях. В рамках исследования также проводилось сравнение Южной Африки с другими участвующими странами по группе показателей, охватывающих характеристики фона обучения, отношения учащихся к обучению, учебных планов, характеристики преподавателей, классов и школьного контекста для обучающей и учебной деятельности. Анализ также включал сравнение средних оценок результатов тестирования с учетом фактора времени.

**Использование результатов.** Результаты TIMSS использовались в парламентских дебатах.

**Темы, представляющие интерес для обсуждения.** У Южной Африки есть 11 официальных языков. Некоторые слова должны были быть переведены на южноафриканский английский язык, в свя-

зи с чем определенный контекст их трактовок претерпевает неизбежные изменения. Также значительное количество времени в исследовании было отведено на решение логистических проблем, которые являются неизбежными составляющими, связанными с неполным соответствием сервисных услуг требованиям исследования, хотя бесперебойная работа таких средств связи, как почта и телефон, вполне может быть гарантирована в индустриальных странах. Национальная исследовательская команда сочла указанные сроки выполнения работ, предложенные в программе TIMSS, крайне трудными для соблюдения. Первоначальный этап работ, связанный с формированием выборки, выявил около 4000 школ, которые не числились в национальной базе данных. Возможности связи между командами, выполняющими три оценки в программе TIMSS, для передачи информации по оцениваемым умениям были ограничены. Только один из сотрудников-участников первой команды в оценке по программе TIMSS участвовал в команде оценки TIMSS 2003 года. Большинство учащихся выполняли тест, написанный на языке, который не был языком их общения в семье.

Второе исследование TIMSS использовалось для подробного анализа результатов оценивания внутри страны (Howie, 2002). Результаты включали следующие выводы.

- Официальные статистические данные по размерам классов отличались (значительно) от данных, полученных при формировании национальной репрезентативной выборки участвующих школ, которые давали неточные данные о зачислении в школы.
- Некоторые учащиеся боялись, что результаты их тестирования будут учитываться как официальные школьные результаты. Некоторые учащиеся боялись обратиться за помощью. Многие учащиеся испытывали трудности при выполнении открытых заданий. Опоздания, прогулы и обман во время процедуры предъявления тестов вызвали дополнительные проблемы.
- Многие учащиеся испытывали трудности по своевременному выполнению тестов и анкет из-за проблем с языком,

на котором они были представлены. Многие преподаватели не владели бегло языком в той степени, которая позволяла им эффективно общаться с учащимися.

- Преподаватели потратили значительное время на учебный материал, который должен был быть изучен в предыдущих классах.
- Почти четверть преподавателей, работающих в 8-х классах, не обладали нужной квалификацией для преподавания математики и не имели высшего специального образования.
- Учащиеся, чей язык семейного общения был или английским, или африкаанс, показали более высокие результаты, чем учащиеся, которые говорили дома на других африканских языках.
- Менее 0,5% учащихся страны достигли высшего уровня по математике в результатах выполнения теста по сравнению с 10% для международной выборки. Средний балл (381) для самых высоких результатов учащихся в девяти провинциях (Западный Мыс) был значительно ниже, чем средний балл по результатам международной оценки в программе TIMSS (487).
- Ни размер школы, ни размер класса не являлись значимыми предикторами учебных достижений по математике.

Чтобы получить основные данные для будущей оценки и предложить стратегические инициативы, была выполнена национальная оценка в 3-х, 6-х и 9-х классах, требуемая Департаментом образования. Каждая из этих программ оценки использовала данные анкет, так же как и результаты по тестам учебных достижений, чтобы обеспечить информационную базу для эвалюации долгосрочных действий по повышению доступности, качества, эффективности и равенства образования. Сравнения между провинциями показали значительные региональные различия в достижениях учащихся. Общие уровни выполнения тестов оказались низкими. Например, в среднем проценты правильных ответов учащихся были такими же небольшими, как оценка в 38% для теста по языку, 27% для теста по математике и 41% для теста по естественным наукам в 6-х классах.



Отдельные отчеты о результатах учащихся в 6-х классах были подготовлены для каждой провинции, а на их основе был сформирован единый национальный доклад.

*Источник:* Howie, 2000, 2002; Kanjee, 2006; Reddy, 2005, 2006.

#### **А.5. Шри-Ланка**

*Цель.* Оценить достижения учащихся, которые закончили 4-й класс в 2003 году.

*Частота.* Предыдущая оценка была проведена для учащихся 3-х классов (1996) и 5-х классов (1994, 1999). Дальнейшие программы оценки были выполнены для учащихся 4-х классов (2007), а также 8-х и 10-х классов (2005).

*Классы.* Учащиеся 4-го класса.

*Оцениваемые достижения.* Оцениваются учебные достижения по первому языку (сингальский или тамильский язык), математике и английскому языку.

*Инструментарий.* Тесты учебных достижений; анкеты, предъявляемые директорам школ, руководителям подразделений, преподавателям и родителям (см. табл. А.5.1).

*Кто выполняет?* Национальный центр исследований в образовании и эвалюации при факультете образования в Университете Коломбо.

*Выборка или генеральная совокупность.* Выборка, обеспечивающая репрезентативность национальной генеральной совокупности учащихся 4-х классов и генеральной совокупности самих 4-х классов в каждой из девяти провинций.

*Анализ.* Сравнения оценок достижений по видам школ, их местоположению, гендерным признакам и уровню педагогического образования. Провинции и районы были проранжированы по результатам учащихся в каждой предметной области. Связи между школой, семейным фоном и факторами учащихся, с одной стороны, и достижениями учащихся, с другой стороны, были выявлены путем проведения анализа.

ТАБЛИЦА А.5.1

Контекстные данные и их источники в программе национального оценивания в Шри-Ланке

Вид информации	Анкета	Раздел	Количество заданий
Основные характеристики школы	Директор	<ul style="list-style-type: none"> <li>• Главная характеристика</li> <li>• Профиль преподавателя</li> <li>• Обустройство школы</li> <li>• Финансовый статус</li> <li>• Мнения</li> </ul>	37
	Руководитель подразделения	<ul style="list-style-type: none"> <li>• Главная характеристика</li> <li>• Обустройство школы</li> <li>• Оценивание процедур обучения</li> </ul>	13
	Классный руководитель	<ul style="list-style-type: none"> <li>• Главная характеристика</li> <li>• Академическая и профессиональная информация</li> <li>• Детальные характеристики класса</li> <li>• Мнения</li> </ul>	41
Дом	Родители	<ul style="list-style-type: none"> <li>• Главная характеристика</li> <li>• Обстановка в доме</li> <li>• Социальноэкономический статус семьи</li> <li>• Поддержка учащегося в семье</li> <li>• Мнения</li> </ul>	51
	Учащиеся	<ul style="list-style-type: none"> <li>• Главная характеристика</li> <li>• Дошкольное образование</li> <li>• Деятельность после школы</li> <li>• Мнения</li> </ul>	26

Источник: Perera and others, 2004, табл. 3.7.

**Использование результатов.** Результаты использовались для анализа сектора образования, чтобы помочь развитию новых стратегий для правительства и спонсоров, поддерживающих образование. В настоящее время они используются с целью установления точек отсчета, по сравнению с которыми будет проводиться мониторинг уровней учебных достижений учащихся в каждой из провинций.

**Темы, представляющие интерес для обсуждения.** Шриланкийская национальная команда по оцениванию выбрала оценку в 80% в качестве порогового балла для определения уровня «мастер» (это определение было, очевидно, основано на уровне, используемом Организацией Объединенных Наций по вопросам обра-

зования, науки и культуры в более раннем исследовании Мониторинга образовательных достижений (ЮНЕСКО, 1990). Проценты учащихся рассматривались как достигшие «мастерства» по каждой из трех проверенных предметных областей. Результаты показали, что этот ожидаемый стандарт (стандарт выполнения теста – критериальный балл (пороговый балл)), делящий группу испытуемых на достигших и не достигших требуемого уровня образовательных достижений (мастерства. – *Примеч. ред.*) был установлен на нереалистично высоком уровне. На основании средних оценок в отчете было сделано заключение о том, что результаты выполнения тестов по первому языку «кажутся удовлетворительно соответствующими стандарту» (Perera and others, 2004, 47), в то время как эти результаты, оцененные по критерию уровня мастерства, показали совсем другую картину. Менее 40% учащихся достигли порогового балла по родному языку и математике и менее 10% – по английскому языку. Результаты показали значительные различия в достижениях среди провинций и районов (табл. 5.2). Были выявлены подгруппы с низкими уровнями достижений. Для каждой из девяти провинций страны были опубликованы отдельные отчеты.

*Источник:* Perera and others, 2004.

**ТАБЛИЦА А.5.2**

**Процент учащихся, достигших уровня мастерства по первому языку, по провинциям**

Группа	Оценка	Провинция	Процент учащихся, достигших уровня мастерства	Целевой процент
Свыше 50%	1	Западная	53,5	80
26–50%	2	Южная	42,6	80
	3	Северо-Западная	42,2	80
	4	Сабарагамува	40,2	80
	5	Север центра	35,6	80
	6	Ува	33,9	80
	7	Центральная	33,8	80
1–25%	8	Восточная	23,7	80
	9	Северная	22,7	80

*Источник:* Perera and others, 2004, табл. 4. 14.

### **А.6. Непал**

**Цель.** Национальная оценка 2001 года была выполнена, чтобы определить, в какой степени изменились достижения учащихся за четырехлетний период серьезных реформ в образовательной политике.

**Частота.** Основные данные были получены по учащимся 3-х классов в 1997 году. (В 1999 году была проведена оценка 5-х классов.)

**Класс.** Учащиеся 3-го класса.

**Оцениваемые достижения.** Математика, непальский язык и общественные науки.

**Инструментарий.** Тесты учебных достижений по математике, непальскому языку и общественным наукам, которые были предъявлены всем выборкам учащихся. Анкеты были предъявлены директорам школ и преподавателям по этим трем предметным областям в каждой выбранной школе. Были проинтервьюированы 25% учащихся и их родителей.

**Кто выполняет?** Сервисный центр в образовании и развитии.

**Выборка или генеральная совокупность.** Выборка из 171 школы.

**Анализ.** Результаты выше 75% правильных ответов были отнесены к категории «удовлетворительных» результатов выполнения теста. Другой вид анализа включал исследование надежности каждого теста и сравнение средних оценок выборок учащихся для 1997 и 2001 годов. Для сравнения средних оценок результатов выполнения тестов учащимися по регионам выполнялся дисперсионный анализ, а также с целью идентификации факторов, связанных с достижениями учащихся, проводился множественный регрессионный анализ.

**Использование результатов.** Результаты использовались для мониторинга учебных достижений в период с 1997 до 2001 года, в частности для эвалюации эффекта политических изменений в образовании, которые включали увеличение бюджетных поступлений, переход к новым учебным планам, использова-

ние новых учебников и обучающих материалов и открытие новых центров для обучения преподавателей. Были определены регионы с самыми высокими результатами выполнения тестов. В 2001 году различия между средними оценками мальчиков и девочек были значимыми только по математике, где мальчики имели более высокие средние оценки. Все средние оценки по общественным наукам были более значимо выше в 2001 году, чем в 1997 году.

*Темы, представляющие интерес для обсуждения.* Полученные данные помогли идентифицировать содержательные области учебного плана, где у учащихся встречались некоторые трудности. По математике учащиеся в целом могли дать словесные аналогии для чисел и совершить обратный переход от слов к числам, они знали меры времени и веса, складывали числа вплоть до четырехзначных и могли просуммировать десятичные дроби. Тем не менее просматривалась тенденция, когда учащиеся не могли выполнить задачу, данную в словесной форме и включающую использование любого из четырех основных арифметических действий (сложение, вычитание, умножение, деление). В результатах по непальскому языку просматривалась тенденция, когда средний учащийся был в состоянии прочитать простую историю и использовать некоторый лексический запас, но он не мог читать и правильно отвечать на вопросы, основанные на отрывках текста, или на вопросы, которые задавались к истории в картинках.

Результаты оценивания показали, что многие из реформ в образовании имели незначительный эффект. Более чем 60% преподавателей указали на то, что их классы никогда не проверялись инспекторами. Как правило, в своей работе они получали относительно небольшую материально-техническую поддержку. Приблизительно одна треть преподавателей были не обучены. Обучение в классе было расценено как неэффективное.

В отчете был сделан вывод о том, что хотя многие из образовательных реформ несомненно имели место, но, вероятно, на момент оценки было слишком рано ожидать явных улучшений в достижениях учащихся. Отчет по национальной оценке

также подчеркивал относительно низкое качество поддержки образования, имеющейся в семьях обучающихся. Более четверти матерей были классифицированы как неграмотные, а менее 7% имели законченное образование на уровне 5-го класса.

*Источник:* Khaniya and Williams, 2004.

#### **А.7. Чили**

*Цель.* Чилийская система de Medicion de la Calidad de la Educacion (SIMCE) первоначально предназначалась для помощи родителям в выборе школы. В настоящее время она выполняет следующие функции: обеспечение обратной связи для информирования о степени, в которой учащиеся достигают целей обучения, рассматриваемых как минимально приемлемые министерством образования страны; обеспечение обратной связи с родителями, преподавателями и властями на муниципальном, региональном и центральном уровнях; обеспечение данных для тактиков, которые дают основания для принятия решений о распределении ресурсов учебников, развитии учебного плана и штатной подготовки преподавателей, особенно в наиболее нуждающихся в помощи областях. Система также предназначается для повышения качества образования путем установления процедур, которые подчеркивают важность эвалюации, информирования и стимулирования в образовании. Она также позволяет подчеркнуть обязательства министерства образования по улучшению как качества, так и доступности системы образования.

Чили также развивает самостоятельную, но связанную с другими подходами оценочную систему SNED (Национальная система оценки результатов деятельности преподавателей в общественно поддерживаемых школах), обеспечивая стимулы для преподавателей и школ в работе по повышению уровней достижений учащихся.

*Частота.* Ежегодно.

*Классы.* Учащиеся 4-х и 8-х классов.

**Оцениваемые достижения.** Испанский язык (чтение и письмо), математика, естественные науки и общественные науки.

**Инструментарий.** Тесты для оценки перцепции, самооценки и итоговых учебных достижений учащихся. Анкеты, которые выполнялись директорами школ, преподавателями и родителями (только один год).

**Кто выполняет?** Первоначально в 1978 году оценку осуществляло внешнее агентство, Pontificia Universidad Catolica de Chile, теперь же она проводится в системе SIMCE с помощью министерства образования.

**Выборка или генеральная совокупность.** Все (или фактически все) учащиеся в соответствующих классах оцениваются по испанскому языку и математике. Тесты по естествознанию, истории и географии предъявляются 10% учащихся. Очень маленькие школы с недоступным местоположением исключаются из тестирования.

**Анализ.** Школы получают свои ранги по сравнению с рангами других школ в той же самой социально-экономической категории, что и ранги по национальной выборке. В системе SIMCE устанавливаются 900 школ, которые имеют оценки ниже 10% выполнения в тестах по математике и языку внутри своих провинций и для которых будут обеспечены специальные ресурсы (программа P-900).

**Использование результатов.** Результаты системы SIMCE в значительной степени используются в политических дискуссиях. Система SIMCE сообщает результаты классов школ в форме, содержащей средний процент правильных ответов по каждому оцениваемому объекту, так же как среднее число правильных ответов по всему тесту. В начале учебного года результаты SIMCE сообщаются на общенациональном уровне, а также по школам, школьным районам и регионам. Руководства по системе SIMCE объясняют результаты, а также то, как преподаватели и школы могли бы использовать эти результаты для повышения качества достижений учащихся. По программе P-900 школы получают поддержку в форме различных мер по улуч-

шению инфраструктуры, учебников и классных библиотек, обучающих материалов внутришкольных семинаров. Школы исключаются из программы P-900, когда средние оценки учащихся в системе SIMCE начинают превышать 10-процентный пороговый балл.

Программа SNED использует оценки системы SIMCE наряду с четырьмя другими мерами качества школьного образования. Преподаватели в школах, лучшие по результатам тестирования внутри региона, получают денежное вознаграждение, примерно эквивалентное ежемесячной зарплате. В целях гарантии равенства министерство отбирает школы, работающие с учащимися из подобных социально-экономических групп, которые классифицируются с точки зрения городского или сельского местоположения и уровня начального или среднего образования. Хотя целый ряд факторов принимается во внимание при вычислении показателя при ранжировании школ, вес школьных достижений в нем эквивалентен почти двум третям его полной величины (табл. 7.1). Система взвешивания регулярно модифицируется для отражения политических приоритетов текущего периода.

*Темы, представляющие интерес для обсуждения.* Система SIMCE использует интенсивную кампанию связей с общественностью, что включает выпуск брошюр для родителей и школ, плакатов

**Таблица А.7.1.**

**Показатель награждения за заслуги для школ в Чили, 1998–1999 годы**

Фактор	Процент
Эффективность (оценки по математике и естественным наукам в системе SIMCE)	37
Прибавляемая величина (средняя достигнутая оценка по SIMCE)	28
Инициативность преподавателей	6
Улучшение условий обучения	2
Равенство возможностей учащихся	22
Сотрудничество родителей и преподавателей	5

Источник: Delannoy, 2000, табл. 1.5.



для школ, видео для семинаров, телевизионных программ и пресс-релизов. Отчеты по результатам оценки распространяются среди директоров школ, муниципальных руководителей, школьных инспекторов и официальных представителей министерства. Родители также получают индивидуализированный отчет по своей школе. Газеты публикуют результаты по каждой школе. Поскольку муниципалитеты получают финансирование от центрального правительства по числу учащихся, у них есть личная заинтересованность в высоких результатах: хорошие результаты по данным SIMCE имеют тенденцию привлекать больше учащихся и, следовательно, способствуют увеличению финансирования таких школ.

Школы, где большое число учащихся в день тестирования отсутствовали, не получают результатов. Некоторые школы переоценили степень бедности семей своих учащихся, чтобы увеличить возможности на получение помощи по программе P-900. Преподаватели склонны проявлять больше внимания к рангу своей школы относительно подобных школ, а не возможности использовать внутришкольный диалог для диагностики тех предметов и содержательных разделов, в которых у учащихся есть трудности в обучении. Некоторые преподаватели критически относятся к чрезмерно методическому характеру отчетов по школам. Система SIMCE уделяет относительно небольшое внимание данным, полученным в ходе анкетирования учащихся, родителей и преподавателей. Отношение к учебе и ценностные ориентации учащихся трудно измерить. Программа SNED предполагает, что финансовые стимулы будут вдохновлять преподавателей для приложения больших усилий к повышению качества результатов обучения учащихся.

*Источник:* Arregui and McLauchlan, 2005; Benveniste, 2000; Himmel, 1996, 1997; McMeekin, 2000; Olivares, 1996; Wolff, 1998.

#### **А.8. Соединенные Штаты Америки**

*Цель.* Программа «Национальная оценка образовательного прогресса (NAEP)», которая была начата в 1969 году, измеряет

образовательные достижения учащихся и осуществляет мониторинг изменений достижений по возрастным группам и классам. Программа NAEP, часто называемая «Национальной отчетной картой», также проверяет достижения субпопуляций учащихся на основе демографических характеристик и определенных базовых условий. Выборка в большинстве штатов в рамках программы NAEP достаточно большая, чтобы делать выводы о достижениях в отдельных штатах страны.

**Частота.** Оценка выполняется, по крайней мере, один раз в два года по математике и чтению и реже – по другим предметам учебного плана.

**Классы.** 4-е, 8-е и 12-е. Раздельные оценки на уровне каждого штата, использующие тесты NAEP, ограничены 4-ми и 8-ми классами.

**Оцениваемые достижения.** Математика, чтение, естественные науки, письмо, искусство, основы гражданственности, экономика, география и история Америки. Новые предметные области: иностранный язык и всемирная история.

**Инструментарий.** Тесты учебных достижений по чтению, математике, естественным наукам, письму, истории Америки, основам гражданственности, экономике, географии и искусству. Анкеты для учащихся (выполняемые на добровольной основе) в конце тестового буклета собирают информацию относительно демографических характеристик учащихся, событий в классе и поддержки образовательного процесса. Анкеты для преподавателей сосредоточены на фоновых характеристиках обучения, самом обучении и учебных методах. Под фоновыми характеристиками (данными) или переменными понимаются те аспекты субъектов оценочного процесса, которые могут влиять на другие характеристики или переменные в образовательной системе, но сами не испытывают никакого влияния со стороны последних. Например, демографические характеристики: возраст, этническая принадлежность, семейный уклад, образование родителей учащегося и т.д. Как правило, в английском языке для фоновых характеристик используются термины: background variables или subject, classification,

individual-difference variable. – *Примеч. ред.*). Школьные анкеты нацелены на получение информации о школьной политике и характеристиках школы. Фоновые данные по учащимся с ограниченными возможностями или с ограниченным знанием английского языка предоставляет преподаватель.

**Кто выполняет?** Управление по национальной оценке, назначенное представителем министерства образования, который несет общую ответственность за выполнение программы NAEP. Управление состоит из губернаторов, членов законодательного собрания штатов, чиновников на уровне района и штата, педагогов, представителей бизнеса и общественности. Различные агентства работают по контрактам для выполнения различных направлений деятельности в рамках программы NAEP. За период с 2003 по 2006 год отдельные агентства несли ответственность за выполнение каждого из следующих направлений деятельности: разработка тестов, анализ заданий, формирование выборки и сбор данных, распределение тестовых форм, оценка результатов и поддержка вебсайта.

**Выборка или генеральная совокупность.** Выборка из учащихся 4-х и 8-х классов на уровне штата (только государственные школы) и учащихся 12-х классов на национальном уровне. Объем выборки для каждого теста программы NAEP составляет приблизительно 2500 учащихся в каждом штате. Отдельные отчеты по исследованию долгосрочной тенденции сообщают о результатах национального уровня по математике и чтению для выборки учащихся возраста 9, 13 и 17 лет, взятых из государственных и частных школ.

**Анализ.** Каждый учащийся проходит только часть заданий тестов из их общего количества в конкретной содержательной области. Данные позволяют проводить групповые сравнения (например, учащиеся женского и мужского пола в отдельном штате). Моделирование ответов по каждому пункту теста используется, для того чтобы оценить особенности измерения каждого вопроса оценки и создать единый масштаб для представления работы. Выборочные веса применяются, чтобы отразить особен-

ности генеральной совокупности. Шкалы строятся таким образом, чтобы обеспечить возможность сравнения оценок, полученных в различные годы для генеральной совокупности. На каждой стадии анализа применяются меры контроля качества. Сообщаются проценты учащихся, достигших каждого из трех уровней профессионализма, – базового (частичное владение предпосылками к усвоению знаний), профессионального (компетентное владение материалом предмета) и продвинутого (превосходный уровень выполнения теста).

*Использование результатов.* Результаты получают широкую огласку. Политические и общественные деятели использовали результаты NAEP для освещения как положительных, так и отрицательных моментов качества американской школьной системы.

*Темы, представляющие интерес для обсуждения.* Программа NAEP ведет мониторинг тенденций в результатах выполнения тестов подгруппами. Особое внимание обращено на экспертные оценки прогресса подгрупп меньшинства, в которых отмечается преимущественное повышение оценок по чтению с 1971 года. В целом оценки по чтению и математике повышаются для учащихся 4-го класса, а разрыв достижений учащихся различных рас сужается. Вообще замедление темпов роста оценок по чтению было зарегистрировано в тот период, когда число латиноамериканских учащихся (у кого традиционно были трудности с освоением чтения на английском языке) удвоилось. Изменяющаяся структура контингента учащихся создает трудности при установлении того, дают ли эффект усилия по улучшению педагогических воздействий и совершенствованию учебных планов.

*Источники:* Johnson, 1992; Американский Национальный центр статистики образования, 2005, 2006.

#### **А.9. Уганда**

*Цель.* Программа по национальной оценке образовательного прогресса (NAPE), которая была проведена в июле 2005 года во второй половине учебного года, была одной из программ

в серии национальной оценки в Уганде. Были выделены следующие цели оценки:

- определить уровень достижений учащихся в изучении английского языка и действиях с числами;
- исследовать связи между достижениями учащихся с учетом гендерных и возрастных факторов, а также местоположения школ (городские, пригородные, сельские) и зонами страны;
- проверить профили учебных достижений;
- сравнить достижения учащихся 3-х и 6-х классов в 1999 и 2005 годах.

**Частота.** Уганда проводит национальную оценку образовательных достижений с 1996 года. Первоначально пары предметов (грамотность и умения по выполнению действий с числами; естественные и общественные науки) оценивались каждые три года. С 2003 года основное внимание сместилось на грамотность и умения по выполнению действий с числами, которые оцениваются ежегодно.

**Классы.** Учащиеся 3-х и 6-х классов.

**Оцениваемые достижения.** Грамотность по английскому языку и умения по выполнять действия с числами. Каждые три года проводится оценка владения разговорной речью на английском языке.

**Инструментарий.** Тесты учебных достижений для проверки грамотности и умений по выполнению действий с числами. Ранние программы национальной оценки использовали анкеты для учащихся, преподавателей и директоров школ. Такая оценка, объединяющая данные анкет, теперь проводится раз в три года.

**Кто выполняет?** Национальная экзаменационная комиссия Уганды (UNEB).

**Выборка или генеральная совокупность.** Первоначально были выбраны районы в каждой из 14 зон страны. Объем выборки был увеличен, чтобы гарантировать оценку минимум трех школ в каждом районе.

*Анализ.* Оценки учащихся по каждому тесту были ранжированы по одному из четырех уровней: продвинутый, адекватный, базовый и недостаточный. Соответствие оценок различным уровням было определено и установлено в период, когда шел процесс конструирования тестов группой экспертов, являющихся официальными представителями Национального центра разработки учебных планов, коллегии преподавателей начальных классов, Агентства по образовательным стандартам, UNEB и профессиональных преподавателей. С помощью теста из 50 заданий по английскому языку для 3-го класса были определены следующие диапазоны оценок результатов выполнения: 38–50 баллов – продвинутый уровень, 20–37 баллов – адекватный уровень, 15–19 баллов – базовый уровень и 0–14 – недостаточный уровень. Группы экспертов приняли решение о том, что адекватный уровень нужно рассматривать как минимальный желательный уровень профессионализма в освоении предмета. Менее 40% учащихся 3-го класса достигли желаемого уровня по английскому языку (табл. А.9.1). Результаты тестов учебных достижений сообщались (в%) согласно возрасту учащихся, местоположению школы (городская или сельская), географической области и зоне.

*Использование результатов.* UNEB опубликовала презентации по каждому из 3-х и 6-х классов в Уганде, перечисляя области учебного плана, где выполнение теста учащимися на национальном уровне можно было считать адекватным (например, «Мы

**ТАБЛИЦА А.9.1.**

**Учащиеся 3-х классов в Уганде, ранжированные по квалификационным уровням по английскому языку, 2005 год, %**

Рейтинг	Юноши	Девушки	Все
Профессиональный (продвинутый + адекватный)	36,9	39,7	38,3
Ниже желаемого профессионального уровня (базовый + недостаточный)	63,1	60,3	61,7

*Источник:* Национальная экзаменационная комиссия Уганды, 2006, табл. 3.02.

умеет считать» или «Мы умеем выполнять сложение и вычитание чисел, представленных в цифрах и словесной форме») и менее адекватным (например, «необходима помощь в развитии лексического запаса», или «необходима помощь при делении чисел правильно», или «необходима помощь при решении примеров по математике»). Также была подготовлена подобная презентация для преподавателей.

У комиссии UNEB есть планы по распространению основных уроков за период с 2005 года при проведении программы NARE в форме отдельных легких для использования сообщений о результатах программы NARE для преподавателей, директоров школ, наблюдателей, инспекторов и тактиков. Также планируется использование экспериментальных данных и подходов в национальной оценке для помощи в повышении оценок и улучшении материально-технической базы классов.

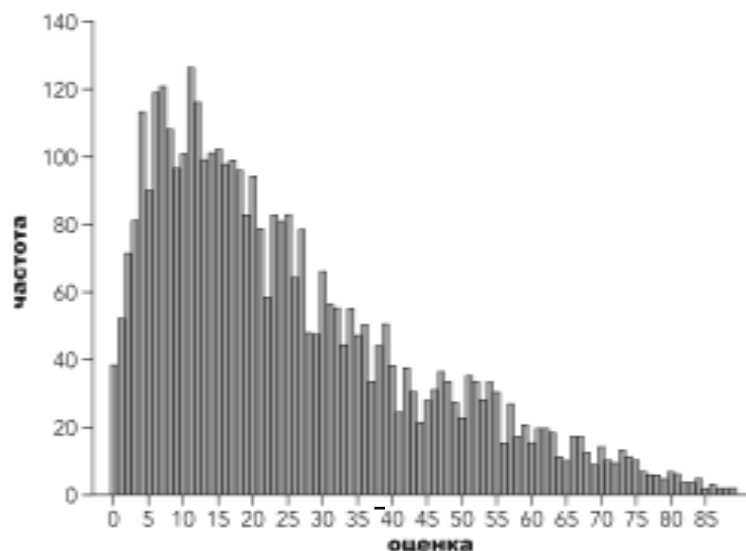
*Темы, представляющие интерес для обсуждения.* Подавляющее большинство учащихся должны были выполнить тесты на втором языке. Определить используемый язык в семье, на котором можно было бы представить тест, оказалось очень сложно. Более четверти основных школ не смогли принять участие в национальной оценке частично из-за общественных беспорядков в отдельных регионах. При выполнении программы UNEB было отмечено, что школы иногда раздували данные об учащихся, чтобы увеличить свои возможности при распределении ресурсов.

Многие из заданий языковых тестов проверяли материал раздела под общим заголовком «Грамматика» (50% заданий для 3-го класса и 30% заданий для 6-го класса). Вообще учащиеся сочли тестовые задания слишком трудными. Многие учащиеся получили относительно низкие баллы (см. рис. А.9.1). Хотя типичный учащийся 3-го класса, как ожидалось, должен был быть возраста приблизительно 8–9 лет, фактически средний возраст учащихся, которые выполняли тест для 3-го класса, составлял 10,2 лет; некоторые были одиннадцатилетнего возраста и старше.

Существенные различия в достижениях наблюдались по зонам страны. Общий процент учащихся (87,5%) 6-го класса

Рис. А.9.1

Распределение тестовых оценок грамотности в 6-х классах в Уганде



Источник: Clarke, 2005.

в зоне Кампалы достигли желательного уровня в грамотности по английскому языку. Соответствующий процент в каждой из шести других зон был менее 30%. Результаты выполнения тестов учащимися 6-х классов субтеста по письму показали существенные различия между ожидаемыми и фактическими уровнями результатов. Примерно половина учащихся достигли желаемого уровня мастерства в написании историй по картинкам, одна четверть – в написании писем, и одна десятая – в составлении композиций и написании сочинений. Методический отчет включает образец написания письма учащимися и списки наиболее часто встречающихся ошибок в тестах по математике. Он также включает ряд рекомендаций и список агентств или структур, которые должны были нести ответственность за реализацию рекомендаций.

Комитет UNEB использовал услуги внешних консультантов для анализа качества его работы, особенно качества статистических характеристик заданий и соответствия между отобран-



ными заданиями и целями учебного плана. Консультант отметил близкое соответствие между заданиями и учебным планом, но рекомендовал уделять больше внимания решению задач по математике. Работа консультанта была несколько ограничена недостатком информации о результатах более ранних программ национальной оценки, касающейся разработки тестов, весов при формировании выборки, дизайна и анализа. Некоторые из проблемы явились следствием того факта, что некоторые аналитические работы программы NARE проводились сторонними организациями, не связанными с UNEB. Консультант рекомендовал, чтобы копии всего инструментария, деталей формирования выборки, аналитические процедуры и другая соответствующая документация были сохранены национальным агентством по оценке (UNEB).

*Источник:* Национальная экзаменационная комиссия Уганды, 2006.



## ПРИЛОЖЕНИЕ В

# МЕЖДУНАРОДНЫЕ ИССЛЕДОВАНИЯ

### **В.1. МЕЖДУНАРОДНОЕ ИССЛЕДОВАНИЕ ТЕНДЕНЦИЙ В ИЗУЧЕНИИ МАТЕМАТИКИ И ЕСТЕСТВЕННЫХ НАУК**

#### **Структура**

Основные цели исследования TIMSS, организованного Международной ассоциацией по эвалюации образовательных достижений, были следующими:

- оценить достижения учащихся по математике и естественным наукам, описанные в терминах понятий, процессов, умений и отношений;
- описать контекст, в котором развиваются достижения учащихся, чтобы установить факторы, связанные с их обучением и поддающиеся управлению посредством изменения политики (связанные, например, с акцентами в учебном плане, распределением ресурсов по школам или используемыми методами обучения).

Исследование TIMSS проводилось три раза: первый – в 45 образовательных системах в 1994–1995 годах на трех категориях учащихся (3-й и 4-й классы; 7-й и 8-й классы; последний класс средней школы), второй – в 38 образовательных систе-

мах в 1999 году в 8-х классах, третий – в 4-м и 8-м классах в 50 системах в 2003 году. Дополнительные исследования были проведены в 2007 и 2008 годах (только последний класс средней школы), а следующее намечено на 2011 год.

Программа TIMSS проводит различия между намеченным, реализуемым и достигнутым учебными планами и в исследованиях анализирует взаимосвязь между ними. *Намеченный учебный план* представляет цели, заявленные в обществе для обучения, которые, как правило, описаны в учебных планах, программах, программных заявлениях и декларациях и отражены в учебниках, информационных ресурсах и экзаменах. *Реализованный учебный план* представляет собой результат интерпретации намеченного учебного плана преподавателями и его доступности для учащихся. Данные по выполнению учебного плана (которые обеспечивают показатель возможности учащегося к усвоению учебного материала), собираются, главным образом, с помощью анкет, предъявляемых преподавателям и учащимся. *Достигнутый учебный план* характеризует то, чему научились учащиеся, и то, что выводится из результатов выполнения ими тестов.

### **Инструментарий**

В тестах TIMSS оцениваются следующие компоненты математики.

- *Содержание.* Числа; измерение; геометрия; пропорции; функции, связи и уравнения; работа с информацией, вероятность, статистика; начала анализа; построение несмещенных выборок и связь между переменными.
- *Ожидаемые результаты.* Знания, использование типичных процедур, исследование и решение проблем, математические рассуждения и коммуникации.
- *Перспективы.* Отношения, карьерный рост, участие, возрастающий интерес и развитие умственных способностей.

Компоненты по естественным наукам в TIMSS включают следующее:

- *Содержание.* Наука о Земле; науки о жизни; физика; естествознание, технология, математика; история науки; проблемы охраны окружающей среды; наука о природе и другие дисциплины.
- *Ожидаемые результаты.* Понимание; теоретизирование, анализ, решение проблем; использование инструментов, обычные процедуры и научные процессы; исследование мира природы; коммуникация.
- *Перспективы.* Отношения, карьерный рост, участие, возрастающий интерес, безопасность и развитие умственных способностей.

С самого начала в программе TIMSS была изменена структура для отражения изменений в учебных планах и педагогической работе в участвующих странах. Проектировщики TIMSS использовали структуру учебного плана, которая была основана на более ранних исследованиях (в частности, по математике во Втором международном исследовании по математике), чтобы разработать тесты на основе достижения согласия среди участвующих стран. Несколько сот заданий (с множественным выбором и с конструируемым ответом) были апробированы и оценены с точки зрения их пригодности и соответствия учебному плану. Максимальный охват учебного плана был достигнут без перегрузки учащихся, которые приняли участие в исследовании, путем распределения тестовых заданий по буклетам. Каждый учащийся отвечал на задания только одного буклета. В табл. В.1.1 представлен пример структуры учебного плана для оценки в программе TIMSS 2007 года.

Были разработаны анкеты, по которым собиралась информация по следующим вопросам:

- общие социальные и образовательные контексты (системный уровень);
- местные коллективные и школьные контексты (школьный уровень);
- личные фоновые факторы (индивидуальный уровень учащегося).

Инструментарии были переведены более чем на 30 языков.

ТАБЛИЦА В.1.1

Планируемые процента содержательных разделов и когнитивных процессов в 4-х и 8-х классах в тестах по математике в программе TIMSS 2007 года, %

Содержательные области для 4-го класса	Доля	
Числа	50	
Геометрические фигуры и меры	35	
Изображение данных	15	
Содержательные области для 8-го класса	Доля	
Числа	30	
Алгебра	30	
Геометрия	20	
Данные и вероятности	20	
Когнитивные процессы	Доля	
	4-й класс	8-й класс
Знания	40	35
Применение	40	40
Рассуждения	20	25

Источник: Mullis and others, 2005. Т. 2. Воспроизводится с разрешения правообладателя.

### Участники

В первоначальной программе TIMSS в 1994–1995 годах принимали участие три совокупности учащихся.

- *Первая совокупность.* Учащиеся в паре смежных классов, в которых было больше всего учащихся 9 лет (как правило, 3-й и 4-й классы).
- *Вторая совокупность.* Учащиеся в паре смежных классов, в которых было больше всего учащихся 13 лет (как правило, 7-й и 8-й классы).
- *Третья совокупность.* Учащиеся последнего класса средней школы. Были выделены две подгруппы: все учащиеся, которые прошли текст по математике и грамотности, и учащиеся, которые обучались в профильных классах по математике или физике и выполняли специализированный тест.

В 1994–1995 годах в TIMSS участвовали 45 систем образования (первая, вторая и третья категории). Среди них одна аф-

риканская (Южная Африка), шесть из Азии / Ближнего Востока (Гонконг (Китай), Исламская Республика Иран, Израиль, Япония, Республика Корея, Кувейт, Сингапур и Таиланд) и одна из Латинской Америки и Карибского бассейна (Колумбия). Названия систем образования в этом приложении перечислены в сообщениях об исследованиях.

В 1999 году в TIMSS участвовали 38 систем образования (вторая категория).

Среди них были три африканские системы (Марокко, Южная Африка и Тунис); 13 – из Азии / Ближнего Востока (Китайский Тайбэй, Гонконг (Китай), Индонезия, Исламская Республика Иран, Израиль, Япония, Иордания, Республика Корея, Малайзия, Филиппины, Сингапур, Таиланд, Турция) и две из Латинской Америки и Карибского бассейна (Аргентина и Чили).

В TIMSS 2003 года было пятьдесят участников (первая и вторая категория). Среди них шесть из Африки (Ботсвана, Арабская Республика Египет, Гана, Марокко, Южная Африка и Тунис), 17 – из Азии / Ближнего Востока (Бахрейн, Китайский Тайбэй, Гонконг (Китай), Индонезия, Исламская Республика Иран, Израиль, Япония, Иордания, Республика Корея, Ливан, Малайзия, Палестина, Филиппины, Саудовская Аравия, Сингапур, Сирийская Арабская Республика и Республика Йемен) и одна из Латинской Америки и Карибского бассейна (Чили).

### **Некоторые результаты**

Табл. В.1.2 представляет результаты 2003 года теста для 8-го класса по математике. Примерно одна треть учащихся в системах с самыми высокими результатами тестирования набрали баллы, соответствующие продвинутому эталонному уровню. Полную противоположность показывают результаты в 19 системах с самыми низкими оценками по тесту, где этому эталонному уровню соответствовали около 1% или меньше учащихся. Сингапур получил первое место в рейтинге по тестам и для 4-го, и для 8-го классов. Некоторые системы продемонстрировали значительно более высокие средние достижения по срав-

**ТАБЛИЦА В.1.2**  
**TIMSS, распределение достижений по математике, 8-й класс**

Страны	Вс. в школе	Средний балл TIMSS	Распределение учащихся по категориям					Средний балл	Результат относительно среднего международного
			5-й перцентиль	25-й перцентиль	50-й перцентиль	75-й перцентиль	95-й перцентиль		
Австралия	0	54.3						551 (2.8)	↑
▶ Австралия (Брисбен)	0	55.5						594 (2.3)	↑
† Аргентина	0	55.8						593 (2.3)	↑
† Австралия (Тасмания)	0	54.3						551 (2.8)	↑
Венгрия	0	54.4						550 (2.8)	↑
Вьетнам	0	54.1						541 (3.4)	↑
† Великобритания	0	54.3						554 (2.8)	↑
† Германия	0	55.2						581 (2.0)	↑
Венгрия	0	54.9						559 (2.8)	↑
Мексика	0	54.5						568 (2.1)	↑
Италия	0	55.0						569 (2.0)	↑
Республика Корея	7-й класс	55.8						599 (1.5)	↑
США	0	55.8						599 (1.5)	↑
† Австралия (Северная Территория)	0	55.7						597 (1.6)	↑
† ОАЭ	0	55.0						594 (2.0)	↑
† Вьетнам	0	55.7						593 (2.0)	↑
† Индия	0	55.0						591 (2.0)	↑
1 Австралия	†	55.7						593 (2.0)	↑
2 Израиль	0	54.0						494 (2.4)	↑
Канада (Торонто)	8-й - 9-й класс	54.1						474 (3.3)	↑
Словения	7-й класс	53.8						493 (2.0)	↑
Индия	0	53.8						494 (2.3)	↑
1 Ирландия	0	54.9						493 (2.0)	↑
† Израиль	0	54.9						477 (3.4)	↑
Венгрия	0	54.9						494 (2.0)	↑
Румыния	0	55.0						491 (2.0)	↑
<b>Среднеазиатский регион</b>	0	53.7						<b>471 (3.7)</b>	↑
Казахстан	7	55.5						587 (1.6)	↑
Молдова	0	55.7						583 (1.8)	↑
Кипр	0	55.8						604 (1.3)	↑
2 Малайзия	0	55.5						581 (1.8)	↑
Испания	0	55.5						581 (1.8)	↑
† Австралия	0	55.7						604 (1.3)	↑
Италия	0	55.8						477 (3.4)	↑
1 Индонезия	0	54.9						477 (3.4)	↑
Турция	0	54.8						470 (3.5)	↑
Бразилия	0	54.4						464 (3.9)	↑
Вьетнам	0	54.1						467 (3.5)	↑
† Австралия (Северная Территория)	0	54.3						491 (2.0)	↑
Италия	0	54.7						597 (1.6)	↑
1† Австралия	0	55.7						597 (1.6)	↑
Малайзия	0	54.8						478 (3.4)	↑
Вьетнам	0	55.1						583 (1.8)	↑
Бразилия (Север)	0	55.1						581 (1.8)	↑
Тайвань	0	55.6						593 (2.0)	↑
Китай (Пекин)	0	55.1						581 (1.8)	↑
† Австралия	†	55.8						599 (1.5)	↑
<b>Среднеазиатский регион</b>									
Казахстан (Астана, Алматы)	0	55.1						497 (2.3)	↑
Киргизия (Бишкек, Суйф)	0	55.1						509 (1.8)	↑
Оман (Сальвадор)	0	53.8						537 (2.9)	↑
Кувейт, Катар	0	55.1						593 (2.0)	↑

- \* Представляет год обучения, отсчитываемый от первого года уровня 1 ISCED.
- \*\* Взято из Второй объединенной программы национального развития, сообщение по развитию человеческого потенциала, 2003. С. 237–240.
- † Отвечает руководящим принципам для норм выборочного участия только после того, как замены школ были включены (см. А.9).
- ‡ Приблизительно удовлетворяет руководящим принципам для норм выборочного участия только после того, как замены школ были включены (см. А.9).
- § Не удовлетворяет руководящим принципам для норм выборочного участия (см. А.9)

<sup>1</sup> Национальная желаемая популяция не охватывает всей желаемой международной популяции (см. А.6)

<sup>2</sup> Национальная желаемая популяция не охватывает менее 90% желаемой международной популяции (см. А.6)

► Корея тестировала тех же самых учащихся, что и другие страны, но позже, в 2003 году, в начале очередного школьного года.

( ) Стандартные ошибки взяты в скобки. Поскольку результаты округлены до ближайшего целого числа, то общие количества могут оказаться непоследовательными.

Прочерк (–) указывает, что сравнение данных не возможно.

Источник: Mullis and others. Раздел 1.1. Воспроизводится с разрешения правообладателя.

нению с их результатами в 1995 и 1999 годах, тогда как другие показали значительное снижения уровня. Среди стран, которые улучшили результаты тестирования в 8-х классах, можно отметить Республику Корею, Гонконг (Китай), Латвию, Литву и Соединенные Штаты Америки.

В целом гендерные различия в достижениях по математике были незначительны. В некоторых образовательных системах девочки получили более высокие результаты по сравнению с мальчиками, в то время как мальчики добились большего успеха в других образовательных системах. Фактически во всех системах высокий уровень родительского образования был связан с более высокими достижениями учащихся. И в 4-х и в 8-х классах в исследовании 2003 года число книг, имеющихся дома, значимо коррелировало с достижениями учащихся по математике.

Степень охвата учебного плана, заложенная в тестах 2003 года, изменялась в разных системах образования. Сообщения преподавателей по 8-м классам указывали, что в процессе преподавания тема «Числа» была охвачена, в среднем, на 95%, тема «Измерения – на 78%, геометрия – на 69%, алгебра – на 66% и работа с данными – на 46%. Более 80% учащихся обучались у преподавателей, которые имели, по крайней мере, некоторую профессиональную подготовку по математике. В качестве базы для преподавания широко использовались учебники. Использование калькулятора, напротив, от системы к системе значительно изменялось. Широко использовать его в 4-м классе было разрешено только в пяти образовательных системах. Школы, в которых обучалось незначительное число учащихся, живущих в бедных семьях, имели оценку в среднем на 57 баллов в 8-м классе и 47 баллов в 4-м классе выше, чем школы, в которых более половины учащихся были выходцами из бедных семей.



## В.2. МЕЖДУНАРОДНОЕ ИССЛЕДОВАНИЕ ПРОГРЕССА ГРАМОТНОСТИ ЧТЕНИЯ

### Структура

Исследование грамотности чтения 1991 года, проведенное Международной ассоциацией по эвалюации образовательных достижений, послужило основой для определения грамотности чтения в PIRLS. Для PIRLS (и в 2001 и в 2006 году) грамотность чтения была определена как:

...способность понять и использовать те формы письменного языка, которые востребованы обществом и/или ценятся индивидуумом. Молодые читатели могут выявлять смысл из множества текстов. Они читают, чтобы учиться, чтобы участвовать в сообществах читателей и для удовольствия (Международная ассоциация по эвалюации образовательных достижений, 2000, 3).

Структура оценки для PIRLS включает две главных цели чтения, которые пересекаются с четырьмя процессами понимания. *Цели* следующие:

- *Литературные.* Чтение для приобретения литературного опыта, в ходе которого читатель включается в текст, будучи вовлеченным в предполагаемые события и характеры, и получает удовольствие от владения языком.
- *Информационные.* Чтение для приобретения и использования информации, которая знакомит читателя с аспектами реального мира, представленными в хронологических текстах (например, когда события описаны в биографиях, рецептах и инструкциях) или в нехронологических текстах, в которых идеи организованы логически, а не хронологически (например, в обсуждении или текстах убеждения).

*Процессы понимания* требуют, чтобы учащиеся могли сделать следующее:

- *Сосредоточиваться и восстанавливать явно представленную информацию.* Например, выделять определенные идеи, нахо-

дить предложения, в которых формулируется тема или явно выражена главная идея.

- *Делать прямые выводы.* Например, сделать вывод о том, что одно из событий является причиной другого, идентифицировать обобщения в тексте.
- *Интерпретировать и объединять идеи и информацию.* Например, выявить обобщенное сообщение или главную тему текста, сравнивать и противопоставлять информацию из текста.
- *Исследовать и оценивать содержание, язык и текстовые элементы.* Описать, как автор разрабатывал неожиданное окончание; рассуждать о законченности или ясности информации в тексте.

PIRLS проводилось в 2001 и 2006 годах.

### **Инструментарий**

По приближенным оценкам для использования аутентичных текстов (т.е. тех, которые являются типичными для чтения учащимися в их повседневной жизни) по каждой цели (чтение для литературного опыта и чтение для приобретения и использования информации) требуется четырехчасовое тестирование. Поскольку ожидать, что любой учащийся сможет выдержать более одного часа при выполнении теста, неразумно, материал оценки был распределен на 10 буклетов, а каждый отдельный учащийся отвечал только на один из них.

Способности учащихся по каждому из четырех процессов понимания были оценены на основе вопросов, которые сопровождали тексты. Использовались две формы заданий: с множественным выбором и конструируемым ответом.

Информация об отношениях учащихся к чтению и их привычках в чтении была получена с помощью анкет. Чтобы собрать информацию о характеристиках семьи учащихся и школьных факторах, которые считались релевантными для развития грамотности чтения, также было проведено анкетирование для родителей учащихся, преподавателей и директоров школ,

### Участники

В качестве целевой категории учащихся для PIRLS были определены старшие из двух смежных классов, большинству учащихся которых 9 лет. В большинстве систем это был 4-й класс. Тридцать пять систем образования участвовали в PIRLS в 2001 году. Среди них одна африканская (Марокко), шесть – из Азии / Ближнего Востока (Гонконг (Китай), Исламская Республика Иран, Израиль, Кувейт, Сингапур и Турция) и три из Латинской Америки и Карибского бассейна (Аргентина, Белиз и Колумбия) (Mullis and others, 2003). Сорок одна система участвовала в PIRLS в 2006 году. Представительство Африки увеличилось на одну систему (Южная Африка), азиатских / ближневосточных стран – на две (с добавлением Китайского Тайбэя, Индонезии и Катара, но выбыла Турция). Что касается Латинской Америки и Карибского бассейна, то к программе присоединилась одна новая страна (Тринидад и Тобаго), в то время как три страны, участвовавшие в PIRLS в 2001 году, выбыли из него.

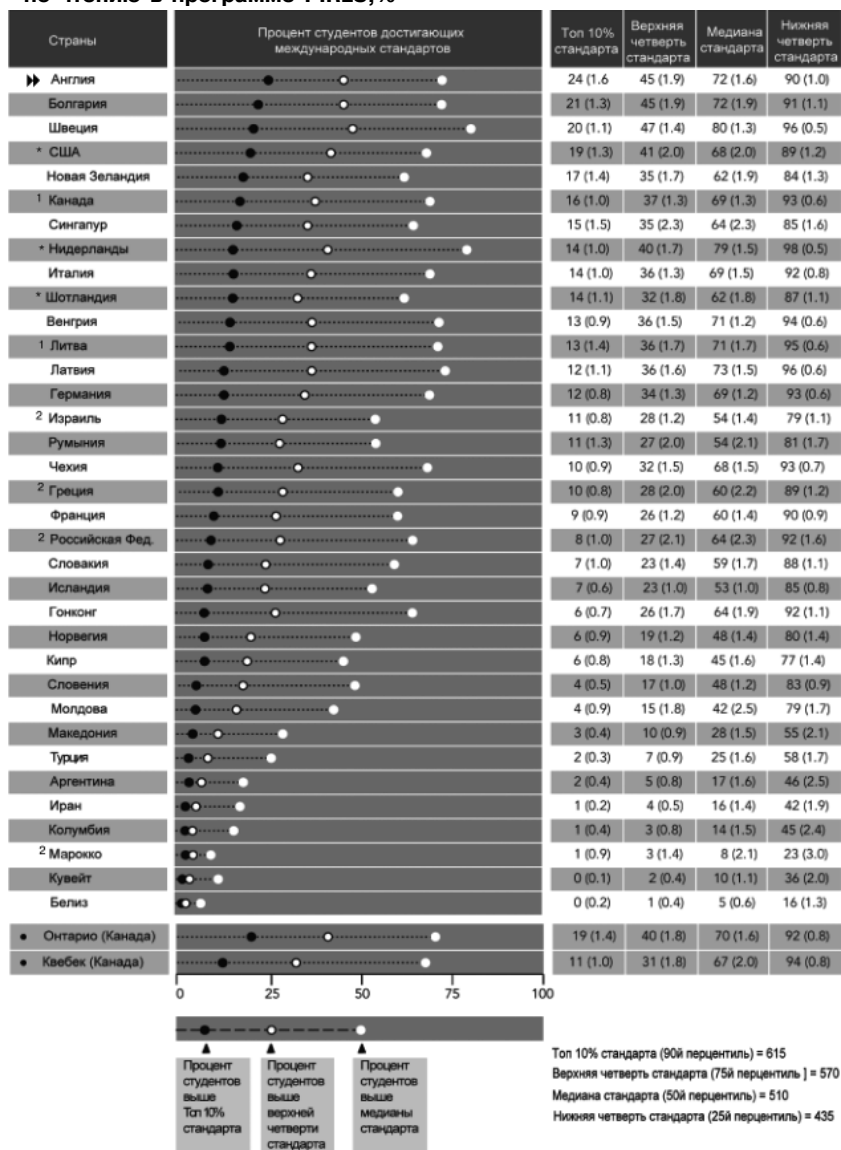
PIRLS намечено провести снова в 2011 году.

### Некоторые результаты

На основе тестовых оценок учащихся были выделены четыре реперных точки. К ним были отнесены: *нижняя четверть*, определенная как 25-й процентиль (точка, выше которой были оценки лучших 75% учащихся); *средняя точка*, определенная как 50-й процентиль и отделявшая половину учащихся с более высокими оценками; *верхняя четверть*, определенная как 75-й процентиль; *точка, отделявшая лучшие 10% учащихся* и определенная как 90-й процентиль. Если оценки достижений по чтению были распределены примерно одинаково в каждой стране, то приблизительно 10% учащихся получили бы ранг в диапазоне, отделенном высшей реперной точкой. Табл. В.2.1 представляет результаты для участвующих стран. Она показывает, например, что 24% английских учащихся были оценены в самой высокой категории и что у 10 образовательных систем к этой категории относились менее чем 5% учащихся.

ТАБЛИЦА В.2.1

Учащиеся 4-го класса, достижения реперных точек в достижениях по чтению в программе PIRLS,%



\* Канада представлена только провинциями Онтарио и Квебек. Средняя оценка в международных исследованиях не включает результаты по этим провинциям отдельно.

† Отвечает руководящим принципам для норм выборочного участия только после того, как замены школ были включены (см. А.7)

‡ Приблизительно удовлетворяет руководящим принципам для норм выборочного участия только после того, как замены школ были включены (см. А.7)

§ Национальная желаемая популяция не охватывает всей желаемой международной популяции. Поскольку охват ниже 65%, Канада аннотируется как Канада (O, Q) только для провинций Онтарио и Квебек.

2a Национальная популяция охватывает менее 95% желаемой международной популяции (см. А.4).

2b Национальная популяция охватывает менее 80% желаемой международной популяции (см. А.4).

( ) Стандартные ошибки взяты в скобки. Поскольку результаты округлены до ближайшего целого числа, то некоторые общие количества могут оказаться непоследовательными.

Девочки получили значительно более высокие средние оценки во всех системах, чем мальчики. По заданиям, которые измеряли чтение для информационных целей, учащиеся из Швеции, Нидерландов и Болгарии получили самые высокие оценки. Раннее развитие грамотности в дошкольный период, такое как чтение книги и рассказ историй, было положительно связано с результатами по чтению в последующий период. Более высокие оценки достижений по чтению получили учащиеся, родители которых положительно относились к чтению. Учащиеся, которые говорили дома на языке, используемом в оценивании, получали, как правило, более высокие оценки, чем учащиеся, которые говорили на других языках. Ответы директоров школ показали, что чтению уделялось большее внимание, чем любой другой области учебного плана с 1-го по 5-й класс.

Преподаватели в среднем сказали, что они просили большинство четвероклассников ежедневно читать вслух целому классу. Библиотеки были практически не задействованы, даже несмотря на их доступность. В среднем большинство преподавателей полагалось на свои собственные оценки больше, чем на объективные тесты, которые использовались для мониторинга прогресса учащихся. Примерно два из каждых трех учащихся сказали, что они читают истории или романы, по крайней мере, один раз в неделю. Во всех образовательных системах отношение учащихся к чтению было положительно связано с их достижениями по чтению.

### **В.3. ПРОГРАММА МЕЖДУНАРОДНОГО ОЦЕНИВАНИЯ УЧАЩИХСЯ**

#### **Структура**

Программа международного оценивания учащихся (PISA) выявляет знания и умения пятнадцатилетних учащихся каждые три года под эгидой Организации экономического сотрудничества и развития (ОЭСР). PISA была разработана, чтобы обеспечить регулярные показатели достижений учащихся в конце

обязательного обучения для международных показателей систем образования под эгидой ОЭСР.

Учащиеся оценивались в трех областях: чтение, математика и естественные науки. К настоящему моменту в рамках программы PISA были выполнены три оценивания. В 2000 году главной областью оценки было чтение, а математика и естественные науки рассматривались как второстепенные области. В 2003 году главной областью стала математика, а чтение и естественные науки заняли второе место. В 2006 году приоритет был отдан естественным наукам, а чтению и математике отводилось второе место.

Программа PISA используется отдельными странами, чтобы: измерить умения по грамотности учащихся отдельной страны по сравнению с достижениями в грамотности учащихся других участвующих стран; установить реперные точки для отсчета в области образовательного усовершенствования в отдельной стране в терминах результатов тестирования учащихся в других странах; оценить способность страны в обеспечении высокого уровня равных возможностей получения образования и высоких результатов. Программа PISA пытается оценить степень, в которой учащиеся к концу периода обязательного образования приобрели некоторые знания и умения, которые важны для полноценного участия в жизни общества.

### **Участники**

В 2000 году в PISA участвовали 32 страны. Два года спустя к задачам оценки PISA 2000 присоединились еще 11 стран. Африканские страны в оценке 2000 года не участвовали. Азиатские / Ближневосточные участники включали две страны – члена ОЭСР (Япония и Республика Корея) и пять стран – партнеров, не являющихся членами ОЭСР (Гонконг (Китай), Индонезия, Израиль, Российская Федерация и Таиланд). Системы Латинской Америки и Карибского бассейна включали Мексику, а также следующие страны, не входящие в ОЭСР: Аргентину, Бразилию, Чили и Перу. В оценке 2003 года приня-

ли участие все 30 государств – членов ОЭСР и еще 11 систем – партнеров.

Среди новых систем – партнеров одна была из Африки (Тунис), одна – из Азии (Макао, Китай) и одна – из Латинской Америки и Карибского бассейна (Уругвай). Три первоначальные системы – партнеры (Аргентина, Чили и Перу) не участвовали в оценке 2003 года. Турция, член ОЭСР, первый раз участвовала в PISA в 2003 году. К 2006 году число участвующих стран возросло до 57. Единственным представителем Африки остался Тунис. Новые системы – партнеры из Азии / Ближнего Востока включали Азербайджан, Китайский Тайбэй, Иорданию, Кыргызстан и Катар. Латиноамериканские системы, которые участвовали в исследованиях 2000 или 2003 года, также участвовали в тестировании 2006 года. К ним присоединилась и одна новая система – партнер (Колумбия).

Совокупность испытуемых, представлявшая интерес в исследовании, включала пятнадцатилетних учащихся. Они выбраны случайным образом по годам обучения в участвующих школах.

### **Инструментарий**

Тест грамотности чтения предполагает, что учащиеся в состоянии читать, и направлен на оценку их способностей к пониманию и размышлению в широком диапазоне письменных материалов в различных ситуациях. Было определено трехмерное измерение: содержание или структура текстов (целых произведений, таких как рассказы и описания, и отрывков, таких как таблицы, диаграммы и формы); процессы, которые были необходимы для выполнения тестов (поиск, интерпретация, размышления и оценка); ситуации, в которых знания и умения применяются (личные, общественные, профессиональные и образовательные).

Математический тест грамотности связан со способностью учащихся анализировать, рассуждать и сообщать свои идеи в процессе формулирования, решения и интерпретации математических проблем во множестве контекстов. Для структуры

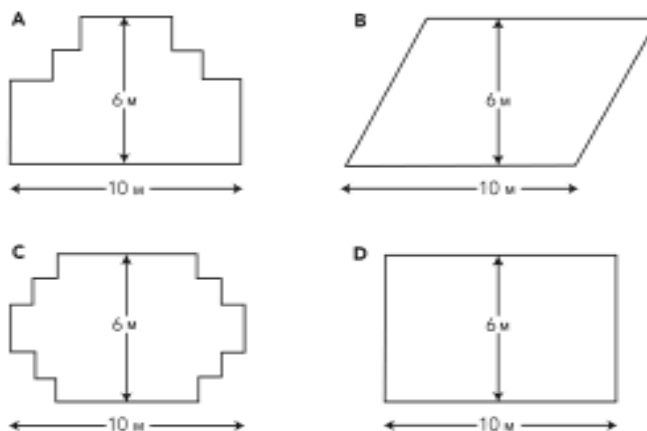
тестов по математике характерна трехмерность: содержание (пространство и форма, изменения и связи, количество и неопределенность), компетенции (кластер по воспроизведению, кластер по установлению связей и кластер по рефлексии) и ситуации (личные, образовательные или профессиональные, общественные и научные). Задания теста преимущественно ориентированы на реальные «жизненные» ситуации в отличие от обычных тестов учебных достижений (см. рис. В.3.1).

Рис. В.3.1

## Пример заданий PISA по математике

## ПЛОТНИК

У плотника есть 32 метра бруса, и он хочет сделать ограду для грядки. Он рассматривает следующие варианты дизайна грядки.



Вопрос 1	
Обведите в кружок «Да или Нет» для каждого дизайна, чтобы отметить, можно ли сделать ограду для грядки из 32 метров бруса	
Дизайн грядки	Используя этот дизайн, можно ли сделать ограду для грядки из 32 метров бруса?
Дизайн А	Да / Нет
Дизайн В	Да / Нет
Дизайн С	Да / Нет
Дизайн D	Да / Нет



Тест грамотности по естественным наукам оценивает способность учащихся делать соответствующие заключения из приведенных фактов и информации, критически рассматривать требования на основе фактов и различать точки зрения, выведенные из фактических данных. Структура теста для естественных наук предполагает трехмерное измерение: научные *понятия* (отобранные из физики, химии, биологии, науки о земле и космосе), *процессы* (описание, объяснение и предсказание научных явлений, понимание научных исследований и интерпретация научных доказательств и заключений) и *применение* (в жизни и здоровье; на земле и в окружающей среде; в технологии).

Наличие гораздо большего числа тестовых заданий, чем отдельный учащийся мог бы выполнить, гарантирует подробное освещение различных областей. Тестовые задания были разбиты на 13 буклетов. Каждый из них состоял из различных комбинаций заданий по математике, чтению, естественным наукам и решению проблем.

Анкеты были предъявлены учащимся (чтобы получить информацию об их учебных обязанностях, используемых стратегиях обучения и уверенности в своих силах; их восприятии среды обучения; их семейном фоне) и директорам школ (чтобы получить информацию о школьной политике, методах и качестве доступных ресурсов) (ОЭСР, 2004b).

### **Некоторые результаты**

Программа PISA приводит средние оценки по странам в «сравнительной таблице» (рис. В.3.2). Также результаты учащихся разбиты по категориям мастерства, которые основаны на тестовых оценках и показывают, что учащиеся, как правило, могут делать. Рис. В.3.3 описывает умения, связанные с каждым из шести уровней мастерства в программе PISA для математики. Рис. В.3.4 подводит итоги того, как учащиеся в каждой стране выполняли тест на разных уровнях мастерства.

Результаты указывают на очень существенные различия между странами, такими как Финляндия, Республика Корея

и Канада, где большинство из учащихся получили оценки выше уровня 2, и Бразилия, Тунис и Индонезия, где только небольшое число учащихся достигли этого уровня мастерства. Другие результаты показали, что меньше 5 % учащихся в странах – членах ОЭСР достигли уровня 6, в то время как приблизительно одна треть учащихся смогла выполнить задачи, связанные с уровнями 4, 5 и 6. 11% учащихся не смогли выполнить задачи по математике на уровне 1. В большинстве стран мальчики получили более высокие оценки, чем девочки, особенно в задачах, связанных с пространством и формой.

В некоторых странах (Австралия, Австрия, Япония, Нидерланды, Норвегия и Польша) гендерные различия в достижениях не были значительными. Девочки проявили более низкий интерес – и удовольствие – к математике. Они утверждали, что испытали больше напряжения, чем мальчики при выполнении заданий в этой области учебного плана. Американские учащиеся были склонны иметь более сильные представления о «Я-концепции» в математике, чем учащиеся в других странах. Напротив, учащиеся из Японии и Республики Корея, которые получили более высокие баллы в тесте по математике, имели относительно слабые представления о самооценке в математике. Работа родителей и их поддержка образования была сильно связана с достижениями учащихся.


Гендерные различия в достижениях по естественным наукам редко были очевидны. Почти одинаковое число процентов мальчиков и девочек набрали наиболее высокие и наиболее низкие баллы. По чтению средняя оценка в Финляндии была больше чем наполовину уровня мастерства выше, чем в странах ОЭСР. Финляндия, наряду с Кореей и Канадой, также показала относительно низкие внутренние различия, предлагая больший уровень образовательного равенства, чем в большинстве участвующих стран. Очень немногие учащиеся из Индонезии, Туниса или Сербии достигли уровня 3 или выше (см. рис. В.3.5).


Рис. В.3.2


## Средние оценки грамотности чтения в программе PISA и подшкалы по чтению

Комплексная оценка грамотности чтения		Восприятие текста		Интерпретация текстов		Рефлексия и оценка текстов	
Страна	Средний балл	Страна	Средний балл	Страна	Средний балл	Страна	Средний балл
Финляндия	546	Финляндия	556	Финляндия	555	Канада	542
Канада	534	Австралия	536	Канада	532	Объединенное королевство Великобритании и Северной Ирландии	539
Новая Зеландия	529	Новая Зеландия	535	Австралия	527	Ирландия	533
Австралия	528	Канада	530	Ирландия	526	Финляндия	533
Ирландия	527	Республика Корея	530	Новая Зеландия	526	Япония	530
Республика Корея	525	Япония	526	Республика Корея	525	Новая Зеландия	529
Объединенное королевство Великобритании и Северной Ирландии	523	Ирландия	524	Швеция	522	Австралия	526
Япония	522	Объединенное королевство Великобритании и Северной Ирландии	523	Япония	518	Республика Корея	526
Швеция	516	Швеция	516	Исландия	514	Австрия	512
Австрия	507	Франция	515	Объединенное королевство Великобритании и Северной Ирландии	514	Швеция	510
Бельгия	507	Бельгия	515	Бельгия	512	Соединенные Штаты Америки	507
Исландия	507	Норвегия	505	Австрия	508	Норвегия	506
Норвегия	505	Австрия	502	Франция	506	Испания	506
Франция	505	Исландия	500	Норвегия	505	Исландия	501

Соединенные Штаты Америки	504	Соединенные Штаты Америки	499	Соединенные Штаты Америки	505	Дания	500
Дания	497	Швейцария	498	Чехия	500	Бельгия	497
Швейцария	494	Дания	498	Швейцария	496	Франция	496
Испания	493	Италия	488	Дания	494	Греция	495
Чехия	492	Испания	483	Испания	491	Швейцария	488
Италия	487	Германия	483	Италия	489	Чехия	485
Германия	484	Чехия	481	Германия	488	Италия	483
Венгрия	480	Венгрия	478	Польша	482	Венгрия	481
Польша	479	Польша	475	Венгрия	490	Португалия	480
Греция		Португалия	455	Греция	475	Германия	478
Португалия		Греция	450	Португалия	473	Польша	477
Люксембург	441	Люксембург	433	Люксембург	446	Мексика	446
Мексика	422	Мексика	402	Мексика	419	Люксембург	442
В среднем по ОЭСР	500	В среднем по ОЭСР	498	В среднем по ОЭСР	501	В среднем по ОЭСР	502
Страны, не являющиеся членами ОЭСР		Страны, не являющиеся членами ОЭСР		Страны, не являющиеся членами ОЭСР		Страны, не являющиеся членами ОЭСР	
Лихтенштейн	483	Лихтенштейн	492	Лихтенштейн	484	Лихтенштейн	468
Российская Федерация	462	Латвия	451	Российская Федерация	468	Латвия	458
Латвия	458	Российская Федерация	451	Латвия	459	Российская Федерация	455
Бразилия	396	Бразилия	365	Бразилия	400	Бразилия	417

 Среднее значительно выше среднего значения по США

 Среднее незначительно отличается от среднего значения по США

 Среднее значительно ниже среднего значения по США

*Примечание.* Нидерланды принимали участие в программе Международного оценивания (PISA) учащихся в 2000 году, технические проблемы с выборками препятствовали включению результатов этой страны в обсуждение, которое проводится здесь. Для получения информации по Нидерландам см. ОЭСР (2001). Среднее значение оценок в ОЭСР является средним по национальным средним оценкам в 27 странах ОЭСР. Поскольку PISA принципиально является программой ОЭСР, то результаты стран, не входящих в ОЭСР, показаны отдельно от результатов стран ОЭСР и не входят в подсчет среднего по ОЭСР.

*Источник:* ОЭСР, 2001, рис. 3. Воспроизводится с разрешения правообладателя.

Рис. В.3.3

## Квалификационные уровни учащихся для математики в программе PISA

		Что могут обычно выполнять учащиеся
Оценка	Уровень	
668	6	Учащиеся могут осмысливать, обобщать и использовать информацию на основе исследования и моделирования сложных проблемных ситуаций. Они могут связывать различные источники информации и представления данных, а также свободно преобразовывать их. Учащиеся этого уровня обладают продвинутым математическим мышлением и аргументацией. Эти учащиеся способны разобраться в существе вопроса и наряду с освоением символических и формальных математических операций и отношений развивать новые подходы и стратегии для решений в новых ситуациях. Учащиеся на этом уровне могут сформулировать и точно сообщить свои действия и размышления относительно своих результатов, интерпретаций, аргументов и уместности их по отношению к исходным условиям
606	5	Учащиеся могут развить и работать с моделями для сложных условий, определяя ограничения и устанавливая предположения Они могут выбирать, сравнивать и оценивать подходящие стратегии решения задач со сложными условиями, связанные с этими моделями Учащиеся могут работать стратегически, используя широкое, хорошо развитое мышление и аргументацию, надлежащим образом связанные представления, символические и формальные характеристики и понимание, относящиеся к этим условиям. Они могут размышлять над своими действиями, формулировать и связывать свои интерпретации и обоснования
544	4	Учащиеся могут работать с определенными моделями для сложных конкретных условий, которые могут включать ограничения или вынуждать делать предположения. Они могут выбирать и интегрировать разные виды представления данных, включая символические, связывая их напрямую с аспектами реальных условий Учащиеся этого уровня могут использовать хорошо развитые умения и свободно аргументировать их использование в рамках контекста. Они могут создавать и передавать объяснения и аргументы, основанные на их интерпретациях, аргументах и действиях
482	3	Учащиеся могут выполнять четко описанные процедуры, включая те, которые требуют последовательных решений. Они могут выбирать и применять простые стратегии решения задач. Учащиеся могут интерпретировать и использовать представления, основанные на разных источниках информации, и рассуждать, основываясь на них. Они могут создавать краткие сообщения отчетности, приводить свои объяснения, результаты и аргументацию
420	2	Учащиеся могут объяснять и распознавать ситуации в условиях, которые не требуют ничего, кроме прямых выводов. Они могут извлекать важную информацию из одного источника и использовать один репрезентативный режим. Учащиеся этого уровня могут использовать базовые алгоритмы, формулы, процедуры или правила. Они могут давать прямые обоснования и буквально интерпретировать результаты
358	1	Учащиеся могут отвечать на вопросы, затрагивающие привычный контекст, где вся важная информация представлена и вопросы четко определены. Они могут определять информацию и выполнять стандартные процедуры, касающиеся прямых инструкций в обычных ситуациях. Они могут выполнять действия, которые непосредственно вытекают из данных стимулов
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37

Рис. В.3.4

**Учащиеся каждого квалификационного уровня по шкале PISA (математика),%**

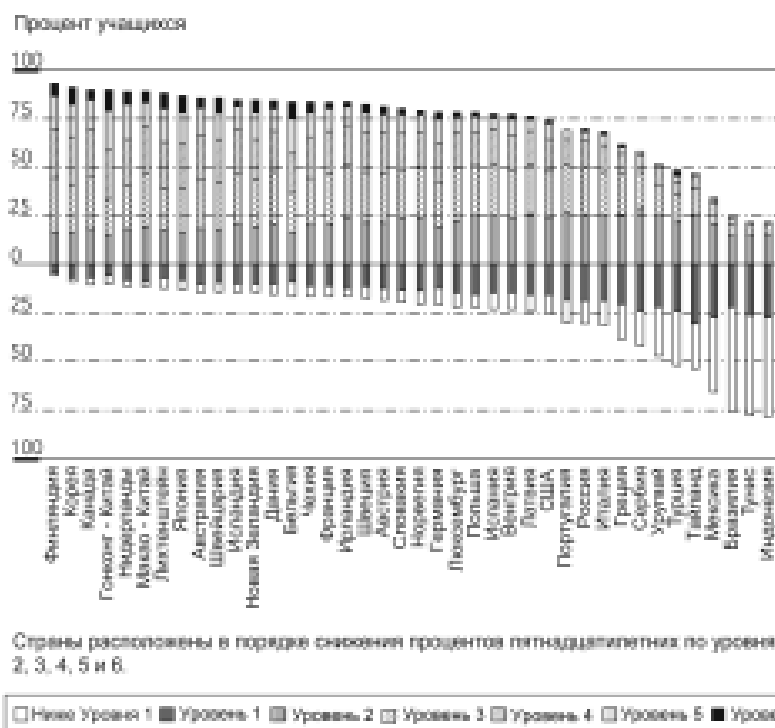
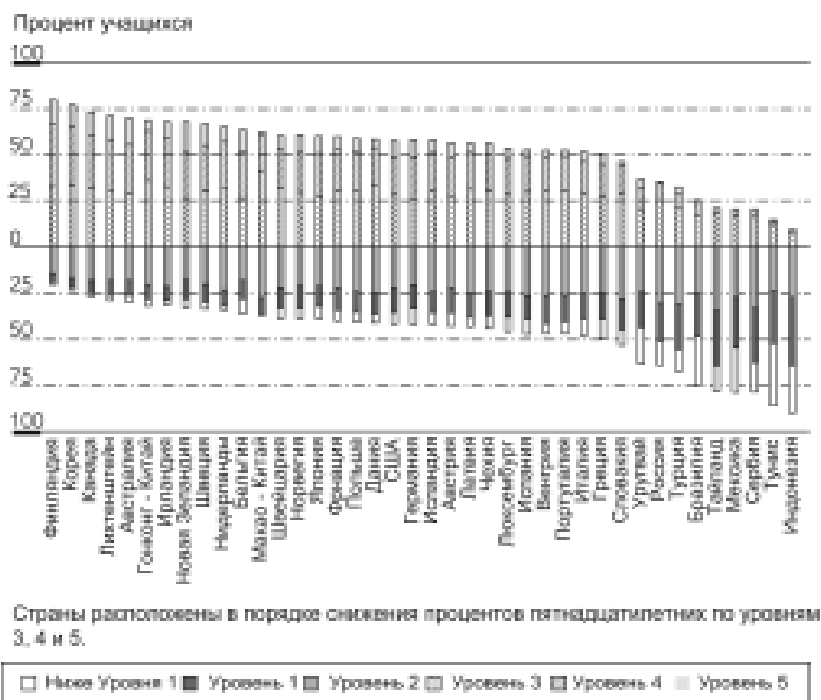


Рис. В.3.5

Учащиеся каждого квалификационного уровня по шкале PISA (чтение), %





## ПРИЛОЖЕНИЕ С

# РЕГИОНАЛЬНЫЕ ИССЛЕДОВАНИЯ

### **С.1. ЮЖНО- И ВОСТОЧНОАФРИКАНСКИЙ КОНСОРЦИУМ ПО МОНИТОРИНГУ КАЧЕСТВА ОБРАЗОВАНИЯ**

#### **Структура**

Южно- и Восточноафриканский консорциум по мониторингу качества образования (SACMEQ) является добровольным объединением министерств образования в Южной и Восточной Африке, включая Ботсвану, Кению, Лесото, Малави, Маврикий, Мозамбик, Намибию, Сейшельские острова, Южную Африку, Свазиленд, Танзанию (материк), Танзанию (Занзибар), Уганду, Замбию и Зимбабве. Основанный в 1995 году при содействии Международного института планирования образования ЮНЕСКО, SACMEQ был создан для: развития институционального потенциала посредством совместного обучения (обучение на практике для специалистов по планированию образования) и совместных исследований политики в области школьного обучения и качества образования (например, выявление слабых мест в системе образования в терминах входных данных и процессов) и для мониторинга изменений в образовательных дос-



тижениях (Международный институт планирования образования, 2007). Примечательной особенностью SACMEQ является его целенаправленная стратегия на консультирование с руководителями тактиков в правительстве для выявления важнейших вопросов, которые должны быть рассмотрены в эмпирических исследованиях. Она также направлена на поощрение участия заинтересованных сторон и большей прозрачности в принятии решений. Первый цикл исследований SACMEQ был проведен в период между 1995 и 1999 годами.

Политические интересы исследований SACMEQ II, которые были проведены между 2000 и 2003, были сгруппированы по пяти основным темам (Mugimba, 2005b; Passos and others, 2005).

- Характеристики учащихся и их учебная среда.
- Характеристики преподавателей и их восприятие (например, по вопросам обучения и ресурсов).
- Характеристики директоров школ и их восприятие (например, по управлению школами и разрешению проблем).
- Равенство в распределении человеческих и материальных ресурсов между регионами и школами.
- Достижения в чтении и математике учащихся и их преподавателей.

SACMEQ был основан на более раннем (1991) исследовании, проведенном в Зимбабве (Ross and Postlethwaite, 1991), и начал свою работу по проведению ряда национальных исследований. Тем не менее, он выполнял международное измерение, так как эти исследования объединяли многие общие характеристики (вопросы исследования, инструментарий, целевые группы испытуемых, процедуры формирования выборок и анализ). Для каждой страны был подготовлен отдельный отчет. Межнациональные сравнения были проведены для SACMEQ II, но не для SACMEQ I.

### **Инструментарий**

Были собраны данные по уровню подготовки учащихся по математике и уровню грамотности с помощью тестов учебных

достижений. Ряд заданий из Международного исследования по оценке тенденций в математическом и естественно-научном образовании (TIMSS) был включен в тесты SACMEQ II для обеспечения сравнимости данных. Анкеты использовались для сбора данных по основным показателям вложений в образование, общим условиям школьного обучения и оценке равенства при распределении материальных и человеческих ресурсов. Информация по семейным контекстным условиям жизни учащихся была получена с помощью анкет для учащихся. Их просили отметить в списках то, что они имеют дома, например ежедневные газеты, еженедельные или ежемесячные журналы, радио, телевизоры, телефоны, мотоциклы, велосипеды, водопровод и электричество.

Тесты SACMEQ II включали задания, отобранные из четырех ранее проведенных исследований: исследования в Зимбабве показателей качества образования, SACMEQ I, TIMSS и исследования Международной ассоциации по эвалюации в образовании, проведенного для оценивания грамотности чтения. Использование этих заданий позволило провести сравнение результатов тестирования учащихся с результатами тестирования в SACMEQ II.

Значительное место в отчетах уделено описаниям характеристик преподавателей (например, квалификации) и условиям в школах (например, наличие мебели в классных комнатах, предметы снабжения, размеры и пространство), а также и тому, как они соотносятся со стандартами министерства и насколько варьируются от школы к школе.

SACMEQ II принял определение грамотности чтения, использованное в исследовании грамотности чтения Международной ассоциации по эвалюации в образовании в 1990 году: «способность понимать и использовать письменные формы языка, требуемые обществом и/или ценные для индивидуума» (Elley, 1992, 3). На этом определении также базировалась и разработка теста по трем областям, установленным в исследовании Международной ассоциации по эвалюации в образовании.

- *Повествование.* Сплошной текст, в котором цель автора рассказать историю, а не факт или выдумку.
- *Пояснение.* Сплошной текст, направленный на описание, объяснение или сообщение фактической информации или мнения автора.
- *Документы.* Структурированная информация, представленная в виде диаграмм, таблиц, карт, графиков, списков или инструкций.

Таблица спецификаций была построена таким образом, что три эти области пересекались с семью уровнями умений в чтении:

- дословное воспроизведение;
- пересказ понятия;
- выявление главной идеи;
- выводы из текста;
- позиционирование информации;
- позиционирование и обработка;
- применение правил.

Математическая грамотность была определена SACMEQ II как «способность понимать и применять математические процедуры и делать логические выводы как личность и как член общества» (Shabalala, 2005, 76). Тест оценивал компетентность по трем областям.

- *Числа.* Операции и числовая ось, квадратные корни, округление, веса, значащие цифры, дроби, проценты, отношения.
- *Измерение.* Операции, связанные с расстоянием, длиной, областью, объемом, деньгами, временем.
- *Пространственные данные.* Геометрические формы, диаграммы, таблицы данных.

Таблица спецификаций связала три эти области с пятью «предполагаемыми» (или ожидаемыми) уровнями умений, начиная от, например, способности выполнять простые одно-сложные операции с использованием двузначных чисел (уровень 1) до способности делать вычисления с применением

нескольких действий и нескольких операций, используя дроби, десятичные дроби и целые числа (уровень 5).

Большинство тестовых заданий – задания с множественным выбором.

Результаты были представлены в трех формах: средние баллы, процент учащихся, достигших минимального и желаемого уровней достижений, и процент учащихся, достигших восьми уровней компетентности на базе моделей шкалирования современной теории тестов (имеется в виду теория IRT и модель Раша (Rasch). – *Примеч. ред.*).

Средние баллы являются усредненными мерами результатов выполнения тестов. Они могут быть использованы для описания того, как различные категории учащихся выполняют тесты (например, мальчики и девочки, учащиеся, живущие в разных провинциях и районах).

Минимальный и желаемый уровни достижений были определены комитетами экспертов (состоящими из специалистов по учебному плану, исследователей и опытных преподавателей). Были идентифицированы два уровня.

- *Минимальный* уровень определяет учащегося, который с трудом будет успевать в течение следующего года обучения.
- *Желаемый* уровень определяет учащегося, который будет в состоянии справиться со следующим годом обучения.

Исследования проводились с целью выявления различных уровней умений учащихся и обеспечения более полного представления о природе достижений учащихся.

Умения в чтении были распределены по восьми уровням и включали следующее.

- *Уровень 1.* Подготовка к чтению: устанавливает соответствие слов и картинок, применяя конкретные понятия и бытовые предметы.
- *Уровень 2.* Становление чтения: устанавливает соответствие слов и картинок, привлекая предлоги и абстрактные понятия; использует реплики и дальнейшее чтение текста для интерпретации фраз.

- *Уровень 3.* Базовые умения по чтению: интерпретирует смысл в коротком и простом тексте (путем сопоставления слов и фраз, составляющих предложение).
- *Уровень 4.* Чтение со смыслом: читает текст вперед и возвращается назад, чтобы установить связь и интерпретировать информацию, расположенную в разных частях текста.
- *Уровень 5.* Интерпретирующее чтение: читает текст дальше и возвращается назад, чтобы собрать и интерпретировать информацию из различных частей текста с применением дополнительной внешней информации, полученной из контекста.
- *Уровень 6.* Дедуктивное чтение: читает более длинные (повествовательные, разъяснительные) тексты для объединения информации из различных частей текста и делает логические выводы о цели автора.
- *Уровень 7.* Аналитическое чтение: определяет информацию в более длинных текстах (повествовательных, разъяснительных), объединяет информацию, чтобы вывести личное отношение автора (системы ценностей, предубеждения, пристрастия).
- *Уровень 8.* Критическое чтение: определяет информацию в более длинных текстах (повествовательных, разъяснительных), определяет и оценивает предполагаемую писателем информацию о предмете интересов и характеристиках читателя (например, возрасте, знаниях, личном отношении, оценках).

Умения по математике были также разделены на восемь уровней.

- *Уровень 1.* Подготовка к количественному мышлению: применяет одношаговые действия или операции по вычитанию; узнает простейшие формы, устанавливает соответствие между числами и картинками; считает в целых числах.
- *Уровень 2.* Становление способности к количественному мышлению: применяет двухшаговые операции сложения или вычитания, включая выполнение и проверку (через основную оценку); оценивает длину знакомых фигур; распознает общие двумерные формы.

- *Уровень 3.* Базовая способность к количественному мышлению: переводит графическую информацию в дроби; интерпретирует место величины в ряду целых чисел до тысячи; интерпретирует повседневные общие единицы измерения.
- *Уровень 4.* Начальные способности к количественному мышлению: использует многократные математические операции с целыми числами, простыми дробями, десятичными дробями или со всеми этими числами.
- *Уровень 5.* Компетентность в количественном мышлении: решает задачи с несколькими действиями, включающие повседневные единицы измерения, целые и смешанные числа или все эти числа.
- *Уровень 6.* Математически подготовленный: решает задачи с несколькими действиями, включающие дроби, отношения и десятичные числа; переводит словесную и графическую информацию в символическую и алгебраическую информацию и в формы уравнения.
- *Уровень 7.* Решение проблем: извлекает информацию из таблиц, диаграмм, визуальных и символических материалов для идентификации и решения задач с несколькими действиями.
- *Уровень 8.* Решение абстрактных проблем: определяет природу неустановленной математической проблемы, представленную в виде словесной или графической информации, переводит ее в алгебраическую форму или в уравнение, чтобы решить проблему.

### **Участники**

В период между 1995 и 1999 годами семь министерств образования собирали информацию согласно SACMEQ I по навыкам чтения у учащихся 6-х классов. Четырнадцать министерств завершили исследования SACMEQ II между 2000 и 2002 годами по изучению умений чтения у учащихся 6-х классов. Условия были весьма различны в зависимости от страны. Например, на Сейшельских островах валовой национальный доход был

почти в 40 раз больше (US \$ 6730), чем в Малави (US \$ 170). Государственные расходы на образование колебалась от 30% в Свазиленде до 7% в Танзании, в то время как доля обучающихся в начальных школах составляет 40% в Мозамбике и более 90% на Маврикии, Сейшельских Островах и в Южной Африке (Murimba, 2005b). В некоторых странах преподаватели наравне с учащимися выполняли тесты достижений.

### **Некоторые результаты**

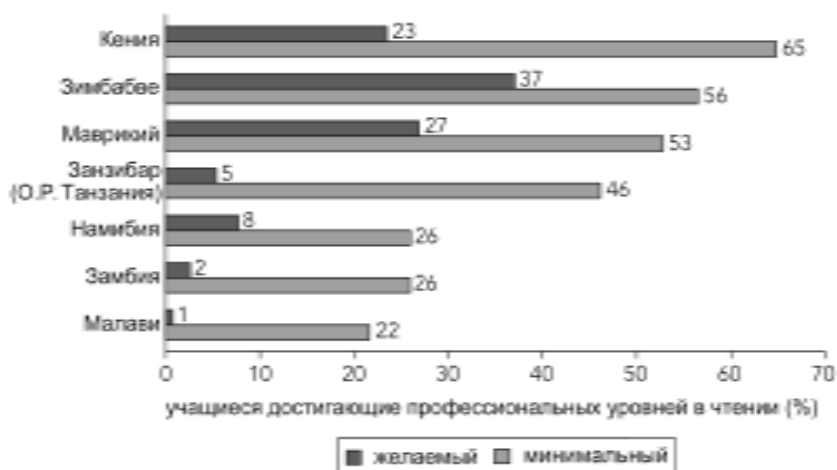
Между странами существовали значительные различия в достижениях в образовании (рис. С.1.1). Только 1% учащихся шестых классов в Малави достиг желаемого уровня по чтению, в то время как в Зимбабве эта цифра составила 37%. Почти 4 из 10 учащихся в странах – участниках SACMEQ II достигли минимального уровня мастерства в чтении, и только 1 из 10 достиг желаемого уровня.

Сравнения оценок по грамотности чтения городских и сельских учащихся показали большие различия в пользу городских учащихся в трех странах (Кения, Намибия, Танзания), хотя в Маврикии и на Сейшельских островах различия не были статистически значимыми. Вероятные причины таких различий между результатами учащихся из городских и сельских районов сложные. По сравнению с городскими учащимися, учащиеся в сельских районах имели более низкий уровень социально-экономического статуса семьи, они были старше, и, вероятно, им приходилось чаще оставаться на второй год, они получали меньше поддержки в семье. Более того, сельские школы имели ресурсы более низкого качества по сравнению с городскими школами, что нашло свое отражение в том, как преподаватели оценивали и исправляли домашнюю работу школьников, как часто они встречались с родителями школьников и какую поддержку обеспечивали инспектора (Zhang, 2006).

Интересной особенностью SACMEQ было использование результатов исследований для сравнения ресурсного обеспечения и тенденций достижений в чтении за период времени, отмеченный быстрым ростом учащихся, поступивших в шко-

Рис. С.1.1

Учащиеся 6-х классов, достигшие квалификационных уровней в SACMEQ, 1995–1998 годы, %



Источник: ЮНЕСКО, 2004, рис. 3.1. Воспроизводится с разрешения правообладателя.

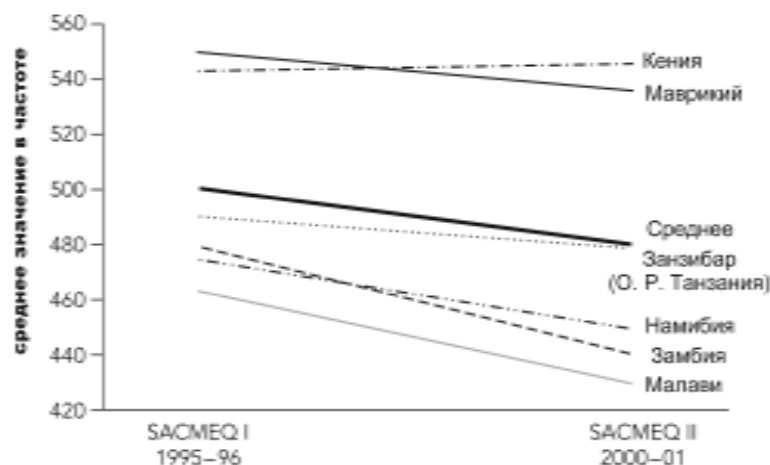
лы в регионе. Все шесть образовательных систем, которые участвовали в исследованиях SACMEQ I (1995) и SACMEQ II (2000), выявили абсолютное увеличение ресурсной обеспеченности школ между двумя проведенными оценками (Murimba, 2005a). В пяти из шести стран национальные средние баллы по грамотности чтения уменьшились (рис. С.1.2). Эти различия были статистически значимыми только в Малави, Намибии и Замбии. В целом в шести странах оценки уменьшились в среднем на 4%.

Каждый национальный доклад предоставлял ряд рекомендаций для тактиков. Например, танзанийский отчет рекомендовал правительству исследовать гендерное неравенство при зачислении в школы и идентифицировать варианты для устранения гендерного разрыва (Mrutu, Ponera and Nkumbi, 2005). Действия правительства должны были включать меры по обеспечению заботы об осиротевших детях, освобождению девочек от тяжелых домашних обязанностей так, чтобы они могли учиться в школе.



Рис. С.1.2

Изменения в оценках грамотности в промежуток времени между SACMEQ I и SACMEQ II



Источник: ЮНЕСКО, 2004, рис. 2.4. Воспроизводится с разрешения правообладателя.

В некоторых странах также оценивалась предметная квалификация преподавателей при помощи тестов, которые предъявляли учащимся. В Танзании менее половины преподавателей достигли высшего уровня (уровень 8) в чтении (46,1 %) и по математике (43,9%).

Результаты исследований SACMEQ были представлены в президентских и национальных комиссиях (в Зимбабве и Намибии), в отчетах премьер-министра и кабинета образовательной политики (в Занзибаре), в национальных исследованиях сектора образования (в Замбии) и в обзорах национального образовательного генерального плана (в Маврикии).

В нескольких странах результаты интерпретировались как показатели потребности в обеспечении стандартов для ресурсов в образовании. Например, в Кении были введены минимальные нормы по оборудованию классной комнаты (столы и учебники для учащихся). В Зимбабве для оборудования классных комнат были введены специальные фонды.

Высокий уровень отсева и низкие его нормы побудили министерство образования Кении усилить свой сектор неофици-

ального образования, чтобы помочь тем, кто не соответствует требованиям официальной системы. Также в Кении результаты исследования SACMEQ по гендерным и региональным различиям и внутренней неэффективности были использованы для руководства планами развития по введению программы «Образование для всех» на национальном, региональном и районном уровнях (Mugimba, 2005a).

## **С.2. ПРОГРАММА АНАЛИЗА СИСТЕМ ОБРАЗОВАНИЯ**

### **Структура**

Программа анализа систем образования (The Programme d'Analyse des Systemes Educatifs de la CONFEMEN, или PASEC) проводится под руководством Конференции министров образования франкоязычных стран мира (Conference des Ministres de l'Education des Pays ayant le Francais en Partage, или CONFEMEN). Начало программы было положено в 1991 году на конференции министров образования франкоязычных стран в Джибути, где было выполнено первое исследование в 1992 году.

Первичной целью программы PASEC являлся сбор информации для принятия решений в образовании и, в частности, для решения важных проблем национальной политики. Она осуществляется путем оценки достижений учащихся и попыток определения ключевых факторов и требуемых расходов, связанных с ними, для установления иерархии потенциальных образовательных мероприятий с точки зрения их эффективности.

Стоит отметить пять особенностей исследований PASEC. Во-первых, они имеют международные масштабы, где предложения по исследованиям в стране рассматривают на встрече государств-членов CONFEMEN. Если предложение одобрено, национальный представитель CONFEMEN становится ответственным за формирование междисциплинарной группы экспертов внутри министерства образования, которая, в свою очередь, будет ответственной за исполнение работы (дизайн

анкет, предъявление тестов и анкет, ввод данных и их анализ, подготовка отчетов). Однако изначально программа PASEC не предусматривала возможность сравнения достижений учащихся по странам.

Во-вторых, учащиеся проходят тестирование как в начале, так и в конце учебного года. Эта система предполагает, что при анализе данных учитываются входные характеристики учащихся, чтобы получить меру прироста их достижений в течение года.

В-третьих, исследования в четырех странах (Гвинее, Мали, Нигерии и Того) были разработаны со специальным замыслом. Например, Гвинея и Того выбрали в качестве своей темы исследования стратегию набора преподавателей (включая программы переподготовки преподавателей), которая была введена в 1983 году в Того и в 1998 в Гвинее. Стратегия предполагала уменьшение затрат по набору большего числа преподавателей, признавая, что эта политика способна повлиять на улучшение качества преподавания.

В-четвертых, начиная с 1995 года один и тот же инструментарий использовался в 5 странах (Буркина-Фасо, Камеруне, Кот-д'Ивуаре, Сенегале [1995/1996] и на Мадагаскаре [1997/1998]), что позволило провести сравнения достижений учащихся на международном уровне.

В-пятых, в двух странах (Кот-д'Ивуаре и Сенегале) репрезентативная выборка, выделенная из числа учащихся вторых классов, тестировалась постоянно в процессе исследования вплоть до 6-го класса (с 1995 по 2000 год) в рамках лонгитюдных исследований.

### **Инструментарий**

Тесты (как с множественным выбором, так и с конструируемыми ответами) были разработаны по математике и французскому языку на основе общих элементов учебных планов франкоязычных стран Африки. Тесты предназначались для предъявления в начале и в конце учебного года во 2-х и 5-х классах. Тесты, проводимые в конце учебного года, включали не-

сколько заданий из предыдущих тестов в дополнение к заданиям, охватывающим годовой учебный курс.

Во вторых классах тесты по французскому языку оценивали словарный запас учащихся по чтению, понимание предложений и текстов и письмо. Тесты, проводимые в 5-х классах, в дополнение к оценке понимания, измеряли также произношение и разделы грамматики у учащихся.

Тесты по математике для 5-х классов включали задания, которые оценивали знания учащихся свойств чисел и их способность выполнять основные вычисления (сложение и вычитание). В тесты были включены также задания, которые требовали от учащихся использования сложения, вычитания, умножения и деления при решении задач, а также задания, которые оценивали знание учащимися десятичных и простых дробей и основных геометрических понятий.

Также на Маврикии проводилось тестирование по арабскому языку и по малагасийскому языку на Мадагаскаре. В Камеруне был предъявлен тест по французскому языку, переведенный на английский язык для англоговорящих учащихся.

Фоновые данные были собраны с помощью анкет, предъявленных учащимся, как по их персональным данным (пол, возраст, питание и язык общения), так и по контекстным факторам (образование родителей, наличие книг дома и расстояние до школы). Также было проведено анкетирование преподавателей (возраст, пол, образование или уровень подготовки), и выявлялось состояние классных комнат.

В анализе фоновые факторы связывались с достижениями учащихся, чтобы попытаться установить соотношения между двумя наборами переменных. Основное внимание было уделено «приросту» или «добавленной величине к достижениям учащихся» в течение годичного курса обучения, вкладу внутришкольных факторов, таких как уровень подготовки преподавателей, размер классов, доступность учебников, и внешних факторов, таких как образование родителей, расстояние от дома до школы, язык, на котором общаются в семье (Bernard, 1999; CONFEMEN, 1999; Kulpo and Coustere, 1999).

### Участники

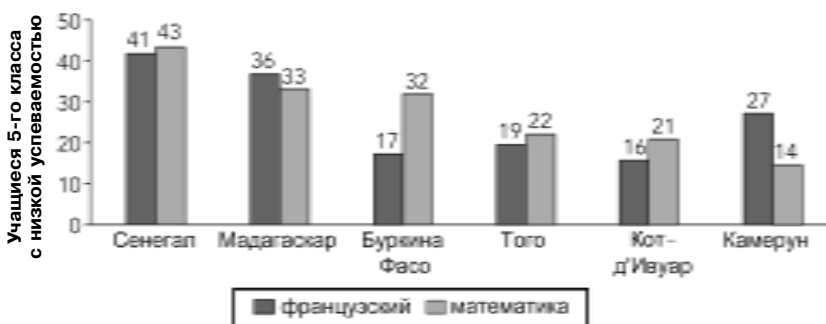
К настоящему времени уже 18 стран приняли участие в деятельности PASEC: Бенин, Буркина-Фасо, Камерун, Центральноафриканская Республика, Чад, Кот-д'Ивуар, Демократическая Республика Конго, Джибути, Габон, Гвинея, Мадагаскар, Мали, Маврикий, Мавритания, Нигер, Республика Конго, Сенегал и Того.

### Некоторые результаты

Результаты показывают низкие уровни достижений по чтению и математике (рис. С.2.1). «Нижний уровень достижений» был определен как оценка ниже 25-го перцентиля в тестах по математике и чтению.

Рис. С.2.1

Учащиеся 5-го класса с низкими достижениями, PASEC, 1996–2001, %



Примечание: Оценка была проведена в Буркина-Фасо, Камеруне, Кот-д'Ивуаре и Сенегале в 1995–1996 годах, на Мадагаскаре – в 1997–1998 годах, а в Того – в 2000–2001 годах. Страны ранжируются по доле плохо успевающих по математике школьников. Низкая успеваемость определяется как оценка ниже 25-го перцентиля по чтению и математике.

Источник: ЮНЕСКО, 2004, рис. 3.32. Воспроизводится с разрешения правообладателя.

Было выполнено несколько видов анализа данных PASEC. В одном из них данные по пяти странам (Буркина-Фасо, Камерун, Кот-д'Ивуара, Мадагаскара и Сенегала) были обработаны с помощью иерархической линейной модели для оценки индивидуального и школьного уровней, а также национальных характеристик, определяющих достижения учащихся пятых

классов по французскому языку и математике (Michaelowa, 2001). Были сделаны следующие выводы.

Во-первых, с достижениями учащихся были связаны некоторые индивидуальные характеристики учащихся и семейные характеристики (включая грамотность родителей и использование в семье учащегося французского языка). Во-вторых, хотя, как может показаться, для учащихся было бы полезно повторно учиться в одном и том же классе, эффект будет временным. В-третьих, как исходное образование преподавателей, так и систематическое повышение квалификации являются очень важными для определения достижений учащихся. В-четвертых, число дней, когда преподаватели отсутствовали в школе, негативно влияет на достижения учащихся. В-пятых, даже несмотря на то что преподаватели, нанимаемые родителями, получали меньшую зарплату, их работа была более эффективна, чем работа преподавателей, являющихся государственными служащими.

В-шестых, членство преподавателей в Союзе учителей имело большое значение и негативно отражалось на достижениях учащихся. В-седьмых, наличие учебников у учащихся оказывало сильное позитивное воздействие на учебные достижения. В-восьмых, размер класса (не более 62 учащихся) был положительно связан с достижениями. В-девятых, обучение во внесезонных классах оказывало положительное воздействие на достижения учащихся. В-десятых, учащиеся в школах, которые посещал инспектор в течение учебного года, имели лучшие результаты выполнения теста, чем те, чьи школы не посещались. И наконец, достижения девочек, кажется, улучшались, когда их преподаватели были женского пола, а достижения мальчиков улучшались, соответственно, с преподавателями мужского пола.

### **С.3. ЛАТИНОАМЕРИКАНСКАЯ ЛАБОРАТОРИЯ ПО ЭВАЛЮАЦИИ КАЧЕСТВА ОБРАЗОВАНИЯ**

#### **Структура**

Первое международное сравнительное исследование по языку и математике в Латинской Америке было проведено Латин-

ноамериканской лабораторией по эвалюации качества образования (Laboratorio Latinoamericano de Evaluacion de la Calidad de la Educacion, или LLECE). Эта сеть национальных систем образования в странах Латинской Америки и Карибского бассейна была создана в 1994 году и координируется региональным офисом ЮНЕСКО в Латинской Америке и странах Карибского бассейна.

Основной целью данного исследования было обеспечение информации о достижениях учащихся и связанных с ними факторов, которые могут быть полезны в разработке и осуществлении политики в области образования внутри этих стран. Это должно было быть сделано с помощью оценивания достижений по популяциям учащихся начальных школ путем ответа на следующие вопросы. Что изучают школьники? Какие уровни обученности встречаются? Какие умения приобрели учащиеся? Когда происходит обучение? При каких условиях происходит обучение? (Casassus and others, 1998).

Сравнительная структура рассматривалась как один из лучших способов повышения понимания состояния образования в странах. О необходимости проведения международного исследования в Латинской Америке свидетельствует тот факт, что благодаря участию некоторых стран было выявлено, что в исследовании не учитывали особенности учебных планов, характерные для этого региона.

### **Инструментарий**

Для проведения исследования были разработаны тесты (два варианта) достижений по математике и языку, в которых было представлено содержание учебных планов каждой из стран – участниц. Тесты включали задания с множественным выбором и открытые задания (только по языку).

Языковые компоненты включали понимание прочитанного, метаязыковую практику и воспроизведение текста на испанском языке, кроме Бразилии, где учащиеся тестировались на португальском.

Математические компоненты включали числа, операции с использованием натуральных чисел, обыкновенные дроби, геометрию и измерения.

Подробная информация была собрана с помощью анкет (заполненных учащимися, преподавателями, родителями или опекунами, директорами школ) по факторам, которые рассматривались как связанные с достижениями учащихся (например, вид школы и ее месторасположение, уровень образования родителей или опекунов, а также представления учащихся и преподавателей о доступности учебных ресурсов в школе).

#### **Участники**

В 1997 году 13 стран приняли участие в исследовании: Аргентина, Боливия, Бразилия, Чили, Колумбия, Коста-Рика, Куба, Доминиканская Республика, Гондурас, Мексика, Парагвай, Перу и Венесуэла. Данные по 11 странам были включены в первый отчет по исследованию. В каждой стране были оценены целевые подгруппы школьников: около 4000 учащихся 3-х классов (восьми- и девятилетнего возраста) и 4-х классов (девяти- и десятилетних). Самая старшая часть учащихся (20%) была исключена из исследования (Casassus and others, 1998, 18).

#### **Некоторые результаты**

Результаты, классифицированные по виду посещаемой школы (частная или государственная) и месторасположению (в городах с населением более одного миллиона или сельской местности), показывают, что образовательные достижения учащихся на Кубе, независимо от расположения школы, намного выше достижений учащихся других стран (табл. С.3.1 и С.3.2). Более 90% кубинских учащихся достигли высшего уровня (уровень III) по владению языком. Исключение составляют сельские школы, где более 75% учащихся достигли того же уровня и по математике. В то время как 72% сельских учащихся Кубы достигли уровня III по математике, меньше чем 10% учащихся из сельских районов других стран достигли такого же уровня.



ТАБЛИЦА С.3.1

Учащиеся, которые достигли каждого классификационного уровня по языку по видам школ и их расположению, LLECE, 1997 год, %

Страна	Общественные			Частные			Мегаполис			Городские			Сельские		
	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III
Аргентина	95	77	57	99	93	78	96	85	72	96	79	59	88	62	42
Боливия	87	55	30	91	70	46	90	66	39	87	58	35	77	40	24
Бразилия	95	80	54	98	93	72	96	88	62	95	82	58	84	62	38
Чили	93	71	49	97	86	67	94	76	53	95	79	60	89	63	41
Колумбия	89	59	35	97	81	56	96	79	53	89	60	36	89	57	33
Куба	100	98	92	н.д.	н.д.	н.д.	100	99	93	100	98	92	100	98	92
Доминиканская республика	77	52	30	83	64	42	84	65	42	73	44	25	73	39	20
Гондурас	87	55	29	94	73	44	92	67	38	87	55	29	78	35	17
Мексика	89	58	38	96	84	65	94	70	50	89	64	43	82	48	30
Парагвай	88	60	37	93	75	54	н.д.	н.д.	н.д.	90	67	44	81	51	32
Перу	86	55	29	94	78	54	92	70	43	85	57	34	71	30	13
Венесуэла	88	59	38	91	70	49	91	68	48	88	60	38	84	58	39

Примечание: н.д. – нет данных.

Источник: ЮНЕСКО, 2001, табл. 8.

**ТАБЛИЦА С.3.2**

**Учащиеся, которые достигли каждого классификационного уровня по математике по видам школ и их расположению, LLECE, 1997 год, %**

Страна	Общественные			Частные			Мегаполис			Городские			Сельские		
	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III	Уровень I	Уровень II	Уровень III
Аргентина	96	54	12	98	71	23	98	70	26	96	54	11	94	43	6
Боливия	93	43	9	96	59	18	95	49	12	94	51	14	89	36	8
Бразилия	93	52	12	97	67	26	96	58	17	94	55	15	84	40	7
Чили	92	46	7	97	57	15	94	49	10	95	52	12	87	38	6
Колумбия	93	42	5	97	55	10	97	53	8	93	43	6	92	50	12
Куба	100	92	79	н.д.	н.д.	н.д.	100	95	82	99	90	76	99	50	72
Доминиканская республика	82	37	4	86	43	7	86	42	6	81	36	4	79	38	7
Гондурас	84	36	7	93	39	5	87	35	3	86	39	8	78	23	13
Мексика	94	55	10	98	69	20	97	62	13	94	58	13	90	46	10
Парагвай	87	29	2	90	49	12	н.д.	н.д.	н.д.	88	42	9	82	34	8
Перу	87	29	2	94	54	11	88	43	8	89	33	4	78	23	2
Венесуэла	76	25	2	76	33	5	75	26	3	77	27	3	68	22	2

Примечание: н.д. – нет данных.

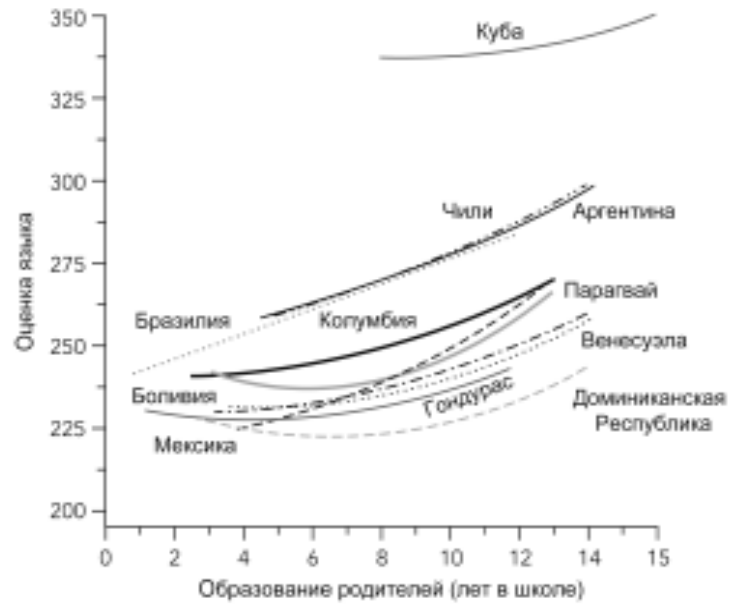
Источник: ЮНЕСКО, 2001, табл. 8.

Дальнейший анализ данных LLECE был направлен на выявление степени, в которой связи между социально-экономическим статусом (основанном на уровне образования и достижениях родителей) менялись от страны к стране (рис. С.3.1). Данные показывают, что социально-экономические градиенты существенно меняются по странам. Связь более явно выражена в Аргентине и Бразилии, чем на Кубе, которая имела сравнительно малую долю дисперсии в уровне образования родителей. Хотя учащиеся в частных школах уступают учащимся в государственных школах, различия между этими группами не были статистически значимыми, когда принимался во внимание социально-экономический статус (Саммит Америки, 2003). На Кубе была наименьшая вариация в образовательных достижениях родителей и учащиеся продемонстрировали самые высокие уровни достижений. Дальнейшие исследования показали, что на Кубе, по сравнению с другими странами, прослеживается больше заботы об учениках в течение дня, более активное обучение дома, обучение проводится в меньших классах, с наличием более квалифицированных преподавателей и меньшим количеством круглогодичных классов или классов, сгруппированных по способностям (Willms and Somers, 2001). В последующих исследованиях результаты LLECE были использованы для определения школ с наиболее выдающимися результатами в семи странах: Аргентина, Боливия, Чили, Колумбия, Коста-Рика, Куба и Венесуэла (LLECE, 2002).

Несмотря на это разнообразие видов анализа, представители целевой группы по реформе системы образования в Центральной Америке (2000, 19) в своем докладе под названием «Завтра будет слишком поздно» отметили, что почти в каждом случае нет четких политических указаний на то, как результаты эвалюации могут и должны быть использованы. Тесты академических достижений еще не стали частью отчетности в политике, которая востребована различными группами. Не было обсуждений по типу решений, которые могут быть основаны на этих результатах, и не достигнуто даже небольшое согласие по выбранной ценности результатов тестирования уча-

Рис. С.3.1

Социо-экономические градиенты для 11 латиноамериканских стран, LLECE



Источник: Willms and Somers, 2005.

щихся. В результате, эти программы являются особенно уязвимыми к изменениям в правительстве и даже в руководстве министерств.





## СПИСОК ЛИТЕРАТУРЫ

Arregui, P., and C. McLauchlan. 2005. "Utilization of Large-Scale Assessment Results in Latin America." Unpublished document prepared for the Partnership for Educational Revitalization in the Americas and the World Bank Institute.

Beaton, A.E., T.N. Postlethwaite, K.N. Ross, D. Spearritt, and R.M. Wolf. 1999. *The Benefits and Limitations of International Educational Achievement Studies*. Paris: UNESCO International Institute for Educational Planning.

Benveniste, L. 2000. "Student Assessment as a Political Construction: The Case of Uruguay." *Education Policy Analysis Archives* 8 (32): 1–41.

———. 2002. "The Political Structuration of Assessment: Negotiating State Power and Legitimacy." *Comparative Education Review* 46: 89–118.

Bernard, J.-M. 1999. "Les Enseignants du Primaire dan Cinq Pays du Programme d'Analyse des Systemes Educatifs de la CONFEMEN: Le Role du Maitre dans le Processus d'Acquisition des Eleves." Report of the Working Group on the Teaching Profession, Francophone Section, of the Association for the Development of Education in Africa (ADEA). Paris: ADEA.

Bhutan, Board of Examinations, Ministry of Education. 2004. *National Educational Assessment in Bhutan: A Benchmark of Student Achievement in Literacy and Numeracy at Class 6, 2003*. Thimphu, Bhutan: Ministry of Education.

Braun, H., and A. Kanjee. 2007. "Using Assessment to Improve Education in Developing Countries." In *Educating All Children: A Global Agenda*, ed. J.E. Cohen, D.E. Bloom, and M.B. Malin, 303–53. Cambridge, MA: MIT Press.

Campbell, J.R., D.L. Kelly, I. V.S. Mullis, M.O. Martin, and M. Sainsbury. 2001. *Framework and Specifications for PIRLS Assessment 2001*. 2nd ed. Chestnut Hill, MA: Boston College.

Casassus, J., J.E. Froemel, J.C. Palafox, and S. Cusato. 1998. *First International Comparative Study of Language, Mathematics, and Associated Factors in Third and Fourth Grades*. Santiago, Chile: Latin American Laboratory for Evaluation of the Quality of Education.

Chinapah, V. 1997. *Handbook on Monitoring Learning Achievement: Towards Capacity Building*. Paris: United Nations Educational, Scientific, and Cultural Organization.

Clarke, M. 2005. *NAPE Technical Analysis and Recommendations*. Kampala: Uganda National Examinations Board.

CONFEMEN (Conference des Ministres de l'Education des Pays ayant le Francais en Partage). 1999. *Les Facteurs de l'Effi cacite dans l'Enseignement Primaire: Les Resultats du Programme PASEC sur Neuf Pays d'Afrique et de l'Ocean Indien*. Dakar: CONFEMEN.

Connecticut Department of Education. 2006. "State Releases Connecticut Mastery Test Results." News, August 9. <http://www.sde.ct.gov/sde/lib/sde/PDF/PressRoom/2006cmtresults.pdf>.

Coulombe, S., J.-F. Tremblay, and S. Marchand. 2004. *International Adult Literacy Survey: Literacy Scores, Human Capital, and Growth across Fourteen OECD Countries*. Ottawa: Statistics Canada.

Crespo, M., J.F. Soares, and A. de Mello e Souza. 2000. "The Brazilian National Evaluation System of Basic Education: Context, Process, and Impact." *Studies in Educational Evaluation* 26: 105–25.

- Delannoy, F. 2000. *Education Reforms in Chile 1980–98: A Lesson in Pragmatism*. Washington, DC: World Bank.
- Eivers, E., G. Shiel, R. Perkins, and J. Cosgrove. 2005. *The 2004 National Assessment of English Reading*. Dublin: Educational Research Centre.
- Elley, W.B. 1992. *How in the World Do Students Read? IEA Study of Reading Literacy*. The Hague, Netherlands: International Association for the Evaluation of Educational Achievement.
- , ed. 1994. *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Oxford, U.K.: Pergamon.
- . 2005. “How TIMSS-R Contributed to Education in Eighteen Developing Countries.” *Prospects* 35 (2): 199–212.
- Ethiopia, National Organisation for Examinations. 2005. *Second National Learning Assessment of Ethiopia*. Addis Ababa: National Organisation for Examinations.
- Ferrer, G. 2006. *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Washington, DC: Partnership for Educational Revitalization in the Americas.
- Ghana, Ministry of Education, Youth, and Sports. 2004. *Results from Ghanaian Junior Secondary 2 Students’ Participation in TIMSS 2003 in Mathematics and Science*. Accra: Ministry of Education, Youth, and Sports.
- Greaney, V., and T. Kellaghan. 1996. *Monitoring the Learning Outcomes of Education Systems*. Washington, DC: World Bank.
- Hanushek, E.A., and D.D. Kimko. 2000. “Schooling, Labor-Force Quality, and the Growth of Nations.” *American Economic Review* 90 (5): 1184–208.
- Hanushek, E.A., and L. Wossmann. 2007. *Education Quality and Economic Growth*. Washington, DC: World Bank.
- Himmel, E. 1996. “National Assessment in Chile.” In *National Assessments: Testing the System*, ed. P. Murphy, V. Greaney, M.E. Lockheed, and C. Rojas, 111–28. Washington, DC: World Bank.
- . 1997. “Impacto Social de los Sistemas de Evaluación del Rendimiento Escolar: El Caso de Chile.” In *Evaluación y reforma educativa: Opciones de*



*politica*, ed. B. Alvarez H. and M. Ruiz-Casares, 125–57. Washington, DC: ABEL/PREAL/U.S. Agency for International Development.

Horn, R., L. Wolff, and E. Velez. 1992. “Educational Assessment Systems in Latin America: A Review of Issues and Recent Experience.” *Major Project of Education in Latin America and the Caribbean Bulletin* 27: 7–27.

Howie, S. 2000. “TIMSS-R in South Africa: A Developing Country Perspective.” Paper presented at American Educational Research Association annual meeting, New Orleans, April 24–28.

———. 2002. “English Proficiency and Contextual Factors Influencing Mathematics Achievement of Secondary School Pupils in South Africa.” PhD thesis, University of Twente, the Netherlands.

Howie, S., and C. Hughes. 2000. “South Africa.” In *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*, ed. D. Robitaille, A. Beaton, and T. Plomp, 139–45. Vancouver, BC: Pacific Educational Press.

Hoxby, C.E. 2002. “The Cost of Accountability.” Working Paper 8855, National Board of Economic Research, Cambridge, MA.

Husen, T. 1973. “Foreword.” In *Science Achievement in Nineteen Countries*, ed. L. C. Comber and J. P. Keeves, 13–24. New York: Wiley.

Husen, T., and T.N. Postlethwaite. 1996. “A Brief History of the International Association for the Evaluation of Educational Achievement (IEA).” *Assessment in Education* 3 (2): 129–41.

IEA (International Association for the Evaluation of Educational Achievement). 2000. *Framework and Specifications for PIRLS Assessment 2001*. Chestnut Hill, MA: International Study Center, Boston College.

IIEP (International Institute for Educational Planning). 2007. “Southern and Eastern Africa Consortium for Monitoring Educational Quality.” IIEP, Paris. <http://www.unesco.org/iiep/eng/networks/sacmeq/sacmeq.htm>.

Ilon, L. 1996. “Considerations for Costing National Assessments.” In *National Assessment: Testing the System*, ed. P. Murphy, V. Greaney, M. E. Lockheed, and C. Rojas, 69–88. Washington, DC: World Bank.

India, National Council of Educational Research and Training, Department of Educational Measurement and Evaluation. 2003. *Learning*

*Achievement of Students at the End of Class V*. New Delhi: Department of Educational Measurement and Evaluation.

Ishino, T. 1995. "Japan." In *Performance Standards in Education: In Search of Quality*, 149–61. Paris: OECD.

Johnson, E. G. 1992. "The Design of the National Assessment of Educational Progress." *Journal of Educational Measurement* 29 (2): 95–110.

Jones, L. V. 2003. "National Assessment in the United States: The Evolution of a Nation's Report Card." In *International Handbook of Educational Evaluation*, ed. T. Kellaghan and D.L. Stufflebeam, 883–904. Dordrecht, Netherlands: Kluwer Academic.

Kanjee, A. 2006. "The State of National Assessments of Learner Achievement." Unpublished paper prepared for the Human Sciences Research Council, Pretoria, South Africa.

Keeves, J.P. 1995. "The Contribution of IEA Research to Australian Education." In *Reflections on Educational Achievement: Papers in Honour of T. Neville Postlethwaite*, ed. W. Bos and R.H. Lehmann, 137–58. New York: Waxman.

Kellaghan, T. 1996. "IEA Studies and Educational Policy." *Assessment in Education* 3 (2): 143–60.

———. 1997. "Seguimiento de los resultados educativos nacionales." In *Evaluacion y reforma educativa: Opciones de politica*, ed. B. Alvarez H. and M. Ruiz-Casares, 23–65. Washington, DC: ABEL/PREAL/U.S. Agency for International Development.

———. 2003. "Local, National and International Levels of System Evaluation: Introduction." In *International Handbook of Educational Evaluation*, ed. T. Kellaghan and D. L. Stufflebeam, 873–82. Dordrecht, Netherlands: Kluwer Academic.

———. 2006. "What Monitoring Mechanisms Can Be Used for Cross-National (and National) Studies?" In *Cross-National Studies of the Quality of Education: Planning Their Design and Managing Their Impact*, ed. K.N. Ross and I.J. Genevois, 51–55. Paris: International Institute for Educational Planning.

Kellaghan, T., and V. Greaney. 2001a. "The Globalisation of Assessment in the 20th Century." *Assessment in Education* 8 (1): 87–102.

———. 2001b. *Using Assessment to Improve the Quality of Education*. Paris: International Institute for Educational Planning.

———. 2004. *Assessing Student Learning in Africa*. Washington, DC: World Bank.

Khaniya, T., and J.H. Williams. 2004. "Necessary but Not Sufficient: Challenges to (Implicit) Theories of Educational Change—Reform in Nepal's Education System." *International Journal of Educational Development* 24 (3): 315–28.

Kirsch, I. 2001. *The International Adult Literacy Study (IALS): Understanding What Was Measured*. Princeton, NJ: Educational Testing Service.

Kulpoo, D., and P. Coustere. 1999. "Developing National Capacities for Assessment and Monitoring through Effective Partnerships." In *Partnerships for Capacity Building and Quality Improvements in Education: Papers from the ADEA 1997 Biennial Meeting, Dakar*. Paris: Association for the Development of Education in Africa.

Lesotho, Examinations Council of Lesotho and National Curriculum Development Centre. 2006. *Lesotho: National Assessment of Educational Progress, 2004*. Maseru: Examinations Council of Lesotho and National Curriculum Development Centre.

LLECE (Latin American Laboratory for Evaluation of the Quality of Education). 2002. *Qualitative Study of Schools with Outstanding Results in Seven Latin American Countries*. Santiago: LLECE.

Lockheed, M.E., and A. Harris. 2005. "Beneath Education Production Functions: The Case of Primary Education in Jamaica." *Peabody Journal of Education* 80 (1): 6–28.

Makuwa, D. 2005. *The SACMEQ II Project in Namibia: A Study of the Conditions of Schooling and Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.

McMeekin, R.W. 2000. *Implementing School-Based Merit Awards: Chile's Experiences*. Washington, DC: World Bank.

Michaelowa, K. 2001. "Primary Education Quality in Francophone Sub-

Saharan Africa: Determinants of Learning Achievement and Efficiency Considerations.” *World Development* 29 (10): 1699–716.

Mrutu, A., G. Ponera, and E. Nkumbi. 2005. *The SACMEQ II Project in Tanzania: A Study of the Conditions of Schooling and the Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Mullis, I.V.S., A.M. Kennedy, M.O. Martin, and M. Sainsbury. 2006. *PIRLS 2006: Assessment Framework and Specifications*. Chestnut Hill, MA: 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: International Study Center, Boston College.

Mullis, I.V.S., M.O. Martin, E.J. Gonzalez, and A.M. Kennedy. 2003. *PIRLS 2001 International Report: IEA’s Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, MA: International Study Center, Boston College.

Mullis, I.V.S., M.O. Martin, G.J. Ruddock, C.Y.O’Sullivan, A. Arora, and E. Erberber. 2005. *TIMSS 2007 Assessment Frameworks*. Chestnut Hill, MA: International Study Center, Boston College.

Murimba, S. 2005a. “The Impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).” *Prospects* 35 (1): 91–108.

———. 2005b. “The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ): Mission Approach and Projects.” *Prospects* 35 (1): 75–89.

Nassor, S., and K.A. Mohammed. 1998. *The Quality of Education: Some Policy Suggestions Based on a Survey of Schools—Zanzibar*. SACMEQ Policy Research 4, International Institute for Educational Planning, Paris.

Naumann, J. 2005. “TIMSS, PISA, PIRLS, and Low Educational Achievement in World Society.” *Prospects* 35 (2): 229–48.

OECD (Organisation for Economic Cooperation and Development). 2001. *Outcomes of Learning: Results from the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy*. Paris:

OECD. <http://nces.ed.gov/pubs/2002/2002115.pdf>.

———. 2003. *The PISA 2003 Assessment Framework: Reading, Mathematics, Science and Problem Solving Knowledge and Skills*. Paris: OECD.

———. 2004a. *First Results from PISA 2003: Executive Summary*. Paris: OECD. <http://www.oecd.org/dataoecd/1/63/34002454.pdf>.

———. 2004b. *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.

———. 2007. "Sample Questions: PISA Mathematics with Marking Guide." OECD, Paris. <http://pisa-sq.acer.edu.au>.

OECD (Organisation for Economic Co-operation and Development) and UNESCO (United Nations Educational, Scientific, and Cultural Organization) Institute for Statistics. 2003. *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris and Montreal: OECD and UNESCO Institute for Statistics.

Olivares, J. 1996. "Sistema de Medicion de la Calidad de la Educacion de Chile: SIMCE, Algunos Problemas de la Medicion." *Revista Iberoamericana de Educacion* 10. <http://www.rieoei.org/oeivirt/rie10a07.htm>.

Passos, A., T. Nahara, F. Magaia, and C. Lauchande. 2005. *The SACMEQ II Project in Mozambique: A Study of the Conditions of Schooling and the Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Perera, L., S. Wijetunge, W.A. de Silva, and A.A. Navaratne. 2004. *Achievement after Four Years of Schooling. National Assessment of Achievement of Grade Four Pupils in Sri Lanka: National Report*. Colombo: National Education Research and Evaluation Centre, University of Colombo.

Postlethwaite, T.N. 2004. "What Do International Assessment Studies Tell Us about the Quality of School Systems?" Background paper for *Education for All Global Monitoring Report 2005*, United Nations Educational, Scientific, and Cultural Organization, Paris.

Prakash, V., S.K.S. Gautam, and I.K. Bansal. 2000. *Student Achievement under MAS: Appraisal in Phase-II States*. New Delhi: National Council of Educational Research and Training.

- Ramirez, F.O., X. Luo, E. Schofer, and J.W. Meyer. 2006. "Student Achievement and National Economic Growth." *American Journal of Education* 113 (1): 1–29.
- Ravela, P. 2005. "A Formative Approach to National Assessments: The Case of Uruguay." *Prospects* 35 (1): 21–43.
- Reddy, V. 2005. "Cross-National Achievement Studies: Learning from South Africa's Participation in the Trends in International Mathematics and Science Study." *Compare* 35 (1): 63–77.
- . 2006. *Mathematics and Science Achievement at South African Schools in TIMSS 2003*. Capetown, South Africa: Human Sciences Research Council Press.
- Robitaille, D.F., A.E. Beaton, and T. Plomp, eds. 2000. *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*. Vancouver, BC: Pacific Educational Press.
- Rojas, C., and J.M. Esquivel. 1998. "Los Sistemas de Medicion del Logro Academico en Latino America." LCSHD Paper 25, Washington, DC: World Bank.
- Ross, K. 1987. "Sample Design." *International Journal of Educational Research* 11 (1): 57–75.
- Ross, K., and T.N. Postlethwaite. 1991. *Indicators of the Quality of Education: A Study of Zimbabwean Primary Schools*. Harare: Ministry of Education and Culture; Paris: International Institute for Educational Planning.
- Shabalala, J. 2005. *The SACMEQ II Project in Swaziland: A Study of the Conditions of Schooling and the Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.
- Shukla, S., V.P. Garg, V.K. Jain, S. Rajput, and O.P. Arora. 1994. *Attainments of Primary School Children in Various States*. New Delhi: National Council of Educational Research and Training.
- Sofroniou, N., and T. Kellaghan. 2004. "The Utility of Third International Mathematics and Science Study Scales in Predicting Students' State Examination Performance." *Journal of Educational Measurement* 41 (4): 311–29.

Straus, M. 2005. "International Comparisons of Student Achievement as Indicators for Educational Policy in Slovenia." *Prospects* 35 (2): 187–98.

Summit of Americas. 2003. *Regional Report: Achieving the Educational Goals*. Santiago: Ministry of Education, Chile; Paris: United Nations Educational, Scientific, and Cultural Organization.

Task Force on Education Reform in Central America. 2000. *Tomorrow Is Too Late*. <http://thediologue.org/publications/preal/tomorrow.pdf>.

UNEB (Uganda National Examinations Board). 2006. *The Achievements of Primary School Pupils in Uganda in English Literacy and Numeracy*. Kampala: UNEB.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 1990. *Final Report of the World Congress on Education for All: Meeting Basic Learning Needs*, Jomtien, Thailand. Paris: UNESCO.

———. 2000. *The Dakar Framework for Action—Education for All: Meeting Our Collective Commitments*. Paris: UNESCO.

———. 2001. *Technical Report of the First International Comparative Study*. Santiago: Regional Office for Latin America and the Caribbean.

———. 2002. *EFA Global Monitoring Report 2002: Is the World on Track?* Paris: UNESCO.

———. 2004. *EFA Global Monitoring Report 2005: The Quality Imperative*. Paris: UNESCO.

U.S. National Center for Education Statistics. 2005. *National Assessment of Educational Progress: The Nation's Report Card, Reading 2005*. Washington, DC: U.S. National Center for Education Statistics.

———. 2006. "NAEP Overview." U.S. National Center for Education Statistics, Washington, DC. <http://nces.ed.gov/nationsreportcard/about/>.

———. n.d. "Comparing NAEP, TIMSS, and PISA in Mathematics and Science." U.S. National Center for Education Statistics, Washington, DC. [http://nces.ed.gov/timss/pdf/naep\\_timss\\_pisa\\_comp.pdf](http://nces.ed.gov/timss/pdf/naep_timss_pisa_comp.pdf).

Wilkins, J. L. M., M. Zembylas, and K. J. Travers. 2002. "Investigating Correlates of Mathematics and Science Literacy in the Final Year of Secondary School." In *Secondary Analysis of the TIMSS Data*, ed.

- D. F. Robitaille and A.E. Beaton, 291–316. Dordrecht, Netherlands: Kluwer Academic.
- Willms, J.D., and M.-A. Somers. 2005. “Raising the Learning Bar in Latin America: Measuring Student Outcomes.” Policy Brief, Canadian Research Institute for Social Policy, University of New Brunswick, Fredericton.
- Winograd, P., and B. Thorstensen. 2004. “Using Large Scale Assessments to Inform the Policies and Practices That Support Student Learning.” Working paper developed for the International Reading Association and the World Bank’s Global National Assessment Training Project, Office of Education Accountability, Santa Fe, NM.
- Wolff, L. 1998. “Educational Assessment in Latin-America: Current Progress and Future Challenges.” Working Paper 11, Programa de Promocion de la Reforma Educativa en America Latina y el Caribe, Partnership for Educational Revitalization in the Americas, Washington, DC.
- World Bank. 2004. *Vietnam: Reading and Mathematics Assessment Study*. Vols. 1–3. Washington, DC: World Bank.
- . 2007. EdStats database. <http://www1.worldbank.org/education/edstats/>.
- World Declaration on Education for All*. 1990. Adopted by the World Conference on Education for All, Meeting Basic Learning Needs, Jomtien, Thailand, March 5–9. New York: United Nations Educational, Scientific, and Cultural Organization. [http://www.unesco.org/education/information/nfsunesco/pdf/JOMTIE\\_E.PDF](http://www.unesco.org/education/information/nfsunesco/pdf/JOMTIE_E.PDF).
- Zhang, Y. 2006. “Urban-Rural Literacy Gaps in Sub-Saharan Africa: The Roles of Socioeconomic Status and School Quality.” *Comparative Education Review* 50 (4): 581–602.



*Научное издание*

Национальная оценка учебных достижений

**Грини Винсент**

**Келлаган Томас**

**ОЦЕНКА ОБРАЗОВАТЕЛЬНЫХ  
ДОСТИЖЕНИЙ НА НАЦИОНАЛЬНОМ  
УРОВНЕ**

Ответственная за выпуск, литературный редактор *С.А. Есина*

Корректор *Г.В. Ежова*

Компьютерная верстка *И.В. Кравченко*

Оформление *И.В. Кравченко*

Подписано в печать 18.02.2011. Формат 70x100/16

Печать офсетная. Бумага офсетная. 13 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](mailto:universitas@mail.ru)

Дополнительная информация на сайте: <http://logosbook.ru>