

**МЕЖДУНАРОДНАЯ ТЕСТОВАЯ КОМИССИЯ**

**Руководство Международной тестовой комиссии  
по контролю качества подсчета баллов, анализа  
результатов и информирования о них**

**ФИНАЛЬНАЯ ВЕРСИЯ**

**Перевод:**

**Коваленко Алексей**

**Рахубовская Кристина**

**Сидорук Евгения**

**Тихоненко Дарья**

**ИЮНЬ 2012**

© 2012, Международная тестовая комиссия.

Следует цитировать как:

**INTERNATIONAL TEST COMMISSION (2011). ITC GUIDELINES FOR QUALITY  
CONTROL IN SCORING, TEST ANALYSIS, AND REPORTING OF TEST SCORES.  
[[HTTP://WWW.INTESTCOM.ORG](http://www.intestcom.org)]**

## **БЛАГОДАРНОСТЬ**

Руководство было разработано Ави Аллалуф для Совета Международной тестовой комиссии. Автор благодарен Марис Борн за ее ценное содействие в осуществлении этого проекта, а также людям, которые анализировали предшествующие версии и чей вклад является важным для разработки данного руководства:

Альваро Арце-Феррер  
Джеймс Остин  
Джо-Анн Бейрд  
Джулия Балбони  
Хелен Барон  
Дэйв Бартрам  
Марис Борн  
Джеймс Бутчер  
Джанет Карлсон  
Айен Койн  
Курт Гайзингер  
Рон Хамблтон  
Джон Хатти  
Фред Леонг  
Джейсон Ламприну  
Том Окленд  
Фред Освальд  
Кристофер Раш

Автор также обязан своим коллегам из NITE (Национальный институт тестирования и оценки), особенно из департамента подсчета баллов и сравнения, где были разработаны и обычно применяются многие процедуры контроля качества.

## Содержание

### 1. ВВЕДЕНИЕ

- 1.1. Цель и задачи
- 1.2. Для кого предназначено Руководство по Контролю Качества (РКК)
- 1.3. Локальные и международные факторы
- 1.4. Ошибки и необходимость в РКК
- 1.5. Область применения РКК
- 1.6. Определение контроля качества
- 1.7. Примеры из других профессий
- 1.8. Структура РКК

### 2. ОБЩИЕ ПРИНЦИПЫ

- 2.1. Проверка стандартов контроля качества, используемых в настоящее время
- 2.2. Основные приготовления и соглашения между вовлеченными лицами
- 2.3. Ресурсы
- 2.4. Потребности и ожидания заинтересованных сторон
- 2.5. Персонал и рабочая атмосфера
- 2.6. Независимый мониторинг процедур контроля качества
- 2.7. Фиксация ошибок

### 3. ПОШАГОВОЕ РАБОЧИЕ РЕКОМЕНДАЦИИ

- 3.1. Разработка отчетов о результатах
- 3.2. Демографические данные
- 3.3. Подсчет баллов
  - 3.3.1. Получение и хранение ответов испытуемых
  - 3.3.2. Подсчет баллов
- 3.4. Анализ теста
  - 3.4.1. Анализ заданий теста
  - 3.4.2. Анализ результатов тестов действия (performance tests), практических заданий, деловых игр, интервью и т.д
  - 3.4.3. Уравнивание (Equating)/Проверка новых форм теста и заданий
  - 3.4.4. Подсчет стандартизированных баллов
  - 3.4.5. Проверка теста на защищенность от фальсификации
- 3.5. Информирование о результатах
  - 3.5.1. Информирование о результатах теста
  - 3.5.2. Документация

### 4. ВЫВОДЫ И ЗАКЛЮЧИТЕЛЬНЫЕ КОММЕНТАРИИ

### 5. ЛИТЕРАТУРА

# 1. ВВЕДЕНИЕ

## 1.1. Цель и задачи

Стандартизация и точность существенны на всех этапах тестирования, начиная с разработки теста и его применения, вплоть до подсчета баллов, анализа, интерпретации результатов и информирования о них. Любой человек, вовлеченный в подсчет, анализ результатов теста и информирование о результатах, несет ответственность за соблюдение профессиональных стандартов, которые могут быть обоснованы соответствующими заинтересованными сторонами, включая заказчиков, психологические ассоциации, колледжи и университеты, правительственные учреждения и юридические лица. Практикующий профессионал должен знать и предвидеть ошибки, которые могут иметь место на любом этапе, и должен действовать в соответствии с существующими на данный момент стандартами, чтобы предвидеть и устранить эти ошибки. Неточный подсчет баллов как следствие неправильного ключа к ответам, некорректной конвертации сырых баллов в стандартизированные, ошибок в вычислении, случайного причисления баллов другому испытуемому или неверной интерпретации указанных баллов – это примеры ошибок, которые не должны произойти. Человеку свойственно ошибаться, но ошибки должны быть сведены к минимуму через использование адекватных процедур контроля качества. Психологи должны иметь обширные знания о путях и способах контроля качества, так как это имеет решающее значение для правильного применения тестов. Мы полагаем, что этот документ также внесет весомый вклад в процесс Непрерывного улучшения качества (НУК, CQI, Continuous Quality Improvement) в области, в которой мы стремимся добиться прогресса, шаг за шагом.

Руководство по Контролю Качества (РКК), представленное ниже, предназначено для улучшения эффективности, четкости и точности подсчета баллов, анализа результатов и информирования о них (ПАИ) в процессе тестирования. Оно имеет двойную функцию: может быть использовано само по себе, как конкретные указания для контроля качества ПАИ тестов; также оно может рассматриваться и использоваться как дополнение определенных частей Руководства Международной тестовой комиссии по использованию тестов (2000). Рекомендуем читателю ознакомиться с руководствами Международной тестовой комиссии, со стандартами AERA, APA и NCME (2011), а также с другими соответствующими международными и национальными стандартами.

## 1.2. Для кого предназначено Руководство по Контролю Качества (РКК)

РКК распространяется на ситуации крупномасштабного тестирования, в первую очередь на те, в которых используются тесты успешности или способностей (в отличие от личностных опросников). Как таковые, они в особенности применимы в процессе крупномасштабного образовательного тестирования и оценки способностей и навыков при приеме на работу. Тем не менее, многие из последующих рекомендаций могут также быть применены в процессе менее масштабного тестирования и в оценке, включающей в себя другие виды тестов.

РКК предназначено для лиц, ответственных за следующие аспекты использования тестов:

- Разработка тестов
- Проведение тестирования
- Подсчет баллов
- Анализ психометрических характеристик теста (включая нормообразование и стандартизацию теста)

- Соблюдение надежности результатов тестов
- Интерпретация результатов тестов
- Информирование о результатах теста (создание отчетов) и обеспечение обратной связи испытуемым
- Обучение и супервизия лиц, проводящих тестирование
- Разработка компьютерных систем и программ для обработки результатов тестирования
- Правовые аспекты тестирования
- Издание тестов

Осведомленность о контроле качества важна для любого профессионала, вовлеченного в процесс тестирования. Несмотря на то, что РКК изначально предназначено для профессионального использования специалистами, работающими в области тестирования, оно также объединяет основные принципы успешной практики и, как таковое, подходит и для тех, кто использует тесты исключительно в исследовательских целях.

### **1.3. Локальные и международные факторы**

РКК предназначено для международной аудитории профессионалов, которые используют тесты в своей работе. Это Руководство может помочь профессионалам в разработке специфических местных стандартов качества. Локальные факторы, такие как местные или государственные законы или стандарты, существующая регламентация или специфические контракты между клиентами и издателями тестов должны учитываться при реализации РКК на локальном уровне или в процессе их практического использования в заданных условиях. К примеру, в некоторых странах конфиденциальность личной информации испытуемых защищена законом.

### **1.4. Ошибки и необходимость в РКК**

Ошибки, которые допускаются во время процессов ПАИ, могут иметь серьезные последствия в любой области тестирования – психологической, образовательной, HR и поведенческой. К примеру, большое количество ошибок при подсчете результатов теста может повлиять на его результаты и их надежность – надежность подсчета результатов теста почти наверняка будет ниже. Это, в свою очередь, может привести к меньшей прогностической валидности. В некоторых случаях возможна и такая ситуация, что человек с патологическим поведением может быть ложно идентифицирован как человек с нормальным поведением. В других случаях ошибки могут помешать квалифицированному кандидату получить работу или привести к неправильному принятию решений, касающихся образования испытуемого. Ошибки также могут привести к неоправданному вмешательству в процесс образования, к примеру, если кому-то будет назначена несоответствующая образовательная программа, или некий человек получит профессиональную лицензию, несмотря на то, что ему/ей недостает требуемых знаний и умений. Ошибки могут вызвать продолжительную отсрочку в сообщении о результатах теста, что, в свою очередь, может повлечь серьезные проблемы для тех, кто ожидает зачисления в учебное заведение. Таким образом, ошибки могут иметь серьезные последствия и даже наносить ущерб. Ошибки также могут привести к потере общественного доверия к образовательным и психологическим тестам, и уменьшить убедительность некоторых тестов в глазах общественности. В некоторых случаях ошибки могут привести к тому, что относительно тестирующей организации, учебного учреждения или компаний, которые ищут квалифицированных работников, будут приняты правовые меры.

Профессионалы, которые осуществляют процесс тестирования, являются объектом потенциального давления со стороны четырех источников: организаций, испытуемых, тестирующих компаний и средств массовой информации. Все они ожидают быстрой и недорогостоящей разработки тестов и быстрого получения результатов. Для соблюдения стандартов качества крайне необходимо, чтобы профессионалы противостояли давлению, которое исходит со стороны тех, кто старается ускорить или сократить длительность этих процессов или пропустить некоторые их этапы. Под значительным давлением, к примеру, находятся организации, которые обязаны по контракту провести ПАИ результатов тестирования в сжатые сроки.

Существует значительная вероятность ошибок в длительных процессах – таких, как разработка тестов, подсчет баллов, анализ результатов и информирование о них, которые содержат последовательность этапов, где каждый последующий строго зависит от предыдущего. Стандарты качества должны тщательно проверяться и регулярно обновляться.

### **1.5. Область применения РКК**

РКК распространяется на крупномасштабные процессы тестирования. Тем не менее, они также могут быть использованы для широкого спектра других ситуаций тестирования (например, индивидуальное тестирование с целью профориентации или определения уровня личностного развития) и техник оценки (например, тесты с множественными вариантами выбора, оценка успешности выполнения, структурированные и неструктурированные интервью, оценка деятельности группы) и практически для любой ситуации, в которой происходит оценка (например, в образовательных целях, или в центрах оценки). Некоторые рекомендации носят конкретный характер и касаются различных индивидуальных или проводимых в группе стандартизированных тестов, а другие – имеют гораздо более широкое применение (например, в области клинического, образовательного и профессионального тестирования). Большое количество профессий используют практику оценки (например, в медицинских и реабилитационных, судебных, образовательных, связанных с трудоустройством целях), и здесь также РКК могут быть очень полезны.

РКК применимо в любой форме проведения тестирования (включая бланковые тесты и все более распространяющуюся компьютеризированную оценку через Интернет или в автономном режиме). Конструирование тестов, их выбор и проведение тестирования не входит в фокус РКК. Тем не менее, полезность или успешное применение РКК в ПАИ тестов будет зависеть от уместности самого теста и от того, являются ли результаты теста надежными и прогностическими относительно четко определенных критериев. Выделение ресурсов на контроль качества представляет собой инвестирование в ответственную практику, подконтрольность и добросовестность - основные элементы любого этического кодекса.

### **1.6. Определение контроля качества**

Контроль качества можно определить как *формальный систематический процесс, направленный на поддержание высоких стандартов качества на всех этапах подсчета, анализа и информирования о результатах тестов, и тем самым на минимизацию ошибок и повышение надежности измерения.*

### **1.7. Примеры из других профессий**

Процедуры контроля качества используются во многих других профессиях, таких как инженерное дело, авиация, разработка программного обеспечения и медицина. Что касается области медицины, интересно отметить некоторые из факторов, провоцирующих ошибки, которые происходят в клиниках. К ним относятся ненадлежащие условия хранения лекарственных средств, сложность медицинских вмешательств, внедрение новых технологий, проблемы коммуникации, плохая командная работа и отсутствие четкой политики в области безопасности. Такой пример имеет аналогию в области тестирования, где акцент делается на тесте, со всеми возможными ошибками, которые возникают в процессе тестирования и подсчета баллов.

## **1.8. Структура РКК**

РКК состоит из двух основных частей:

1. Общие принципы - общие моменты, которые должны быть учтены и согласованы до подсчета, анализа результатов и информирования о них.
2. Пошаговые указания для работы.

За этими разделами следуют две дополнительные секции: краткие выводы и список литературы.

## **2. ОБЩИЕ ПРИНЦИПЫ**

### **2.1. Проверка стандартов контроля качества, используемых в настоящее время**

- 2.1.1 Определите, какие принципы контроля качества в настоящее время существуют для тестирования в вашей организации или в вашей стране. В случае необходимости, разработайте специфические процедуры контроля качества до проведения тестирования. Пересматривайте, обновляйте и модифицируйте принципы, когда происходят изменения в процессе, а также время от времени в качестве обычной проверки.
- 2.1.2 Убедитесь, что до проведения тестирования имеются соответствующие процедуры контроля качества.
- 2.1.3 При работе с новым тестом, проведите пилотное исследование, включающее все процессы ПАИ. Там, где пилотное исследование не совершается, рассматривайте первое исследование как пробное и будьте готовы внести поправки в процессы ПАИ до последующих тестирований.
- 2.1.4 Создайте специфические стандарты для каждого теста в тех случаях, если они еще не существуют.
- 2.1.5 Создавайте специфические стандарты для каждого нового теста во время его конструирования.

### **2.2. Основные приготовления и соглашения между вовлеченными лицами**

Прежде чем провести тестирование, необходимо согласовать основные принципы между специалистами, ответственными за тест, в том числе между лицами, ответственными за конструирование теста, проведение тестирования, подсчет баллов, стандартизацию, интерпретацию, валидизацию и информирование о результатах. Хотя они имеют разные обязанности и роли, работа всех вовлеченных специалистов должна быть согласована; хорошо настроенная коммуникация между людьми, которые участвуют в этом процессе, должна повысить качество и упростить использование тестов.

Специалисты (см. п. 1.2) должны:

- 2.2.1 - определить все заинтересованные стороны в процессе тестирования и согласовать, кто отвечает за принятие решений в отношении различных этапов процесса тестирования.
- 2.2.2 - определить и указать цель или цели использования теста (например, подбор персонала, оценка достижений, научное исследование).
- 2.2.3 - согласовать расписание для процессов ПАИ.
- 2.2.4 - придумать средства связи между людьми или группами (там, где вовлечено более одной команды).
- 2.2.5 - придумать способы предоставления обратной связи клиенту о процессе тестирования.
- 2.2.6 - определить методы передачи данных лицам, ответственным за процессы ПАИ, например, данные, полученные с помощью оптического носителя или сканированные бланки, или данные в электронном виде, полученные в результате компьютеризированного тестирования.
- 2.2.7 - определить показатели по субтестам (если они есть), а также обосновать эти показатели. Необходимо также быть готовыми к изменению этих значений после получения данных.
- 2.2.8 - согласовать инструкции по подсчету баллов, то есть количество баллов, начисляемых за каждый правильный ответ, и решить, как поступать с

- неправильными ответами. Необходимо также быть готовыми изменить инструкции после получения данных.
- 2.2.9 - выбрать шкалу подсчета баллов и определить диапазон показателей в ней.
  - 2.2.10 - решить, как поступать с недостающими данными (в случаях, если испытуемые пропустили задание или по ошибке перескочили на другую строку, отмечая ответ, или в случаях, если исследователь забыл оценить конкретного испытуемого или же сделал это неправильным способом без возможности повторить тестирование).
  - 2.2.11 - определить и описать модель стандартизации, необходимый состав и объем выборки, а также используемые методы стандартизации.
  - 2.2.12 - определить и описать модель установления стандартов, а также необходимый состав и объем выборки в соответствии с этими стандартами.
  - 2.2.13 - согласовать степень детализации, с которой результаты должны быть представлены респондентам и заинтересованным организациям, а также какую дополнительную информацию, касающуюся баллов, необходимо предоставить в отчетах.
  - 2.2.14 - определить, какие конкретные лица, органы или организации должны получить результаты тестов, обеспечив при этом соблюдение правовых ограничений в отношении конфиденциальности данных.
  - 2.2.15 - определить, могут (должны) ли отчеты о результатах содержать другую информацию (например, было ли изменено содержание теста, сколько заданий было выполнено, были ли предложены особые условия для людей с ограниченными возможностями).
  - 2.2.16 - согласовать необходимость письменной фиксации данных.
  - 2.2.17 - согласовать необходимость повторения с целью проверки качества важных процессов (например, создание таблиц конвертации сырых баллов в стандартизированные).

### **2.3. Ресурсы**

- 2.3.1 Убедитесь, что имеются достаточные ресурсы (материальные средства, время и персонал) для эффективного и надлежащего подсчета, анализа результатов и информирования о них.
- 2.3.2 Проверьте доступность резервных ресурсов (например, если специалист по стандартизации не может провести процедуру, определите, кто сделает это вместо него; или если сканер для бланков ответов вышел из строя, найдите альтернативный сканер).
- 2.3.3 Будьте в курсе проблем с расчетом времени, которые могут возникнуть, если используются резервные ресурсы. Запланируйте действия на случай непредвиденных обстоятельств, чтобы покрыть неожиданное отсутствие ключевых сотрудников.
- 2.3.4 Распределите задачи между членами команды: Кто будет заниматься подсчетом баллов, анализом результатов и информированием о них? Кто несет ответственность за весь процесс? Специалисты, отвечающие за тестирование, должны определить, например, имеют ли те, кто участвует в каждом этапе процесса, навыки, необходимые для работы, которую они будут выполнять, а также должны определить требования и уровень автоматизации процессов.
- 2.3.5 Определите необходимые временные ресурсы: установите расписание для каждого этапа процессов ПАИ. Крайний срок окончания процесса и информирования о результатах должен быть реалистичен.

- 2.3.6 Убедитесь в наличии необходимого программного обеспечения, компьютерных и сетевых ресурсов: защищенное авторским правом и разработанное на заказ ПО, ноутбуков, ПК, центральных процессоров, доступного места на жестком диске и на сервере, и т.д.
- 2.3.7 Определите необходимые ресурсы рабочего пространства – достаточно ли рабочего места (достаточно ли комнат, столов, стульев и т.д.) для персонала и испытуемых.
- 2.3.8 Определите действия, которые необходимо предпринять для безопасного сохранения данных в электронном виде.
- 2.3.9 Удостоверьтесь, что все необходимое дополнительное оборудование (ключи для подсчета вручную, калькуляторы) доступно.

## **2.4 Потребности и ожидания заинтересованных сторон**

Те, кто использует результаты тестов - испытуемые, родители / опекуны, преподаватели / консультанты, а также те, кто проводит тестирование (например, специальные агентства) – имеют конкретные потребности и ожидания в отношении тестирования и времени, отведенного на информирование о результатах. Эти потребности и ожидания должны быть обоснованными и согласованными между сторонами (см. также Руководство МТК по использованию тестов, 2000, Приложение В – Руководство по выработке соглашений между сторонами, вовлеченными в процесс тестирования)

- 2.4.1 В случае необходимости разработайте соглашение между вовлеченными сторонами, согласованное со специалистами, ответственными за подсчет баллов, стандартизацию и информирование о результатах. Учтите, что нужно время от времени вносить изменения в соглашение.
- 2.4.2 Определите, кто несёт главную ответственность и имеет право решать, как действовать в случае возникновения проблем.

*Например, когда на вопрос с разными вариантами ответа нет правильного ответа, интервьюер ведет себя нетактично, или испытуемых отвлекает шумная обстановка. Другой случай - когда был предложен вопрос, на который предполагался только один правильный ответ, но позже испытуемый доказывает, что и другой ответ тоже является правильным.*

- 2.4.3 Заранее определите алгоритм действий в случае, когда после подсчета баллов обнаружена ошибка.
- 2.4.4 Обеспечьте испытуемым возможность поставить под сомнение правильность предлагаемых ответов и оспорить свои оценки, или задавать вопросы и предлагать свои варианты.
- 2.4.5 Необходимо иметь документ, который обосновывает систему подсчета баллов по заданиям теста.

## **2.5 Персонал и рабочая атмосфера**

Убедитесь в том, что люди, ответственные за процессы подсчета баллов, анализа результатов, стандартизации и информирования о результатах, являются профессионалами, которые обладают всеми необходимыми навыками и знаниями о процессах ПАИ. Убедитесь, что эти люди имеют все необходимые компетенции для

работы. Если в процессе тестирования участвует группа людей, важно обеспечить им комфортное взаимодействие. Таким образом, при приеме на работу новых сотрудников, их способность работать в команде и уживаться друг с другом должна быть ключевым фактором.

- 2.5.1. Избегайте необоснованного давления на людей с целью повышения скорости выполнения работы.
- 2.5.2. Избегайте излишних часов работы.
- 2.5.3. Попробуйте выработать в себе тщательное, внимательное к деталям отношение к работе (особенно с точки зрения предотвращения ошибок), в то же время не перенапрягайтесь. Спокойная, но целенаправленная рабочая атмосфера является наиболее эффективной средой для поддержания высоких стандартов.
- 2.5.4. Поддерживайте работников, обеспечивая профессиональное развитие и тренинги, а также в некоторых случаях также и личностный рост и развитие социальных навыков .

## **2.6 Независимый мониторинг процедур контроля качества**

Назначьте одного или нескольких специалистов (в зависимости от величины и сложности проекта) для мониторинга соблюдения процедуры контроля качества и обеспечения отслеживания того, чтобы все вопросы, проблемы или ошибки были учтены. Наблюдатели должны работать независимо от тех, кто участвует в процедурах подсчета баллов, анализа и информирования о результатах. Мониторинг должен проводиться в сотрудничестве со всеми заинтересованными сторонами с целью учёта специфических процессов.

## **2.7 Фиксация ошибок**

- 2.7.1 Все, кто участвует в процессе тестирования, должны следовать согласованным процедурам в отношении фиксации деятельности и ошибок или проблем, которые могут возникать.
- 2.7.2 Согласуйте заранее, какой сотрудник за какой этап отвечает.
- 2.7.3 Регистрируйте все виды деятельности. Проверяйте все процессы на предмет их качества.
- 2.7.4 Фиксируйте подробно все ошибки (независимо от того, известны ли их причины), начиная с характера ошибки, того, кто обнаружил её и когда, какие были последствия, и какие шаги были / будут предприняты для ее устранения. Также фиксируйте случаи, в которых были обнаружены ошибки до того, как проявились их последствия.
- 2.7.5 Консультируйте других специалистов касательно возможных ошибок для их своевременного предупреждения, иногда даже на специальных собраниях, посвященных предотвращению ошибок.
- 2.7.6 Делайте выводы о том, как предотвратить ошибки в будущем.

## **3. ПОШАГОВЫЕ РАБОЧИЕ РЕКОМЕНДАЦИИ**

Вам предлагаются конкретные шаги, которые должны быть предприняты в процессе подсчета баллов, анализа результатов и информирования о них. При проведении широкомасштабного тестирования каждый этап должен быть обдуман и проведен с осторожностью. Перед началом работы с реальными данными следует провести пилотное исследование, чтобы результаты были более точными. В ситуациях масштабного тестирования, где участвуют тысячи людей, эти принципы должны особо четко соблюдаться. В ситуациях же менее масштабного тестирования, где участвуют только десятки людей, эти рекомендации должны быть реализованы, но при этом некоторые этапы могут опускаться или упрощаться. Это необходимо потому, что некоторые процедуры являются ресурсоемкими и основаны на модели с большими выборками. Эти процедуры должны быть адаптированы в соответствии с количеством испытуемых.

### **3.1 Разработка отчетов о результатах**

До проведения основных этапов тестирования должно быть достигнуто соглашение в отношении информирования о результатах тестирования, что является конечным продуктом всего процесса. Необходимо решить, что именно следует сообщить, насколько подробно, кому, когда и т.д. Недостаточно лишь сообщить результат учреждению или испытуемому в виде чисел или стандартизированных баллов (стендов и т.д.). Очень важна правильная интерпретация результатов. При разработке тестов, подсчете баллов и анализе результатов мы должны принимать во внимание конечный продукт - предоставление обратной связи о результатах. В этом смысле, основной целью является обеспечение того, чтобы результаты были правильно поняты. Различные аспекты интерпретации должны быть объяснены в самом начале отчета. Должно быть достигнуто соглашение со всеми участниками в отношении предоставления обратной связи по более частным результатам (например, по баллам подшкал) в дополнение к основным. Должны ли они быть представлены? Будут ли они использоваться?

### **3.2 Демографические данные**

Демографические данные могут быть очень полезными в процессе контроля качества для таких целей, как подтверждение личности испытуемого, понимание неожиданных результатов и распределение испытуемых по группам в целях стандартизации. Рекомендуются следующие шаги:

- 3.2.1 Если позволяют правовые нормы, соберите демографические и биографические данные (возраст, пол, национальность, образование, результаты предыдущих тестов и т.д.) с помощью предварительной регистрации, регистрации на месте или после тестирования. Собранные демографические данные должны быть релевантными конкретному исследованию и необходимо защищать конфиденциальность испытуемых настолько, насколько это возможно.
- 3.2.2 Если это возможно, систематически время от времени проверяйте демографические данные испытуемых; обращайте внимание на несоответствия в демографических данных испытуемых, которые проходят тестирование повторно.
- 3.2.3 Проведите исследование для определения корреляции между демографическими данными и результатами тестирования, поищите несоответствия результатов тестирования с результатами предыдущих тестирований, результатами научных исследований и т.д. Например, может быть, что в конкретном исследовании взрослые люди демонстрируют более высокие результаты по конкретному тесту,

чем более молодые. Если согласно результатам исследований более молодые люди должны проходить тест лучше, то это повод изучить результаты на предмет ошибок.

### **3.3 Подсчет баллов**

#### 3.3.1 Получение и хранение ответов испытуемых

При необходимости, все ответы испытуемых на бумаге должны храниться в электронном виде, как правило, с идентификационным номером каждого испытуемого. Хранение таких материалов в бумажном и электронном виде осуществляется на протяжении установленного местным законодательством и профессиональными стандартами срока. Это касается и бланков ответов, и тестовых ответов и результатов в электронном виде.

3.3.1.1 Если бумажные копии существуют, они должны быть сохранены в течение периода времени, определенного законами страны, штата или провинции, в которых происходит тестирование.

3.3.1.2 Что касается электронной версии, используйте системы бесперебойного электропитания и резервные батареи для компьютеров, чтобы уменьшить вероятность прерывания работы и потери данных.

3.3.1.3 Сканеры, если они используются, должны регулярно проверяться на наличие неисправностей.

3.3.1.4 Проверяйте вручную результаты сканирования данных.

3.3.1.5 Проверяйте базы данных испытуемых с целью поддержания устойчивой системы идентификации испытуемых. Например, обратите внимание на случаи, когда идентификационный номер относится более чем к одному имени.

3.3.1.6 Все данные должны быть защищены. Защищайте личную информацию испытуемых путем отделения их личных данных (например, имена) от результатов тестирования, когда это возможно. Например, заведите отдельные файлы: один с биографическими данными, и другой с результатами, которые могут быть идентифицированы по идентификационному номеру. Все эти действия должны соответствовать национальным законам, касающимся информационной безопасности и хранения данных.

3.3.1.7 Выполняйте проверку правильности ключей подсчета баллов и надлежащего использования норм и таблиц преобразования сырых баллов в стандартизированные.

#### 3.3.2 Подсчет баллов

После того как тестовые данные были обработаны и надежно сохранены в базе данных, ответы испытуемых, как правило, используются для подсчета сырых баллов. В классической теории тестирования (Classical Test Theory, СТТ), например, сырые оценки обычно равны количеству правильных ответов (когда в тесте есть правильные и неправильные варианты ответов); иногда применяется поправка на угадывание, иногда в тесте есть задания, которые «вешат» больше, чем другие. В теории «Задание-Ответ» (Item

Response Theory, IRT) сырой балл отображает скрытые (латентные) способности - они часто называются «тета» или «оценка черты». На результаты теста могут оказать негативное воздействие различные виды ошибок, например, ошибки в ключе. Иногда ошибки приводят к крайне низким результатам. Используйте следующие процедуры контроля качества, чтобы находить и устранять такие ошибки:

- 3.3.2.1 Проверьте, соответствует ли структура данных требованиям документа, в который эти данные вносятся, например, порядок элементов в файле.
- 3.3.2.2 Применяйте заранее утвержденные правила для удаления невалидных протоколов, перекодирования недостающей информации и обработки повторяющихся случаев.
- 3.3.2.3 Сравнивайте данные выборки с ожидаемым диапазоном баллов, а также сравнивайте описательные статистики с нормами оригинального теста, описанными издателями (в случае адаптации теста). Статистики выборки могут несколько отклоняться (от того, что ожидается), но большие отличия необходимо отмечать и исследовать.
- 3.3.2.4 Рассматривайте крайние значения отдельных лиц или определенных групп – как высокие, так и низкие. Экстремальные оценки могут указывать на три возможные проблемы: ошибки в подсчете баллов, фальсификация ответов или ошибка при получении данных.
- 3.3.2.5 Пересмотрите данные по отдельным испытуемым в случаях, когда различия в корреляциях между баллами по субшкалам больше, чем ожидалось. Для этого необходимо заранее определить критическую величину различий.
- 3.3.2.6 Проанализируйте задания теста и рассмотрите описательные статистики заданий - ошибку в ключе для одного задания трудно обнаружить, если описательные статистики заданий недоступны (задания с неправильным ключом часто кажутся сложными и могут негативно коррелировать с критерием).
- 3.3.2.7 Проверяйте результаты на наличие пропущенных ответов. Ответ мог быть удален из расчётов некоторых из испытуемых по ошибке.
- 3.3.2.8 Обратите особое внимание на группы, которые проходили тестирование в разных условиях, и выполните дополнительную проверку полученных в этих группах данных. Например, испытуемые могли проходить тестирование в разное время или с другой версией теста, или с использованием разных методов ответа.
- 3.3.2.9 Вычислите и рассмотрите описательные статистики основных групп, участвующих в тестировании, распределенных, например, по месту проведения тестирования, по специалисту, проводящему тестирование, или по Интернет-ресурсу, с помощью которого проводилось тестирование. Ошибки могут возникать в результате использования неправильной формы теста на конкретной группе испытуемых.
- 3.3.2.10 Если позволяют ресурсы, случайным образом отобранная часть результатов тестирования может быть передана отдельной группе специалистов для

подсчета и анализа результатов. Позже можно сравнить основные результаты с результатами независимой группы специалистов.

### **3.4. Анализ теста**

#### 3.4.1. Анализ заданий теста (обычно для крупномасштабного тестирования)

Анализ заданий теста обеспечивает базовые статистические показатели для принятия решений по характеристикам заданий теста и тому, как они все вместе формируют общий балл. Рекомендуется проводить анализ заданий теста во всех случаях, кроме тех, в которых количество испытуемых небольшое. Результаты статистического анализа заданий теста включают в себя уровень сложности задания (или тенденцию вызывать установку на согласие – это касается личностных опросников) и дискриминативность. Параметры каждого задания теста могут рассчитываться по-разному в зависимости от модели, которая использовалась в процессе разработки теста. Кроме того, анализ заданий представляет общие статистические данные теста (надежность и/или стандартная ошибка, среднее, стандартное отклонение, информация о тесте, распределение ответов испытуемых и т.д.). Следующие процедуры должны быть рассмотрены в случае, если количество испытуемых превышает минимальное, в зависимости от используемой модели.

- 3.4.1.1 Используйте надежные программы для проведения анализа заданий теста и убедитесь, что они имеют соответствующую техническую документацию.
- 3.4.1.2 Если у Вас есть основания сомневаться в качестве программы, которую Вы используете для анализа заданий теста, или если Вы используете новую программу, используйте две программы одновременно и сравните их результаты.
- 3.4.1.3 Проведите анализ заданий теста после его проведения или проанализируйте накопленные данные (например, за 3-5 лет его проведения), если тест проводится периодически. Рассмотрите возможность проведения анализа заданий теста на неполном наборе данных (пока данные не собраны полностью), чтобы можно было быстро найти ошибки.
- 3.4.1.4 Пересмотрите результаты анализа заданий теста прежде чем делать какие-то выводы об испытуемых.
- 3.4.1.5 Анализ заданий теста даст возможность определить проблемы, связанные с ключами теста. Например, то, что кажется очень популярным «дистрактором» (отвлекающим вариантом ответа), может оказаться правильным ответом (ключом); или же негативная корреляция между заданиями может обнаружить задание, которое должно иметь обратный ключ. Если результаты определенного задания проблемные, ключ и содержание этого задания должны быть пересмотрены.
- 3.4.1.6 Повторите анализ заданий теста, если ключ к тесту меняется или какие-то задания удаляются. Обновляйте документацию (например, таблицы подсчета баллов и характеристики стандартизации) на протяжении всего процесса.

#### 3.4.2 Анализ результатов тестов действия (performance tests), практических заданий, деловых игр, интервью и т.д.

В то время как анализ заданий с множественным выбором ответов (МВ) является объективным (базируется на конкретном ключе) и надежным, анализ заданий с открытыми вопросами (ОВ) (тесты действия, анкеты с открытыми вопросами, практические задания, деловые игры и т.д.) обычно имеет субъективный компонент. По

этой причине такой вид оценки, как правило, менее надежен, чем МВ, потому что часто требует оценки со стороны человека и зависит от его индивидуальных особенностей. Тем не менее, можно использовать различные методы для снижения субъективности анализа ОВ заданий, а также для повышения надежности и точности анализа.

3.4.2.1 Убедитесь, что результаты тестов, практических заданий и интервью оцениваются обученными ассессорами, которые обладают необходимыми знаниями и опытом, а также полномочиями, профессиональной подготовкой или соответствующим академическим образованием.

3.4.2.2 Инструкции по оценке ОВ ответов должны быть четкими и хорошо структурированными. Необходимо провести предварительное тестирование открытых вопросов для более точной формулировки инструкций.

3.4.2.3 Определите диапазон ответов с помощью распределения примеров ответов по рубрикам.

3.4.2.4 Требуйте от ассессоров участия в тренинговых мероприятиях до проведения оценки. Обучение дает им возможность ознакомиться с инструкциями и оценкой результатов, прежде чем они приступят к оценке фактических ответов испытуемого.

3.4.2.5 Оценивайте ассессоров на основании их подготовки, прежде чем допустить их к оценке результатов конкретного тестирования.

3.4.2.6 Старайтесь использовать не менее двух ассессоров для каждого тестирования, в зависимости от стоимости и наличия необходимых ресурсов.

3.4.2.7 При наличии только одного ассессора для всех испытуемых (в связи с финансовыми или другими соображениями) используйте еще одного ассессора для оценки небольшой части выборки (например, для анализа 10% всех данных), чтобы оценить надежность результатов.

3.4.2.8 Если используется компьютерный подсчет открытых вопросов, убедитесь, что результаты будут рассмотрены человеком позже. Необходимо эмпирически обосновать использование компьютеризированного подсчета, прежде чем использовать его.

3.4.2.9 Убедитесь, что ассессоры работают независимо друг от друга.

3.4.2.10 Применяйте статистические методы для оценки надежности процесса оценки путем вычисления соответствующих мер согласия между ассессорами, а также различий между ассессорами.

3.4.2.11 Осуществляйте периодический мониторинг качества оценки в режиме реального времени, так, чтобы была доступна обратная связь.

3.4.2.12 Если ассессор не отвечает ожиданиям, (результаты оценки ненадежны или недостаточно близки к результатам оценки других ассессоров) информируйте ассессора об этом соответствующим образом и рекомендуйте ему переподготовку; замените ассессора, если проблема не будет решена.

3.4.2.13 Разработайте правила работы с большими расхождениями между оценками ассессоров. Когда различия невелики, они должны быть усреднены или суммированы. Когда есть большие расхождения, опытный ассессор может выступить посредником для их решения.

### 3.4.3 Уравнивание (Equating)/Проверка новых форм теста и заданий

Иногда уравнивание не требуется, если испытуемые соревнуются только с теми, с кем они одновременно проходят тестирование. Если не проводить уравнивание, результаты испытуемых по предыдущим тестированиям нельзя переносить и сравнивать результатами по новой форме теста с другими испытуемыми. Если баллы по разным версиям теста должны быть на одной шкале, новые формы теста должны быть эквивалентны предыдущим, чтобы можно было сравнивать новые формы теста со старыми, которые могут иметь другие психометрические характеристики. Результатом уравнивания является то, что баллы по всем формам теста имеют одно и то же значение. Уравнивание может быть проведено до тестирования (пре-уравнивание), и/или после тестирования (пост-уравнивание). Уравнивание может проводиться на уровне заданий, шкал или всего теста. Существуют разные подходы и методы уравнивания (например, линейное уравнивание, эквиперцентильное уравнивание, уравнивание, основанное на теории «Задание-ответ», уравнивание по респондентам или по заданиям). Уравнивание обычно требует больших выборок, их размер зависит от метода и модели уравнивания (см. Kolen & Brennan, 2004; Lamprianou, 2007).

- 3.4.3.1 В случаях необъяснимых проблем в уравнивании (например, полученные баллы ниже, чем ожидалось), убедитесь, что все формы теста были применены в одинаковых стандартизированных условиях. Если условия не были эквивалентны, попробуйте оценить влияние разных условий.
- 3.4.3.2 Убедитесь в том, что специфические процедуры и модели уравнивания были соблюдены.
- 3.4.3.3 Исследуйте теоретические предпосылки процедуры уравнивания и/или определите, могут ли разные процедуры уравнивания с разными теоретическими предпосылками давать похожие результаты. Проведите анализ стабильности параметров одинаковых заданий теста (уравнивание по заданиям). Если Вы используете набор одних и тех же заданий для уравнивания, необходимо обосновать, почему Вы удаляете какие-то задания из одной из версий теста, и влияние таких решений на баллы и на стандартизированные баллы. Также зафиксируйте содержание и статистические характеристики общего для двух версий набора заданий теста после анализа данных. Такой стандарт также применим к модели общей выборки (уравнивание по респондентам); но анализ фокусируется на испытуемых.
- 3.4.3.4 Сравните реальные баллы, полученные по конкретным испытуемым, с теми, которые предполагались на основе особенностей выборки (см. 3.2.1). Если имеются различия, перепроверьте баллы.
- 3.4.3.5 Проведите сравнение баллов и проходных оценок во времени. В хорошо разработанных крупномасштабных исследованиях колебания между тестированиями, проводящимися в разное время, часто малы. Большие различия могут свидетельствовать о проблемах в уравнивании баллов, например, об изменениях в характеристиках популяции.
- 3.4.3.6 Когда проводится большое количество тестирований с небольшим количеством испытуемых в каждом, необходимо применять специфические инструменты контроля качества для отслеживания стабильности результатов теста.

Некоторые из них: графики Shewhart и CUSUM по данным тестирования, модели временных серий, модели смены точек и инструменты интеллектуального анализа данных (см. Von Davier, 2011).

- 3.4.3.7 Если в тесте есть граничные показатели, отделяющие респондентов, прошедших тестирование, от тех, кто его не прошел или по другим критериям, проверьте эти показатели. Сравните их с предполагаемыми на основании предыдущих тестирований, особенностей испытуемых и результатов других тестов показателями.
- 3.4.3.8 Убедитесь в последовательности граничных показателей в разных группах, используйте обоснованные методы и фиксируйте процесс. Также фиксируйте случаи, в которых Вы, по каким-то причинам, отклоняетесь от стандартного процесса.
- 3.4.3.9 Если Вы используете разные форматы проведения тестирования (например, компьютеризированное применение теста, который был разработан как тест для бумаги и карандаша), важно сравнить характеристики нового теста с характеристиками старого, а также иногда уравнивать новый тест со старым.
- 3.4.3.10 Для тестов, имеющих важные последствия для испытуемого (high-stake tests), необходимо стараться как можно чаще повторять процедуру уравнивания и привлечь для контроля третью сторону, не имеющую отношения к процессу уравнивания.

#### 3.4.4 Подсчет стандартизированных баллов

Во многих случаях стандартизированные баллы используются, чтобы сделать результаты более понятными. В таких случаях сырые баллы могут подсказать нам, какие именно стандартизированные баллы лучше использовать (например, станайны, децили). Показатели или таблицы преобразования используются для подсчета баллов по шкалам и для отображения стандартизированных баллов или перцентилей. Обычно сырые баллы (количество правильных ответов или количество правильных ответов с поправкой на угадывание) или тета баллы (в тестах, основанных на теории «Задание-ответ») должны быть преобразованы в специфическую шкалу. Преобразование проводится с помощью специальных таблиц или математической функции (например, при линейном преобразовании).

- 3.4.3.1 Проведите соответствующие преобразования сырых баллов для получения определенных стандартизированных шкал.
- 3.4.3.2 Проверьте процедуры преобразования сырых баллов в стандартизированные и другие процедуры на предмет ошибок.
- 3.4.3.3 Проверьте, правильно ли было осуществлено преобразование.
- 3.4.3.4 Убедитесь в том, что низкие стандартизированные баллы соответствуют низким сырым баллам и что высокие стандартизированные баллы соответствуют высоким сырым баллам.
- 3.4.3.5 В некоторых случаях могут потребоваться дополнительные процедуры после преобразования (например, определение единого минимума и максимума для всех результатов).
- 3.4.3.6 Сравните свойства новой формы теста с параметрами других форм теста на предмет необычных различий или сходства.
- 3.4.3.7 Учитывайте изменения, которые могут возникать в шкале со временем.
- 3.4.3.8 Посчитайте некоторые показатели вручную и сравните результаты с теми, которые были подсчитаны на компьютере.
- 3.4.3.9 Проверьте статистическое соотношение между сырыми и стандартизированными баллами с помощью диаграммы рассеяния.
- 3.4.3.10 Используйте две разных компьютерных программы для подсчета стандартизированных баллов и их сравнения.

- 3.4.3.11 В техническом пособии или других связанных с ним материалах предоставьте подробное описание процедур, используемых для преобразования сырых баллов в стандартизированные. Так как эта техника может быть разной для разных форм теста, необходимо описать процедуру, используемую для каждой формы теста.

#### 3.4.5 Проверка теста на защищенность от фальсификации

Фальсификация результатов теста создает серьезную проблему, которая подвергает риску безопасность и целостность теста и системы тестирования. К сожалению, фальсификация не может быть полностью предотвращена, даже учитывая постоянный контроль и другие средства предотвращения фальсификации; соблазн фальсифицировать результаты может быть огромным, особенно в тестах с важными последствиями для испытуемого. В качестве еще одного инструмента борьбы с фальсификацией стоит проконсультироваться с юристами с целью рассмотрения методов предотвращения фальсификации и подтверждения их пригодности. В тестах ЕГЭ (ЗНО), имеющих важные последствия для испытуемых, обман может быть осуществлен не только на уровне индивидуального испытуемого, а и на уровне класса, школы, района или места, где проводится тестирование. Фальсификация может иметь место там, где непосредственно происходит тестирование, с помощью мобильных телефонов или с помощью интернета. В сфере оценки персонала тестирование через интернет становится все более популярным, и это также повышает риск фальсификации результатов. Побочным преимуществом проверки теста на защищенность от фальсификаций является то, что она также может помочь обнаружить проблемы в процедурах тестирования, сбора данных и их хранения.

Рекомендуются следующие меры предосторожности:

- 3.4.5.1 Проверяйте личность испытуемых тогда, когда они допускаются в зал, где проходит тестирование, или когда они проходят тестирование дома, используйте удостоверение личности или биометрические показатели – такие, как сканирование радужной оболочки глаза или отпечатков пальцев. Используйте продвинутые методы для идентификации удаленных испытуемых.
- 3.4.5.2 Лучше использовать разные формы теста. Когда используется одна форма, не следует садить испытуемых, которые знакомы между собой (например, живут в одном районе, ходят в одну школу), рядом друг с другом. Например, можно садить испытуемых в алфавитном порядке.
- 3.4.5.3 Записывайте номера мест и составляйте список того, кто где сидел для того, чтоб облегчить анализ списываний.
- 3.4.5.4 Если это необходимо (например, когда подозревается списывание), используйте статистические показатели для определения случаев списывания, основанные на сходстве бланков ответов у различных испытуемых в одном зале и во всем месте, в котором проводится тестирование.
- 3.4.5.5 Нанимайте обученных, благонадежных надзирателей и осуществляйте за ними тщательный надзор. Убедитесь в том, что у надзирателей нет конфликта интересов.
- 3.4.5.6 Проверяйте отклоняющиеся или неожиданные паттерны ответов, (например, когда на сложные вопросы даются правильные ответы и на простые вопросы даются неправильные ответы).
- 3.4.5.7 Возьмите образец почерка у каждого испытуемого до тестирования и непосредственно перед ним, чтобы определить самозванцев или испытуемых с подозрительным поведением. Эту процедуру можно пропустить, если нет никаких проблем с идентификацией.
- 3.4.5.8 Проанализируйте противоречивые результаты при повторном тестировании

(если можно идентифицировать повторы), используя статистическое распределение приемлемых различий между результатами на данный момент и результатами, полученными в предыдущих тестированиях. Сильные различия могут свидетельствовать о том, что место испытуемого занял самозванец, или о том, что кандидат получил информацию про задания теста до проведения тестирования. С другой стороны, респондент конечно может отвечать по-другому на повторном тестировании.

- 3.4.5.9 Необходимо иметь документированную процедуру (если нужно, юридически утвержденную) для работы с подозрениями в фальсификации. Проинформируйте всех испытуемых заранее о том, что применяются различные методы борьбы с фальсификацией.
- 3.4.5.10 Иногда учителя заинтересованы в повышении оценок студентов в стандартизированных тестах. По этой причине учителям нельзя давать доступ к результатам стандартизированных тестов.
- 3.4.5.11 Используйте закрывающиеся кабинеты и безопасные сервера, чтоб обезопасить материалы и результаты теста. Убедитесь, что все, кто участвует в разработке заданий и тестов, заслуживают доверия и полностью проинструктированы по вопросам защиты от фальсификации. Конфиденциальность заданий теста также должна быть полностью обеспечена на всех этапах, начиная с разработки самих заданий. Задания теста должны быть перенесены с помощью надежных средств от разработчика теста к реализатору, и все файлы должны храниться и применяться на запоминающем устройстве USB и/или на автономных ноутбуках, а не на стационарных компьютерах или серверах, где они могут быть легко доступны каким-то посторонним людям.
- 3.4.5.12 На компьютерах, на которых происходит тестирование, должно быть предотвращено сохранение и отправка материалов теста. Необходимо избегать подключения к интернету, если это даст возможность отправить материалы теста.
- 3.4.5.13 Для того, чтоб сохранять материалы теста в секрете, необходимо убедиться в том, что испытуемый не снимает ничего на фотоаппарат, телефон и т.д.
- 3.4.5.14 Для того, чтобы убедиться в справедливом отношении к каждому испытуемому, необходимо обеспечить конфиденциальность всем испытуемым на всех этапах тестирования и подсчета баллов.

## **3.5 Информирование о результатах**

### 3.5.1 Информирование о результатах теста

Результаты тестов выдаются как испытуемым, так и заказчикам. В идеале, результаты должны выдаваться в печатной форме. В некоторых случаях и контекстах наиболее приемлемым вариантом сообщения о результатах является интернет. Сообщение о результатах теста должно происходить таким образом, чтобы как заказчику, так и испытуемому было понятно значение результатов.

- 3.5.1.1 Применяйте метод фокус-групп, состоящих из испытуемых, или возможно методы “мозгового штурма”, экспериментальные исследования или индивидуальные интервью для получения информации, которая может быть полезна в разработке понятных и емких объяснений результатов тестирования и руководства по интерпретации.
- 3.5.1.2 Убедитесь в том, что все, кто получает отчет о результатах теста, располагает соответствующими знаниями для их интерпретации и сможет понимать эти результаты. Подтверждением этому должно выступать то, что отчеты по результатам теста дают возможность специалисту, использующему тест, давать

адекватные интерпретации по нему.

- 3.5.1.3 Создавайте компьютерно сгенерированные отчеты с основной информацией о тесте и значении результатов, что поможет получателю наиболее быстро и адекватно понять его результаты.
- 3.5.1.4 Храните данные испытуемых в бумажном и электронном виде для последующего подсчета международных, национальных и областных результатов теста, если это потребуется.
- 3.5.1.5 Выясните, до какого уровня можно применять разные результаты (например, результаты по субшкалам теста могут быть слишком ненадежными для их использования в принятии решений, имеющих важные последствия). Решение о том, стоит ли сообщать результаты по субшкалам (субтестам), должно также основываться на (а) теории данного теста, (b) цели тестирования и психометрических характеристиках субтестов.
- 3.5.1.6 Пользуйтесь помощью экспертов в области связей с общественностью (PR) в случаях, когда результаты и баллы по тестам сообщаются СМИ и государственным деятелям.

### **Принимайте меры для поддержки надежности сообщения о результатах теста**

- 3.5.1.7 Принимайте меры для того, чтобы убедиться, что индивидуальный отчет о результатах не может быть подделан испытуемым.
- 3.5.1.8 Избегайте редактирования отчета о результатах всего учреждения: это может привести к серьезным проблемам. Если необходимо изменить один или несколько результатов, используйте предназначенное для этих целей ПО, или создайте отчет снова.
- 3.5.1.9 Кодируйте названия электронных файлов отчетов для хранения и передачи.
- 3.5.1.10 Убедитесь, что отчеты отправляются только соответствующим лицам. Не отправляйте отчеты, содержащие больше информации, чем необходимо. Конечно, легче отправить один и тот же полный отчет всем заказчикам, но для поддержания конфиденциальности только уместная информация должна отправляться каждому получателю.
- 3.5.1.11 Информировать учреждения о том, что только тот отчет, который был отправлен напрямую в учреждение – а не та копия, которая была отослана испытуемому (которая может быть впоследствии сфальсифицирована) – может использоваться для официальных целей. Также рекомендуйте учреждениям произвести проверку отчета по учреждению.

### 3.5.2 Документация

Тщательная документация всего процесса подсчета баллов, включая описательные статистики ключей (среднее, стандартное отклонение, медиана, диапазон баллов, надежность), и сравнение текущих групп испытуемых с предыдущими группами испытуемых должно быть проведено и завершено до или сразу после выдачи результатов по тесту. Хорошая культура документации может способствовать большей надежности и точности вышперечисленных процессов в будущем. Предоставление публичного доступа к информации, касающейся результатов теста, служит дополнительным методом контроля всех процессов подсчета баллов, анализа результатов и информирования о них. Очень важно:

- 3.5.2.1 - документировать весь процесс шаг за шагом, с тщательной фиксацией всего процесса, включая основные статистики и сравнения по группам.
- 3.5.2.2 – убедиться, что новые формы теста проводятся только после того, как документация по старой версии теста была завершена.

- 3.5.2.3 – предоставьте желающим доступ к описательным статистикам, распределенным по полу и возрасту. Должны быть также предоставлены краткие объяснения этих статистик. Суммарные статистики защищают анонимность индивидуальных испытуемых.

#### **4. ВЫВОДЫ И ЗАКЛЮЧИТЕЛЬНЫЕ КОММЕНТАРИИ**

Вдобавок ко всем рекомендациям, сделанным в этом документе, мы хотели бы предложить Вам некоторые общие рекомендации. Каждый раз, когда новый тест предлагается заказчиком, необходимо провести детальное жизненное моделирование всего процесса, шаг за шагом (см. Texas Education Agency et al., 2004); далее могут быть опробованы новые процедуры тестирования. Такое моделирование поможет в формулировании новых процедур контроля качества. Кроме того, процесс подсчета баллов, анализа результатов и информирования о них состоит из последовательных этапов, каждый из которых основан на успешном прохождении предыдущего этапа. Таким образом, рекомендуется организовать контроль качества таким образом, чтоб можно было перейти на следующий этап без прохождения предыдущего. Компьютеризация процессов стандартизации, модификации и контроля процедур контроля качества позволяет делать это просто, прозрачно и эффективно. Однако, хотя преимущества компьютеризации широко признаются, необходимо опробовать и адаптировать их к конкретной ситуации.

## 5. ЛИТЕРАТУРА

- AERA/APA/NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.
- Allalouf, A. (2007). Quality Control Procedures in the Scoring, Equating, and Reporting of Test Scores. *Educational Measurement: Issues and Practice*, 26: 36-43.
- Bartram, D., Hambleton, R.K. (Eds.) (2006). *Computer-Based Testing and the Internet*. West Sussex: John Wiley & Sons.
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.
- ITC (2001). International Guidelines on Test Use. *International Journal of Testing*, 1: 95-114.
- ITC (2006). International Guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6: 143-172.
- Kolen, M. J., and Brennan, R. L. (2004). *Test equating, linking and scaling: Methods and practices*. New York: Springer.
- Lamprianou, I. (2007). Comparability methods and public distrust: an international perspective. In Newton, P., Baird J., Goldstein, H., Patric, H., & Tymms, P. (Eds.) *Techniques for monitoring the comparability of examination standards*. Qualifications and Curriculum Authority, London.
- Nichols, S. L. & Berliner, D. C. (2005). The inevitable corruption of indicators and educators through high-stakes testing, *Educational Policy Studies Laboratory*, College of Education, Arizona State University.
- Rhoades, K., & Madaus, G. (2003). *Errors in standardized tests: A systemic problem*. (NBETPP Monograph). Boston, MA: Boston College, Lynch School of Education.
- Texas Education Agency, Pearson Educational Measurement, Harcourt Educational Measurement & Beta, Inc. (2004) *Chapter 9: Quality control procedures. Texas Student Assessment Program*. Technical Digest (2003-2004) <http://www.tea.state.tx.us/student.assessment/resources/techdig04/>
- Toch, T. (2006). *Margins of error: The testing industry in the No Child Left Behind era*. Washington: Education Sector Report.
- Von Davier, A. (2011) *Statistical Models for Test Equating, Scaling, and Linking*. Springer
- Wild, C. L., & Rawasmany, R. (Eds.) (2007). *Improving testing: Applying process tools and techniques to assure quality*. Mahwah, NJ: Erlbaum.
- Zapf, D. & Reason, J. (1994). Introduction: Human Errors and Error Handling. *Applied Psychology: An International Review*, 43: 427-432.