



Міжнародна тестова комісія

Міжнародні тестові домовленості

Керівництво Міжнародної тестової комісії з контролю якості підрахунку балів, аналізу результатів та інформування про них

Фінальна версія

Червень 2012

Переклад:

Коваленко Олексій
Рубаховська Крістіна
Сідорук Євгенія
Тихоненко Дар'я

© 2012, Міжнародна тестова комісія.

Варто цитувати як:

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)]

Подяка

Керівництво було розроблене Аві Аллалуф для Ради Міжнародної тестової комісії. Автор висловлює подяку Маріс Борн за її цінне сприяння у здійсненні цього проекту, а також людям, які аналізували попередні версії і чий вклад є важливим для розробки даного керівництва:

Альваро Арце-Феррер
Джеймс Остін
Джо-Анн Бейрд
Джулія Балбоні
Хелен Барон
Дейв Бартрам
Маріс Борн
Джеймс Бутчер
Джанет Карлсон
Айен Койн
Курт Гайзінгер
Рон Хамблтон
Джон Хатті
Фред Леонг
Джейсон Лампріну
Том Окленд
Фред Освальд
Крістофер Раш

Автор також зобов'язаний своїм колегам з NITE (Національного інституту тестування та оцінки), особливо департаменту підрахунку балів і порівняння, де були розроблені і зазвичай застосовуються більшість процедур контролю якості.

Зміст

1. Вступ

- 1.1. Цілі та задачі
- 1.2. Для кого призначене Керівництво з Контролю Якості (ККЯ)
- 1.3. Локальні та міжнародні фактори
- 1.4. Помилки і необхідність в ККЯ
- 1.5. Область застосування ККЯ
- 1.6. Визначення контролю якості
- 1.7. Приклади з інших професій
- 1.8. Структура ККЯ

2. Загальні принципи

- 2.1. Перевірка стандартів контролю якості, що використовуються в даний час
- 2.2. Основні приготування та домовленості між залученими особами
- 2.3. Ресурси
- 2.4. Потреби і очікування зацікавлених сторін
- 2.5. Персонал і робоча атмосфера
- 2.6. Незалежний моніторинг процедур контролю якості
- 2.7. Фіксація помилок

3. Покрокові робочі рекомендації

- 3.1. Розробка звітів про результати
- 3.2. Демографічні дані
- 3.3. Підрахунок балів
 - 3.3.1. Отримання та зберігання відповідей респондентів
 - 3.3.2. Підрахунок балів
- 3.4. Аналіз тесту
 - 3.4.1. Аналіз завдань тесту
 - 3.4.2. Аналіз результатів тестів дії (performance tests), практичних завдань, ділових ігор, інтерв'ю і т.д.
 - 3.4.3. Прирівнювання (Equating)/Перевірка нових форм тесту і завдань
 - 3.4.4. Підрахунок стандартизованих балів
 - 3.4.5. Перевірка тесту на захищеність від фальсифікації
- 3.5. Інформування про результати
 - 3.5.1. Інформування про результати тесту
 - 3.5.2. Документація

4. Висновки та заключні коментарі

5. Література

1. Вступ

1.1. Цілі та задачі

Стандартизація і точність вагомі на всіх етапах тестування, починаючи з розробки тесту і його застосування, закінчуючи підрахунком балів, аналізом, інтерпретацією результатів та інформуванням про них. Будь-яка особа, включена в підрахунок, аналіз результатів та інформування про результати тесту, несе відповідальність за дотримання професійних стандартів, які можуть бути обґрунтовані відповідними зацікавленими сторонами, включаючи замовників, психологічні асоціації, коледжі та університети, урядові установи та юридичних осіб. Практикуючий професіонал повинен знати та передбачувати помилки, які можуть мати місце на будь-якому етапі, і повинен діяти відповідно з існуючими на даний момент стандартами, щоб попередити та усунути ці помилки. Неточний підрахунок балів як наслідок неправильного ключа до відповідей, некоректна конвертація сирих балів в стандартизовані, помилки в обчисленні, випадкове нарахування балів іншому респонденту або невірна інтерпретація вказаних балів – це приклади помилок, яких необхідно уникнути. Людині властиво помилятися, однак помилки повинні бути зведені до мінімуму шляхом використання адекватних процедур контролю якості. Психологи повинні мати широкі знання про шляхи і способи контролю якості, оскільки це має вирішальне значення для правильного застосування тестів. Ми вважаємо, що цей документ також зробить вагомий внесок в процес Безперервного покращення якості (БПЯ, CQI, Continuous Quality Improvement) в області, в якій ми прагнемо досягти прогресу, крок за кроком.

Керівництво з Контролю Якості (ККЯ), викладене нижче, призначене для підвищення ефективності, чіткості та точності підрахунку балів, аналізу результатів та інформування про них (ПАІ) в процесі тестування. Воно має подвійну функцію: може використовуватися самостійно, як конкретні вказівки для контролю якості ПАІ тестів, або може розглядатися і використовуватися в якості доповнення певних частин Керівництва Міжнародної тестової комісії з використання тестів (2000). Рекомендуємо читачу ознайомитися з керівництвами Міжнародної тестової комісії, зі стандартами AERA, APA и NCME (2011), а також з іншими відповідними міжнародними та державними стандартами.

1.2. Для кого призначене Керівництво з Контролю Якості (ККЯ)

ККЯ поширюється на ситуації великомасштабних тестувань, в першу чергу на ті, в яких використовуються тести успішності або здібностей (на відміну від особистісних опитувальників), як такі, вони широко застосовуються в процесі великомасштабного освітнього тестування і оцінки здібностей, навичок при працевлаштуванні і прийомі на роботу. Тим не менше, більшість з наступних рекомендацій можуть також бути застосовані в процесі менш масштабного тестування та в оцінці, що включає в себе інші види тестів.

ККЯ призначене для осіб, відповідальних за наступні аспекти використання тестів:

- Розробка тестів
- Проведення тестування
- Підрахунок балів
- Аналіз психометричних характеристик тесту (включаючи нормоутворення та стандартизацію тесту)
- Дотримання надійності результатів тесту
- Інтерпретація результатів тесту
- Інформування про результати тесту (створення звітів) та забезпечення зворотного зв'язку респондентам

- Навчання та супервізії осіб, що проводять тестування
- Розробка комп'ютерних систем і програм для обробки результатів тестування
- Правові аспекти тестування
- Видання тестів

Поінформованість з приводу контролю якості важлива для будь-якого професіонала, включеного в процес тестування. Незважаючи на те, що ККЯ спочатку було призначене для професіонального використання спеціалістами, які працюють в сфері тестування, воно також об'єднує основні принципи успішної практики і, як таке, підходить для тих, хто використовує тести виключно в дослідницьких цілях.

1.3. Локальні та міжнародні фактори

ККЯ призначене для міжнародної аудиторії професіоналів, які використовують тести в своїй роботі. Це Керівництво може допомогти професіоналам в розробці специфічних локальних стандартів якості. Локальні фактори, такі як місцеві або державні закони чи стандарти, існуюча регламентація або специфічні контракти між клієнтами та видавцями тестів, повинні враховуватися для реалізації ККЯ на локальному рівні або в процесі їх практичного застосування в заданих умовах. Наприклад, в деяких країнах конфіденційність особистої інформації респондентів захищена законом.

1.4. Помилки та необхідність в ККЯ

Помилки, які допускаються під час процесів ПАІ, можуть мати серйозні наслідки в будь-якій сфері тестування – психологічній, освітній, HR, поведінковій. Наприклад, велика кількість помилок у підрахунку результатів тесту може вплинути на його результати та надійність – надійність підрахунку результатів тесту найімовірніше буде нижчою. Це, в свою чергу, може призвести до меншої прогностичної валідності. В деяких випадках можлива ситуація, коли людина з патологічною поведінкою може бути помилково ідентифікована як людина з нормальною поведінкою. В інших випадках помилки можуть завадити кваліфікованому кандидату отримати роботу або призвести до неправильного прийняття рішень щодо освіти досліджуваного. Помилки також можуть призвести до невиправданого втручання в процес освіти, наприклад, якщо комусь буде призначена невідповідна освітня програма, або якась особа одержить професійну ліцензію, незважаючи на те, що їй бракуватиме необхідних знань та умінь. Помилки можуть призводити до тривалої відстрочки у повідомленні результатів тесту, що в свою чергу, може призвести до серйозних проблем для тих, хто очікує на зачислення в учбові заклади. Таким чином, помилки можуть мати серйозні наслідки і навіть завдавати шкоди. Помилки можуть призвести до втрати довіри до освітніх та психологічних тестів, зменшити їх переконливість в очах громадськості. В деяких випадках помилки можуть призвести до застосування правових заходів по відношенню до тестуючої організації, навчального закладу або компанії, яка шукає кваліфікованих працівників.

Професіонали, які здійснюють процес тестування, є об'єктом потенційного тиску з боку чотирьох джерел: організацій, досліджуваних, тестуючих компаній та засобів масової інформації. Всі вони очікують швидкої та недорогої розробки тестів, швидкого отримання результатів. Для дотримання стандартів якості вкрай важливо, щоб професіонали протидіяли тиску з боку тих, хто намагається прискорити або скоротити тривалість цих процесів або пропустити певні їх етапи. Під значним тиском, наприклад, знаходяться організації, котрі зобов'язані контрактом провести ПАІ результатів тестування у стиснуті строки.

Існує значна ймовірність помилок в тривалих процесах – таких, як розробка тестів, підрахунок балів, аналіз результатів та інформування про них, котрі містять послідовність

етапів, де кожний наступний прямо залежить від попереднього. Стандарти якості повинні ретельно перевірятися і регулярно оновлюватися.

1.5. Сфера застосування ККЯ

ККЯ розповсюджується на великомасштабні процеси тестування. Тим не менше, вони також можуть бути використані для широкого спектру інших ситуацій тестування (наприклад, індивідуальне тестування з ціллю профорієнтації або визначення рівня особистісного розвитку) та технік оцінки (наприклад, тести з множинним вибором, оцінка успішності виконання, структуровані та неструктуровані інтерв'ю, оцінка діяльності групи) і практично для будь-якої ситуації, в якій відбувається оцінка (наприклад, освітні цілі або в центрах оцінки). Деякі рекомендації носять конкретний характер і стосуються різноманітних індивідуальних або стандартизованих тестів, що проводяться в групі, інші – мають набагато ширше застосування (наприклад, в сфері клінічного, освітнього та професійного тестування). Більшість професій використовують практику оцінки (наприклад, в медичних та реабілітаційних, судових, освітніх, пов'язаних з працевлаштуванням цілях), і тут також ККЯ можуть бути корисними.

ККЯ можна застосовувати для різноманітних форм проведення тестування (включаючи бланкові тести і комп'ютеризовану оцінку через інтернет, яка набуває все більшого розповсюдження, в автономному режимі). Конструювання тестів, їх вибір та проведення не входить в фокус ККЯ. Тим не менш, корисність або успішне застосування ККЯ в ПАІ тестів буде залежати від доцільності самого тесту і від того, чи є результати тесту надійними та прогностичними відносно чітко встановлених критеріїв. Виділення ресурсів на контроль якості представляє собою інвестування у відповідну практику, підконтрольність та сумлінність – основні елементи будь-якого етичного кодексу.

1.6. Визначення контролю якості

Контроль якості можна визначити як *формальний систематичний процес, направлений на підтримку високих стандартів якості на всіх етапах підрахунку, аналізу та інформування про результати тесту, і, відповідно, на мінімізацію помилок і підвищення надійності вимірів.*

1.7. Приклади з інших професій

Процедури контролю якості використовуються в багатьох інших професіях, таких як інженерна справа, авіація, розробка програмного забезпечення і медицина. Що стосується галузі медицини, цікаво відзначити деякі з факторів, що провокують помилки, які відбуваються в клініках. До них відносяться невідповідні умови зберігання лікарських засобів, складність медичних втручань, впровадження нових технологій, проблеми комунікації, погану командну роботу та відсутність чіткої політики в області безпеки. Такий приклад має аналогію в області тестування, де акцент робиться на тесті, з усіма можливими помилками, які виникають в процесі тестування і підрахунку балів.

1.8. Структура ККЯ

ККЯ складається з двох основних частин:

1. Загальні принципи – загальні моменти, які повинні бути враховані і узгоджені до підрахунку, аналізу результатів та інформування про них.
2. Покрокові вказівки для роботи.

За цими розділами слідують дві додаткові секції: короткі висновки та список літератури.

2. Загальні принципи

2.1. Перевірка стандартів контролю якості, що використовуються в даний час

- 2.1.1. Визначте, які принципи контролю якості в даний час існують для тестування у вашій організації або у вашій країні. У випадку необхідності, розробіть специфічні процедури контролю якості до проведення тестування. Переглядайте, оновлюйте та модифікуйте принципи у разі змін в процесі, а також час від часу в якості рутинної перевірки.
- 2.1.2. Переконайтесь до проведення тестування в наявності відповідних процедур контролю якості.
- 2.1.3. При роботі з новим тестом, проведіть пілотне дослідження, що включатиме в себе процеси ПАІ. Там, де пілотне дослідження не здійснюється, розгляньте перше дослідження як пробне та будьте готові зробити поправки в процеси ПАІ для подальших досліджень.
- 2.1.4. Створіть специфічні стандарти для кожного тесту у випадку їх відсутності.
- 2.1.5. Створіть стандарти для кожного нового тесту під час його конструювання.

2.2. Основні приготування та домовленості між залученими особами

Перш ніж провести тестування, необхідно узгодити основні принципи між спеціалістами, що відповідають за конструювання тесту, проведення тестування, підрахунок балів, стандартизацію, інтерпретацію, валідизацію та інформування про результати. Хоча вони мають різні обов'язки та ролі, робота всіх включених спеціалістів повинна бути узгоджена; гарно налаштована комунікація між людьми, які приймають участь в процесі, повинна підвищити якість і спростити використання тестів.

Спеціалісти (див. п. 1.2) повинні:

- 2.2.1. Визначити всі зацікавлені сторони в процесі тестування і узгодити, хто відповідає за прийняття рішень відносно різних етапів процесу тестування.
- 2.2.2. Визначити та вказати ціль або цілі використання тесту (наприклад, підбір персоналу, оцінка досягнень, наукові дослідження).
- 2.2.3. Узгодити розклад для процесів ПАІ.
- 2.2.4. Продумати засоби зв'язку між людьми або групами (у випадках, де залучено більше однієї команди).
- 2.2.5. Продумати засоби надання зворотного зв'язку клієнту про процес тестування.
- 2.2.6. Визначити методи передачі даних особам, відповідальним за процеси ПАІ, наприклад, дані, отримані за допомогою оптичного носія або відскановані бланки, або дані в електронному вигляді, отримані в результаті комп'ютеризованого тестування.
- 2.2.7. Визначити показники по субтестам (якщо вони є), а також обґрунтувати ці показники. Необхідно також бути готовим до змін цих значень після одержання даних.
- 2.2.8. Узгодити інструкції по підрахунку балів, тобто кількість балів, які нараховуються за кожну правильну відповідь, і вирішити, як обходитися з неправильними відповідями. Необхідно також бути готовим змінити інструкцію після одержання даних.
- 2.2.9. Обрати шкалу підрахунку балів і визначити діапазон показників в ній.
- 2.2.10. Вирішити, як обходитися з відсутніми даними (у випадках, якщо досліджувані пропустили завдання або помилково перескочили на іншу строку відзначаючи

відповідь, або у тих випадках, коли дослідник забув оцінити конкретного досліджуваного або зробив це неправильним чином, без можливості повторити дослідження).

- 2.2.11. Визначити і описати модель стандартизації, необхідний склад і об'єм вибірки, а також методи стандартизації, що будуть застосовуватися.
- 2.2.12. Визначити та описати модель встановлення стандартів, а також необхідний склад та об'єм вибірки відповідно цим стандартам.
- 2.2.13. Узгодити рівень деталізації, з якою результати повинні бути представлені респондентам та зацікавленим організаціям, а також, яку додаткову інформацію стосовно балів, необхідно надати у звітах.
- 2.2.14. Визначити, які конкретно особи, органи або організації повинні одержувати результати тестів, забезпечивши при цьому дотримання правових обмежень стосовно конфіденційності даних.
- 2.2.15. Визначити, чи можуть (повинні) звіти про результати дослідження містити іншу інформацію (наприклад, чи були зміни в змісті тесту, скільки завдань було виконано, чи були запропоновані особливі умови для людей з обмеженими можливостями).
- 2.2.16. Узгодити необхідність письмової фіксації даних.
- 2.2.17. Узгодити необхідність повторення з ціллю перевірки якості важливих процесів (наприклад, створення таблиць конвертації сирих балів у стандартизовані).

2.3. Ресурси

- 2.3.1. Переконайтесь, що наявні достатні ресурси (матеріальні засоби, час і персонал) для ефективного та належного підрахунку, аналізу результатів та інформування про них.
- 2.3.2. Перевірте доступність резервних ресурсів (наприклад, якщо спеціаліст з стандартизації не може провести процедуру, визначте, хто зробить це замість нього; або, якщо сканер для бланків відповідей вийшов з ладу, знайдіть альтернативний пристрій).
- 2.3.3. Будьте в курсі проблем з розрахунком часу, які можуть виникнути, якщо використовуються резервні ресурси. Заплануйте дії на випадок непередбачуваних обставин, щоб покрити несподівану відсутність ключових співробітників.
- 2.3.4. Розподіліть задачі між членами команди: Хто буде займатися підрахунком балів, аналізом результатів та інформуванням про них? Хто несе відповідальність за весь процес? Спеціалісти, які відповідають за тестування, повинні визначити, наприклад, чи мають ті, хто приймає участь в кожному етапі процесу, необхідні навички для виконання покладеної на них роботи, а також повинні визначити вимоги та рівень автоматизації процесів.
- 2.3.5. Визначте необхідні часові ресурси: встановіть розклад для кожного етапу процесів ПАІ. Крайній термін завершення процесу та інформування про результати тесту повинен бути реалістичним.
- 2.3.6. Переконайтесь в наявності необхідного програмного забезпечення, комп'ютерних та мережевих ресурсів: захищене авторським правом та розроблене на замовлення ПЗ, ноутбуки, ПК, центральні процесори, вільне місце на жорсткому диску і на сервері тощо.
- 2.3.7. Визначте необхідні ресурси робочого простору – чи достатньо робочого місця (кімнат, столів, стільців тощо) для персоналу та досліджуваних.

- 2.3.8. Визначте дії, які необхідно виконати для безпечного зберігання даних в електронному вигляді.
- 2.3.9. Переконайтесь, що все необхідне додаткове обладнання (ключі для підрахунку вручну, калькулятори) доступне.

2.4. Потреби і очікування зацікавлених сторін

Ті, хто використовує результати тестів – досліджувані, батьки/опікуни, викладачі/консультанти, а також ті, хто проводить тестування (наприклад, спеціальні агенції) – мають конкретні потреби і очікування стосовно тестування і часу, відведеного на інформування про результати. Ці потреби та очікування повинні бути обґрунтованими та узгодженими між сторонами (див. також Керівництво МТК по використанню тестів, 2000, Додаток В – Керівництво з розробки угоди між сторонами, включеними в процес тестування).

- 2.4.1. У випадку необхідності, розробіть угоду між включеними сторонами, узгоджену зі спеціалістами, які відповідають за підрахунок балів, стандартизацію та інформування про результати. Враховуйте, що необхідно час від часу вносити зміни в угоду.
- 2.4.2. Визначте, хто несе головну відповідальність і має право вирішувати, як діяти у випадку виникнення проблем.

Наприклад, коли на питання з різними варіантами відповідей немає правильної відповіді, інтерв'юер поводить себе нетактовно або досліджуваних відволікає сторонній галас. Інший випадок – коли було запропоноване запитання, на яке передбачалась лише одна правильна відповідь, але потім досліджуваний доводить, що й інша відповідь також є правильною.
- 2.4.3. Попередньо визначте алгоритм дії у випадку, коли після підрахунку балів виявлена помилка.
- 2.4.4. Забезпечте досліджуваних можливістю поставити під сумнів правильність запропонованих відповідей і оскаржити свої оцінки або поставити власні запитання і запропонувати власні відповіді.
- 2.4.5. Необхідно мати документ, який обґрунтовує систему підрахунку балів по завданням тесту.

2.5. Персонал і робоча атмосфера

Переконайтеся в тому, що люди, відповідальні за процес підрахунку балів, аналіз результатів, стандартизацію та інформування про результати, є професіоналами, які володіють всіма необхідними навичками та знаннями про процеси ПАІ. Переконайтесь, що ці люди володіють всіма необхідні компетенціями для роботи. Якщо в процесі тестування приймає участь група людей, важливо забезпечити їм комфортну взаємодію. Таким чином, при прийомі на роботу нових співробітників, їх здатність працювати в команді і уживатися один з одним повинна стати ключовим фактором.

- 2.5.1. Уникайте необґрунтованого тиску на людей з метою підвищення швидкості виконання роботи.
- 2.5.2. Уникайте зайвих робочих годин.
- 2.5.3. Спробуйте виробити в собі ретельне, уважне ставлення до деталей в роботі (особливо з точки зору попередження помилок), у той же час не перенапружуйтесь. Спокійна, але цілеспрямована робоча атмосфера, є найбільш ефективним середовищем для підтримки високих стандартів.

- 2.5.4. Підтримуйте працівників, забезпечуючи їх професійний розвиток та тренінги, а також, в деяких випадках, їх особистісне зростання та розвиток соціальних навичок.

2.6. Незалежний моніторинг процедур контролю якості

Призначте одного або декількох спеціалістів (в залежності від величини і складності проекту) для моніторингу процедури контролю якості і забезпечення відстеження того, щоб всі питання, проблеми та помилки були враховані.

Спостерігачі повинні працювати незалежно від тих, хто приймає участь в процедурах підрахунку балів, аналізу та інформування про результати. Моніторинг повинен проводитися в співпраці з усіма зацікавленими сторонами з метою врахування специфічних процесів.

2.7. Фіксація помилок

- 2.7.1. Всі, хто приймає участь в процесі тестування, повинні слідувати узгодженим процедурам відносно фіксації діяльності і помилок чи проблем, які можуть виникати.
- 2.7.2. Попередньо узгодьте, хто із співробітників за який етап відповідає.
- 2.7.3. Реєструйте всі види діяльності. Перевіряйте всі процеси на предмет їх якості.
- 2.7.4. Фіксуйте детально всі помилки (незалежно від того, відомі їх причини чи ні), починаючи з характеру помилки, того, хто її виявив, коли, які були наслідки і які кроки були/будуть здійснені задля її усунення. Також фіксуйте випадки, в яких були виявлені помилки до того, як з'явилися їх наслідки.
- 2.7.5. Консультуйтеся в інших спеціалістів стосовно можливих помилок для їх своєчасного попередження, іноді навіть на спеціальних нарадах, присвячених попередженню помилок.
- 2.7.6. Робіть висновки про те, як попередити помилки в майбутньому.

3. Покрокові робочі рекомендації

Вам пропонуються конкретні кроки, які повинні бути здійснені в процесі підрахунку балів, аналізу результатів та інформування про них. При проведенні великомасштабного тестування, кожний етап повинен бути продуманий і проведений з обачністю. Перед початком роботи з реальними даними, варто провести пілотне дослідження, щоб результати були більш точними. В ситуаціях масштабного тестування, де приймають участь тисячі людей, ці принципи повинні дотримуватися особливо чітко. В ситуаціях менш масштабного тестування, де приймають участь десятки людей, ці рекомендації повинні бути реалізовані, однак деякі етапи можуть опускатися або спрощуватися. Це необхідно, оскільки процедури вимагають значних ресурсних затрат і засновані на моделі з великими вибірками. Всі процедури повинні бути адаптовані у відповідності до кількості досліджуваних.

3.1. Розробка звітів про результати

До проведення основних етапів тестування, необхідно узгодити деталі інформування про результати тестування, яке є кінцевим продуктом всього процесу. Необхідно вирішити, що саме варто повідомити, наскільки детально, кому, коли тощо. Недостатньо повідомити результат установі чи досліджуваному лише у вигляді чисел та стандартизованих балів (стенів тощо). Дуже важлива правильна інтерпретація результатів. При розробці тестів, підрахунку балів та аналізі результатів, повинен бути прийнятий до уваги кінцевий продукт – надання зворотного зв'язку про результати. В цьому сенсі, основною ціллю є забезпечення правильного розуміння результатів. Різноманітні аспекти інтерпретації повинні бути роз'яснені на початку звіту. Необхідно узгодити з усіма учасниками умови надання зворотного зв'язку стосовно більш приватних результатів (наприклад, по балам субшкал) в додаток до основних. Чи повинні вони бути представлені? Чи будуть вони використовуватися?

3.2. Демографічні дані

Демографічні дані можуть бути дуже корисними в процесі контролю якості для таких цілей, як підтвердження особистості досліджуваного, розуміння неочікуваних результатів і розподілу досліджуваних по групах з метою стандартизації.

Рекомендуються наступні кроки:

- 3.2.1. Якщо правові норми дозволяють, зберіть демографічні та біографічні дані (вік, стать, національність, освіта, результати попередніх тестувань тощо) за допомогою попередньої реєстрації, реєстрації на місці або після тестування. Зібрані демографічні дані повинні бути релевантними конкретному дослідженню, необхідно захищати конфіденційність досліджуваного настільки, наскільки це можливо.
- 3.2.2. Якщо це можливо, систематично час від часу перевіряйте демографічні дані досліджуваних; звертайте увагу на невідповідність в демографічних даних досліджуваного, який проходить тестування повторно.
- 3.2.3. Проведіть дослідження для виявлення кореляції між демографічними даними та результатами тестування, пошукайте невідповідність результатів тестування з результатами попередніх тестувань, результатами наукових досліджень тощо. Наприклад, може статися, що в конкретному дослідженні дорослі люди демонструють більш високі результати по конкретному тесту, ніж молоді. Якщо згідно результатам дослідження більш молоді люди повинні проходити тест краще, це привід вивчити результати на предмет помилок.

3.3. Підрахунок балів

3.3.1. Отримання та зберігання відповідей респондентів

За необхідності, всі відповіді досліджуваних на папері повинні зберігатися в електронному вигляді, як правило, з ідентифікаційним номером кожного досліджуваного. Зберігання таких матеріалів в паперовому та електронному вигляді здійснюється протягом встановленого місцевим законодавством та професійними стандартами терміну. Це стосується і бланків відповідей, і тестових відповідей, і результатів в електронному вигляді.

- 3.3.1.1. Якщо паперові копії існують, вони повинні зберігатися протягом періоду часу, визначеного законами країни, штату чи провінції, в межах яких проводиться тестування.
- 3.3.1.2. Що стосується електронної версії, використовуйте систему безперебійного електроживлення і резервні батареї для комп'ютерів, щоб зменшити ймовірність переривання роботи і втрати даних.
- 3.3.1.3. Сканери, якщо вони застосовуються, повинні регулярно перевірятися на наявність неполадок.
- 3.3.1.4. Перевіряйте вручну результати сканування даних.
- 3.3.1.5. Перевіряйте бази даних досліджуваних з метою підтримки стійкої системи ідентифікації досліджуваних. Наприклад, зверніть увагу на випадки, коли ідентифікаційний номер співвідноситься більше, ніж є одним ім'ям.
- 3.3.1.6. Всі дані повинні бути захищені. Захищайте особисту інформацію досліджуваних шляхом виявлення їх особистих даних (наприклад, імена) від результатів тестування, коли це можливо. Наприклад, заведіть окремі файли: один з біографічними даними, інший – з результатами, які можуть бути ідентифіковані по ідентифікаційному номеру. Всі ці дії повинні відповідати державним законам, що стосуються інформаційної безпеки і зберігання даних.
- 3.3.1.7. Виконуйте перевірку правильності ключа підрахунку балів і належне використання норм та таблиць перетворення сирих балів в стандартизовані.

3.3.2. Підрахунок балів

Після того, як тестові дані оброблені та надійно збережені в базі даних, відповіді досліджуваних, як правило, використовуються для підрахунку сирих балів. В класичній теорії тестування (Classical Test Theory, СТТ), наприклад, сирі бали зазвичай дорівнюють кількості правильних відповідей (коли в тесті є правильні та неправильні відповіді); іноді застосовуються поправки на вгадування, іноді в тесті є завдання, які "важать" більше, ніж за інші. В теорії "Завдання-Відповідь" (Item Response Theory, IRT) сирій бал відображає приховані (латентні) здібності – вони часто називаються "тета" або "оцінка риси". На результати тесту можуть негативно впливати різноманітні види помилок, наприклад, помилки в ключі. Іноді помилки призводять до вкрай низьких результатів. Використовуйте наступні процедури контролю якості, щоб знаходити та усувати такі помилки:

- 3.3.2.1. Перевірте відповідність структури даних вимогам документу, в який вносяться дані, наприклад порядок елементів в файлі.

- 3.3.2.2. Застосовуйте попередньо узгоджені правила для видалення невалідних протоколів, перекодування відсутньої інформації і обробки повторних випадків.
- 3.3.2.3. Порівнюйте дані вибірки з очікуваним діапазоном балів, а також порівнюйте описові статистики з нормами оригінального тесту, описаними видавцями (у випадку адаптації тесту). Статистики вибірки можуть дещо відхилитися (від очікуваних), але великі відхилення варто відзначати та перевіряти.
- 3.3.2.4. Розгляньте крайні значення окремих осіб або певних груп – як високі, так і низькі. Екстремальні оцінки можуть вказувати на три можливі проблеми: помилки в підрахунку балів, фальсифікацію відповідей або помилку в отриманні даних.
- 3.3.2.5. Перегляньте дані по окремим досліджуваним у випадках, коли різниця в кореляції між балами по субшкалам більше, ніж очікувалось. Для цього необхідно попередньо визначити критичну величину різниці.
- 3.3.2.6. Проаналізуйте завдання тесту і розгляньте описові статистики завдань – помилку в ключі для одного завдання складно виявити, якщо описові статистики завдань недоступні (завдання з неправильним ключем часто здаються складними і можуть негативно корелювати з критерієм).
- 3.3.2.7. Перевіряйте результати на наявність пропущених відповідей. Відповідь може бути видалена з розрахунків деяких досліджуваних помилково.
- 3.3.2.8. Особливу увагу зверніть на групи, які проходили тестування в різних умовах, і виконайте додаткову перевірку отриманих в цих групах даних. Наприклад, досліджувані могли проходити тестування в різний час або з іншою версією тесту, або використанням різних методів відповідей.
- 3.3.2.9. Вирахуйте і розгляньте описові статистики основних груп, що приймають участь в тестуванні, розподілених, наприклад, за місцем тестування, спеціалістом, що проводить тестування, або за інтернет-ресурсом, за допомогою якого здійснювалось тестування. Помилки можуть виникати в результаті використання неправильної форми тесту для конкретних груп досліджуваних.
- 3.3.2.10. Якщо дозволяють ресурси, випадковим чином відібрана група результатів тестування може бути передана окремій групі спеціалістів для підрахунку і аналізу результатів. Пізніше можна порівняти основні результати з результатами незалежної групи спеціалістів.

3.4. Аналіз тесту

3.4.1. Аналіз завдань тесту (зазвичай для великомасштабного тестування)

Аналіз завдань тесту забезпечує базові статичні показники для прийняття рішень за характеристиками завдань тесту і тим, як вони всі разом формують загальний бал. Рекомендується проводити аналіз завдань тесту у всіх випадках, крім тих, в яких кількість досліджуваних невелика. Результати статистичного аналізу завдань тесту включають в себе аналіз рівня складності завдання (або тенденцію викликати установку на погодження – стосується особистісних опитувальників) і дискримінативність. Параметри кожного завдання тесту можуть розраховуватися по-різному в залежності від моделі, яка використовувалася в процесі розробки тесту. Крім того, аналіз завдань представляє собою загальні статистичні дані цього тесту (надійність і/або стандартну похибку, середнє, стандартне відхилення, інформацію про тест, розподіл відповідей досліджуваних тощо). Наступні процедури повинні бути

розглянуті у випадку, якщо кількість досліджуваних перевищує мінімальну, в залежності від моделі, що використовується.

- 3.4.1.1. Використовуйте надійні програми для проведення аналізу завдань тесту і переконайтесь, що вони мають відповідну технічну документацію.
 - 3.4.1.2. Якщо у вас є підстави для сумнівів у якості програми, яку ви використовуєте для аналізу завдань тесту, або якщо ви використовуєте нову програму, використовуйте дві програми одночасно і порівняйте їх результати.
 - 3.4.1.3. Проведіть аналіз завдань тесту після його проведення або проаналізуйте накопичення даних (наприклад, за 3-5 років його використання), якщо тест проводиться періодично. Розгляньте можливість проведення аналізу завдань тесту на неповному наборі даних (поки дані не зібрані повністю), щоб можна було швидко знайти помилки.
 - 3.4.1.4. Перегляньте результати аналізу завдань тесту перед тим, як робити які-небудь висновки про досліджуваних.
 - 3.4.1.5. Аналіз завдань тесту дає можливість визначити проблеми, пов'язані з ключами тесту. Наприклад, те, що здається дуже популярним "дистрактором" (відволікаючий варіант відповіді), може виявитися правильною відповіддю (ключем); або ж негативна кореляція між завданнями може виявити завдання, яке має обернений ключ. Якщо результати певного завдання проблемні, ключ і зміст цього завдання повинні бути переглянуті.
 - 3.4.1.6. Повторіть аналіз завдань тесту, якщо ключ до тесту змінюється або якісь завдання видаляються. Обновлюйте документацію (наприклад, таблиці підрахунку балів і характеристики стандартизації) протягом всього процесу.
- 3.4.2. Аналіз результатів тестів дії (performance tests), практичних завдань, ділових ігор, інтерв'ю тощо.

В той час як аналіз завдань з множинним вибором відповідей (МВ) є об'єктивним (базується на конкретному ключі) і надійним, аналіз завдань з відкритими питаннями (ВП) (тести дії, анкети з відкритими питаннями, практичні завдання, ділові ігри тощо) зазвичай має суб'єктивний компонент. З цієї причини такий вид оцінки, як правило, менш надійний, ніж МВ, оскільки часто вимагає оцінки окремої людини і залежить від її індивідуальних особливостей. Тим не менш, можна використовувати різноманітні методи для зниження суб'єктивності аналізу ВП завдань, а також підвищувати надійність і точність аналізу.

- 3.4.2.1. Переконайтесь, що результати тестів, практичних завдань та інтерв'ю оцінюються навченими асесорами, які володіють необхідними знаннями та досвідом, а також мають повноваження, професійну підготовку або відповідну академічну освіту.
- 3.4.2.2. Інструкції по оцінці ВП відповідей повинні бути чіткими та добре структурованими. Необхідно провести попереднє тестування відкритих питань для більш точного формулювання інструкцій.
- 3.4.2.3. Визначте діапазон відповідей за допомогою розподілу прикладів відповідей по рубрикам.
- 3.4.2.4. Вимагайте від асесорів участі в тренінгових заходах до проведення оцінки. Навчання дасть їм можливість ознайомитися з інструкціями та оцінкою

результатів, перш ніж вони приступлять до оцінки фактичних відповідей досліджуваного.

- 3.4.2.5. Оцінюйте асесорів на основі їх підготовки, перш ніж допустити їх до оцінки результатів конкретного тестування.
- 3.4.2.6. Намагайтеся використовувати не менше двох асесорів для кожного тестування, в залежності від вартості та наявності відповідних ресурсів.
- 3.4.2.7. При наявності одного асесора для всіх досліджуваних (у зв'язку з фінансовими або іншими міркуваннями), використовуйте ще одного асесора для оцінки невеликої кількості частини вибірки (наприклад, для аналізу 10% всіх даних), щоб оцінити надійність результатів.
- 3.4.2.8. Якщо використовується комп'ютерний підрахунок відкритих відповідей, переконайтеся, що результати будуть розглянуті людиною також. Необхідно емпірично обґрунтувати використання комп'ютерного підрахунку, перш ніж використовувати його.
- 3.4.2.9. Переконайтеся, що асесори працюють незалежно один від одного.
- 3.4.2.10. Застосовуйте статистичні методи для оцінки надійності процесу шляхом підрахунку відповідних вимірів погодження серед асесорів, а також протиріч між ними.
- 3.4.2.11. Здійснюйте періодичний моніторинг якості оцінки в режимі реального часу так, щоб була можливість дати зворотній зв'язок.
- 3.4.2.12. Якщо асесор не відповідає очікуванням (результати оцінки ненадійні або недостатньо близькі до результатів інших асесорів), проінформуйте асесора про це відповідним чином і рекомендуйте йому перепідготовку. Замініть асесора, якщо проблема не буде вирішена.
- 3.4.2.13. Розробіть правила роботи з великими відмінностями в оцінках асесорів. Коли відмінності невеликі, вони повинні бути усереднені або підсумовані. Коли є великі відмінності, більш досвідчений асесор може виступати посередником для їх вирішення.

3.4.3. Прирівнювання (Equating)/Перевірка нових форм тесту і завдань

Іноді прирівнювання не потрібне, якщо досліджувані змагаються лише між собою, проходячи тестування одночасно. Якщо не проводити прирівнювання, результати досліджуваних за попередніми тестуваннями неможна переносити та порівнювати з результатами за новою формою тесту з іншими досліджуваними. Якщо бали за різними версіями тесту повинні бути на одній шкалі, нові форми тесту повинні бути еквівалентні попереднім, щоб можна було порівнювати нові форми тесту зі старими, які можуть мати інші психометричні характеристики. Результатом прирівнювання є те, що бали по всім формам тесту мають одне й те саме значення. Прирівнювання може бути проведене до тестування (пре-прирівнювання) і/або після тестування (пост-прирівнювання). Прирівнювання може проводитися на рівні завдань, шкал та всього тесту. Існують різні підходи і методи прирівнювання (наприклад, лінійне прирівнювання, еквіперцентильне прирівнювання, прирівнювання, основане на теорії "Завдання-Відповідь", прирівнювання за респондентами або завданнями).

Прирівнювання зазвичай вимагає великих вибірок, їх розмір залежить від методу та моделі прирівнювання (див. Kolen & Brennan, 2004; Lamprianou, 2007).

- 3.4.3.1. У випадку незрозумілих проблем у прирівнюванні (наприклад, отримані бали нижче, ніж очікувалось), переконайтесь, що всі форми тесту були застосовані в однаково стандартизованих умовах. Якщо ці умови не були еквівалентними, спробуйте оцінити вплив різних умов.
- 3.4.3.2. Переконайтесь в тому, що специфічні процедури і моделі порівняння були дотримані.
- 3.4.3.3. Досліджуйте теоретичні передумови процедури прирівнювання і/або визначте, чи можуть різні процедури прирівнювання з різними теоретичними передумовами давати схожі результати. Проведіть аналіз стабільності параметрів однакових завдань тесту (прирівнювання по завданням). Якщо ви використовуєте набір одних й тих самих завдань для прирівнювання, необхідно обґрунтувати, чому ви видаляєте певні завдання з однієї версії тесту, і вплив таких рішень на сирі і стандартизовані бали. Також фіксуйте зміст і статистичні характеристики спільного для двох версій набору завдань тесту після аналізу даних. Такий стандарт також може застосовуватися до моделі загальної вибірки (прирівнювання за респондентами), однак тоді аналіз фокусується на досліджуваних.
- 3.4.3.4. Порівняйте реальні бали, отримані за конкретними досліджуваними, з тими, які передбачалися на основі особливостей вибірки (див. 3.2.1.). Якщо присутні відмінності, перевірте бали повторно.
- 3.4.3.5. Проведіть порівняння балів та перехідних оцінок в часі. В добре розроблених великомасштабних дослідженнях коливання між результатами тестувань, що проводяться в різний час, доволі малі. Великі відмінності можуть вказувати на проблеми в прирівнюванні балів, наприклад, про зміни в характеристиках популяції.
- 3.4.3.6. Якщо проводиться велика кількість тестувань з невеликою кількістю досліджуваних в кожному, необхідно застосовувати специфічні інструменти контролю якості для відстеження стабільності результатів тесту. Деякі з них: графіки Shewhart и CUSUM за даними тестування, моделі часових серій, моделі зміни точок та інструментів інтелектуального аналізу даних (див. Von Davier, 2011).
- 3.4.3.7. Якщо в тесті є граничні показники, що відділяють респондентів, які пройшли тестування, від тих, хто його не пройшов, або за іншими критеріями, перевірте ці показники. Порівняйте їх з очікуваними показниками, виходячи з попередніх тестувань, особливостей досліджуваних і результатів інших тестів.
- 3.4.3.8. Переконайтеся в послідовності граничних показників в різних групах, використовуйте обґрунтовані методи і фіксуйте процес. Також фіксуйте випадки, в яких ви, з певних причин, відхиляєтеся від стандартного процесу.
- 3.4.3.9. Якщо ви використовуєте різні формати проведення тестування (наприклад, комп'ютеризоване використання тесту, який був розроблений як тест для паперу та олівця), важливо порівнювати характеристики нового тесту з характеристиками старого, а також іноді прирівнювати новий тест зі старим.
- 3.4.3.10. Для тестів, що мають важливі наслідки для досліджуваного (high-stake tests), необхідно намагатися якомога частіше проводити процедуру прирівнювання і залучати для контролю третю сторону, яка не має відношення до процесу прирівнювання.

3.4.4. Підрахунок стандартизованих балів

В багатьох випадках стандартизовані бали використовуються, щоб зробити результати більш зрозумілими. В таких випадках сирі бали можуть підказати нам, які саме стандартизовані бали краще використовувати (наприклад, станаїни, децилі). Показники або таблиці перетворення використовуються для підрахунку балів за шкалами і для відображення стандартизованих балів або прецентиль. Зазвичай сирі бали (кількість правильних відповідей або кількість правильних відповідей з поправкою на вгадування) або тета бали (в тестах, розроблених на теорії "Завдання-Відповідь") повинні бути перетворені в специфічну шкалу. Перетворення проводиться за допомогою спеціальних таблиць або математичної функції (наприклад, при лінійному перетворенні).

- 3.4.4.1. Проведіть відповідні перетворення сирих балів для одержання певних стандартизованих шкал.
- 3.4.4.2. Проведіть процедуру перетворення сирих балів в стандартизовані та інші процедури на предмет помилок.
- 3.4.4.3. Перевірте, чи правильно було здійснено перетворення.
- 3.4.4.4. Переконайтеся в тому, що низькі стандартизовані бали відповідають низьким сирим балам, і що високі стандартизовані бали відповідають високим сирим балам.
- 3.4.4.5. В деяких випадках можуть потребуватися додаткові процедури після перетворення (наприклад, визначення єдиного мінімуму і максимуму для всіх результатів).
- 3.4.4.6. Порівняйте властивості нової форми з параметрами інших форм тесту на предмет незвичайних відмінностей або подібностей.
- 3.4.4.7. Врахуйте зміни, які можуть виникнути в шкалі з часом.
- 3.4.4.8. Порахуйте деякі показники вручну і порівняйте результати з тими, які були підраховані на комп'ютері.
- 3.4.4.9. Перевірте статистичне співвідношення між сирими та стандартизованими балами за допомогою діаграми розсіювання.
- 3.4.4.10. Використовуйте дві різні комп'ютерні програми для підрахунку стандартизованих балів та їх порівняння.
- 3.4.4.11. В технічному посібнику або інших пов'язаних з ним матеріалами, надайте детальний опис процедур, що використовувалися для перетворення сирих балів в стандартизовані. Оскільки ця техніка може бути різною для різних форм тесту, необхідно описати процедуру, яка використовувалась для кожної форми тесту.

3.4.5. Перевірка тесту на захищеність від фальсифікації

Фальсифікація результатів тесту створює серйозну проблему, яка піддає ризику безпеку та цілісність тесту та системи тестування. На жаль, фальсифікація не може бути повністю попереджена, навіть враховуючи постійний контроль та інші засоби попередження фальсифікації; спокуса фальсифікувати результати може бути величезною, особливо в тестах з важливими наслідками для досліджуваного. В якості ще одного інструменту боротьби з фальсифікацією варто проконсультуватися з юристами з ціллю розгляду методів попередження фальсифікації і підтвердження придатності. В тестах ЗНО, які мають важливі наслідки для досліджуваного, обман

може бути здійснений не лише на рівні індивідуального досліджуваного, а й на рівні класу, школи, району, міста, де проводять тестування. Фальсифікація може мати місце там, де безпосередньо відбувається тестування, за допомогою мобільних телефонів або за допомогою інтернету. В сфері оцінки професіоналів, тестування через інтернет стає все більш популярним, що також підвищує ризик фальсифікації результатів. Побічною перевагою перевірки тесту на захищеність від фальсифікації є те, що вона також може допомогти виявити проблеми в процедурах тестування, збору даних та їх зберігання.

Рекомендуються наступні запобіжні заходи:

- 3.4.5.1. Перевіряйте особистість досліджуваних тоді, коли вони заходять в приміщення, де буде проходити тестування, або, коли вони проходять тестування вдома, використовуйте посвідчення особи або біометричні показники – такі, як сканування райдужки оболонки ока або відбитків пальців.
- 3.4.5.2. Краще використовувати різні форми тесту. Коли використовується одна форма, не варто садити досліджуваних, що знайомі між собою (наприклад, живуть в одному районі або ходять в одну школу), поряд один з одним. Наприклад, можна садити досліджуваних в алфавітному порядку.
- 3.4.5.3. Записуйте номер місця і складіть список місць, хто де сидів, щоб полегшити аналіз списування.
- 3.4.5.4. Якщо це необхідно (наприклад, при підозрі списування), використовуйте статистичні показники для окремих випадків списування, основані на схожості бланків відповідей в різних досліджуваних в одному залі та одному місці, в якому проводиться тестування.
- 3.4.5.5. Залучіть навчених, благонадійних спостерігачів і здійсніть ретельне спостереження за ними. Переконайтеся, що у спостерігачів немає конфлікту інтересів.
- 3.4.5.6. Перевіряйте відхилення або неочікувані патерни відповідей (наприклад, на складні запитання дають правильні відповіді, а на прості запитання даються неправильні відповіді).
- 3.4.5.7. Візьміть зразок почерку у кожного досліджуваного до тестування та безпосередньо перед ним, щоб визначити шахраїв або досліджуваних з підозрілою поведінкою. Цю процедуру можна пропустити, якщо немає ніяких проблем з ідентифікацією.
- 3.4.5.8. Проаналізуйте суперечливі результати при повторному тестуванні (якщо можливо, ідентифікуйте повтори), використовуючи статистичний розподіл допустимих відмінностей між результатами на даний момент і результатами, отриманими в попередніх тестуваннях. Сильні відмінності можуть свідчити, що місце досліджуваного зайняла інша особа, або про те, що кандидат міг отримати інформацію про завдання тесту до проведення тестування. З іншого боку, респондент може відповідати інакше на повторному тестування.
- 3.4.5.9. Необхідно мати задокументовану процедуру (якщо необхідно, юридично затверджену) для роботи з підозрами в фальсифікації. Попередньо проінформуйте всіх досліджуваних про застосування різноманітних методів боротьби з фальсифікацією.
- 3.4.5.10. Іноді викладачі зацікавлені в підвищенні оцінок студентів в стандартизованих тестах. З цієї причини вчителям неможна давати доступ до результатів стандартизованих тестів.

- 3.4.5.11. Використовуйте кабінети, що закриваються, та безпечні сервери, щоб убезпечити матеріали та результати тесту. Переконайтеся, що всі, хто приймає участь в розробці завдань та тестів, заслуговують довіри і повністю проінструктовані з питань захисту від фальсифікації. Конфіденційність завдань тесту також повинна бути повністю забезпечена на всіх етапах, починаючи з розробки завдань. Завдання тесту повинні переноситися за допомогою надійних засобів від розробника тесту до реалізатора, всі файли повинні зберігатися та застосовуватися на пристрої USB та/або автономних ноутбуках, а не на стаціонарних комп'ютерах або серверах, де вони можуть стати легкодоступними стороннім особам.
- 3.4.5.12. На комп'ютерах, на яких проходить тестування, повинно попереджатися зберігання та відправка матеріалів тесту. Необхідно уникати підключення до інтернету, якщо це дає можливість передавати матеріали тесту.
- 3.4.5.13. Для того, щоб зберігати матеріали тесту в секреті, необхідно переконатися, що досліджуваний не знімає нічого на фотоапарат, телефон тощо.
- 3.4.5.14. Для того, щоб переконатися в справедливому відношенні до кожного досліджуваного, необхідно забезпечити конфіденційність всім досліджуваним на всіх етапах тестування та підрахунку балів.

3.5. Інформування про результати

3.5.1. Інформування про результати тесту

Результати тесту видаються як досліджуваним, так і замовникам. В ідеалі, результати повинні видаватися в друкованому вигляді. В деяких випадках і контекстах найбільш доцільним варіантом повідомлення про результати є інтернет. Повідомлення про результати тесту повинно відбуватися таким чином, щоб як замовнику, так і досліджуваному було зрозуміле значення результатів.

- 3.5.1.1. Застосовуйте метод фокус-групи, що складається з досліджуваних, або "мозкового штурму", експериментальних досліджень чи індивідуальних інтерв'ю для одержання інформації, яка може бути корисна в розробці зрозумілих та ємких пояснень результатів тестування і створення керівництва по інтерпретації.
- 3.5.1.2. Переконайтесь в тому, що всі, хто отримує звіт про результати тестування, володіють відповідними знаннями для їх інтерпретації і можуть зрозуміти ці результати. Підтвердженням цього повинно виступати те, що звіти за результатами тесту дають можливість спеціалісту, який використовує тест, давати адекватні інтерпретації за ним.
- 3.5.1.3. Створіть комп'ютерно згенеровані звіти з основною інформацією про тест та значення результатів, що допоможе отримувачу найбільш швидко та адекватно зрозуміти його результати.
- 3.5.1.4. Зберігайте дані досліджуваних в паперовому та електронному вигляді для наступного підрахунку міжнародних, національних та обласних результатів тесту, якщо це необхідно.
- 3.5.1.5. З'ясуйте, до якого рівня варто застосовувати різні результати (наприклад, результати за субшкалами тесту можуть бути надто ненадійними для їх використання в прийнятті рішень, що матимуть важливі наслідки). Рішення про те, чи варто повідомляти результати по субшкалам (субтестам), повинно

також ґрунтуватися на (а) теорії даного тесту, (б) цілі тестування і психометричних характеристиках субтестів.

- 3.5.1.6. Користуйтеся допомогою експертів в області зв'язку з громадськістю (PR) у випадку, коли результати і бали за тестами повідомляються в ЗМІ та державними діячами.

Вживайте заходів підтримки надійності повідомлення про результати тесту!

- 3.5.1.7. Вживайте відповідні заходи, щоб індивідуальний звіт про результати тесту не міг бути підробленим досліджуваним.
- 3.5.1.8. Уникайте редагування звіту про результати всієї установи: це може призвести до серйозних проблем. Якщо необхідно змінити один або декілька звітів, використовуйте призначене для цього ПЗ або створіть звіт повторно.
- 3.5.1.9. Кодуйте назви електронних файлів для зберігання та передачі.
- 3.5.1.10. Переконайтеся, що звіти відправляються лише відповідним особам. Не відправляйте звіти, що містять більше інформації, ніж це необхідно. Звичайно, легше відправити один і той же повний звіт всім замовникам, однак для підтримки конфіденційності, лише доречна інформація повинна відправлятися кожному отримувачу.
- 3.5.1.11. Інформуйте установи про те, що лише звіт, який був відправлений напряму в установу, - а не копія, яка була відправлена досліджуваному (яка може бути сфальсифікована), - може використовуватися для офіційних цілей. Також рекомендуйте проводити перевірку звіту по установі.

3.5.2. Документація

Ретельна документація всього процесу підрахунку балів, включаючи описові статистики ключів (середнє, стандартизоване відхилення, медіану, діапазон балів, надійність), і порівняння поточних груп досліджуваних з попередніми групами досліджуваних, необхідно провести та завершити до або одразу після видачі результатів тестування. Висока якість документації може сприяти більшій надійності та точності вищезазначених процесів в майбутньому. Надання публічного доступу до інформації, що стосується тесту, слугує додатковим методом контролю всіх процесів підрахунку балів, аналізу результатів та інформування про них.

Дуже важливо:

- 3.5.2.1. Документувати весь процес, крок за кроком, з ретельною фіксацією всіх етапів, включаючи основні статистики та порівняння по групам.
- 3.5.2.2. Переконатися, що нові форми тесту проводять лише після того, як документація за старою версією тесту буде завершена.
- 3.5.2.3. Надати бажаним доступ до описових статистик, розподілу за віком та статтю. Необхідно надавати короткі пояснення цих статистик. Сумарні статистики захищають анонімність індивідуальних досліджуваних.

4. Висновки та заключні коментарі

На додачу до всіх рекомендацій, зроблених в цьому документі, ми б хотіли запропонувати вам ще деякі поради. Щоразу, коли новий тест пропонується замовникам, необхідно провести ретельне життєве моделювання всього процесу, крок за кроком (див. Texas Education Agency et al., 2004); далі можуть бути випробувані нові процедури тестування. Таке моделювання допоможе у формулювання нових процедур контролю якості. Крім того, процес підрахунку балів, аналізу результатів та інформування про них, складається з послідовних етапів, кожний з яких заснований на успішному проходженні попереднього. Таким чином, рекомендується організувати контроль якості так, щоб можна було перейти на наступний етап без проходження попереднього. Комп'ютеризація процесів стандартизації, модифікації і процедур контролю якості дозволяють зробити це просто, прозоро та ефективно. Однак, хоча переваги комп'ютеризації широко визнаються, необхідно спробувати та адаптувати їх до конкретної ситуації.

5. Література

- 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.