

Гнедых Д. С.

ПРОБЛЕМА ОЦЕНКИ ЭКВИВАЛЕНТНОСТИ КОМПЬЮТЕРНЫХ И ТРАДИЦИОННЫХ ВЕРСИЙ...

Российский психологический журнал, 2020, Т. 17, № 1, 44–59. doi: 10.21702/rpj.2020.1.4

ОБЩАЯ ПСИХОЛОГИЯ, ПСИХОЛОГИЯ ЛИЧНОСТИ, ИСТОРИЯ ПСИХОЛОГИИ

---

УДК 159.9.07 doi: [10.21702/rpj.2020.1.4](https://doi.org/10.21702/rpj.2020.1.4)

*Систематический обзор*

## Проблема оценки эквивалентности компьютерных и традиционных версий психодиагностических методик

**Дарья С. Гнедых**

Санкт-Петербургский государственный университет, г. Санкт-Петербург, Российская Федерация

E-mail: [d.gnedyh@spbu.ru](mailto:d.gnedyh@spbu.ru)

ORCID ID: <https://orcid.org/0000-0003-4955-4779>

---

### Аннотация

**Введение.** В статье рассматриваются теоретико-методологические основания доказательства эквивалентности компьютерной и традиционной форм психодиагностических методик. Представлен анализ исследований психометрических показателей компьютерных версий традиционных тестов, продемонстрирована противоречивость точек зрения в подходах к проверке эквивалентности двух форм методик. Научная новизна заключается в попытке структурирования основных проблем, возникающих в процессе установления эквивалентности компьютерной и бумажной версий методик, а также поиске путей и способов их преодоления. Поднимается вопрос о минимальном, но достаточном наборе математико-статистических методов проверки двух форм на эквивалентность.

**Теоретическое обоснование.** Приведены основные проблемы, связанные с оценкой эквивалентности компьютерной и традиционной форм теста: уровень общекультурной и информационной компетентности респондентов, тревожность, социальное окружение, мотивация к проведению тестирования, сложность в организации одинаковых условий при заполнении компьютерных и «бланковых» форм методик (Л. Н. Бабанин, Y. P. Chua, M. Russell, P. Květon и др.). В качестве методов оценки эквивалентности двух форм теста исследователи чаще всего выбирают: сравнение основных статистических показателей (средних значений, дисперсий и др.), проверку конструктивной валидности и надежности компьютерной версии психодиагностической методики.

**Результаты и их обсуждение.** В результате анализа методологии исследований, посвященных оценке эквивалентности компьютерных и традиционных форм тестов, было выявлено разнообразие подходов: к применению математико-статистических методов для проверки психометрических показателей компьютерных версий традиционных тестов; к выбору дизайна исследования; по отношению к учету специфики ситуации компьютерной психодиагностики. Делается вывод о необходимости формулировки конкретных и структурированных требований к процедуре проверки эквивалентности компьютерной и традиционной форм психодиагностических методик. Предложены рекомендации касательно основных разделов, которые должны войти в данные требования: организация эмпирического исследования, математико-статистические методы, контроль факторов, специфичных для компьютерной психодиагностики и способных оказать влияние на результаты оценки эквивалентности.

### Ключевые слова

психодиагностические методики, психометрические показатели, эквивалентность тестов,

Гнедых Д. С.

ПРОБЛЕМА ОЦЕНКИ ЭКВИВАЛЕНТНОСТИ КОМПЬЮТЕРНЫХ И ТРАДИЦИОННЫХ ВЕРСИЙ...

**Российский психологический журнал**, 2020, Т. 17, № 1, 44–59. doi: 10.21702/rpj.2020.1.4

ОБЩАЯ ПСИХОЛОГИЯ, ПСИХОЛОГИЯ ЛИЧНОСТИ, ИСТОРИЯ ПСИХОЛОГИИ

---

математико-статистические методы, компьютеризация тестов, компьютерное тестирование, компьютерная психодиагностика, методология исследования, валидность, надежность теста

### Основные положения

► одной из методологических проблем проверки психометрической эквивалентности компьютерной и бумажной форм методик является отсутствие стандартизации данного процесса;

► для унификации процесса оценки эквивалентности компьютерной и традиционной форм психодиагностических методик необходимо разработать единые требования к дизайну исследования, а также к методам математико-статистической обработки данных;

► для учета специфических факторов ситуации компьютерного тестирования исследователям рекомендуется пользоваться картами наблюдения или анкетами на выявление уровня компьютерной грамотности респондентов, их мотивации и установок по отношению к компьютерной психодиагностике.

---

### Для цитирования

Гнедых, Д. С. (2020). Проблема оценки эквивалентности компьютерных и традиционных версий психодиагностических методик. *Российский психологический журнал*, 17(1), 44–59. doi: 10.21702/rpj.2020.1.4

Дата получения рукописи: 28.01.2020

Дата окончания рецензирования: 29.02.2020

Дата принятия к публикации: 04.03.2020

### Введение

С ростом темпа жизни, увеличением объема информации и технических возможностей возникает необходимость в поиске способов экономии времени при проведении психологических исследований. Как следствие, происходит перевод «бланковых» (бумажных, или традиционных) психодиагностических методик в компьютерные версии, что позволяет собирать информацию дистанционно, а также автоматизировать процесс обработки данных. Для этих целей используются различные сервисы – Google-формы, конструкторы онлайн-опросов или специализированные программы. В большинстве случаев перевод (из бланка в компьютер) сводится к простому действию – текст опросника (или иного рода стимульный материал) копируется в электронный формат и рассылается всем желающим для заполнения. При этом с интерпретацией результатов поступают схожим образом – нормы «бланковых» методик применяются к их компьютерным «аналогам».

В научном сообществе уже давно поднимается вопрос, насколько эквивалентными являются компьютерная и традиционная формы психодиагностической методики. Mazzeo, Druesne, Raffeld, Checketts, & Muhlstein (1992) придерживаются мнения о том, что необходимо проводить специальные исследования сопоставимости показателей данных форм. При этом за всю историю вопроса наблюдаются противоречия во взглядах на то, какие именно психометрические показатели теста следует подвергать проверке. Существует мнение, что если доказать эквивалентность двух форм, то и валидности их будут эквивалентны (George, Lankford, & Wilson, 1992; Ford, Vitelli, & Stuckless, 1996). Однако L. M. Honaker прямо указывает на то, что показатели валидности, определенные для бумажной формы, не могут автоматически

переноситься на компьютерную (Honaker, 1988), и в таком случае следует проводить дополнительные исследования валидности (Russell, Goldberg, & O'Connor, 2003). С ним солидарны Анастаси и Урбина (2009, с. 93): «...надежность и валидность теста могут варьировать в зависимости от формы предъявления». Батури и Мельникова (2011а, 2011б) говорят о том, что процедура создания компьютерной версии методики – не простое копирование стимульного материала из одной формы в другую, а *модификация* методики, что подразумевает процесс ее переработки. Такую точку зрения поддерживает ряд других ученых: «компьютерная версия методики – это вполне самостоятельная методика, использование которой возможно только после установления групповых норм и других психометрических характеристик» (Вассерман, Иовлев и Червинская, 2010, с. 23). Более того, некоторые исследователи предполагают, что совпадение баллов, набранных по бумажной и компьютерной формам теста, не говорит о том, что обе формы измеряют один и тот же психологический конструкт (что было показано на примере личностных методик) (Meade, Michels, & Lautenschlager, 2004). Несмотря на то, что на данную проблему обратили внимание еще в конце XX в., не все исследователи, как будет показано ниже, проводят полную проверку психометрических параметров компьютерной формы традиционной методики, математико-статистические методы анализа их эквивалентности довольно разнообразны, а выбор данных методов часто не обосновывается авторами.

Наличие разнообразия мнений относительно набора психометрических показателей компьютеризированного теста, требующих проверки, отчасти объясняется отсутствием четких универсальных инструкций по процедуре установления эквивалентности компьютерной и бумажной форм психодиагностических методик. Существуют рекомендации по созданию психодиагностического инструментария в целом, которые позволяют структурировать накопленную в этой области информацию и облегчить работу психологам и другим специалистам, встающим на этот путь (хорошим примером, на наш взгляд, является серия статей Батурина и Мельниковой (2009–2011); American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999) (AERA, APA, NCME)). Проверка эквивалентности компьютерной версии теста его бумажному аналогу – это отдельная специфическая область психодиагностики, которая также требует определенности и стандартов.

Очевидно, что компьютеризация бланковых методик в будущем будет принимать всё большие масштабы. Одновременно нарастает обеспокоенность научного сообщества по поводу валидности компьютерных версий «бланковых» тестов. Таким образом, возникает необходимость стандартизации и контроля процесса перевода бумажных методик в компьютерные формы.

В статье рассматривается опыт проверки психометрических показателей компьютерных форм традиционных психодиагностических методик, поднимается вопрос о минимальном наборе математико-статистических методов проверки двух форм на эквивалентность, а также о необходимости формализации процедуры фиксации специфических условий компьютерного тестирования (непосредственно при проведении исследования на эквивалентность) и их последующего учета при интерпретации результатов.

### **Теоретическое обоснование**

Проблема влияния различных факторов на процесс и процедуру заполнения компьютерной формы «бланковой» методики активно обсуждается в научном сообществе. Бабанин (2010), указывая на факторы, которые могут повлиять на эквивалентность двух форм методик (наличие компьютерного опыта, тревожность, социальное окружение и др.), приходит к выводу,

что основным из них является уровень общекультурной и информационной компетентности (уровень освоения информационных технологий). Но учет компьютерной грамотности респондентов всё же не снимает проблемы различия в данных при проверке эквивалентности двух форм методик (Russell et al., 2003). При переносе «бумажных» тестов в компьютерный формат необходимо максимально приблизить условия заполнения методики к таковым для бумажной версии (соблюдать ограничения во времени предъявления стимульного материала, если это требуется в оригинале методики, учесть возможность исправления ответов, строгого/нестрогого порядка выполнения заданий и т. п.). Это один из способов, который может повысить вероятность эквивалентности данных форм.

Следующий момент, который требует внимания, – это выбор математико-статистических методов анализа данных. Какие действия для проверки психометрической эквивалентности компьютерной и бумажной форм методики можно считать необходимыми и достаточными?

За рубежом в 1986 г. вышло руководство по оценке тестов (APA, 1986), где перечислены основные статистические методы и показатели, позволяющие определить психометрическую эквивалентность компьютерной и традиционной форм теста: описательная статистика (средние значения, дисперсии, распределения и ранговые порядки оценок), конструктивная валидность, надежность. Тем не менее, в исследованиях на проверку эквивалентности авторами не всегда соблюдаются все эти пункты, а в частных случаях проводятся дополнительные виды анализа (например, Bartram, 1994; van de Looij-Jansen, Goldschmeding, & Jan de Wilde, 2006; Květon, Jelínek, Vobořil, & Klimusová, 2007; Chua, 2012). Чем же всё-таки обусловлен выбор методов проверки психометрических характеристик компьютерных версий традиционных психодиагностических методик? Каков минимальный перечень методов, которые являются достаточными для доказательства эквивалентности компьютерной формы бумажной форме, или ее самостоятельности как инструментария?

В таблице 1 представлен обзор исследований эквивалентности компьютерных и бумажных форм тестов за последние 20 лет. В качестве критериев анализа выступали: дизайн исследования; применение тех или иных статистических методов для обработки результатов; обоснование их выбора или цели их применения; учет специфики процедуры компьютерного тестирования (проводилось ли наблюдение за респондентами во время выполнения тестов на компьютере, заполнялись ли дополнительные анкеты на выявление отношения участников исследования к компьютерному тестированию в случае, если оно проходило через Интернет, и т. п.).

## Результаты и их обсуждение

Как видно из таблицы 1, при сравнении бумажных и компьютерных форм одного и того же теста, исследователи задаются разными вопросами. Одни проверяют только надежность компьютерного теста, по умолчанию принимая факт, что он валиден; другие считают достаточным провести корреляционный анализ и сравнение средних значений, чтобы доказать эквивалентность; третьи предпочитают сразу проверить все психометрические параметры компьютерной формы. Возможно ли, чтобы выбор статистических методов зависел от специфики стимульного материала? Кажется, что если методика состоит из вопросов (утверждений) и вариантов ответов (по типу опросника), то при переносе ее в компьютерную форму серьезных изменений произойти не должно. К такому выводу приходят некоторые исследователи: надежность многошкальных тестов сохраняется при их переносе с бланка в компьютеризированную форму (Ромек и Сатин, 2000).

Таблица 1		Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методов				
Автор, год	Название методики, краткая характеристика стимульного материала*	Дизайн исследования**	Математико-статистические методы обработки данных	Цель математико-статистического анализа	Процедура учета специфики компьютерного тестирования (наблюдение или опрос)	Основные выводы об эквивалентности двух форм теста
В. Г. Ромек, Д. К. Сатин, 2000	Тест Айзенка (ОПР; без ЛВ; ВЕРБ; ЛИЧ); Тест уверенности в себе (ОПР; без ЛВ; ВЕРБ; ЛИЧ)	Межгрупповое сравнение Дистанционно (через Интернет); в присутствии специалиста	Сравнение факторных структур теста; индекс надежности (α-Кронбаха); сравнение средних значений и дисперсий; ретестовая надежность	Сопоставление надежности; выявление различий между нормами	Данные о процедуре учета и специфике ситуации компьютерного тестирования не приводятся	Психометрические характеристики методики многошкальных тестов сохраняются; надежность не снижается
И. А. Кибальченко и др., 2004	Методика по изучению интересов младших школьников (ОПР; без ЛВ; ВЕРБ; ЛИЧ); «Карта одаренности» (ОПР; без ЛВ; ВЕРБ; ЛИЧ)	Внутригрупповое сравнение	Коэффициент корреляции Спирмена	Проверка ретестовой надежности	Данные о процедуре учета и специфике ситуации компьютерного тестирования не приводятся	Репрезентативная надежность компьютерных вариантов традиционных методик высокая, можно использовать вместо бумажной версии

Таблица 1 Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методик		Название методики, краткая характеристика ристика стимуляционного материала*	Дизайн исследования**	Математико-статистические методы обработки данных	Цель математико-статистического анализа	Процедура учета специфики компьютерного тестирования (наблюдение или опрос)	Основные выводы об эквивалентности двух форм теста
Автор, год							
Б. В. Иовлев и др., 2006 (Иовлев, Новожилова, Червинская и Щелкова, 2006)	Методика «Невротические черты личности» (ОПР; без АВ; ВЕРБ; ЛИЧ)	Межгрупповое и внутригрупповое сравнение	Ранжирование средних значений шкал опросника в каждой группе; корреляционный анализ	Не приводится	Данные о специфике ситуации компьютерного тестирования не приводятся	Бланковый и компьютерный версии методики эквивалентны	
Р. М. van de Looij-Jansen et al., 2006	YMR questionnaire (опросник мониторинга здоровья молодежи)	Межгрупповое сравнение В присутствии учителя	Двухфакторный дисперсионный анализ (ANCOVAs)	Сравнение результатов компьютерного и бумажного тестирования	Данные о специфике ситуации компьютерного тестирования не приводятся	Для большинства показателей по тестам значимых различий не обнаружено	

Таблица 1		Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методик				
Автор, год	Название методики, краткая характеристика, стимульного материала*	Дизайн исследования**	Математико-статистические методы обработки данных	Цель, математико-статистического анализа	Процедура учета специфики компьютерного тестирования (наблюдение или опрос)	Основные выводы об эквивалентности двух форм теста
R. Květon et al., 2007	Тест Бурдона (ГС; ЛВ; НВЕРБ; КОГ)	Межгрупповое сравнение	Однофакторный дисперсионный анализ; попарные сравнения (тест Тьюки); а-Кронбаха	Сравнение результатов компьютерного и бумажного тестирования; сопоставление надежности	Для учета влияния условий на результаты использовался разный дизайн компьютерной версии теста	Бланковый и компьютерные версии методики значительно отличаются друг от друга
R. Květon et al., 2007	Тест на концентрацию внимания (ТСА) (ГС; ЛВ; НВЕРБ; КОГ)	Межгрупповое и внутригрупповое сравнение	Общая линейная модель (GLM); тест межсубъектных эффектов (многомерная общая линейная модель); попарные сравнения (тест Тьюки); корреляционный анализ	Выявление влияния порядка заполнения теста (К-Б; Б-К; Б-Б; К-К)*** и его формы на результаты; определение надежности	Данные о специфике ситуации компьютерного тестирования не приводятся	Формы теста не являются эквивалентными. Обнаружены специфические эффекты формы и порядка тестирования, влияющие на результаты теста

<p>Таблица 1 Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методик</p>	<p>Название методики, краткая характеристика, стимulusного материала*</p>	<p>Дизайн исследования**</p>	<p>Математико-статистические методы обработки данных</p>	<p>Цель математико-статистического анализа</p>	<p>Процедура учета специфики компьютерного тестирования (наблюдение или опрос)</p>	<p>Основные выводы об эквивалентности двух форм теста</p>
<p>У. Р. Чуа, 2012</p>	<p>Тест Яна Пиа «Креативные и критические стили мышления» (YBRAINS) (ОПР; без АВ; ВЕРБ; КОГ)</p>	<p>Межгрупповое и внутригрупповое сравнение</p>	<p>t-тест для независимых выборок; split-plot ANOVA; тест-ретестовая надежность (Pearson product moment coefficients); <math>\alpha</math>-Кронбаха</p>	<p>Выявление влияния эффекта тестирования и формы теста на результаты; ретестовая надежность; сопоставление надежности двух форм теста</p>	<p>Данные о процедуре учета и специфики ситуации компьютерного тестирования не приводятся</p>	<p>Высокая валидность обеих форм теста; бумажный вариант оказался более чувствительным к времени тестирования и мотивации к тестированию</p>

Таблица 1		Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методик				
Автор, год	Название методики, краткая характеристика стимульного материала*	Дизайн исследования**	Математико-статистические методы обработки данных	Цель математико-статистического анализа	Процедура учета специфики компьютерного тестирования (наблюдение или опрос)	Основные выводы об эквивалентности двух форм теста
P. Žitný et al., 2012 (Žitný, Halama, Jelínek, & Květon, 2012)	Тест интеллектуального потенциала (Test of Intellect Potential (PIP)) (ГС; ЛВ; НВЕРБ; КОГ);	Межгрупповое и внутригрупповое сравнение	Корреляционный анализ; z-преобразование Фишера; расчет величины эффекта (Cohen's d) на основе t-теста для независимых выборок	Сравнение средних значений тестовых баллов в зависимости от пола и места проживания; проверка критерияльной и конструкторной валидности	Данные о процедуре учета и специфики ситуации компьютерного тестирования не приводятся	Доказана конструкторная валидность всех форм тестов, результаты компьютерной версии сопоставимы с традиционной формой методик
	Тест «Венские матрицы» (Vienna Matrices Test) (ГС; ЛВ; НВЕРБ; КОГ)	Межгрупповое и внутригрупповое сравнение	Фри-коэффициент корреляции Гилфорда, хи-квадрат; корреляционный анализ	Анализ соотношения тестовых баллов; ретестовая надежность	Индивидуальное взаимодействие с психологом во время тестирования	Формы тестирования не являются полностью эквивалентными
V. H. Конова, И. В. Нахалева, 2013	«Цветовое зеркало» (ПМ; МП; без ЛВ; НВЕРБ; ЛИЧ)	В присутствии специалиста				

Таблица 1 Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методик						
Автор, год	Название методики, краткая характеристика, стимульного материала*	Дизайн исследования**	Математико-статистические методы обработки данных	Цель математико-статистического анализа	Процедура учета специфики компьютерного тестирования (наблюдение или опрос)	Основные выводы об эквивалентности двух форм теста
О. В. Митина, В. В. Сорокина, 2015 (Митина и Сорокина, 2015)	Методика ценностных предпочтений Шварца (ОПР; без АВ; ВЕРБ; ЛИЧ)	Межгрупповое сравнение	α-Кронбаха; t-критерий Стьюдента; критерий Манна-Уитни; сопоставление дисперсий (критерий Ливина); корреляционный анализ	Проверка надежности; сравнение результатов подвыборок по полу; проверка валидности компьютерной версии методики	Данные о специфике ситуации компьютерного тестирования не приводятся	Доказана психометрическая надежность и валидность компьютерной методики
С. С. Ермаков, 2016	Стандартные прогрессивные матрицы плюс Равена (ГС; АВ; НВЕРБ; КОГ)	Внутригрупповое сравнение	Коэффициент корреляции; t-критерий Стьюдента для связанных выборок	Проверка ретестовой надежности; выявление различий между группами	Данные о специфике ситуации компьютерного тестирования не приводятся	Компьютерный вариант методики может считаться аналогом ее бумажного варианта

Таблица 1		Зарубежные и отечественные исследования, посвященные сравнению компьютерных и бумажных (традиционных) форм психодиагностических методик				
Автор, год	Название методики, краткая характеристика стимульного материала*	Дизайн исследования**	Математико-статистические методы обработки данных	Цель, математико-статистического анализа	Процедура учета специфики компьютерного тестирования (наблюдение или опрос)	Основные выводы об эквивалентности двух форм теста
К. В. Суго-няев и др., 2018	Методика КОТ-30 (укороченная и модифицированная версия Краткого ориентировочного теста) (ОПР; ГС; АВ; ВЕРБ; НВЕРБ; КОГ)	Межгрупповое сравнение Интернет-версия (дистанционная); диалоговая и бланковая версии методики – в присутствии специалиста	Корреляционный и факторный анализ; сравнение средних, мер вариативности и одномоментной надежности; сравнения коэффициентов трудности и дискриминативности заданий, их факторных нагрузок на генеральный фактор	Исследование сопоставимости результатов тестирования в контролируемом и неконтролируемом интернете-формате	Данные о специфике ситуации компьютерного тестирования не приводятся	Психометрические характеристики оценок выполнения теста в Интернете не уступают таковым, рассчитанным в условиях контроля-руемого тестирования

Примечание: \* ОПР – опросник; ГС – графические символы, знаки, рисунки; ПМ – проективная методика; МП – манипуляционная методика; АВ – имеет лимит времени; без АВ – без лимита времени; ВЕРБ – вербальный тест; НВЕРБ – невербальный тест; КОГ – когнитивные особенности; ЛИЧ – личностные особенности;

\*\* описание дизайна исследования по следующему параметрам: внутригрупповое или межгрупповое сравнение; дистанционное исследование или в присутствии психолога/специалиста (если приводятся данные в статье);

\*\*\* К – компьютерная версия методики; Б – бумажная версия методики.

С другой стороны, если стимульный материал представляет собой творческие задания, или методика предполагает работу с картинками, графическими символами, лимит времени и т. д. (как, например, в некоторых когнитивных тестах), то можно предположить, что к проверке психометрических параметров такой компьютерной формы нужно подходить более ответственно. Как показывает анализ исследований, приведенных в таблице, данная логика далеко не всегда лежит в основе выбора статистических методов для обоснования эквивалентности двух форм.

Более того, наблюдается разнообразие подходов к дизайну исследования, призванного проверить эквивалентность двух форм методики (компьютерной и традиционной). Например, в исследованиях Ромека и Сатина (2000), van de Looij-Jansen et al. (2006) используется межгрупповое сравнение (когда одна группа заполняет только бумажную версию методики, а другая – только компьютерную) (см. также Andersson, Kaldo-Sandström, Ström, & Strömgren, 2003; Vecchione, Alessandri, & Barbaranelli, 2012), у Ермакова (2016) и Кибальченко, Устинова и Шаповалова (2004) – внутригрупповое (группа заполняет как бумажную, так и компьютерную формы через определенный промежуток времени) (также Hays & McCallum, 2005), а где-то – как межгрупповое, так и внутригрупповое сравнение (Květon et al., 2007; Кононова и Нахаева, 2013). Каждый из описанных выше вариантов исследований имеет свои преимущества и недостатки. При этом, на наш взгляд, наличие внутригруппового сравнения в большей степени позволяет контролировать влияние личности респондента на заполнение как бумажной, так и компьютерной версий теста. Такого рода влияние может проявиться в ситуации межгруппового дизайна, когда разница в результатах может быть обусловлена не спецификой формы теста, а различиями в личностных особенностях двух групп респондентов (например, в том случае, когда в одну группу случайно попадает больше тревожных испытуемых, чем в другую).

Наиболее тщательно проверить эквивалентность позволяет также учет дополнительных факторов при процедуре заполнения компьютерной версии методики (Chua, 2012). В связи с этим в таблицу был включен такой критерий анализа содержания статей, как «процедура учета специфики компьютерного тестирования (наблюдение или опрос)». Даже относительно незначительная модификация дизайна методики в компьютерном варианте (например, изменение цветовой схемы) может существенно повлиять на результаты (Květon et al., 2007). Также существует мнение, что когнитивная нагрузка при прохождении компьютерного теста иная, чем при заполнении теста на бумаге, и это приводит к «расслоению выборки», когда худшие показывают результаты еще хуже, а лучшие еще лучше» (Ермаков, 2016, с. 203). К тому же, примирение дизайна пре-тест и пост-тест может привести к проявлению эффекта тестирования (опыт прохождения пре-теста влияет на результаты пост-теста), а исследователь иногда, упуская это из виду, делает вывод о том, что влияние на результаты последнего тестирования было оказано условием прохождения (например, компьютером, если он был вторым по порядку). Во избежание неверных выводов, связанных с влиянием эффекта тестирования, Chua (2012) предлагает использовать экспериментальный план Р. Л. Соломона для четырех рандомизированных групп (подробнее о плане см. Solomon, 1949; Campbell & Stanley, 1963; и др.). Кроме того, существует еще одна проблема, связанная с проведением компьютерного тестирования, предполагающего удаленный доступ (через Интернет), – это фальсификация результатов (Сугоняев, Радченко и Соколов, 2018). В таком случае рекомендуется перепроверять их в контролируемых условиях (The International Testing Commission, 2006).

Учет всех этих моментов позволяет более тщательно подойти к продумыванию дизайна исследования – выбору количества групп респондентов, последовательности проведения серий тестирования и т. д. На наш взгляд, также будет полезной организация наблюдения за поведением респондентов во время работы с компьютерной версией теста – это поможет исключить из анализа результаты респондентов, демонстрирующих высокий уровень тревожности во время тестирования, низкий уровень мотивации или свою некомпетентность при взаимодействии с компьютерным интерфейсом (информационную некомпетентность). Дополнительно результаты такого наблюдения помогают скорректировать процедуру компьютерного тестирования или же подачу стимульного материала, если большинство респондентов выказывают затруднения при ее прохождении. Всё это можно организовать на этапе создания компьютерной версии традиционного теста, чтобы исключить влияние некоторых факторов на результаты эквивалентности двух его форм.

Обращаясь к проблеме психометрических свойств компьютерной версии традиционной методики, К. В. Сугоняев с соавт. предлагает конкретную процедуру статистического анализа, позволяющего, по их мнению, судить об эквивалентности двух форм теста. Исследователями высказывается суждение о том, что сравнение «мер центральной тенденции, вариативности и одномоментной надежности интегральных тестовых оценок» (Сугоняев и др., 2018, с. 8) является недостаточным для проверки сопоставимости результатов тестирования. В качестве метода, призванного доказать эквивалентность двух форм методик, предлагается анализ выполнения теста на пунктовом уровне – сходство/различие «распределения коэффициентов трудности и дискриминативности тестовых заданий, а также паттернов факторных нагрузок пунктов на генеральный фактор» (Сугоняев и др., 2018, с. 18). Таким образом, авторами предпринята попытка обоснования и доказательства необходимого минимума статистических методов для вынесения суждения об эквивалентности компьютеризированной версии методики и ее бумажного аналога. В остальных статьях (табл. 1), как правило, исследователями не поясняется выбор методов математико-статистического анализа (а также дизайн исследования (межгрупповой/внутригрупповой)) с точки зрения достаточности для проверки эквивалентности. И всё же, поскольку условия заполнения компьютерной формы теста отличаются от таковых при заполнении бумажной, и с учетом всех перечисленных выше факторов, способных оказать влияние на процесс выполнения и результат компьютеризированного теста, мы предлагаем не отказываться от идеи проверки валидности и надежности компьютерной версии традиционной методики.

Проведенный анализ результатов исследований, посвященных эквивалентности компьютерной и бумажной форм тестов, также позволяет сделать вывод о том, что большинство из них опубликованы зарубежными специалистами. В отечественной практике, несмотря на обилие программ для компьютерной психодиагностики (Мельничук и Сергеев, н. д.; Будко, Мишин и Трегубова, 2007), данные о процедуре проверки психометрических свойств компьютерных аналогов традиционных тестов приводятся достаточно редко. Также представлено небольшое количество статей, позволяющих оценить качество проверки эквивалентности компьютерных версий методик их бумажным аналогам.

### **Заключение**

Вне всякого сомнения, психометрические показатели компьютерных версий традиционных методик должны подвергаться проверке. При этом *необходимы четкие требования*

к процедуре такой проверки и обоснованные инструкции для специалистов, разрабатывающих компьютерные версии традиционных методик.

Очевидно, что кратких рекомендаций, представленных в Guidelines for computer-based tests and interpretations (APA, 1986), или подробных, но лишенных конкретности инструкций из Standards for educational and psychological testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), недостаточно.

На наш взгляд, данные рекомендации должны содержать четкие и обоснованные инструкции по каждому из следующих разделов:

1. Требования к дизайну исследования, необходимые для тщательной проверки эквивалентности бумажной и компьютерной версий психодиагностической методики (количество групп, минимальный размер выборки, последовательность проведения разных форм теста, временной интервал между повторными измерениями, способы проверки конструктивной валидности и т. д.).
2. Перечень необходимых и достаточных методов математико-статистической обработки данных, позволяющих констатировать эквивалентность бумажной и компьютерной форм методики.
3. Описание факторов, которые могут повлиять на результаты компьютерного и бумажного тестирования, и рекомендации для организации определенных условий в процессе исследования эквивалентности. В качестве методов, обеспечивающих учет данных факторов, выступают карты наблюдения за поведением респондентов во время выполнения теста, а также стандартизированные анкеты, позволяющие выявить мотивацию, установки по отношению к компьютерной диагностике, уровень компьютерной грамотности и т. п. (например, опросник на выявление мотивации к тестированию (Testing Motivation Questionnaire) (Chua, 2012), опросник «Знакомство с компьютером» (Computer Familiarity Questionnaire) (Mazzeo et al., 1992)). Это поможет исключить влияние дополнительных переменных на результаты тестирования и получить более достоверные данные об эквивалентности традиционных и компьютерных форм психодиагностических тестов.

## Литература

- Анастаси, А. и Урбина, С. (2009). *Психологическое тестирование*. Санкт-Петербург: Питер.
- Бабанин, Л. Н. (2010). Проблема эквивалентности традиционных и компьютеризированных психологических методик. *Экспериментальная психология в России: традиции и перспективы*. Доступ 27 октября 2019, источник [http://psyjournals.ru/exp\\_collection/issue/32974.shtml](http://psyjournals.ru/exp_collection/issue/32974.shtml)
- Батурин, Н. А. и Мельникова, Н. Н. (2009). Технология разработки тестов. Часть I. *Вестник ЮУрГУ. Серия Психология*, 6(30), 4–14.
- Батурин, Н. А. и Мельникова, Н. Н. (2009). Технология разработки тестов. Часть II. *Вестник ЮУрГУ. Серия Психология*, 7(42), 11–25.
- Батурин, Н. А. и Мельникова, Н. Н. (2010). Технология разработки тестов. Часть III. *Вестник ЮУрГУ. Серия Психология*, 8(4), 4–18.
- Батурин, Н. А. и Мельникова, Н. Н. (2010). Технология разработки тестов. Часть IV. *Вестник ЮУрГУ. Серия Психология*, 11(40), 13–28.
- Батурин, Н. А. и Мельникова, Н. Н. (2011а). Технология разработки тестов. Часть V. *Вестник ЮУрГУ. Серия Психология*, 12(5), 4–14.

Гнедых Д. С.

ПРОБЛЕМА ОЦЕНКИ ЭКВИВАЛЕНТНОСТИ КОМПЬЮТЕРНЫХ И ТРАДИЦИОННЫХ ВЕРСИЙ...

**Российский психологический журнал**, 2020, Т. 17, № 1, 44–59. doi: 10.21702/rpj.2020.1.4

ОБЩАЯ ПСИХОЛОГИЯ, ПСИХОЛОГИЯ ЛИЧНОСТИ, ИСТОРИЯ ПСИХОЛОГИИ

---

- Батурин, Н. А. и Мельникова, Н. Н. (2011б). Технология разработки тестов. Часть VI. *Вестник ЮУрГУ. Серия Психология*, 13(18), 48–59.
- Будко, В. Н., Мишин, Д. Ю. и Трегубова, Т. А. (2007). Тестирующий программный комплекс готовности студентов и школьников к обучению в вузе. *Инфокоммуникационные технологии*, 5(3), 167–169.
- Вассерман, Л. И., Иовлев, Б. В. и Червинская, К. Р. (2010). Компьютерная психодиагностика в теории и практике медицинской психологии: этапы и перспективы развития. *Сибирский психологический журнал*, 35, 20–24.
- Ермаков, С. С. (2016). Особенности компьютерной диагностики интеллектуальных способностей учащихся VI–IX классов. *Психологическая наука и образование*, 8(4), 199–207. doi: [10.17759/psyedu.2016080419](https://doi.org/10.17759/psyedu.2016080419)
- Иовлев, Б. В., Новожилова, М. Ю., Червинская, К. Р. и Щелкова, О. Ю. (2006). Методологические аспекты изучения эффективности компьютерной психодиагностики. *Вестник Санкт-Петербургского университета. Сер. 6. Философия, политология, социология, психология, право, международные отношения*, 2, 115–124.
- Кибальченко, И. А., Устинов, Д. А. и Шаповалов, С. Н. (2004). Компьютерная психодиагностика как условие развития одаренности. *Известия ТРТУ*, 6, 301–303.
- Кононова, В. Н. и Нахаева, И. В. (2013). Исследование психометрической эквивалентности двух версий проективной психодиагностической методики – традиционной («цветовое зеркало») и компьютеризированной («цветомер личности»). *Вестник Московского университета. Серия 14. Психология*, 1, 136–151.
- Мельничук, А. и Сергеев, В. (н. д.). Современные компьютерные системы для эффективной оценки персонала. *Психология и бизнес*. Доступ 13 декабря 2019, источник <https://psycho.ru/library/93>
- Митина, О. В. и Сорокина, В. В. (2015). Ценности старшеклассников: разработка компьютерного диагностического инструментария. *Вестник Московского университета. Серия 14. Психология*, 1, 42–59.
- Ромек, В. Г. и Сатин, Д. К. (2000). Сохранение надежности многофакторных тестов при их использовании в сети Интернет. *Психологический Журнал*, 21(2), 70–75.
- Сугоняев, К. В., Радченко, Ю. И. и Соколов, А. А. (2018). Добровольное Интернет-тестирование как источник валидных оценок группового психометрического интеллекта. *Сибирский психологический журнал*, 69, 6–32. doi: [10.17223/17267080/69/1](https://doi.org/10.17223/17267080/69/1)
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Andersson, G., Kaldö-Sandström, V., Ström, L., & Strömberg, T. (2003). Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. *Journal of Psychosomatic Research*, 55(3), 259–262. doi: [10.1016/S0022-3999\(02\)00575-5](https://doi.org/10.1016/S0022-3999(02)00575-5)
- APA (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- Bartram, D. (1994). Computer-based assessment. In C. L. Cooper (Ed.), *International review of industrial and organizational psychology* (pp. 31–69). London: Wiley.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company.
- Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, 28(5), 1580–1586. doi: [10.1016/j.chb.2012.03.020](https://doi.org/10.1016/j.chb.2012.03.020)

- Ford, B. D., Vitelli, R., & Stuckless, N. (1996). The effects of computer versus paper-and-pencil administration on measures of anger and revenge with an inmate population. *Computers in Human Behavior*, 12(1), 159–166. doi: [10.1016/0747-5632\(95\)00026-7](https://doi.org/10.1016/0747-5632(95)00026-7)
- George, C. E., Lankford, J. S., & Wilson, S. E. (1992). The effects of computerized versus paper-and-pencil administration on measures of negative affect. *Computers in Human Behavior*, 8(2–3), 203–209. doi: [10.1016/0747-5632\(92\)90004-X](https://doi.org/10.1016/0747-5632(92)90004-X)
- Hays, S., & McCallum, R. S. (2005). A comparison of the pencil-and-paper and computer-administered Minnesota Multiphasic Personality Inventory–Adolescent. *Psychology in the Schools*, 42(6), 605–613. doi: [10.1002/pits.20106](https://doi.org/10.1002/pits.20106)
- Honaker, L. M. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8(6), 561–577. doi: [10.1016/0272-7358\(88\)90081-5](https://doi.org/10.1016/0272-7358(88)90081-5)
- Květon, P., Jelínek, M., Vobořil, D., & Klimusová, H. (2007). Computer-based tests: The impact of test design and problem of equivalency. *Computers in Human Behavior*, 23(1), 32–51.
- Mazzeo, J., Druesne, B., Raffeld, P. C., Checketts, K. T., & Muhlstein, A. (1992). Compatibility of computer and paper-and-pencil scores for two CLEP® general examinations. *ETS Research Report Series*, 1. doi: [10.1002/j.2333-8504.1992.tb01446.x](https://doi.org/10.1002/j.2333-8504.1992.tb01446.x)
- Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2004, April). Are online and paper-and-pencil personality tests truly comparable? *Symposium presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology*. Chicago, IL.
- Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer-based Testing and Validity: A look back and into the future. *Assessment in Education: Principles, Policy & Practice*, 10(3), 279–293. doi: [10.1080/0969594032000148145](https://doi.org/10.1080/0969594032000148145)
- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, 46(2), 137–150. doi: [10.1037/h0062958](https://doi.org/10.1037/h0062958)
- The International Testing Commission. (2006). International guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6(2), 143–171. doi: [10.1207/s15327574ijt0602\\_4](https://doi.org/10.1207/s15327574ijt0602_4)
- van de Looij-Jansen, P. M., Goldschmeding, J. E. J., & Jan de Wilde, E. (2006). Comparison of anonymous versus confidential survey procedures: Effects on health indicators in Dutch adolescents. *Journal of Youth and Adolescence*, 35, 652–658. doi: [10.1007/s10964-005-9027-0](https://doi.org/10.1007/s10964-005-9027-0)
- Vecchione, M., Alessandri, G., & Barbaranelli, C. (2012). Paper-and-pencil and web-based testing: The measurement invariance of the Big Five Personality Tests in applied settings. *Assessment*, 19(2), 243–246. doi: [10.1177/1073191111419091](https://doi.org/10.1177/1073191111419091)
- Žitný, P., Halama, P., Jelínek, M., & Květon, P. (2012). Validity of cognitive ability tests – comparison of computerized adaptive testing with paper and pencil and computer-based forms of administrations. *Studia Psychologica*, 54(3), 181–194.

**Конфликт интересов отсутствует**