
Ю. А. Тюменева, А. А. Яременко

РУКОВОДСТВО ДЛЯ АВТОРОВ, ПУБЛИКУЮЩИХ РЕЗУЛЬТАТЫ РАЗРАБОТКИ ОЦЕНОЧНОГО ИНСТРУМЕНТА

Статья поступила
в редакцию
в декабре 2012 г.

Изложены требования журнала «Вопросы образования» к публикациям, в которых докладываются результаты разработки какого-либо инструмента оценки. Термин «инструмент» используется здесь в качестве обобщающего понятия для общераспространенных форм стандартизированных процедур оценки, таких как тест, опросник, шкала, методика и т.д.

Аннотация

Ключевые слова: тестирование, разработка инструментария, тесты, опросники, шкалы, валидность, надежность, нормы.

Цель руководства — предоставить авторам общую структуру для описания процедур, использованных при разработке инструмента, и полученных результатов. Оно может использоваться авторами и редакцией как чек-лист, помогающий заметить и восполнить возможные дефициты проведенной работы.

В руководстве указываются основные содержательные элементы, которые обязательно должны быть представлены в рукописи. Рецензент имеет право не рассматривать рукопись и вернуть ее автору на доработку, если какой-либо из данных элементов отсутствует. Безусловно, любая другая информация, которая, по мнению автора, имеет значение для его работы и представления полученных результатов, может быть включена в текст рукописи.

Редакция журнала рассчитывает, что авторы, разрабатывающие инструмент, владеют знаниями о сути процедур, необходимых для установления валидности и надежности инструмента оценки, и осведомлены об основных правилах интерпретации получаемых коэффициентов.



Обязательные рубрики и их краткое содержание

1. Теоретические основания разрабатываемого инструмента

Представить основные теоретические положения, в рамках которых разрабатывается инструмент. Показать теоретические связи измеряемого конструкта с иными важными с точки зрения планируемой процедуры валидизации конструктами. Описать ожидаемые результаты валидизации с точки зрения теоретических ожиданий. Указать, как эти теоретические представления могут направить процесс разработки инструмента, подбора заданий, начисления баллов, включения тех или иных шкал и т.д.

Например, если автор валидизирует инструмент оценки критического мышления и сравнивает данные, полученные на субгруппах мужчин и женщин, необходимо пояснить, каких межгрупповых различий мы вправе ожидать исходя из существующей теории, связывающей критическое мышление и пол.

2. Предшествующие или уже существующие инструменты оценки этого же конструкта или близких конструктов. Их психометрические характеристики (если известны). Недостатки предшествующих инструментов, которые обуславливают необходимость разработки нового. Ожидаемые преимущества нового инструмента оценки.

Например, автор может указать, что существующие инструменты оценки этого конструкта довольно затратные в смысле времени их заполнения, тогда как предлагаемый инструмент, имея ту же надежность, требует для своего заполнения в 2 раза меньше времени.

Необходимость создания нового инструмента может быть связана также с недостатками прежнего дизайна валидизации, невысокими коэффициентами надежности, риском влияния на ответы социальной желательности, отсутствием национальной версии теста и т.д., и т.п.

3. Описание разработанного инструмента

3.1. Структура инструмента

Необходимо описать количество шкал и субшкал (при наличии), которые содержит инструмент. Количество вопросов (заданий) в полном инструменте и по отдельным шкалам. Привести содержательное название каждой шкалы и примеры пунктов методики, которые к ней относятся.

Кроме формулировок заданий нужно описать используемую шкалу ответов.

3.2. Подсчет сырых баллов (скоринг)

Показать алгоритм подсчета сырых баллов. Если в инструменте представлены задания с разными способами оценки, то следует указать их количество и номера.

Например, тест может состоять из двух частей:



- часть А содержит 35 заданий, которые оцениваются дихотомически (1 балл — правильный ответ, 0 баллов — неправильный ответ);
- часть Б содержит 5 заданий, которые оцениваются политомически (0 баллов — неправильный ответ, допущено две и более ошибки; 1 балл — частично правильный ответ, допущена одна ошибка; 2 балла — полностью правильный ответ, ошибок нет). Максимально возможный балл за тест — 45 баллов.

Если по замыслу разработчика сырые баллы следует переводить в стандартизированную шкалу, об этом нужно написать и указать вид шкалы (например, T-баллы, процентиля, z-баллы и др.).

4. Метод установления надежности и валидизация инструмента

4.1. Участники и процедура

Описывается логика привлечения к исследованию тех или иных групп участников. Обязательно указываются численность и демографические характеристики участников: половой состав (количество мужчин и женщин); возрастной состав (разброс, среднее, стандартное отклонение).

Иные особенности выборки должны быть указаны, если они важны с точки зрения целей разработчика и их влияния на полученные результаты. Например, «сотрудники такого-то предприятия» или «участники такой-то тренинговой программы».

Следует описать ситуацию обследования, форму участия (индивидуальная или групповая), способы мотивации и привести другую важную для понимания процедуры проведения апробации информацию.

4.2. Используемые инструменты оценки

Если применяли другие инструменты кроме разрабатываемого, необходимо обосновать их использование и дать их краткое описание. Если доступна информация о психометрических особенностях этих инструментов, нужно предоставить эти сведения с соответствующей ссылкой на первоисточник.

4.3. Стратегия анализа

Описать использовавшиеся при анализе математико-статистические методы и кратко обосновать их выбор. Кроме того, указать программное обеспечение, которое привлекалось для обработки и анализа данных.

Пример описания процедуры анализа: «Для проверки первичной структуры опросника проводился подтверждающий факторный анализ с использованием программы EQS 6.1 на основе сырых данных с использованием метода *Maximum Likelihood*».

Далее раскрывается последовательность проведения статистического анализа. Например: «Для выделения вторичной структуры опросника будет проверена надежность субшкал и после этого будет использован эксплораторный и подтверждающий факторный анализ».



5. Результаты

Описать результаты установления надежности и валидности инструмента.

5.1. Надежность получаемых баллов

Указать вид надежности, который устанавливался, полученные коэффициенты и дать их интерпретацию. В зависимости от назначения инструмента обоснованным может быть приведение показателей следующих видов надежности: ретестовая, эквивалентных половин, внутренняя согласованность (альфа Кронбаха). Если инструмент содержит несколько шкал и устанавливались разные виды надежности, рекомендуется сводить результаты такой проверки в обобщающую таблицу (пример — см. табл. 1).

Таблица 1 Показатели надежности Методики

Название шкалы/субшкалы	Количество пунктов	α Кронбаха	Ретестовая надежность
Шкала 1	15	0,90	0,96
Субшкала 1	5	0,82	
Субшкала 2	5	0,81	
Субшкала 3	5	0,78	
Шкала 2	15	0,88	0,98
Субшкала 1	5	0,80	
Субшкала 2	5	0,68	
Субшкала 3	5	0,77	
Методика в целом	30	0,90	0,99

5.2. Валидность

После того как было показано, что разработанный инструмент способен измерять некоторую характеристику на приемлемом уровне надежности, следует обосновать, что он измеряет именно то, что планировали разработчики.

На практике широко используются несколько основных видов валидности, подлежащих оценке: содержательная, критериальная и конструктивная. В зависимости от особенностей разрабатываемого инструмента может быть избрана та или иная стратегия валидации или их сочетание.

Например, необходимость оценки факторной структуры теста интеллекта как части исследования конструктивной валидности может быть вызвана тем, что теория, стоящая за разрабатываемой шкалой, подразумевает неоднородность конструкта.

Использование метода контрастных групп при установлении критериальной валидности может быть связано с теоретическим предположением о различной выраженности исследуемой черты или способности у этих групп.

