

ТЕСТТИК ТАПШЫРМАЛАРДЫ ДАЯРДООДО ШКАЛАЛШТЫРУУНУН ТҮРДҮҮ ЖОЛДОРУ

РАЗЛИЧНЫЕ ПОДХОДЫ К ШКАЛИРОВАНИЮ ТЕСТОВЫХ РЕЗУЛЬТАТОВ

Kaidieva N.K

DIFFERENT APPROACHES TO SCALING TEST RESULTS

Аннотация: В данной статье показана необходимость выработки надежного, управляемого, объективного инструментария оценки качества образования. Это приводит к внедрению в практику образования новой системы педагогического измерения – тестирования, которая широко начинает использовать тесты. Тесты являются педагогическими измерителями, а целью измерения является получение численных (количественных) эквивалентов, отождествляемых с оценками измеряемой переменной – уровнем учебных достижений. Также в статье дано определение термина шкалирование и определены виды шкал. Раскрыты цели проведения процедуры шкалирования и различные подходы к нему в разработке тестов.

Annotation: In this article, the necessity of reliable, operated, objective tools development of an education quality assessment is shown. It leads to introduction into practice the formation of new pedagogical measurement system – testing which widely starts using tests. Tests are pedagogical measuring instruments; the purpose of measurement is getting the numerical (quantitative) equivalents identified with estimation of the measured variable – the level of educational achievements. In addition, the definition of the term scaling and the types of scales are given in the article. The carrying out scaling procedure purposes and various approaches to it in development of tests are opened.

Ключевые слова: образование, измерение качества образования, тестирование, шкалирование, виды шкал, подходы к шкалированию.

Key words: education education quality measurement, testing, scaling, types of scales, approaches to scaling.

В настоящее время одной из важнейших задач образования становится выработка надежного, управляемого, объективного инструментария оценки качества образования, его соответствие меняющимся концепциям обучения и контроля, практическим подходам, образовательным потребностям личности. Изменения, происходящие в обществе, определили необходимость инструментального измерения качества образования с целью его повышения.

Поэтому в практику образования, наряду с традиционной системой контроля и оценки качества обучения учащихся и студентов, внедряется новая система педагогического измерения – тестирование, которая широко начинает использовать тесты на основе математических логистических моделей Г. Раша, А. Бирнбаума и др.

Конструируемые на основе этих моделей тесты являются педагогическими измерителями, а целью измерения является получение численных (количественных) эквивалентов, отождествляемых с оценками измеряемой переменной – уровнем учебных достижений, который в какой-то степени отображает латентный параметр обучающегося – подготовленность.

Получение численных эквивалентов происходит из трансформации качественных характеристик и для этого применяются различные методы шкалирования. Дадим определение термина «шкалирование» и видов шкал уровней измерения.

Шкалирование – метод моделирования реальных процессов с помощью числовых систем. В психологии шкалирование является одним из важнейших средств математического анализа изучаемого явления, а также способом организации эмпирических данных, получаемых с помощью наблюдения, анкетирования, экспериментов, тестирования. Общий процесс шкалирования состоит в конструировании по определенным правилам самой шкалы.

Стивенс (1946 г.) определил виды шкал четырех уровней измерения, отличающихся по степени, в которой принадлежащие им шкальные оценки сохраняют свойства множества вещественных чисел. Это – номинальный, порядковый, интервальный уровни и уровень отношений. Номинальные шкалы – это те, в которых числа используются как метки для элементов системы данных и не имеют свойств, позволяющих ввести упорядочивание элементов, равные расстоя-

ния между единицами измерения или фиксированное начало шкалы.[1], [3]

Величины в порядковой шкале могут быть преобразованы в другие величины с помощью любого правила, которое сохраняет первоначальную информацию о порядке данных элементов. Такое преобразование называют монотонным.

Интервальная (метрическая) шкала позволяет количественно оценивать расстояние между объектами или субъектами и начало («ноль») этой шкалы не определено и может выбираться произвольно.

Шкала отношений имеет свойства упорядоченности, равных расстояний между единицами и фиксированного начала (или абсолютного нуля) и объекты или субъекты классифицируются пропорционально степени выраженности измеряемого свойства. В шкалах отношений производятся все измерения физических объектов: вес, длина, возраст в годах или днях и т.д.

Давайте *ответим на вопрос*: Для чего и когда следует использовать процедуру шкалирования?

Для обоснованного сопоставления результатов учащихся между собой, тестовые баллы в соответствии с рядом критериев и норм (число правильно выполненных заданий при дихотомической оценке результатов выполнения каждого задания, сумма оценок по отдельным заданиям при политомической, или взвешенной, оценке) переводятся в производные показатели.

Таким образом, процесс шкалирования состоит в преобразовании сырых баллов в производные показатели, обеспечивающие адекватную интерпретацию и сравнение результатов выполнения педагогических тестов. Перевод сырых баллов в производные показатели и их размещение на готовой шкале не могут повысить надежность и валидность данных по тесту. Согласно расширенной трактовке, шкалирование включает ряд последовательных этапов, охватывающих все компоненты педагогических измерений, и имеет связь с качеством результатов.

Рассмотрим построение шкал для педагогических измерений. При трактовке процесса шкалирования можно выделить четыре основных этапа построения измерительных шкал в образовании для ситуации бланкового тестирования и обобщенного случая измерений:

Этап 1 – определение цели измерения, выбор конструкта, размерности и содержательной области, адекватно описывающей конструкт.

Этап 2 – разработка заданий и экспертное обоснование их качества, экспертное оценивание адекватности содержания заданий конструкту, определение первоначальной длины теста.

Этап 3 – апробация, эмпирический анализ качества теста, чистка и коррекция измерителя для

повышения надежности и валидности шкалы, проверка размерности пространства измерений или доказательство одномерности теста,

Этап 4 – подтверждение качества шкалы и анализ возможности ее использования для представления результатов учащихся по тесту.

Последний этап начинается с построения устойчивой шкалы, выбранной в соответствии с целями измерения и подходом к созданию теста. При последующем использовании теста сырые баллы учеников отображаются на готовой шкале. Особую важность на данном этапе имеет процедура выравнивания результатов педагогических измерений, полученных учащимися по разным вариантам теста.[2]

Необходимость выравнивания может быть не совсем понятна педагогу-практику, поскольку в школе принято выдавать существенно различающиеся по трудности варианты контрольных работ, а затем присваивать одинаковые оценочные эквиваленты разным, зачастую несопоставимым, результатам учащихся. В практике педагогических измерений утвердилась другая норма сравнения и интерпретации результатов испытуемых, основанная на выравнивании, которое представляет собой статистический метод преобразования оценок испытуемых по различным вариантам для обеспечения их сопоставимости.

Рассмотрим общие цели шкалирования. Процесс шкалирования реализует разные цели в зависимости от подхода, выбранного к разработке теста. При нормативно-ориентированном подходе шкалированные показатели позволяют уточнить место, занимаемое результатом испытуемого относительно норм, или сравнить результаты испытуемых, установив место результата каждого учащегося по отношению к результатам остальных учащихся, выполнявших этот тест.

При критериально-ориентированном подходе шкалированный балл показывает процент освоенного содержания и место результата учащегося в сравнении с критериальным баллом. Перечисленным целям отвечают разные шкалы, которые можно построить по результатам выполнения теста.

При выборе метода шкалирования часто обращаются к стандартным показателям, указывающим отличие индивидуального результата испытуемого от среднего балла по выборке в единицах стандартного отклонения. Эти показатели используются для установления места первичного балла каждого испытуемого в сравнении с результатами других на основе подсчета нормированных отклонений и называются z-оценками. Результат отображения z-оценок на числовую ось образует Z-шкалу.

В педагогических измерениях чаще всего обращаются к *стобальной* или *тысячебальной шкале*, полученным на основе преобразования z-оценок. Хотя тысячебальная шкала обладает высокими дифференцирующими возможностями, обычно ее концы оказываются не работающими в силу специального подбора по трудности заданий теста для приближения частотных распределений оценок трудности к виду нормальной кривой. Поэтому, как правило, оценки испытуемых распределяются в интервале от 200 до 800 баллов. Но даже использование менее протяженного диапазона оценок, чем тысячебальная шкала, требует специальных профессиональных навыков по интерпретации баллов учащихся.

Как осмыслить свой результат, если он, например, равен 570 или 650 баллам? Как отнести его к категории плохих или хороших результатов на столь широком диапазоне баллов? Другое дело, если результат испытуемого составляет 5 или 6 баллов по девятибалльной шкале. Поэтому к растянутым шкалам обычно обращаются профессиональные тестовые службы для массового тестирования в образовании, когда большое число испытуемых требует повышения дифференцирующей способности шкалы.

Во многих центрах тестирования при шкалировании данных была выбрана стобальная шкала, в которую переводятся оценки учащихся. Конечно, стобальная шкала — это своего рода компромисс между потребностью в хорошем дифференцирующем эффекте шкалы из-за зна-

чительного числа тестируемых во всех регионах и постепенным переходом от пятибалльной шкалы, существовавшей на протяжении многих лет, к более растянутому шкалам.

Для получения надежных и обоснованных результатов итоговой аттестации учащихся тестовый балл иногда дополняют развернутой содержательной интерпретацией, описывающей характеристики уровня подготовки их в терминах освоенных элементов содержания.

Такие шкалы, позволяющие совместить интерпретацию оценки испытуемого по отношению к результатам остальных тестируемых и к уровням освоения содержания, выделенным по критериальному принципу, получили название *уровневых*.

В заключении можно сказать что, при разработке педагогической процедуры оценивания исследователь должен выбрать метод шкалирования, который позволяет оценить уровень образовательных достижений учащихся.

Литература:

1. Балыхина Т.М. Словарь терминов и понятий тестологии. -М: МГУП, 2000.
2. Гусев А. Н., Уточкин И. С. Психологические измерения: Теория. Методы: Учеб.пособие для студентов вузов. – М.: Аспект Пресс, 2011. – 319 с.
3. Крокер Л. Введение в классическую и современную теорию тестов: учебник. – М.: Логос, 2010. – 668 с.