

Теория

ШКАЛИРОВАНИЕ ТЕСТОВЫХ РЕЗУЛЬТАТОВ

Вадим Аванесов

testolog@mail.ru

Шкалирование определяется как трансформация тестовых результатов из одной числовой системы в другую числовую систему.

Выделены шкалы классификационная (номинальная), поряд-ка, шкала упорядоченной классификации, интервальная и про-порциональная.

Интервальная шкала представляется шкалой нормированных отклонений (Z -шкалой) и производной от неё T -шкалой. Главны-ми условиями корректного применения Z и T шкал признаны на-дёжность исходных результатов тестирования и нормальное рас-пределение данных.

Шкалирование на научной основе предполагает трансформа-цию исходных тестовых баллов в баллы т.н. истинных значений результатов испытуемых (θ). Истинные значения тестовых бал-лов испытуемых определяются на латентной шкале измеряемого свойства личности.

Определение

В литературе приводится довольно много определений шкалирования. Это сделано в трудах Maranell¹, Torgerson² и других. В настоящей работе даётся самое короткое определение шкалирования: это трансформация исходных тестовых результатов из одной числовой системы в другую числовую систему.

Роль элементарной единицы учёта наличия интересующего признака выполняют контрольные задания, подобранные для определения уровня подготовленности. Объектом измерения является носитель интересующего свойства личности. Наиболее корректное средство измерения — педагогический тест. Получаемые при измерении числа позволяют глубже проникнуть в суть изучаемых явлений, что особенно важно для научно обоснованного тестового контроля.

Цель и задачи

В педагогических измерениях при разработке тестов обычно шкалируются уровни подготовленности испытуемых (θ_i) уровни трудности заданий β_j и уровни дифференцирующей способности заданий (a_j).

Шкалирование проводится с целью объективного сравне-

ния результатов испытуемых, полученным по разным тестам, с разным числом заданий, и разных систем оценивания результатов испытуемых.

Шкалирование необходимо также для решения задач формирования комплексной оценки уровня подготовленности испытуемых по результатам нескольких тестов, для определения места испытуемых, для выравнивания баллов в нескольких вариантах теста и др.

Проблемная ситуация

В настоящий момент вопросы шкалирования приобрели отчётливо спорный, если не сказать проблемно-конфликтный, характер в рамках проведения т.н. Единого государственного экзамена (ЕГЭ).

Проблемную ситуацию чётко охарактеризовал Ю.Н. Каргин. Он считает, что разработчики и организаторы проведения ЕГЭ кардинально изменили систему шкалирования результатов в 2011 году. Вследствие чего существующая в России система оценивания знаний из метрической области оказалась сдвинутой в область описательных характеристик и неметрических оценок³.

Ю.Н. Каргин делает вывод, что основной источник проблем действующей сейчас методики шкалирования заложен

1 —
Maranell O.M.
Scaling: A Sourcebook for
Behavioral Scientists.
Chicago: Aldine, 1974.

2 —
Torgerson W.S.
Scaling. (1958) Theory
and Method of Scaling.
N-Y, Wiley.

3 —
Каргин Ю.Н.
Метод независимого
шкалирования результа-
тов ЕГЭ // Педагогичес-
кие измерения. 2013 №1.

в отсутствии какой-либо научной основы. Официальная методика носит скорее умозрительный, искусственный характер и, по-видимому, не имеет под собой ни теоретического обоснования, ни достаточного эмпирического подтверждения⁴. Этот автор предложил новую систему шкалирования, которая существенно лучше существующей.

Критика существующей практики шкалирования содержится также в статье Л.М. Нуриевой и С.Г. Киселёва, публикуемой в настоящем номере журнала «ПИ». Они справедливо отмечают, что учащиеся, делающие лишь единичные ошибки, МАКСИМАЛЬНО НАКАЗЫВАЮТСЯ потерей тестовых баллов⁵.

Например, за работу наугад на экзамене по русскому языку «непредвзятая диагностика ЕГЭ» одарит лоботряса 7–8 первичными или 15–17 тестовыми баллами. Сильных школьников, потерявших 7–8 первичных баллов из 80 возможных (10%), она оборвует на пятую часть тестовых (20%). Отсюда понятной становится уже почти закономерная картина: неожиданно высокие результаты у детей, учившихся в школе на «3–4», и относительно низкие — у отличников, призеров олимпиад и различных конкурсов по предметам. **Именно нелепая система оценивания**

позволяет обвинять школы в «липовых» медалистах, а олимпиады и конкурсы — в слабых победителях. Собственно в этом кривом зеркале ЕГЭ и отражается состояние школьной подготовки⁶.

Как отметил в своём интервью газете «РУ» учитель Григория Перельмана и Станислава Смирнова, двух лауреатов Филдсовской премии, педагог, воспитавший 90 призеров и победителей международных олимпиад по математике, профессор РГПУ им. Герцена, замдиректора по науке, физмат лицея №239 Сергей Рукшин, «Ломоносовых в России уже больше не будет»⁷. Основную причину такого положения он видит в бюрократизации управления образованием.

Виды шкал

С.С. Стивенсу мы обязаны первой, понятной, а потому популярной классификацией шкал⁸. Он выделил четыре шкалы. Каждая из шкал связана с допустимостью использования цифр, чисел, с различиями в интерпретации возможного смысла чисел.

Номинальная шкала

Например, номера футболистов ничего не говорят об уровне мастерства. Потому что но-

Теория

4

4

Там же.

5

Нуриева Л.М.,
Киселёв С.

Единый государственный экзамен нуждается в существенных изменениях // Педагогические измерения. 2013. №4.

6

Там же.

7

«Ломоносовых больше не будет». Сергей Рукшин рассказал «Газете.Ru», в чем ошибки реформы российского образования. 30 ноября 2013 г.

http://www.zavuch.info/news/news_main/456/

8

Стивенс С.С.

Экспериментальная психология. Т. 1. Иност. лит-ра, 1960. С. 19–89.

ПЕД	
	измерения

мера — это цифры, которые не содержат числового и операционального смысла. Цифры нельзя складывать или делить, на основе таких операций невозможно сформировать новые конструктивные суждения. Так же примерно надо относиться к нумерации комнат в гостиницах. Большой номер не обязательно указывает на более качественное помещение для проживания.

Из этих двух примеров видно, что в номинальной шкале используются цифры, а не числа. Вот почему в классификации С.С. Стивенса *номинальная* шкала рассматривается как необходимая прелюдия к измерениям, но не к результатам измерения.

Цифры, используемые для названия, являются обычно оцениванием объектов по признаку наличия или отсутствия интересующего признака. В психометрике уже давно сложилась традиция: наличие признака оценивать числом один, а отсутствие — числом ноль. Это случай не измерения, а оценки. Примеры оценивания: «отличники», «хорошо успевающие», «удовлетворительно подготовленные» «и получившие неудовлетворительные оценки».

При оценивании с опорой на понятие «номинальная шкала», в соответствующем векторе-столбце матрицы испытуемые одного уровня — напри-

мер, отличники оценивают обычно одним баллом. Всех прочих в этом столбце оценивают нулём. Таким образом, их отделяют от испытуемых других классов. В следующем столбце матрицы одним баллом оценивают тех, кто получил балл «хорошо»; все остальные опять получают ноль. И т.д. Каждый вектор-столбец представляет испытуемых одного класса (уровня) подготовленности. Это делается на основе сходных учебных показателей.

Однако если в процессе исследования возникает задача определить тех, у кого чаще (или больше) выражен интересующий признак, то те же единица и ноль могут быть истолкованы как числа. Тогда открывается возможность складывать проявления признака у каждого испытуемого и делать вывод о больших или меньших различиях между испытуемыми по этому признаку. Таким образом, можно выйти на раскрытие смысла понятия «уровень подготовленности». Процесс превращения цифр номинальной шкалы в числа уровневой шкалы наглядно представлен в последнем столбце табл. 1.

Порядковая шкала

По значениям чисел последнего столбца табл. 1 можно судить о некотором уровне (порядке)

Таблица 1

Пример матрицы результатов тестирования

Номера заданий							
Номера испытаний	1	2	3	4	5	6	Сумма баллов
1	1	1	1	1	1	1	6
2	1	1	1	1	1	0	5
3	1	1	1	1	0	1	5
4	1	1	1	0	1	0	4
5	1	1	1	0	0	0	3
6	1	1	0	0	0	0	2
7	1	1	0	0	0	0	2
8	1	1	0	0	0	0	2
9	1	0	0	1	0	0	2
10	1	0	0	0	0	0	1
Сумма	10	8	5	4	3	2	33

баллов подготовленности испытуемых. Поэтому использование единиц, нулей и других чисел в рамках номинальной шкалы могут рассматриваться как числа, имеющие смысл для определения места испытуемого. Именно содержательный смысл в гуманитарных науках открывает возможности трансформации исходных данных в порядковую шкалу.

Порядковая шкала имеет два варианта. В подлинно по-

рядковой шкале каждый испытуемый (или интересующий объект) имеет свой, отличный от всех других испытуемых ранг. Ранг — это число, указывающее на место, занимаемое каждым испытуемым в порядковой шкале.

Пример оценки испытуемых на порядковой шкале приведен в табл. 2.

Здесь каждому испытуемому поставлен в соответствие свой ранг. У каждого испытуе-

Таблица 2

Номера заданий							
Номера испытаний	1	2	3	4	5	6	Сумма баллов
1	1	1	1	1	1	1	6
2	1	1	1	1	1	0	5
3	1	1	1	0	1	0	4
4	1	1	1	0	0	0	3
5	1	1	0	0	0	0	2
6	1	0	0	0	0	0	1
Сумма	6	5	4	2	3	1	21

ПЕД	
	измерения

мого свой ранг, отличный от других испытуемых. Имеет смысл и сумма рангов. Но такая ситуация полного упорядочения случается редко.

А потому возникает необходимость введения ещё одной шкалы, не существующей в шкалах С.С. Стивенса — шкалы упорядоченной классификации.

Шкала упорядоченной классификации

Расположение результатов испытуемых в табл. 2, на порядковой шкале, случается очень редко. Чаще бывает так, что многие, особенно средние, результаты повторяются у нескольких испытуемых. Такие повторяющиеся результаты являются неразличимыми по уровню подготовленности, а потому часть из них включается в один общий уровень (класс) испытуемых. Отсюда возникает статистическое понятие *частоты (f)*. В табл. 3 частоты указывают, сколько человек получили одинаковый исходный тестовый балл.

Пример распределения результатов девяти, взятых для примера, испытуемых можно увидеть в табл. 3. Из неё видно, что четыре исходных тестовых балла получили двое испытуемых, а пять исходных баллов

получили трое испытуемых. Здесь нет отношений строгого порядка между испытуемыми.

В таких случаях некоторые испытуемые упорядочиваются, а некоторые классифицируются — а здесь это те, кто получили баллы 4, 5 и 6. И только затем все тестируемые упорядочиваются. При этом попавшие в один класс испытуемые получают один общий (связанный) ранг.

Результат упорядочения на основе связанных рангов представлен в последней колонке табл. 3. Тех, кто получил в данном примере четыре исходных тестовых балла, можно упорядочить с остальными, дав им девятый с половиной ранг. Тем, кто имеет 5 исходных баллов (а их трое), логично дать седьмой ранг. И наконец, тем, кто получил 6 баллов, даётся четвёртый с половиной ранг. Только в этом случае все испытуемые становятся различимыми (расставленными) на порядковой шкале, или иначе, на шкале рангов.

Интервальная шкала

В интервальной шкале фиксируется не только общее количество интересующего свойства у испытуемых, но и различия между ними по данному свойству. В этой шкале возникает необходимость иметь единицу

Таблица 3

Баллы	Частоты f	Сум. F	Ранг
1	1	13	13
2	1	12	12
3	1	11	11
4	2	10	9–10
5	3	8	6–8
6	2	5	4–5
7	1	3	3
8	1	2	2
9	1	1	1

измерения. В психологических и педагогических тестах это баллы теста, что даёт возможность ответить на вопрос: на сколько единиц измерения результат одного испытуемого выше результата другого испытуемого? Вследствие этого, числа, используемые в интервальной шкале, не только имеют свойства однозначности и упорядоченности, отмечавшиеся для порядковых шкал, но и свойство различий между значениями используемых чисел. В интервальной шкале уместен вопрос: на сколько баллов результат одного испытуемого отличается от баллов другого испытуемого?

В качестве средней арифметической интервальной шкалы нередко выбирается значение нуля, что имеет некоторые преимущества. Примеры использования интервальной шкалы дают нам баллы интеллектуальных (психологических) и педагогических тестов, шкала Г. Раша и др.

Как отмечают Дж. Гласс и Дж. Стэнли, измерение любой непрерывной переменной величины должно сопровождаться определением точности процесса измерения⁹. Этим объясняется потребность введения более дробных значений используемых шкал.

Пропорциональная шкала

Главное свойство пропорциональной шкалы — наличие бесспорного нулевого значения интересующего признака. В психофизике конца XIX и начала XX веков много обсуждался вопрос о возможности наличия человека с нулевым значением интеллекта. И тогда же был поставлен важный вопрос: а можно ли называть человеком лицо, имеющее нулевой интеллект? В итоге дискуссий была принята идея о невозможности создания шкалы интеллекта с нулевым значением,

Теория

ИЗДАНИЕ

9

Дж.Гласс,
Дж. Стэнли.
Статистические методы
в педагогике и психологии.
М.: Прогресс, 1976.
С. 20.

как противоречащую сути психологических измерений.

Примерно такая же логика присуща и педагогическим измерениям. Здесь декларируемый нулевой уровень знаний может оказаться слишком субъективным, поскольку испытуемые студенты и школьники что-нибудь всё-таки знают. Оценка «ноль» в таких случаях сомнительная и спорная.

В отсутствии бесспорного нуля шкалы становятся невозможными утверждения типа «во сколько раз один испытуемый знает больше другого испытуемого». Поэтому в педагогических измерениях уже много лет назад утвердилось мнение о неперспективности создания пропорциональных шкал. Недавняя публикация Ю.Н. Каргина по альтернативному методу шкалирования результатов ЕГЭ вновь возвращает нас к вопросу о возможности создания новых шкал, соотносящихся, в той или иной мере, со свойствами пропорциональной шкалы.

Четыре науки

Почти в каждой науке существует круг исследователей, занимающихся вопросами шкалирования изучаемых там объектов, по интересующим свойствам. Например, в физике

шкалируются данные по таким величинам, как температура, длина, скорость, масса и другим. При этом температура может измеряться по шкале Фаренгейта или Цельсия, длина — в сантиметрах или футах, вес — в килограммах или фунтах. Это примеры разных шкал измерения одного и того же свойства (признака).

Вопросами шкалирования занимаются четыре науки:

1. Собственно шкалирование как прикладная специальная наука, развивающаяся в технике, педагогике, психологии, социологии и т. п. Эта наука имеет общее английское название Scaling с добавлением названия науки.

2. Педагогические, психологические и иные измерения, связанные с разработкой тестов и интерпретацией тестовых результатов. При этом обычно шкалируются баллы испытуемых, уровень трудности заданий теста и показатели дифференцирующей способности каждого задания теста.

В двухпараметрической модели математической теории измерений (Item Response Theory) в результате шкалирования получаются значения θ_i — баллов испытуемых на латентной шкале, значения меры трудности заданий теста (β_j) и значения уровня дифференцирующей способности каждого задания теста (a_j), где

индекс j означает номер задания теста.

3. Статистика, как наука, тоже занимается шкалированием исследуемых объектов по интересующему признаку. Именно в статистике используются т.н. центрирование баллов испытуемых, что означает операцию вычитания среднего арифметического из балла каждого испытуемого ($X_i - M$), а также статистические методы трансформации данных в различные шкалы.

4. Математика и математические методы шкалирования, вопросы непрерывности (континуума) измерения, вопросы размерности шкал, допустимости различных преобразования, аксиомы и пр.

В психометрике уже давно в качестве графического аналога континуума принимается числовая ось, простирающаяся до бесконечности, вправо и влево, от $-\infty$ до $+\infty$.

Условия качественного шкалирования

Задачи шкалирования обычно возникают на заключительных этапах тестового процесса, когда возникает потребность объективного сравнения результатов одного испытуемого по разным тестам, имеющим разное количество заданий.

Распространённую практику сложения всех баллов испытуемого, по всем тестам, и выведения суммы как критерия успешности, как это делается в ЕГЭ, надо признать ошибочной ввиду несопоставимости самих тестов, имеющих различающиеся средние арифметические и показатели вариации, а также разные формы распределения, выражаемые показателями эксцесса и асимметрии. Это теперь уже установлено фактически, в статье, печатаемой в данном номере¹⁰.

Для сложения таких исходных данных требуется процесс стандартизации результатов тестирования.

Предпосылкой для качественного шкалирования результатов рассматриваются качественные педагогические тесты, позволяющие получать измерения со свойствами интервальной шкалы и нормального распределения результатов, а также алгоритмы шкалирования, основанные на науке. В специальной литературе эти свойства справедливо расцениваются как признаки фундаментальности проводимых измерений.

Шкала нормированных отклонений

В педагогических измерениях нет шкалы более популярной и полезной, чем шкала нормиро-

Теория

73

10

Нуриева Л.М.,
Киселёв С.

Единый государственный экзамен нуждается в существенных изменениях. Педагогические измерения. № 4. 2013.

ванных отклонений тестовых баллов испытуемых от среднего арифметического тестового балла. Это так называемая Z-шкала. Популярность объясняется функциями и свойствами этой шкалы, лёгкостью расчёта и возможностью корректного сравнения тестовых баллов испытуемых по нескольким тестам, имеющим разное число заданий. Различия в количестве заданий каждого теста обычно препятствуют объективному сравнению в таких ситуациях. Перевод данных в Z-шкалу снимает данное препятствие.

Пример расчёта z-шкалы

Для перевода результатов тестирования по любому тесту в z-шкалу используются две операции:

- определяются значения отклонений тестовых баллов X_i от среднего арифметического значения M_x . Получается значение отклонения тестового балла, $x_i = X_i - M_x$. Эту операцию иногда называют центрированием данных. Деление x_i/s даёт значение нормированного отклонения тестового балла испытуемого под номером i .

Далее каждое значение отклонения x_i делится на стандартное отклонение этого же теста s_x . Получается $Z_i = x_i/s_x$.

Эту операцию называют нормированием данных. В качестве единицы нормирования тестовых данных выступает стандартное отклонение соответствующего теста. В результате центрирования и нормирования значения любого теста переводятся в значения стандартной z-шкалы.

Для центрирования данных необходимо знать значение среднего арифметического балла (M_x). Последний вычисляется по формуле

$$M_x = \frac{\sum X_i}{N},$$

где символы при знаке Σ , в числителе, означают так называемые пределы суммирования тестовых баллов всех испытуемых X_i , с первого до седьмого включительно; N — число испытуемых.

Находится значение среднего арифметического балла:

$$M_x = \frac{28}{7} = 4 \text{ балла.}$$

Для нормирования данных требуется знание значения стандартного отклонения (S_x) вычисляется по формуле:

$$S_x = \sqrt{\frac{SS}{N}},$$

где SS_x является суммой квадратов отклонений от среднего арифметического значения. SS_x считается по формуле.

$$\sum X^2 - \frac{(\sum X)^2}{N}$$

В нашем примере $\sum X^2 = 140$, $\sum X = 28$, $(\sum X)^2 = 784$. Подставим в наш пример:

$$SS_x = 140 - \frac{784}{7} = 140 - 112 = 28. \quad \text{Свойства z-шкалы}$$

$$S_x = \sqrt{\frac{SS}{N}}, \quad s_x = \sqrt{\frac{28}{7}} = 2.$$

$$Z_1 = \frac{1-4}{2} = -1,5,$$

$$Z_2 = \frac{2-4}{2} = -1,0,$$

$$Z_3 = \frac{3-4}{2} = -0,5,$$

$$Z_4 = \frac{4-4}{2} = 0,$$

$$Z_5 = \frac{5-4}{2} = 0,5,$$

$$Z_6 = \frac{6-4}{2} = 1,0,$$

$$Z_7 = \frac{7-4}{2} = 1,5.$$

Результаты расчётов представлены в табл. 4

Значения z-шкалы удобно вычислять в электронных таблицах Excel.

Z-шкала обладает следующими свойствами:

1. Средняя арифметическая z-шкалы равна нулю: $M_z = 0$. Доказательство:

$$M_z = \frac{\sum Z}{N};$$

$$\frac{\sum Z}{N} = \frac{\sum (X_i - M)}{x} / N.$$

Но по свойству средней арифметической

$$\sum (X_i - M) = 0.$$

Следовательно,

$$\frac{\sum Z}{N} = \frac{0}{N} = 0.$$

Таблица 4

Баллы X	Z – значения баллов X
1	-1,5z
2	-1,0z
3	-0,5z
4	0z
5	0,5z
6	1,0z
7	1,5z
$\sum X = 28$	$\sum z = 0$

Теория

Это означает, что при переводе в z-шкалу среднее арифметическое значение любого теста равно нулю. Среднее арифметическое в z-шкале даёт общую точку для сравнения баллов, что очень удобно для сравнения тестовых баллов различных испытуемых по различным тестам.

2. Все значения меньше среднего арифметического в z-шкале отрицательны, а значения больше среднего арифметического — положительны. Это вытекает из свойства среднего арифметического значения и из свойства операции вычитания $X_i - M_x$.

3. Распределение Z точно такое же по форме, как и распределение исходных баллов. Это линейное преобразование, не меняющее форму распределения исходных данных.

4. Дисперсия и стандартное отклонение равны единице:

$$S_z^2 = S_z = 1,0 .$$

Т-шкала

Потребность в Т-шкале возникла, главным образом, из трудностей общения исследователей с родителями. Последние не допускали даже мысли о том, что их дети могут получить отрицательные баллы. Поэтому родителям была предложена иная шкала, позволяющая трансформировать данные

Z-шкалы в другую шкалу, которая обеспечивает перевод всех данных в область положительных значений. Это стало возможным посредством формулы, переводящей данные Z-шкалы в Т-шкалу:

$$T_i = 50 + 10z_i \quad [1]$$

Одиннадцатибалльная шкала

Для повышения дифференцирующей способности педагогических оценок автором этой статьи в своё время была предложена одиннадцатибалльная шкала, которая получается из известной в литературе шкалы стандартных девяти единиц за счёт добавления по одному интервалу в 0,5 стандартного отклонения справа и слева¹¹.

В отличие от шкалы стандартных девяти единиц (stapine)¹², здесь появляется возможность более тонкой дифференциации испытуемых. Если значениям стандартных баллов приписать оценочные эквиваленты, то соотношение между значениями нормированных отклонений Z, процентом испытуемых и оценочными эквивалентами можно представить в табл. 2.

В случае перевода системы образования с нынешней пятибалльной на более дифференцированную шкалу, одиннадцатибалльная шкала

11

Аванесов В.С.

Дис. на соискание учёной степени д-ра пед. наук. СПб. 1994. 339 стр.

12

Guilford J.P.

Psychometric Methods. 2-ed, N-Y, McGraw-Hill, 1954. 597 p.;

Guilford J.P. Fundamental Statistics in Psychology and Education. 3 ed. N-Y, McGraw-Hill, 1956. 565 p.

Guilford J.P.,

Fruchter B.

Fundamental Statistics in Psychology and Education. 5-ed. N-Y, McGraw-Hill. 1973. 564 p.

является наиболее подходящей и рекомендуемой как по своим статистическим свойствам, так и по наличию лексических эквивалентов каждой числовой оценки¹³. Преимущество последней обосновано с педагогической, статистической и управленческой точек зрения¹⁴.

Соотношение между балльными оценочными эквивалентами баллами представлено в табл. 6.

Утверждение одиннадцатибалльной шкалы в качестве нормативной может дать ряд преимуществ. По сравнению с пятибалльной, повышается дифференцирующая способность педагогической оценки. Не надо вычислять до сотых долей средний балл, получен-

ный из баллов, исходная точность которых была заведомо ниже. Особенно чётко в одиннадцатибалльной шкале дифференцируется один процент лучших и худших¹⁵.

Лучшие — это чаще всего самые талантливые студенты, которых надо выявлять и затем тщательно с ними работать. Если согласиться с тем, что один хороший специалист лучше ста плохих и средних, то ясно, что в век технологий от этого одного процента зависит довольно много. С другой стороны, мы имеем 1% наименее подготовленных студентов, с которыми высшая школа может безболезненно расстаться, также как, впрочем, и с тремя процентами тех, кто в одиннадцатибалльной шкале имеет два балла.

Таблица 5

Соотношение Z-значений и процентов испытуемых

Значения 11-балльной шкалы	Лексико-оценочные эквиваленты	Значения Z	Процент испытуемых
1	Низшая оценка	$Z_i < -2,25$	1
2	Неудовлетворительно	$-2,25 < Z_i < -1,75$	3
3	Малоудовлетворительно	$1,75 < Z_i < -1,25$	7
4	Удовлетворительно	$-1,25 < Z_i < -0,75$	12
5	Ниже среднего	$-0,75 < Z_i < -0,25$	17
6	Средняя оценка	$-0,25 < Z_i < 0,25$	20
7	Выше среднего	$0,25 < Z_i < 0,75$	17
8	Хорошо	$0,75 < Z_i < 1,25$	12
9	Очень хорошо	$1,25 < Z_i < 1,75$	7
10	Отлично	$1,75 < Z_i < 2,25$	3
11	Высшая	$2,25 < Z_i$	1

Теория

¹³ Стр. 196 указанной диссертации.

¹⁴ Дис.

¹⁵ Со стр. 195 докт. дис. В.С. Аванесова. Ук соч. С-Пб, 1994.

ПЕД
измерения

В дополнение к этому, у одиннадцатибальной шкалы, по сравнению с ранее описанной автором девятибальной и десятибальной шкалами¹⁶, есть преимущества психологического характера. В качестве опорных точек шкалы используются три привычных каждому человеку сравнительных понятия — низшее значение «единица», хуже которой нет, среднее значение «шесть», приписываемое легко различимому и наиболее распространённому в практике среднему уровню знаний и высшая оценка «одиннадцать», посредством которой появляется возможность выделить из числа отличников самого отличного студента. Это важно для дальнейшего мотивирования учёбы тех, кто уже учится отлично, но может, при старании, показать более выдающиеся результаты, особо ценные как для развития личности самого студента, так и для пополнения интеллектуального потенциала общества.

Одна из таких важных для высшей школы сфер, где измерения ещё не использовались мало-мальски заметно — это педагогический контроль, научная организация которого на современном этапе развития науки и техники не мыслится без тестов. Применение тестов позволяет быстро и качественно собрать информацию об уровне подготовленности абитуриентов, студентов и выпускников вузов, обработать эту информацию, сравнить достижения, выявить тенденции, достижения и недостатки, осуществлять прогноз, внести коррекции в учебный процесс.

Критерии оценки качества шкал

Качества шкал можно оценивать по следующим критериям:

- уровень шкалы — номинальная, порядковая, интервальная и пропорциональная. Предпочтительны две последние;
- наличие общей единицы измерения, что обеспечивает сравнимость результатов различных тестов;
- размах значений, пределы значений оценок и измерений, получаемых по разным шкалам, дисперсия — желательно иметь их сопоставимыми, что обеспечивает равноценность баллов, получаемых по той или иной шкале;
- совпадающие средние значения шкальных баллов, показатели асимметрии и эксцесса позволяют корректно сравнивать распределения результатов по разным тестам.

Наиболее научно обоснованной шкалой надо признать такую, которая отражает идею измерения на латентной переменной величине. Ввиду обширности этого вопроса, его лучше изложить в отдельной статье.

16

Аванесов В.С.
Основы педагогического контроля в высшей школе // Основы педагогики и психологии высшей школы. Учебное пособие / Под ред. А.В.Петровского. М.: МГУ, 1986. 303 с.;
Аванесов В.С. Основы научной организации педагогического контроля в высшей школе: Учеб. пособ. для слушателей Учебного Центра. М.: МИСиС, 1987. 107 с.