

Калдыбаев С.К.

## СТАТИСТИЧЕСКАЯ ТЕОРИЯ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

*Педагогическое измерение возникло в конце XIX-го – в начале XX-го веков. В его становление и развитие внесли существенный вклад такие выдающиеся ученые, как Ф.Гальтон, Э.Л.Торндайк, К.Пирсон, Ч.Спирмен. Идея Ф.Гальтона о необходимости сравнения результатов тестирования с независимым измерением поставлена на научный путь исследования в начале XX века. Исследования Ч.Спирмена и В.Брауна позволили заложить фундамент статистической теории педагогических измерений. Основная идея их работ заключалась в установлении связи между тестами. Положительная связь между тестами выявляет положительные отношения между понятиями. Такая связь характеризуется наличием общего и специфического факторов (g и S). Учеными предложена формула вычисления коэффициента корреляции, названная формулой Спирмена-Брауна.*

В начале XX века возникла необходимость установить точку отсчета и дозированную меру, чтобы с их помощью оценивать результаты тестирования. Такая точка, полученная статистическим путем, определялась в качестве статистической нормы.

Важной характеристикой педагогических измерений стало определение их качества. В качестве основных были признаны надежность и валидность. Надежность – критерий, который информирует о точности измерения, который позволяет судить, насколько можно доверять результатам измерения.

В статистической теории педагогических измерений прочно утверждены следующие подходы к определению надежности:

1. Ретестовая надежность – повторное измерение с одним и тем же измерительным инструментом спустя некоторое время.

2. Надежность взаимозаменяемых (параллельных) форм – сначала применяется одна форма измерительных инструментов, затем применяется эквивалентная форма.

3. Надежность по методу расщепления – измерительные средства подразделяются на две равноценные половины, результаты испытания сравниваются между собой.

4. Надежность по Кьюдеру-Ричардсону – проведение однократного испытания и определение внутренней согласованности инструмента по формуле Кьюдера-Ричардсона.

Понятие валидности появилось из предположения о годности теста и тестовых заданий для решения поставленных задач. Измерение считается валидным, если на его результаты влияет лишь измеряемое свойство (погрешность должна быть минимальной).

В педагогическом измерении разработано множество видов валидности. В качестве основных были признаны:

1. Содержательная валидность. Проводится сопоставление результатов измерения с экспертными

оценками учителей. В результате экспертизы устанавливается: а) насколько репрезентативно измеряемое свойство в измерительных средствах; б) насколько измерительный инструмент пригоден для измерения свойств; в) насколько правильные ответы испытуемых свидетельствуют об усвоении учебного материала.

2. Прогностическая валидность. Проверяется способность человека к тому или иному виду деятельности. Она оценивается по результатам измерений, но информация по нему собирается спустя некоторое время. Например, измерение умственных способностей детей предсказывает его будущие профессиональные успехи, вступительные испытания предсказывают успешное обучение в вузе и т.д.

3. Конструктная валидность. Понятия, характеризующие не наблюдаемые свойства человека или «идеальное представление о реальности, сконструированное с целью ее познания», являются конструктами [1]. Для педагогического измерения конструктами могут быть «знания», «умения», «способности», «уровни усвоения» и т.д. С целью изучения (в том числе и для оценивания) конструкты должны быть операционализированы, т.е. воплощены (переведены, транслированы) в конкретной методике измерения. С этой точки зрения конструктная валидность означает наличие количественной меры, которая характеризует «эффективность трансляции в актуальную методику или программу действий» [1].

Исследования по проверке качества педагогических измерений привели к обоснованию понятия о погрешности измерений. Были приняты следующие утверждения, получившие название постулатов [2]:

1. Результат измерения (X) состоит из суммы истинного (T) и ошибочного (e) компонента:  $X = T + e$ . Ошибочный компонент (e) появляется в результате влияния случайных ошибок.
2. Истинный компонент измерения (T) выражается в виде математического ожидания:  $T = M(X)$ .
3. Среднее арифметическое ошибочного компонента (e) должно быть равным нулю. Дисперсия ошибочных компонентов измерения параллельными тестами равны.
4. Между истинными и ошибочными компонентами нет корреляционной связи:  $\rho_{Te} = 0$ .
5. Ошибочные компоненты двух тестов не коррелируют:  $\rho_{e_i e_j} = 0$
6. Между ошибочными компонентами одного теста и истинными компонентами другого теста нет корреляционной связи:  $\rho_{e_i T_j} = 0$

Разработка шкал существенно расширила область исследования педагогических измерений. Ши-

роко применяется корреляция для проверки надежности и валидности. Способным предоставлять особо ценные информации оказался факторный анализ.

В 1946 году была опубликована работа С.Стивенса «О теории шкал измерений» [3]. По утверждению С.Стивенса, создание шкалы станет возможным, если существует изоморфизм (соответствие) между свойствами чисел и объектов, преобразований между ними. При этом вид преобразований определяет тип шкалы, который, в свою очередь, определяет уровень использования математико-статистических методов. Обоснованы и нашли широкое применение такие типы шкал, как шкала наименований, шкала порядка, шкала интервалов и шкала отношений.

Особый размах в своем развитии получила процедура шкалирования результатов измерений. Помимо отмеченных выше четырех типов шкал, создано множество видов шкал в образовании. В процессе шкалирования первичные показатели (сырые тестовые баллы) подвергаются различным преобразованиям, в результате чего обеспечиваются истинные результаты (оценки) в той или иной шкале. С помощью этих шкал определяется истинное место испытуемого среди групп, сравниваются результаты испытуемых между собой.

Начиная с 20-х годов XX века широко распространяется процедура получения стандартных показателей. В ней используется понятие «единица стандартного отклонения», которая определенно выражает отличие результата испытуемого от среднего результата выборки. В основе процедуры получения стандартных показателей лежит нормирование первичных показателей, т.е. нахождение разности между первичными баллами ( $X_i$ ) и средним значением выборки ( $M$ ), затем – деление этой разницы на стандартное отклонение ( $SD$ ). В итоге получают нормальное распределение результата со средним  $M_{\theta 0}$  и со стандартным отклонением  $SD_{\theta 1}$ , получившее название «Z-шкала» (иногда ее называют - «Z-оценка» или «Z-показатель») [4, с. 108-114; 5, с. 236-239; 6, с. 77-79].

В Z-шкале используются как положительные, так и отрицательные числа. Испытуемый, получивший оценку  $Z_{\theta 0}$ , характеризуется, что он по результатам измерения превосходит 50% испытуемых. Показатель  $Z_{\theta -1}$  означает, что результат испытуемого лучше, чем 16% всех остальных. А  $Z_{\theta 1}$  – испытуемый превосходит на 84 % всех испытуемых. Диапазон значений Z колеблется от -4 до +4.

С целью придания удобной формы стандартным показателям была предложена формула  $Z_{1\theta} M_1 + SD \cdot Z$ .

При  $M_{1\theta 50}$  и  $SD_{\theta 10}$  переходят на так называемую «Т-шкалу». При  $M_{1\theta 100}$  и  $SD_{\theta 15}$  получают шкалу IQ. А при  $M_{1\theta 500}$  и  $SD_{\theta 100}$  – шкалу СЕЕВ. Была разработана и широко апробирована шкала станайнов ( $M_{1\theta 5}$  и  $SD_{\theta 2}$ ) [4, с. 114-115; 5, с. 244; 6, с. 79-83].

Таким образом, в измерении широко применен математико-статистический аппарат. Проверка и анализ соответствия концептуальной (конструкт) и реальной переменных измерения, выявление размерности измерений, определение соответствующего ранга испытуемых прочно основываются на статистических методах. Посредством статистических методов стало возможным доказывать правильность и обоснованность используемых методов и средств измерений, экспериментально проверять соответствие результатов измерения поставленной цели, обобщать данные, полученные в результате измерения. Статистические методы успешно решали вопросы о нахождении зависимостей между полученными данными, степени их влияния, о выявлении различий между испытуемыми, прогнозировании на основе статистических данных, предупреждении ошибок в оценке свойств личности. Именно статистический аппарат педагогического тестирования обогатил содержание и теории педагогического измерения. Установление механизма тестирования, статистическая обработка результатов и обстоятельная интерпретация позволили вести речь о статистической теории педагогических измерений [2, с. 84-100; 4, с. 19-26]. Нередко эта теория называется классической теорией тестов.

В создание статистической теории педагогических измерений внесли значительный вклад в основном зарубежные исследователи А.Бине, Э.Л.Торндайк, А.Отис, Дж.Райс, Л.Термен, С.Стоун, Г.Годдард и др.

Измерение учебных достижений осуществлено преимущественно с ориентацией на статистическую норму. Все показатели испытуемых оценивались относительно этой нормы. Нормальное распределение предполагает, что левая часть графика указывает на слабую группу испытуемых, показывающих относительно слабую успеваемость, а правая часть – небольшую группу испытуемых, показывающих наиболее хорошую успеваемость. А данные, расположенные около середины графика (около нормы) показывают средний результат испытуемых. Таким образом, относительно нормы определяется место испытуемого, сопоставление результата одного испытуемого среди других.

В 60-е годы учеными исследователями предлагается другой подход в конструировании заданий и интерпретации результатов. В учебном процессе наиболее важное значение имеет не определение ранга испытуемых, не сравнение одних результатов с другими, а выяснение того, знает или не знает испытуемый, что он умеет или допускает ошибку. Результат испытуемого должен быть оценен на основе некоторого критерия, который может стать мерилем успешности обучения. Подход, предполагающий оценку знаний испытуемых на основе выбранных критериев, получил название – тестирование с критерияльно ориентированной интерпретацией. Этот

термин предложен американским ученым-исследователем Р.Глейзером [7].

Не менее важным аспектом в исследованиях было подведение итогов тестирования и их интерпретация. В литературе приводятся следующие способы обсуждения результатов измерения [8; 9; 10; 11]:

- 1) Подсчет баллов и их интерпретация.
- 2) Определение коэффициента усвоения

$K \frac{A}{N}$ , где  $K$  – коэффициент усвоения,  $A$  – количество правильных ответов,  $N$  – количество заданий.

- 3) Расчет процентного ранга с целью определения места учащегося в классе  $P_R = 100 - \frac{100R - 50}{N}$ ,

где  $R$ - место учащегося по списку,  $N$  – количество учащихся в классе.

- 4) Составление таблиц для каждого учащегося, и в целом для всего класса, с указанием успехов и недостатков учащихся.

- 5) Учет количества попыток дать ответы на задания с помощью специально составленного графика, в котором учитывается количество попыток решения заданий.

Статистическая теория педагогических измерений сыграла для системы образования существенную роль. В первую очередь, предприняты огромные усилия для объективной оценки таких педагогических явлений, как уровень подготовленности испытуемых. Педагогической общественностью осознана необходимость в том, что измерительные средства должны быть разработаны не произвольно, а с со-

блюдением определенных правил и этапов, в соответствии с поставленной целью и с учетом содержания контролируемого материала.

#### Литература:

1. <http://www.dictionary.fio.ru>
2. Аванесов В.С. Методологические и теоретические основы тестового педагогического контроля. Дисс. ... д-ра пед. наук: 13.00.01. –Москва, 1994.
3. Stevens S.S. On the theory of scales of measurement. - 1946, V.103, p.677.
4. Звонников В.И. Измерения и шкалирование в образовании: Учеб. пособие. –М.: Университетская книга; Логос, 2006.
5. Майоров А.Н. Теория и практика создания тестов для системы образования: Как выбирать и использовать тесты для целей образования. –М.: Народное образование, 2000.
6. Психологическое тестирование. 7-е изд. /А. Анастаси, С.Урбина. –СПб.: Питер, 2003.
7. Glaser R. Introductory technology and measurement of learning outcomes: Some questing //American Psychologist. 1963. №18. P.519-521.
8. Воскерчян С.И. Об использовании метода тестов при учете успеваемости школьников //Советская педагогика, 1963, №10. –с. 28-37.
9. Вендровская Р.Б. Проверка и оценка знаний учащихся в историческом опыте советской школы //Советская педагогика, 1982, №11. –с.103-108.
10. Ильина Т.А. Тестовая методика проверки знаний и программированное обучение //Советская педагогика. 1967, №2. -с.122-138.
11. Калдыбаев С.К. Программированное обучение и педагогическое тестирование //Школьные технологии. 2006, №6. -с.173-176.