

МЕТОДОЛОГИЯ КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ В ПЕДАГОГИЧЕСКОМ ПРОЦЕССЕ

Военно-медицинский факультет в УО «БГМУ»

В последние десятилетия мы являемся свидетелями интенсивного развития компьютерных технологий и их внедрения в педагогический процесс. Как и в любом другом продукте научно-технического прогресса, в данном явлении следует выделять как положительные, так и отрицательные стороны. К положительным сторонам, несомненно, следует отнести доступность и скорость поиска и извлечения информации, возможности визуализации сложновизуализируемых процессов, заочного общения, дистанционного обучения и др. Однако, не следует забывать тот факт, что это лишь вспомогательные средства, которые не могут заменить живое слово и живое общение студента с преподавателем.

Особую тревогу вызывает модное увлечение компьютерным тестированием, которое все чаще применяется для текущего и итогового контролей, а также для проверки «выживаемости» знаний студентов. Здесь нам следует согласиться с тем фактом, что знания студента являются продуктом его психической деятельности, т. е. психическим феноменом. Для изучения (оценки, измерения) данного феномена у конкретного человека или в человеческой популяции уже более полувека существует научное направление – психологическое тестирование, которое включает две процедуры: психометрику (процесс измерения и его условия) и психодиагностику (оценку результатов тестирования). Процесс создания, апробации и стандартизации психометрических методик требует большого времени, участия больших массивов испытуемых, проверки по критериям валидности и надежности [1]. Существует много других требований, которые демонстрируют сложность и серьезность проблемы. Это целая наука со своими законами и требованиями, пренебрегая которыми исследователь рискует получить ложные результаты, потратить впустую личное и учебное время, а также подорвать авторитет данной науки.

Наш опыт работы в данном направлении позволяет сформулировать несколько принципиальных положений и этапов психометрических и психодиагностических процедур.

Во-первых, компьютерному тестированию может подвергаться только тот студент, для которого компьютер является привычным рабочим местом. К сожалению, еще не все студенты обладают этим качеством, вследствие чего они реагируют не на вопрос теста, а на ситуацию обследования, что существенно искажает результаты. В этом случае мы обязаны предоставить возможность выполнения теста в традиционном предъявлении «бланк-карандаш». Во-вторых, большинство авторов тестов применяют «закрытые» методики, где необходимо выбрать верный ответ из 2-5 предлагаемых. В этом случае оказывается высоким процент угадывания верных ответов и ложная оценка знаний. Целесообразнее применение «открытых» тестов, где ответ испытуемый должен придумать сам. В-третьих, использовать данные технологии целесообразно для текущего контроля и «выживаемости» знаний. Для итогового контроля (зачет, экзамен) лучше оставить традиционные формы, результаты которых должны служить одним из вариантов проверки тестов на валидность. В-четвертых, время тестирования должно занимать не более 15-20% от бюджета учебного времени.

Далее предлагаем читателю ознакомиться с этапами процедуры психометрики. Перечень вопросов создается экспертами-преподавателями конкретной дисциплины по конкретной теме или разделу. Количество вопросов должно ориентировочно соответствовать выделенному времени. Далее идет процедура апробации. Время ответа на 1 вопрос на начальном этапе не ограничивается, но учитывается для последующего установления среднего времени от массива испытуе-

Таблица 1. Методика преобразования «сырых баллов» в Р% (n=561)

Сырой балл	Число испытуемых	Нарастающий итог	Р%	Группа успешности обучения	
10	21	21	3,74	IV	
11	14	35	6,24		
12	18	53	9,45		
13	22	75	13,37		
14	26	101	18,00		
15	30	131	23,35		
16	27	158	28,16	III	
17	22	180	32,09		
18	34	214	38,15		
19	17	231	41,18		
20	24	255	45,45		
21	25	280	49,91		
22	23	303	54,01	II	
23	20	323	57,58		
24	18	341	60,78		
25	13	354	63,10		
26	20	374	66,67		
27	14	388	69,16		
28	18	406	72,37		
29	16	422	75,22		
30	15	437	77,90		I
31	13	450	80,21		
32	12	462	82,35		
33	9	471	83,96		
34	13	484	86,27		
35	9	493	87,88		
36	8	501	89,30		
37	5	506	90,20		
38	4	510	90,91		
39	3	513	91,44		
40	3	516	91,98		
41	8	524	93,40		
42	5	529	94,30		
43	5	534	95,19		
44	5	539	96,08		
45	5	544	96,97		
46	3	547	97,50		
47	2	549	97,86		
48	6	555	98,93		
49	5	560	99,82		
50	1	561	100,00		

мых и фиксации его в компьютерной программе. Одновременно, для установления «веса» (сложности) вопросов, должен учитываться процент верных ответов на каждый вопрос. В дальнейшем от тех вопросов, на которые ответили верно только 10% и менее испытуемых, необходимо отказываться поскольку они либо очень сложные, либо некорректно сформулированы. Те вопросы, на которые дают верные ответы 90% и более испытуемых также должны исключаться из пакета как слишком простые. За счет отказа от этой части вопросов сокращается их количество и время тестирования. На этом этапе мы имеем возможность обратной связи и корректировки времени на рассмотрение отдельных вопросов в программе подготовки: на простые вопросы выделять меньше времени, на сложные – больше.

Следующим этапом адаптации теста является процедура стандартизации, т. е., определение стандарта, согласно которому испытуемый получит академическую оценку, а точнее – узнает как он «выглядит» по данному критерию на фоне референтной ему субпопуляции. Для этих целей используются различные производные взаимопреобразуемые единицы измерений: процентиля (Р%), Z-оценки, Т-баллы, стэны (St) и др. [1, 5]. Наиболее удобными в данном случае являются Р% и St. Главным достоинством St, является то, что они выражаются в 10-балльной шкале, которая соответствует 10-балльной академической оценке и не нуждается в дальнейших преобразованиях.

Дальнейшее рассмотрение процедур продолжим на конкретном примере. Тест содержит 50 вопросов, участвует 561 испытуемый (табл. 1). Исходным показателем является «сырой балл» (СБ), т. е., количество верных ответов на каждый из вопросов, которые ранжируются в порядке возрастания. Далее проводим подсчет числа испытуемых, получивших конкретное число верных ответов. Например, на 10 из 50 вопросов ответили верно 21 человек, на 11–14 и т. д. Арифметическая сумма испытуемых должна равняться числу 561. При этом обнаруживаем, что максимальное количество людей (34 чел.) получили значение 18 баллов (верных ответов). В соответствии с законом нормального распределения это значение соответствует «моде». Если нас интересует, какой процент людей получил данное значение (т. е., процент «моды»), то для этого проводим процедуру нахождения процента: $34 : 561 = 0,06 \times 100 = 6\%$. Однако процент является первичным показателем и представляет собой долю людей, получивших конкретное значение методики. Процент нам интересен и в дальнейшем может понадобиться для построения кривой распределения. Нам же необходимо получить не процент, а Р%.

Для получения Р% необходимо провести суммирование числа испытуемых по принципу «нарастающего итога» от минимального до максимального значения: $21+14=35$; $35+18=53$ и т. д. Сумма «нарастающего итога» также должна равняться числу 561. Далее по той же процедуре, как мы получили процент, вычисляем Р%, однако используем для этого не абсолютное число, а сумму «нарастающего итога». Например, берем сумму «нарастающего итога» (214), соответствующую значению «моды» (34), и проводим процедуру: $214:561=0,38 \times 100=38,15\%$.

В результате имеем: $18 \text{ СБ} = 6\% = 38,15\text{Р}\%$. При проведении такой процедуры первые (минимальные) и последние (максимальные) значения процента и процентиля всегда совпадают, остальные – существенно

отличаются. Таким образом, P% – это производный показатель, указывающий долю лиц, результат которых соответствует данному СБ и ниже, от общего числа членов выборки. P% указывают на положение индивида в выборке стандартизации. Их также можно рассматривать как ранговые градации, общее число которых равно 100. Но при этом предполагается, что в конкретное значение P% (например, 38,15) входят и СБ от 18 и ниже. Другими словами, совокупность этих людей уже прошла этапы значений 0–18 СБ, но еще не достигала значений 19 и более.

50-й P% соответствует медиане – одному из показателей центральной тенденции. Необходимо обратить внимание на соотношение медианы и «моды», которые при нормальном распределении должны совпадать. В данном случае «мода» находится на уровне 18 СБ, а медиана – на 21. Следовательно, если построить кривую распределения по СБ или процентам, она не будет соответствовать нормальному. Если ту же кривую построим по P% – в основном будет соответствовать.

Таким образом, с помощью стандартизации (перевода в стандартные единицы) мы имеем возможность, во-первых, привести любую выборку к относительно нормальному распределению. Во-вторых, при тестировании 562-го человека нам не нужно будет повторять всю процедуру. При наличии готового стандарта мы сразу определим «вид» испытуемого в континууме репрезентативной выборки, однако в 100-балльной шкале P%. При необходимости ее уже можно использовать как академическую оценку, приняв, например, 38,15 P% за 4 академических балла. С помощью таблицы 2 [2] мы имеем возможность переводить P% в 10-балльные St, т. е., получить аналог академической оценки.

Таблица 2. Таблица преобразований единиц измерения в психометрике [2]

Диапазон колебаний P%	Соответствующие значения St
2,28 и менее	1
2,29-6,67	2
6,68-15,68	3
15,69-39,84	4
39,85– 49,00	5
49,10–51,00	5,5
51,01–69,14	6
69,15–84,12	7
84,13–93,31	8
93,32–97,71	9
97,72 и более	10

При условии нормального распределения применяется линейное преобразование СБ в 10-балльную шкалу St с помощью формулы [6]:

$$St = \frac{X - M}{\sigma} \cdot A + B$$

где: X – значение показателя (СБ), M – среднее арифметическое показателей, σ – среднее квадратическое отклонение показателей, A – заданное среднее квадратическое отклонение шкалы St (2,0), B – заданное среднее арифметическое шкалы St (5,5).

При наличии компьютерной базы данных и соответствующего программного обеспечения все приведенные расчеты и преобразования могут производиться автоматически. По мере накопления случаев наблюдений они могут автоматически пересматриваться (пересчитываться, уточняться).

Далее следует процедура психодиагностики. Для дальнейшей работы обратимся еще раз к таблице 1. При ее рассмотрении мы остановились на переводе СБ в P%. Получив условную кривую нормального распределения в процентильных единицах, мы можем дать характеристику кривой распределения, а значит, характеристику (диагноз) данной популяционной выборке. В норме, как мы уже отметили, 50-я P% должна быть медианой. Это правило соблюдается: медиана находится на 49,91. P% свыше 50 представляют показатели выше среднего, а ниже 50 – ниже среднего (II и III центральные квартили). Показатели ниже 25-го и выше 75-го P% являются IV-м (значительно ниже среднего) и I-м (значительно выше среднего) квартилями, которые выделяют нижнюю и верхнюю четверти распределения. Как и медиана, они удобны для описания распределения показателей и сравнения с другими распределениями.

При переводе (преобразовании) P% в St мы получаем аналог 10-балльной академической оценки. Однако, далее нам необходимо узнать академические эквиваленты данных оценок: за получение какой оценки студент может быть аттестован. Т. е., мы должны классифицировать студентов на группы по критерию прогноза на дальнейшее обучение. При условии нормального распределения выборки классификация может быть проведена на основании 4 квартилей процентильных единиц. В нашем случае распределение, очевидно, не является нормальным. Поэтому ориентироваться необходимо на среднее значение и среднее квадратическое отклонение (сигма) шкалы St (5 ± 2) [5]. На этом основании мы имеем возможность выделить 4 группы успешности обучения с центральной тенденцией 5 ± 2 St (4-5,5 и 6-7) и полярными тенденциями (1-3 и 8-10).

Работа по апробации и стандартизации теста (психометрической методики) проводится, например, около года и требует участия большого массива испытуемых до получения нормального распределения. При нормальном распределении шкала может рассматриваться как стандарт для данной популяции испытуемых. Если в очередном учебном году нам необходимо провести тестирование в условиях дефицита времени и на малой выборке, то возникает необходимость создания готовых шкал показателей, полученных на репрезентативной выборке. Обычно авторы психометрических методик приводят только «нормативные» или среднестатистические значения. Большинство исследователей работают, ориентируясь на них, дают грубые оценки и ложные прогнозы, что чаще всего компрометирует психометрику и психодиагностику. Нами [3] предложен метод шкалирования (стандартизации) показателя (таблица 3).

В данном случае представлена прямая шкала (чем выше СБ – тем лучше знания). В ней представлены только начальный этап (диапазон колебаний СБ) и конечный результат – стеновые значения академической оценки. Т. е., она является итогом всей описанной работы. В ней скрыты все тонкости проведенной работы не нужные ни преподавателю, ни студенту в период реального тестового контроля знаний. Именно в таком виде она должна быть представлена студенту с пояснением, за что и

Таблица 3. Прямая шкала стандартизации СБ

Колебания СБ	St	Группы успешности обучения
0 – 9 10 – 11 12 – 13	1 2 3	IV
14 – 16 17 – 20 21	4 5 5,5	III
22 – 26 27 – 33	6 7	II
34 – 40 41 – 46 47 – 50	8 9 10	I

какую оценку он может получить (индивидуальная психодиагностика). Она же может быть использована при проведении теста традиционным методом «бланк-карандаш». В любом случае нам понадобится выделение групп успешности обучения по шкале St: и популяционная (групповая) и индивидуальная диагностика (оценка знаний) должна завершиться заключением о дальнейших перспективах учебы данной группы или отдельного студента. Успешность обучения лиц, отнесенных к IV-й группе (1-3 St) или 1-3 балла академической оценки, оцениваются как «неудовлетворительная». Их прогноз неблагоприятен, они нуждаются в дополнительной подготовке с последующей пересдачей данной темы, т. е., не аттестованы. Получившие 4-5,5 St, т. е., отнесенные к III-й группе, соответствуют 4-5 баллам академической оценки, оцениваются как «удовлетворительно», их прогноз сомнителен, они аттестованы, но нуждаются в дополнительной подготовке. Студенты, отнесенные ко II-й группе, получившие 6–7 St, т. е., 6–7 баллов, оцениваются как «хорошо» с положительным прогнозом. И лишь лица, отнесенные к I-й группе (8-10 St), соответствуют 8–10 баллам академической оценки, оцениваются как «отлично» с устойчивым благоприятным прогнозом на дальнейшую учебу.

У внимательного читателя должен возникнуть вопрос: почему в таблице 1 отсутствует значение P% от 1 до 3,73? Почему группы успешности обучения в шкале St (таблица 3) не соответствуют квартилям процентильной шкалы (таблица 1)? Ответ следующий: это и есть начало популяционной психодиагностики. Именно по таким фактам мы можем сделать следующие выводы. Во-первых, в данной выборке отсутствуют испытуемые с нулевым уровнем знаний (0-9 СБ), но очень большое количество испытуемых (21 чел.) имеют СБ 10, т. е. у них очень низкий уровень знаний. Этим определяется отсутствие значения 1 St и резкий (крутой) подъем левого плеча воображаемой кривой распределения.

Во-вторых, P%-значения 1-го квартиля достигают только 13,37, когда должно быть 25, а значения 4-го квартиля выходят за рамки 75 и достигают 86,27. Эти факты способствуют резкому росту левого и резкому спаду правого плеч кривой распределения. Если представить такую кривую, то она будет высокой и узкой за счет того, что основная кумуляция (накопление) наблюдений (396 человек или около 71 %) приходится на II и III квартили. Таким образом, несмотря на то, что медиана была нормальной (на 50 P%), кривая распределения резко отличается от нормального. Причины этого явле-

ния должны анализироваться и служить поводом для обратной связи: качества субпопуляции, условия обучения, качество преподавания и т. п. Это один из примеров популяционной диагностики. Если этот тест применить в другом вузе, то картина может существенно измениться. Различия или их отсутствие должны стать поводом для размышления в масштабе всей системы образования.

В заключение следует подчеркнуть, что на этом работа с тестом не заканчивается. По мере накопления случаев новых наблюдений (реального тестирования студентов) мы имеем возможность достичь нормального распределения, что позволит проводить прямой перевод СБ в шкалу St с помощью формулы. До этого вся описанная работа должна периодически повторяться для следующих целей. Во-первых, для уточнения критериев оценки по мере накопления наблюдений: чем выше число наблюдений, тем точнее будет оценка. Во-вторых, для сравнения аналогичных результатов за разные периоды времени с возможностью получения обратной связи и корректировки программ подготовки. В-третьих, мы не можем игнорировать тот факт, что академическая оценка за устный ответ студента всегда имеет субъективную компоненту. Применяя данную методологию, мы имеем возможность максимально объективизировать учебный процесс: оценки за устные ответы и за компьютерные тесты должны совпадать. При этом оценка за устный ответ является критерием валидности для компьютерного теста, а оценка за компьютерный тест является критерием объективности оценки устного ответа.

Пользуясь возможностью, мы предлагаем читателю рассмотреть еще одну актуальную проблему. В последние годы в нашей стране получили широкое распространение так называемые «психологические тесты» (в том числе и их компьютерные версии), которые используют для оценки психического здоровья в клиниках, при проведении профессионального психологического отбора и с другими целями. Многим из них можно доверять, так как они созданы именитыми авторами и базируются на серьезных научных исследованиях. Что касается психометрики (процедуры измерения и получения СБ), то она описана авторами и должна неукоснительно соблюдаться. Однако, для психодиагностики (интерпретации полученных измерений) необходимо проведение всех описанных процедур с целью рестандартизации (реадаптации). Дело в том, что большинство данных тестов создавались десятки лет тому и на иных популяционных выборках. В доступной современной литературе имеются единичные публикации [4] по реадаптации психологических тестов к нынешней популяции нашей республики, которые свидетельствуют о невозможности применения старых стандартов. Мало того, психодиагностика является настолько деликатной процедурой, что требует реадаптации методик не просто к популяции, а к ее субпопуляциям по половому, возрастному, профессиональному и другим признакам.

Таким образом, использование описанных процедур стандартизации в интересах психометрики и психодиагностики является единственным способом получения объективной информации как при проведении компьютерного тестирования в учебном процессе, так и при изучении психических особенностей популяции (субпопуляций) с применением психологических тестов.

Вопросы совершенствования учебного процесса ☆

Литература

1. Анастаси, А. Психологическое тестирование / А. Анастаси. В 2-х т.; пер. с англ. / под ред. К. М. Гуревича, В. И. Лубовского. – М.: Педагогика, 1982.

2. Венцлав, С. В. Применение математических методов в задачах профессионального отбора и распределения кадров / С. В. Венцлав, М. А. Данилов, А. Ф. Богачев. – М., 1987. – С. 42.

3. Власенко, В. И. Психофизиология: методологические принципы профессионального психологического отбора: монография / В. И. Власенко; под ред. В. А. Пе-

реверзева. – Минск, БГМУ, 2005. – 244 с.

4. Власенко, В. И. Опыт изучения личностной тревожности у студентов / В. И. Власенко // Военная медицина. – 2012. – № 4. – С.65-68.

5. Кулагин, Б. В. Основы профессиональной психодиагностики / Б. В. Кулагин. – Л.: Медицина, 1984. – 216 с.

6. Основные этапы подбора и адаптации психологических методик в целях профессионального психологического отбора : метод. пособие. – М., 1991. – 72 с.

Поступила 31. 04. 2013 г