

Основы теории тестирования

Тестированием заменяют измерение, когда изучаемый объект недоступен прямому измерению.

Например, практически невозможно точно определить производительность сердца спортсмена во время напряженной мышечной работы; применяют косвенное измерение: измеряют частоту сердечных сокращений, кардиопоказатели.

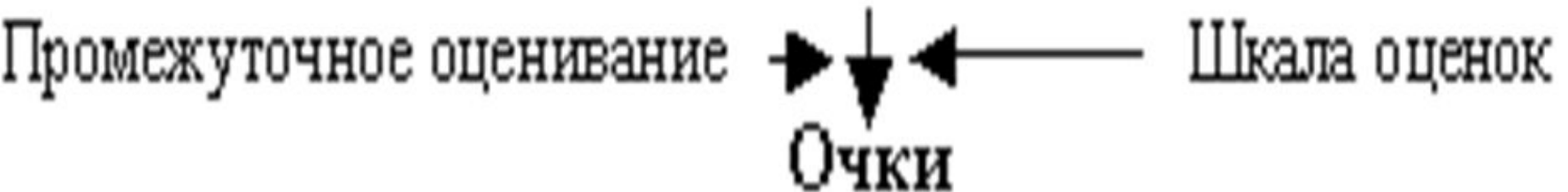
Тесты используют и в тех случаях, когда изучаемое явление не вполне конкретно. Например, правильнее говорить о тестировании ловкости, гибкости, чем об их измерении.

Схема оценивания спортивных результатов и результатов тестов

Название действия

Используется

тестирование



(промежуточная оценка)



Тест (англ. test — проба, испытание) - измерение или испытание, проводимое с целью определения состояния или способностей человека, отвечает метрологическим требованиям:

- определенная цель применения теста;
- стандартность (методика, процедура и условия тестирования должны быть одинаковыми во всех случаях применения теста);
- достаточная надежность и информативность;
- адекватная система оценок.

Процесс испытаний называется *тестированием*, а полученное в итоге измерения числовое значение является *результатом тестирования*.

В зависимости от области применения тесты: педагогические, психологические, интеллекта, индивидуально-ориентированные, специальных способностей и т.д.

Тесты, результаты которых зависят от двух и более факторов - *гетерогенные*, а если от одного фактора — *гомогенные*.

Тестом в спортивной практике может быть названо измерение или испытание, отвечающее следующим метрологическим требованиям:

- определена цель применения теста;
- стандартность (методика, процедура и условия тестирования должны быть одинаковыми во всех случаях применения теста);
- определена надежность и информативность теста;
- разработана система оценок;
- определён вид контроля (оперативный, текущий

Стандартизация проведения тестирования в спортивной практике достигается:

1) режим дня, предшествующего тестированию, должен строиться по одной схеме, исключаются средние и большие нагрузки, проводятся занятия восстановительного характера, что обеспечивает равенство исходного уровня спортсменов перед тестированием;

2) стандартная для всех (по длительности, подбору упражнений, последовательности их выполнения)

- 3) тестирование по возможности должны проводить одни и те же подготовленные люди;
- 4) схема выполнения теста остается постоянной от тестирования к тестированию;
- 5) интервалы между повторениями одного и того же теста должны ликвидировать утомление, возникшее после первой попытки;
- 6) спортсмен должен стремиться показать в тесте максимально возможный результат.

Разновидности двигательных тестов

- . Контрольные упражнения (показать максимальный результат)
- . Стандартные функциональные пробы (одинаковое задание для всех, дозируются: а) по величине выполненной работы; б) по величине физиологических сдвигов)
- . Максимальные функциональные пробы Результат – изменение физиологических или биохимических показателей
(Регистрация ЧСС при стандартной работе, определение максимального кислородного долга).

Общепринятые тесты в спортивной практике

№	Содержание теста	Что измеряется
1	Бег на короткие дистанции (30 - 60м) (рез-т: время забега)	Быстрота
2	Бег на длинные дистанции (рез-т: время забега при фиксир. дистанции или расстояние за фикс. время)	Выносливость (Тест Купера)
3	Челночный бег с указанием прямых участков и кол-ва поворотов	Ловкость
4	Подтягивание или отжимание от пола, скамейки (кол-во повторений)	Сила
5	Подъем туловища из положения сидя или лежа (кол-во повторений)	Сила

6	Наклон вперед из положения сидя или стоя на скамейке (величина наклона)	Гибкость
7	Прыжок в длину с места или с разбега (длина)	Скоростно-силовые
8	<p>Подъем на скамейку опред. высоты в опред. темпе за опред. время (индекс</p> $i = \frac{100t}{f_1 + f_2 + f_3}, \text{ где } t - \text{ время восхождения (сек);}$ <p>f_1, f_2, f_3 – ЧСС за 30 сек на 2, 3 и 4-й минуте восстановления)</p>	<p>Восстановление</p> <p>Степ-тест (снабжен спец. Таблицей величин i).</p>
9	Уровень нагрузки по достижении ЧСС 170 уд/мин.	Физ. Работоспособность (Тест PWC ₁₇₀).

Спортсмен как и всякая живая система является нетривиальным объектом измерения; отличия от классических объектов измерения: изменчивость, многомерность, квалитативность, адаптивность и подвижность.

Изменчивость - непостоянство переменных величин, характеризующих состояние спортсмена и его деятельность.

Многомерность - одновременно измеряют большое число переменных для точной характеристики состояния и деятельности.

Квалитативность — качественный характер, т.е. отсутствие точной количественной меры (физические качества спортсмена, свойства личности, коллектива и др. факторы спортивного результата не поддаются точному измерению).

Адаптивность - свойство адаптироваться (приспосабливаться) к окружающим условиям.

Подвижность — особенность спортсмена, основанная на том, что в большинстве видов спорта его деятельность связана с непрерывными перемещениями.

Не всякие измерения могут быть использованы как тесты. Для этого они должны удовлетворять специальным требованиям:

1) надежность; 2) информативность; 3) наличие системы оценок (см. далее); 4) стандартность — процедура и условия тестирования должны быть одинаковыми во всех случаях применения теста.

Тесты, удовлетворяющие требованиям надежности и информативности, называют добротными или **а у т е н т и ч н ы м и**.

НАДЕЖНОСТЬ ТЕСТОВ

Надежностью тестов называют степень совпадения результатов при повторном тестировании одних и тех же людей (или объектов) в одинаковых условиях.

В идеале один и тот же тест, примененный к тем же испытуемым в тех же условиях, должен давать одинаковые результаты. Однако даже при самой строгой стандартизации испытаний и точной аппаратуре результаты тестирования всегда несколько варьируют.

Например, спортсмен, только что выжавший на кистевом динамометре 55 кг, через несколько минут покажет лишь 50 кг и т.д.

Подобную вариацию называют внутри-индивидуальной или *внутриклассовой*.

Ее вызывают четыре причины:

- 1) изменение состояния испытуемых (утомление, изменение мотивации, концентрации внимания и т. п.);
- 2) неконтролируемые изменения внешних условий и аппаратуры (температуры и влажности воздуха, напряжения в электросети, ветер...);
- 3) изменение состояния лица, проводящего или оценивающего тест, замена судьи другим;
- 4) несовершенство теста (тесты заведомо малонадежные, например штрафные броски в баскетбольную корзину до первого промаха; даже спортсмен, имеющий высокий процент попадания, может случайно ошибиться при первых бросках).

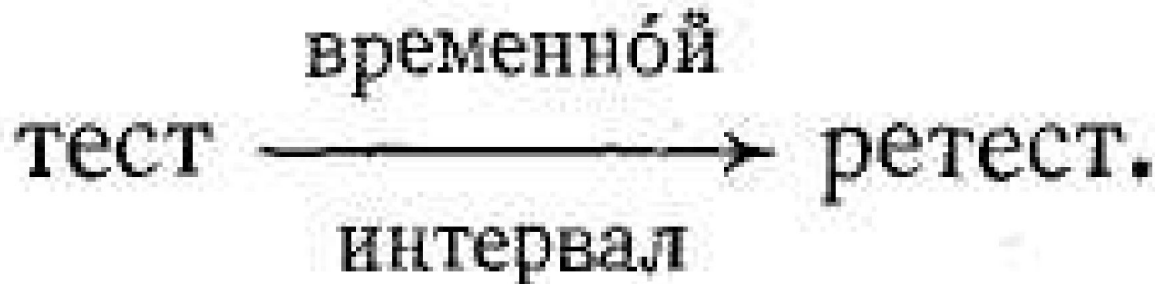
Говоря о надежности тестов, различают их стабильность, согласованность, эквивалентность.

Стабильность теста

Под стабильностью теста понимают воспроизводимость результатов при его повторении через определенное время в одинаковых условиях.

Повторное тестирование называют ретестом.

Схема оценки стабильности теста такова:



Стабильность теста зависит от:

- 1) вида теста,
- 2) контингента испытуемых,
- 3) временного интервала между тестом и ретестом.

Например, наименьшую, стабильность имеют тесты на точность движений (например, броски в цель).

У взрослых результаты тестирования более стабильны, чем у детей; у спортсменов — более стабильны, чем у не занимающихся спортом.

Согласованность теста - независимость результатов от личных качеств лица, проводящего, оценивающего тест. Согласованность определяют по степени совпадения результатов, получаемых на одних испытуемых разными экспертами. При этом возможны 2 варианта:

1. Эксперт только оценивает результаты теста, не влияя на его выполнение. Например, одну и ту же контрольную разные экзаменаторы могут оценить по-разному. Нередко различаются оценки судей в гимнастике, фигурном катании, оценка ЭКГ врачами...
2. Лицо, проводящее тест, влияет на результаты. Например, некоторые эксперты более настойчивы и требовательны, лучше мотивируют испытуемых, что сказывается на результатах.

Эквивалентность теста

Нередко тест - результат выбора из определенного числа однотипных тестов. Например, броски в баскетбольную корзину можно выполнять с разных точек, спринт. бег может проводиться на дистанции 50, 60, 100 м...

В таких случаях используют метод п а р а л л е л ь н ы х форм, когда предлагают выполнить две разновидности одного и того же теста, затем оценивают степень совпадения результатов:



Надежность тестов - степень совпадения результатов при повторном тестировании одних и тех же людей в одинаковых условиях.

Значения коэффициентов Надежность

0,99-0,95	Отличная
0,94-0,90	Хорошая
0,89-0,80	Средняя
0,79-0,70	Приемлемая
0,69-0,60	Низкая

Факторы, влияющие на надежность тестовых измерений

Контекст оценки
(окружение)

Объект оценки
(кого обследуют)

Субъект оценки
(кто обследует)

Инструмент оценки
(методы, процедура)



**Величина
коэффициента
надежности**

Интерпретация

0.90 и выше

Высокая надежность

Необходима в случаях, когда:

- на данных теста предполагается делать серьезные выводы;
- экзаменуемые разделены на множество разных категорий на основании относительно небольших индивидуальных различий, например, интеллекта.

0.80 и выше	Средняя или высокая надежность (16% изменчивости в тестовом балле приходится на долю ошибки).
около 0.70	Низкая надежность <i>Приемлема, если:</i> <ul style="list-style-type: none">•тест используется для получения предварительных выводов;•тест используется для сортировки людей на небольшое количество групп на основании больших индивидуальных различий, например, роста
менее 0.60	Неприемлемо низкая надежность

Пути повышения надежности теста

Надежность тестов до определенной степени может быть повышена путем:

- а) более строгой стандартизации тестирования,
- б) увеличения числа попыток,
- в) увеличения числа оценщиков (судей, экспертов) и повышения согласованности их мнений,
- г) увеличения числа эквивалентных тестов,
- д) лучшей мотивации испытуемых.

Информативность т е с т а — степень точности, с какой он измеряет свойство (качество, характеристику).

Информативность наз. также валидностью (англ. validity — обоснованность, законность).

Пусть для определения уровня спец. силовой подготовленности спринтеров - бегунов и пловцов хотят использовать показатели:

1) кистевая динамометрия, 2) сила подошвенного сгибания стопы, 3) сила разгибателей плечевого сустава, 4) сила мышц-разгибателей шеи.

Информативны ли они? Даже не проводя экспериментов, можно догадаться, что 2-й тест информативен у спринтеров-бегунов, 3-й — у пловцов, а 1-й и 4-й, наверное, ни у пловцов, ни у бегунов не покажут ничего интересного (хотя могут быть полезными например в борьбе).

Если тест используется для определения состояния спортсмена в момент обследования, то говорят о *диагностической информативности*. Если же на основе результатов тестирования хотят сделать вывод о возможных будущих показателях спортсмена, тест должен обладать *прогностической информативностью*.

Факторная информативность — одна из моделей теоретической информативности.

Информативность тестов по отношению к скрытому критерию, который искусственно

Логический метод определения информативности теста - суть этого метода заключается в логическом (качественном) сопоставлении биомеханических, физиологических, психологических и других характеристик критерия и тестов, используется в таких видах спорта, где нет четкого количественного критерия. Например, в спортивных играх логический анализ фрагментов игры позволяет вначале сконструировать специфический тест, а затем проверить его

Эмпирический метод определения информативности теста:

Тесты, содержательная информативность которых признана высокой, должны пройти дополнительную эмпирическую проверку. Для этого результаты теста сопоставляют с критерием. В качестве критерия используют:

- 1) результат в соревновательном упражнении;
- 2) наиболее значимые элементы соревновательных упражнений;

- 3) результаты тестов, информативность которых для спортсменов данной квалификации была установлена ранее;
- 4) сумму очков, набранную спортсменом при выполнении комплекса тестов;
- 5) квалификацию спортсменов.

Общая схема определения информативности теста:

1. Измерение количественных значений критериев.

Если в качестве критерия предполагается использовать элемент соревновательного упражнения, необходимо, чтобы он был наиболее информативным.

2. Проведение тестирования и оценка его результатов.

3. Вычисление коэффициентов корреляции между значениями критерия и тестов. Наибольшие

Пример: соотношения между результатами в лыжной гонке на 15 км, длиной шагов и скоростью на подъеме

Длина шага, м	Скорость, м/с	Место в гонке	Ранги		Длина шага, м	Скорость, м/с	Место в гонке	Ранги	
			длины шагов	скорости				длины шагов	скорости
2,19	3,84	4-е	2	2	2,05	3,79	3-е	5	4
2,02	3,73	7-е	7	6	2,17	3,81	2-е	3	3
2,20	3,93	1-е	1	1	2,02	3,73	6-е	6	5
2,07	3,63	5-е	4	7	1,89	3,57	8-е	8	8

Расчет ранговых коэффициентов корреляции:

между местом на соревнованиях и длиной шага:

$$r = 0,88;$$

между местом на соревнованиях и скоростью на подъеме $r = 0,86$.

Следовательно, оба эти показателя обладают высокой информативностью. Необходимо

отметить, что их значения также взаимосвязаны:

$r = 0,86$. Значит, длина шага и скорость бега на подъеме - эквивалентные тесты и для контроля

БЛАГОДАРЮ
ЗА
ВНИМАНИЕ!