

EDITORIAL LEADERSHIP

Editor-in-Chief:

Zukhra Nishonova – DSc (Psychology), Professor

Deputy Editor:

Ranokhon Oripova – PhD (Pedology), Associate Professor
Uzbek National Institute of Musical Arts named after Yunus Rajabi

Editorial Board:

Amina Ruzikulova – PhD (Psychology), Associate Professor
Navoi State Pedagogical Institute

Amina Sultanova – PhD (Psychology), Associate Professor
Uzbekistan National Pedagogical University named after Nizami

Azam Gaybullaev – PhD (Psychology), Associate Professor
Bukhara State Pedagogical Institute

Dilnoza Kodirova – PhD (Psychology)
Bukhara State University

Gulbakhor Yadgarova – PhD (Psychology), Associate Professor
Bukhara State Pedagogical Institute

Khabiba Yusupova – PhD (Pedology), Associate Professor
University of Tashkent for Applied Sciences

Lola Iliyeva – PhD (Psychology), Associate Professor
Tashkent Institute of Chemical Technology

Mirzokhid Yunusov – PhD (Pedology), Senior Lecturer
Fergana State University

Nasibakhon Akhmedova – PhD (Pedology), Associate Professor
Fergana Region Pedagogical Skills Center

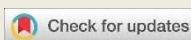
Nilufar Akhmedova – PhD (Pedology)
Customs Institute of the State Customs Committee

Ozodakhon Urinova – PhD in (Pedology), Associate Professor
Fergana State University

Rano Ismaylova – PhD (Psychology), Associate Professor
Alfraganus University

Sarvinoz Kosimova – PhD (Psychology), Associate Professor
Bukhara State University


Shuhrat Kudayberganov – PhD (Pedology), Associate Professor
Urgench State Pedagogical Institute



THE ROLE OF PSYCHOMETRIC INDICATORS IN ASSESSING THE QUALITY OF PSYCHODIAGNOSTIC METHODS

Shakhzoda Mamadjonova¹.

¹ Lecturer, Tashkent International University of Education.

 <https://doi.org/10.5281/zenodo.18666089>

Key words: *psychometrics, test item difficulty, discrimination index, item discrimination coefficient, reliability, validity, representativeness, norm-referenced testing, psychodiagnostic methods.*

ABSTRACT

The article examines the principal psychometric indicators of test item quality, including item difficulty, the discrimination coefficient, the discrimination index, reliability, validity, and representativeness, as applied in the development and adaptation of psychodiagnostic instruments. Particular attention is given to discrimination as one of the key criteria for evaluating test item quality. Discrimination is defined as the ability of a test item to differentiate between examinees with high overall test scores and those with low scores. In other words, a high-quality item should contribute to a statistically significant distinction among respondents based on the level of the measured construct. The article further substantiates that, in norm-referenced educational testing—where the primary objective is to classify learners according to their level of knowledge—the discrimination index of each item constitutes a fundamentally important component of test quality assurance. It is the item's capacity for differentiation that determines its diagnostic value and practical significance.

ЗНАЧЕНИЕ ПСИХОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ В ОЦЕНКЕ КАЧЕСТВА ПСИХОДИАГНОСТИЧЕСКИХ МЕТОДОВ

Ключевые слова:

психометрия, сложность тестового задания, индекс дискриминации, коэффициент дискриминации задания, надежность, валидность, репрезентативность, нормативно-ориентированное тестирование, психодиагностические методы.

АННОТАЦИЯ

В статье рассматриваются основные психометрические показатели качества тестовых заданий, включая сложность, коэффициент дискриминации, индекс дискриминации, надежность, валидность и репрезентативность, применяемые в процессе разработки и адаптации психодиагностических методик. Особое внимание уделено дискриминации как одному из ключевых критериев качества тестовых заданий. Под дискриминацией понимается способность задания дифференцировать испытуемых с высокими индивидуальными показателями по тесту от испытуемых с низкими показателями. Иными словами, качественное задание должно способствовать статистически значимому разграничению респондентов по уровню выраженности измеряемого признака. В статье также обосновано, что для нормативно-ориентированных педагогических тестов, основной целью которых является распределение обучающихся по уровню знаний, показатель дискриминации каждого задания выступает принципиально важным элементом обеспечения качества тестирования. Именно способность задания к дифференциации определяет его диагностическую ценность и практическую значимость.

Основной целью проводимых в Республике Узбекистан реформ является формирование здорового, образованного поколения, обладающего высокими духовно-нравственными качествами. В условиях стремительного глобального развития конкурентные преимущества получают государства, активно внедряющие высокотехнологичную продукцию, современные профессиональные знания и новейшие информационно-коммуникационные технологии.

В информационном обществе стратегическим ресурсом выступают информация, знания и творческий потенциал личности. Формирование человеческого капитала требует создания благоприятной культурной и образовательной среды, обеспечивающей развитие и реализацию способностей и талантов.

В контексте реализации Указа Президента Республики Узбекистан Ш.М. Мирзиёева «О мерах по дальнейшему совершенствованию системы управления проектами в сфере информационно-коммуникационных технологий» от 29 августа 2017 года № УП-3245 особое значение приобретает совершенствование нормативно-правовой базы в сфере информационной безопасности. Усиление мер по обеспечению информационной и общественной безопасности, поддержанию правопорядка с использованием современных технологий, а также защите государственных информационных систем, ресурсов и национального сегмента сети Интернет направлено на снижение зависимости от внешних информационных источников и укрепление информационного суверенитета страны.

Анализ литературы и методология

Обращаясь к истории развития тестологии, можно выделить период 1920-1930-х годов, когда научно-теоретические основания тестирования не рассматривались как приоритетные. Основное внимание уделялось практической эффективности теста – его способности оперативно выявлять и отбирать подготовленных лиц. Эмпирические показатели успешности выполнения заданий выступали в качестве единственного критерия оценки качества теста.

Использование диагностических методик без четко сформулированной теоретической концепции и без достаточной эмпирической обоснованности нередко приводило к ошибочным научным выводам и необоснованным практическим рекомендациям. Выявляемые тестами характеристики зачастую не имели ясного психологического содержания. По сути, подобные инструменты можно охарактеризовать как «слепые» тесты.

Подобный подход к пониманию валидности сохранялся вплоть до начала 1950-х годов. Однако уже во второй половине XX века как в США, так и в других странах усилилась критика узкоэмпирического подхода. Стало очевидным, что разработка тестов должна опираться не только на практические результаты, но и на теоретические психологические концепции. Сформировалось понимание, что практика без теоретического основания является ограниченной, тогда как теория без эмпирической проверки – несостоятельной. В настоящее время валидность психодиагностических методик рассматривается как результат комплексной теоретической и эмпирической оценки.

Показатели качества тестовых заданий

К числу основных психометрических показателей качества тестовых заданий относятся показатель сложности и коэффициент дискриминации. Показатель сложности является важной характеристикой задания и во многих случаях определяет его место в структуре теста. Различают субъективную и статистическую сложность.

Субъективная сложность связана с индивидуальными психологическими барьерами тестируемого. Ее уровень определяется рядом факторов, среди которых:

1. Условия проведения тестирования (*временные ограничения, четкость инструкций и др.*);
2. Степень сформированности знаний и навыков, необходимых для выполнения задания;
3. Функциональное и эмоциональное состояние испытуемого.

С целью минимизации влияния указанных факторов процедура тестирования стандартизируется, что позволяет обеспечить сопоставимость результатов.

В тестах учебных достижений, как правило, учитывается прежде всего правильность выполнения задания, тогда как способы решения, характер возникающих трудностей и энергетические затраты испытуемого анализируются в меньшей степени. В этой связи применяется показатель статистической сложности. Статистическая сложность определяется соотношением количества испытуемых, успешно выполнивших задание, и общего числа участников тестирования.

Наиболее простым способом выражения сложности является абсолютное количество испытуемых, правильно выполнивших задание. Например, если из 250 человек первое задание выполнили 50 человек, а второе – 150 человек, то первое задание считается более сложным.

Однако подобная форма представления имеет существенные ограничения. Во-первых, необходимо указывать общий объем выборки, участвовавшей в апробации.

Во-вторых, такой способ неудобен при сравнении заданий, проверенных на выборках различной численности. В связи с этим в современной тестологии используются относительные показатели (*доли или проценты*), позволяющие обеспечить корректное сопоставление результатов.

Результаты

Этот вариант указания сложности задания можно использовать в реальной практике, но и он имеет недостаток. Увеличение значения этого показателя не указывает на увеличение сложности, а, скорее, на ее уменьшение. То есть для нашего примера: задача сложностью 60% менее сложная, чем задача сложностью 20%. Поэтому было предложено использовать взаимную величину сложности, то есть долю тех, кто не смог выполнить задание. Этот показатель называется **индексом сложности**, который можно найти как взаимную величину сложности:

Для нашего примера:

$$U_1 = 100 - P_1 = 100 - 20 = 80\%$$

$$U_2 = 100 - P_2 = 100 - 60 = 40\%$$

Здесь: P_1 — доля выполнивших первое задание.

P_2 — доля выполнивших второе задание.

$$U = 100 (1 - n/N), \text{ где:}$$

U — индекс сложности в процентах.

n — количество студентов, выполнивших задание

N — общее количество студентов, принявших участие в тестировании.

Для нашего примера:

$$U_1 = 100 (1 - 50/250) = 100 (1 - 0,2) = 80\%$$

$$U_2 = 100 (1 - 150/250) = 100 (1 - 0,6) = 40\%$$

Подсчитать сложность задания в любом случае очень просто, поскольку она отражает процент респондентов, выполнивших (или не справившихся) тестовое задание.

Например, если задачу решили только 20% прошедших тест, то для выборки ее можно оценить как сложную, если 80% — легкую. В данном случае важен только тот факт, выполнена задача или нет.

Статистическая сложность позволяет определить место задания в тесте. Таким образом, если задачу решило большинство респондентов, она ставится в начало как легкая задача. Если задание выполняет небольшая часть респондентов, она ставится в конец теста, так как она сложная. Результатом разделения задач по уровню сложности постепенно должна стать «лестница» усложняющихся задач, каждый из ее этапов будет представлен процентом респондентов, выполнивших соответствующее задание¹.

Педагогическое содержание и важность сложности тестовых заданий заключаются в следующем:

1. Сложность задачи является характеристикой относительной, так как зависит от особенностей выборки и может отличаться от образца к образцу. Невозможно указывать уровень сложности задачи без указания контингента лиц, с которыми эта сложность определяется.

2. Основной целью задания сложности является выявление заданий, непригодных к использованию из-за их чрезмерной простоты или сложности, и определение места задания в тесте. Вопрос о том, включать или нет задания в тест в связи с их крайней простотой или сложностью, должен решаться не автоматически, а с учетом следующих обстоятельств: целей теста, содержательного значения задания, возможности переформулирования задания, простоты или сложности.

3. Апробация должна охватывать все категории лиц, которые могут принимать участие в тестировании, уровень, количество простых и сложных заданий в тесте должны соответствовать возможному распределению подготовки контингента респондентов к тесту. Тесты следует использовать только для той категории лиц, для которой они созданы.

4. Сложность задания может зависеть не только от содержания учебного материала, но и от выбранной формы задания и его формулировки.

5. Сложность — главная особенность не только задания, но и всего теста. Это могут быть простые или сложные задачи [1].

Поэтому слова о том, что студент выполнил 80% заданий теста или что он выполнил только 5 заданий из 40, не имеют педагогического значения до тех пор, пока не будет ясна сложность задач, для которых составлен данный тест.

Еще одним показателем качества тестовых заданий является дискриминация тестовых заданий. Дискримитивность – это способность отделить учащихся с высокими индивидуальными баллами по тесту от учащихся с низкими индивидуальными баллами или респондентов с высокими показателями от респондентов с низкими показателями.

Поскольку основной целью нормативно-ориентированных педагогических тестов является разделение учащихся по уровню знаний, соответственно, дифференциация каждого задания является важной характеристикой качества тестирования.

Поясним понятие дискриминации на примере. Представьте, что мы протестировали группу студентов. В эту группу вошли отличники, хорошисты и неудачники.

Предположить, что отличники должны сдать тест лучше, чем двоечники, логично не правильно. Но оказалось, что и неудачники, и отличники с одинаковым успехом или неудачей выполнили одно из заданий. В чем может быть причина такого нелогичного явления? Оказывается, это происходит только тогда, когда в задании есть существенные недостатки. Типичными недостатками задач с низкой дискриминацией являются:

1. Чрезмерная сложность, запутанные предложения.
2. Неточность опеределения.
3. Доказательство решения.
4. Неправильность вариантов ответа.
5. Появление двух и более правильных ответов, не указанных в инструкции.

В действительности, если задания имеют перечисленные недостатки, с ними с равной вероятностью справятся (или не справятся) как отличники, так и неудачники. Таким образом, определение дискриминации необходимо для предотвращения выполнения некачественных задач.

Самый простой и точный способ дискримитивного расчета – использовать метод экстремальных групп, то есть учитываются результаты наиболее и наименее успешных студентов из всего теста.

Как правило, они получают лучший и худший результат за весь тест от 10 до 30% (*часто 27% на основе статистических соображений*).

Индекс дискриминации задания рассчитывается как разница между долей тестируемых из высокоэффективных и низкоэффективных групп, правильно выполнивших задание.

$$D = \frac{N_{n_{max}}}{N_{max}} - \frac{N_{n_{min}}}{N_{min}}$$

$N_{n_{max}}$ - количество учеников верхней группы, правильно выполнивших задание

$N_{n_{min}}$ - количество учеников худшей группы, правильно выполнивших задание

N_{max} - общее количество респондентов в верхней группе

N_{min} - общее количество респондентов в худшей группе

Например: проведена опробация в группе, состоящей из 200 студентов. Чтобы определить дискриминацию, нам необходимо через оценки результата всего теста упорядочить студентов. Затем выбрать лучшую группу студентов (27%) из 54 и столько же худших студентов.

После этого нам необходимо выяснить, как справились с заданием студенты, отобранные в лучшую группу, и студенты, отобранные в худшую группу. Предположим, что 50 студентов первой группы и 25 студентов второй группы выполнили первое задание. А со вторым заданием 30 учащихся первой группы и 25 учащихся второй группы [2].

Так для первой задачи:

$$D = 50/54 - 25/54 = 0,93 - 0,46 = 0,47$$

Для второй задачи:

$$D = 30/54 - 25/54 = 0,56 - 0,46 = 0,10$$

Индекс дискриминации может варьироваться от +1 (когда с заданием справились все ученики лучшей группы и никто из худшей группы) до -1 (когда ситуация обратная – в лучшей группе никто не справился, а все в худшей группе сделали).

Задания с отрицательным значением индекса дискриминации или значением, близким к нулю, не могут считаться удовлетворительными и в них следует искать серьезные ошибки. Такие задания следует переформулировать или исключить из теста.

Индекс дискриминации выше 0,3 следует считать удовлетворительным. [3].

Есть и другие способы определения дискриминации в задании. Они основаны на нахождении различных коэффициентов корреляции.

Для определения коэффициента корреляции необходимы два набора чисел. Один из них представляет собой результаты выполнения респондентами данного задания, а второй – результаты всего теста.

Поясним, что коэффициент корреляции является количественной мерой силы и направления вероятностной связи между двумя переменными. Коэффициент корреляции принимает значения от -1 до +1.

Обсуждение

Максимальная сила корреляционной связи наблюдается при функциональной зависимости типа «один к одному», когда каждому значению одной переменной соответствует единственное значение другой переменной и наоборот. В данной ситуации эмпирическая зависимость предельно приближается к линейной функциональной связи, а коэффициент корреляции по абсолютной величине стремится к единице. Количественной мерой тесноты связи выступает абсолютное значение коэффициента корреляции, отражающее степень согласованности изменений переменных независимо от направления связи.

Направление корреляционной зависимости определяется характером совместной вариации переменных. Если увеличение значений одной переменной сопровождается увеличением значений другой, фиксируется прямая (положительная) корреляция. В случае, когда рост одной переменной ассоциирован со снижением другой, имеет место обратная (отрицательная) корреляция. Соответственно, знак коэффициента корреляции указывает на направление статистической зависимости, тогда как его модуль характеризует её интенсивность.

В психометрических исследованиях для анализа качества тестовых заданий широко применяется точечно-бисериальный коэффициент корреляции (point-biserial correlation), предназначенный для оценки связи между дихотомической переменной (например, правильность выполнения задания) и количественной переменной (общий тестовый балл). Различают две основные модификации данного показателя:

1. корреляция между результатом выполнения отдельного задания и суммарным баллом по тесту;
2. корреляция между результатом задания и общим тестовым баллом, рассчитанным без включения данного задания (исправленная корреляция).

Следует учитывать, что в первом случае коэффициент корреляции может быть статистически завышен вследствие включения анализируемого задания в структуру общего балла, что приводит к эффекту «часть–целое». Вторая модификация позволяет минимизировать данный эффект и обеспечивает более корректную оценку дискриминативной способности задания.

Наряду с точечно-бисериальной корреляцией, при анализе дискриминации заданий используются коэффициент корреляции Пирсона и его производные, особенно в ситуациях, когда переменные представлены в интервальной шкале. Дополнительно для комплексной оценки распределительной и дифференцирующей способности теста в целом может применяться индекс Фергюсона (формула Фергюсона), позволяющий определить степень вариативности распределения суммарных тестовых баллов и чувствительность инструмента к различиям между испытуемыми.

Заключение

Большинство психометрических показателей качества тестовых заданий – индекс сложности, индекс дискриминации, точноно-бисериальная корреляция и иные коэффициенты – при корректном расчёте и интерпретации демонстрируют сопоставимые тенденции в оценке эффективности задания. Базовым аналитическим параметром традиционно выступает индекс дискриминации, отражающий способность задания дифференцировать испытуемых с различным уровнем выраженности измеряемого признака. После его определения дополнительные характеристики, включая показатели сложности и корреляционные коэффициенты, могут быть рассчитаны с использованием стандартизированных статистических таблиц либо специализированного программного обеспечения.

При анализе заданий закрытого типа, особенно формата множественного выбора, целесообразно оценивать не только долю правильных ответов, но и функционирование каждого дистрактора. Качественная психометрическая модель предполагает, что неверные варианты ответа преимущественно выбираются испытуемыми с низким уровнем общего тестового результата. Соответственно, коэффициенты корреляции между выбором дистрактора и суммарным баллом должны иметь отрицательные значения либо быть статистически незначимыми (близкими к нулю). Положительная корреляция неверного варианта с общим баллом указывает на некорректную конструкцию задания и снижает его дискриминативную ценность. Подобный анализ позволяет выявить нефункциональные или слабые дистракторы и повысить диагностическую точность тестового инструмента.

Современные информационные технологии существенно расширяют возможности статистической обработки, моделирования и интерпретации психометрических данных. Применение специализированных программных пакетов позволяет автоматизировать расчёт ключевых индексов, проводить многофакторный анализ и осуществлять проверку параметров модели, что повышает методологическую строгость исследования и снижает вероятность вычислительных ошибок.

Вместе с тем наличие большого числа психодиагностических методик не свидетельствует об их научной состоятельности. Не все инструменты отвечают фундаментальным психометрическим требованиям. До внедрения методики в диагностическую практику необходимо подтвердить её соответствие критериям валидности, надёжности, точности и общей измерительной состоятельности. Только систематическая эмпирическая верификация указанных параметров обеспечивает научную обоснованность и практическую значимость психодиагностического инструментария.

Список литературы

1. Зароченцев К.Д., Худяков А.И. **Основы психометрии**. – М., 2006. – 23 с.
2. Абрамова Г.С. **Введение в практическую психодиагностику**. – Брест, 2003. – 8 с.
3. **Общая психодиагностика** / Под ред. А.А. Бодалева, В.В. Столина. – М.: Изд-во МГУ, 2007. – 154 с.
4. Бодалев А.А., Столин В.А. **Общая психодиагностика**. – СПб., 2000. – 154 с.
5. Бурлачук Л.Ф. **Психодиагностика: учебник для вузов**. – СПб., 2005. – 300 с.
6. Вассерман Л.И., Дюк В.А., Иовлев Б.В., Червинская К.Р. **Психологическая диагностика и новые информационные технологии**. – СПб., 2007. – 203 с.
7. Гайда В.К., Захаров В.П. **Психологический тест**. – Л., 2002. – 298 с.
8. G'oziyev E.G'. **Psixologiya**. – Т.: O'zMU «Universitet», 2003. – 35 b.
9. G'oziyev E.G'. **Umumiy psixologiya**. – Т.: O'qituvchi, 2010. – 11 b.
10. Бурлачук Л.Ф., Морозов С.М. **Словарь-справочник по психологической диагностике**. – Киев: Наук. думка, 1989. – 200 с.
11. Анастаси А., Урбина С. **Психологическое тестирование**. – СПб.: Питер, 2005. – 688 с.
12. Анастаси А. **Психологическое тестирование**. Кн. 1. – М.: Педагогика, 1982. – 320 с.

НАУЧНЫЙ ОТЗЫВ

Представленная статья посвящена актуальной и методологически значимой проблеме современной психодиагностики – анализу и обоснованию психометрических показателей качества тестовых заданий в процессе разработки и адаптации диагностических методик. В условиях расширения сферы применения тестовых технологий в системе образования и практической психологии обращение к вопросам валидности, надежности, сложности и дискриминативности заданий является научно обоснованным и социально востребованным.

Автор последовательно раскрывает содержание ключевых психометрических параметров, уделяя особое внимание показателю сложности задания и индексу дискриминации. Корректно отражена относительная природа сложности, её зависимость от характеристик выборки и условий тестирования, а также значение стандартизации процедуры диагностики. Теоретические положения аргументированы и сопровождаются расчётными примерами, что придаёт работе прикладной характер и усиливает её методическую ценность.

Отдельного положительного внимания заслуживает анализ дискриминации тестовых заданий. Автор обоснованно рассматривает дискриминативность как центральный показатель качества нормативно-ориентированного тестирования, направленного на дифференциацию испытуемых по уровню выраженности измеряемого признака. Метод экстремальных групп представлен корректно, с последовательным разъяснением алгоритма расчёта индекса дискриминации. Обоснованно подчёркнута необходимость исключения или переработки заданий с отрицательными либо близкими к нулю значениями индекса.

Существенным достоинством статьи является обращение к корреляционным методам анализа, включая использование точечно-бисериального коэффициента корреляции. Автор демонстрирует понимание того, что современный психометрический анализ требует применения статистически строгих инструментов и не может ограничиваться исключительно процентными показателями успешности выполнения заданий.

Особую научную ценность работе придаёт её ориентация на методологическое укрепление психодиагностической практики в условиях цифровизации образовательной среды и повышения требований к качеству измерительных инструментов. Представленный анализ способствует формированию культуры научно обоснованного конструирования тестов, повышению объективности оценки и снижению риска диагностических ошибок. Работа фактически формирует стандарты критического подхода к разработке и адаптации тестовых методик, что значительно расширяет её практическую значимость.

Статья отличается логической структурированностью, последовательностью изложения и корректным использованием профессиональной терминологии. Автор демонстрирует владение базовыми принципами классической тестологии и понимание ключевых психометрических требований к диагностическим инструментам. В перспективе целесообразным представляется расширение теоретической базы за счёт включения современных зарубежных исследований в области теории ответа на задание (IRT), а также более детального сравнительного анализа различных статистических методов оценки дискриминации и надежности.

В целом представленная работа является завершённым научным исследованием, обладающим теоретической и практической значимостью. Статья соответствует требованиям, предъявляемым к публикациям в профильных научных изданиях по психологии и педагогической диагностике, и может быть рекомендована к опубликованию.

Disclaimer ©

This editorial review has been prepared by the Editorial Board of the *Journal of Pedagogical & Psychological Studies* for the purposes of internal editorial assessment and quality assurance within the journal's publication process. This review is intended to provide an analytical overview of the scientific content, methodological approach, and thematic relevance of the submitted work. It does not constitute peer review, does not replace independent expert evaluation, and should not be interpreted as reflecting the personal views of the author(s) or as representing the official position of the journal. The Editorial Board assumes no responsibility for the implementation, interpretation, or consequences of any observations, comments, or analytical conclusions contained in this review. The review may include content generated with the assistance of artificial intelligence tools used for editorial support purposes.

This editorial review is provided solely to enhance transparency in the editorial process and to support the maintenance of academic and publication standards.