УДК 378

Об использовании статистических методов при оценке качества тестов в подготовке менеджеров

Канд. филолог. наук **Гребенников А.О.** agrebennikov@mail.ru
Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., д. 7/9

Оценка эффективности тестирования студентов в целом, и обучающихся по направлению менеджмент в частности, является одной из актуальных проблем при преподавании различных дисциплин. Использование статистических методов является одной из основных методик такого рода оценки. В статье исследуется применение однопараметрической модели Раша для оценки эффективности тестов. Показана применимость данной модели к различным вариантам тестовых заданий. Приводится общий вид и методика вычисления модели, в том числе при помощи различного рода прикладных программ. Показана объективность полученных оценок. Рассматривается практический опыт использования данной модели при работе с тестами, предлагаемыми менеджерам. Анализируются результаты эксперимента с тестами в группах различного уровня подготовленности. Полученные на априори валидных тестах экспериментальные данные подтверждают работоспособность модели.

Ключевые слова: тест, однопараметрическая модель Раша, оценка эффективности теста, объективная оценка, методика преподавания.

Statistical methods in evaluation of tests for managers

Ph. D. (Linguistics) **Grebennikov A.O.** agrebennikov@mail.ru
Saint- Petersburg State University
Russia, St. Petersburg, Universitetskaya Emb., 7/9, 9199034,

Evaluation of testing is one of the most important tasks of teaching methodology nowadays. The possibility of using Rash model for the statistical processing of testing results for managers is explored in the article. A detailed justification of the model choice and example of its calculation are given. The experimental results for managers different level are analyzed. The data obtained justifies the validity of the model in question for evaluation of tests.

Keywords: test, Rash model, evaluation of tests, objective evaluation, teaching techniques.

Эффективную подготовку менеджеров в наши дни невозможно представить без обучения иностранному языку на должном уровне (1). Многообразие тестов, используемых в различных методиках преподавания, неизбежно ставит вопрос об их эффективности, т.е. адекватности теста уровню обучаемого, эффективности его использования и т.п. Необходимо получить объективные оценки, т.е. оценки, не зависящие от множества случайных факторов, возникающих при тестировании, как то: предъявления несложного теста группе сильных обучающихся и наоборот, распределения учащихся по способностям

внутри группы и т.п. Роль статистических методов при решении проблем подобного рода неуклонно возрастет (2).

Значительную популярность в последние полвека приобрел метод, названный в честь датского математика Георга Раша (Georg Rasch, 1901 — 1980), — так называемая однопараметрическая модель Раша (3). В своем основном виде модель применяется для анализа дихотомических (т.е., имеющих только два возможных значения) переменных, что как нельзя лучше подходит к тестам, где мы легко можем обозначить правильный ответ на вопрос как 1, а неправильный – как 0. После проведения теста легко составить матрицу, в строчках которой были бы испытуемые, в столбцах — вопросы теста, а на их пресечении — баллы (1 или 0), полученные испытуемым за ответ на тот или иной вопрос. Легко заметить, что подобная матрица не зависит ни от числа вопросов в тесте, ни от количества студентов, что делает первоначальное применение модели практически универсальным и в высшей степени нетрудоемким. Суммировав баллы в каждой строке (т.е., подсчитав число правильных ответов), мы получаем первичный балл испытуемого (аналогично можно получить первичный балл задания, суммируя в столбцах число правильных ответов на задание всеми испытуемыми, чем больше правильных ответов на задание, тем легче оно оказывается для данной группы испытуемых). Отсюда, разделив полученный балл на общее число заданий, легко перейти к вероятности правильных (p), и, соответственно, вероятности неправильных (q) ответов (4).

Необходимо отметить, что подобный метод может применять не только к традиционным тестам с множественным выбором ответов, как это может показаться на первый взгляд. Действительно, путем несложных преобразований любое тестовое задание становится доступным для анализа с использованием модели Раша. Ведь давая задание, скажем, на перевод, любой преподаватель четко знает, владение какими именно лексическими единицам и грамматическими конструкциями он хочет проверить у обучающихся того или иного уровня. Вряд ли мы увидим ошибки в использовании окончания 3-его лица ед. числа у глаголов в английском языке в группе upper-intermediate, одновременно, никто не предложит перевод конструкций с инфинитивом группе начинающих. Соответственно, любое предложение может быть представлено как набор из N вопросов, предполагающих единственно правильный ответ, т.е. как набор 1 и 0 с точки зрения использования нашей модели, где 1 – правильно переведённое слово (словосочетание) или верно использованное правило грамматики.

Можно ввести понятия уровня подготовленности (ability) S студента:

$$S = \frac{p}{q},$$

и трудности заданий t:

$$t = \frac{1-p}{p}$$

Модель Раша зависит только от отношения S к t поэтому носит название однопараметрической.

Можно обозначить $\ln t = \delta$.

Предлагается считать задания, с $\delta > 2.6$ очень сложными, с δ от 1,5 до 2.59 – трудными, -14.9 — 1.49 – заданиями среднего уровня сложности, -2,59 — -1,5 легкими, < -2,6 – очень легкими . Задания, оценивающиеся нулями (предельно трудные) и единицами (предельно легкими), считаются «нетекстовыми», т.е. – недифференцирующими испытуемых по уровню подготовленности.

Вводя дополнительно $\ln S = \theta$, получаем общий вид модели, или вероятность того, что участник с уровнем подготовки S, правильно выполнит задание трудности t:

$$P(\delta,\theta)=\frac{\exp(\theta-\delta)}{1+\exp(\theta-\delta)},$$

где θ — уровень обучающегося, а δ — сложность вопроса, модель верна для любого уровня трудности заданий и для любой пары участников тестирования.

Другое название модели – функция успеха. Единица измерения модели называется логитом. Как и в случае со многими другими статистическими параметрами, логит является условной единицей, смысл численного значения которой полностью выявляется только при сравнении результатов анализа нескольких тестов. При одном логите вероятность успеха равна 0,5, т.е. сложность задания соответствует уровню обучающегося. Вероятность правильного ответа обучающегося на вопрос простой для его уровня подготовки больше 0,5, в противном случае – меньше 0,5. Задание считается более трудным, если вероятность правильного ответа на него, меньше, чем на другое, независимо от уровня выполняющего его. Одновременно, более подготовленный студент имеет большую вероятность правильного ответа на все задания.

Для обработки полученной матрицы имеются разнообразные коммерческие пакеты программ (5), тем не менее, она достаточно легко вычисляется при последовательном применении стандартизованных (т.е. имеющихся в меню программы) формул в таблице Microsoft Excel. Это позволяет быстро обработать результаты любого тестирования практически независимо от количества испытуемых даже пользователями начального уровня.

Подобный анализ качественных данных количественными методами позволяет получить оценки знаний испытуемых, которые не зависят от уровня трудности отдельного теста, т.е. — объективные оценки (6).

В наши дни уже существуют значительный практические наработки по использованию модели Раша для оценки, например, результатов ЕГЭ (7, 8).

Для проверки работы модели в сфере тестирования по иностранному языку был проведён следующий эксперимент. В эксперименте принимали участие 12 групп студентов разного уровня владения английским языком общей численностью 139 человек. При проведении итогового тестирования в группах полученные результаты были обработаны с использованием модели Раша. Необходимо отметить, что студентам были предложены стандартизованные, т.е. априори валидные тесты, что делает полученные результаты еще более показательными. В силу ряда причин административного характера одна из групп осталась без наблюдения во время тестирования и, таким образом, получила возможность использования любых дополнительных материалов при ответе на вопросы теста. В итоговой таблице она обозначена как «группа списавших» без указания уровня обучения. Следовательно, случайным образом, мы получили прекрасную возможность проверить работоспособность модели и для случая, когда тестовые задания теряют всякую трудность для испытуемых.

Результаты приводятся в Таблице 1.

Таблица 1

Значения модели Раша для результатов теста в группах испытуемых

Уровень изучения языка	Количество человек	Среднее значение модели Раша
группа pre- intermediate	22	0,46
группа intermediate -	22	0,42
группа intermediate +	27	0,65
группа upper-intermediate	24	0,76
группа advanced	31	0,76
группа «списавших»	13	0,79

Необходимо отметить, что мы не приводили значение модели Раша для каждого студента, а использовали усредненное, так как, по данным наблюдений дисперсия в каждой группе не превышает 0,03 (для групп с более низким уровнем) и 0,01 (для групп с более высоким уровнем и «списавших»). Подобные отклонения чрезвычайно малы и дают нам полное право рассматривать усредненную величину в качестве основной.

Даже беглый анализ приведенных результатов показывают, что более подготовленный студент имеет большую вероятность правильного ответа на все задания, по сравнению с менее подготовленным. Самая высокая вероятность у группы списавших студентов, что вполне закономерно.

Таким образом, выбор модели Раша подтверждает целесообразность своего использования для обеспечения профессионального подхода к созданию высокого уровня надежности тестовых заданий с устойчивыми значениями параметра трудности (9). Одновременно необходимо отметить, что при анализе тестов все же не следует пренебрегать таким важным вопросом любой статистической обработки как репрезентативность выборки, т.е., количеством испытуемых, обследуемых для оценки того или иного теста и соотношением этого количества с генеральной совокупностью, т.е. общим количеством обучающихся (10).

Список литературы

1. *Рябухина Ю.В.* Преподавание иностранных языков: от обучения менеджеров к менеджменту обучения // Научный журнал НИУ ИТМО. Серия «Экономика и экологический менеджмент». 2013. № 2.
2. *Дмитренко Н.А.* Групповая форма обучения при подготовке менеджеров в современном вузе // Экономика и экологический менеджмент. 2012. №2.
3. Rasch, G. Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition. — Chicago: The University of Chicago Press. — 1980.
4. Wright B. D., Masters G. N. Rating Scale Analysis: Rasch Measurement. — Chicago: Mesa Press. — 1982.
5. RUMM: Rasch Unidimensional Measurment Models for analysing assessment and attitude questionaire data – Режим доступа: http://www.rummlab.com.au, свободный.
6. Гребенников А.О. Метод анализа Раша в применении к тестовым данным. // Актуальные вопросы языкового тестирования. — СПб.: Изд-во СПбГУ, — 2007 — С. 73 — 77.
7. Летова Л.В. Модель Раша как инструмент для объективного измерения уровня учебных достижений по математике // Инновации в образовании. 2014. № 2. С. 139 — 148.

8. Летова Л.В. Исследование качества теста единого государственного экзамена по физике с помощью модели Раша // Управление образованием: теория и практика. 2013. № 3 (11). С. 52 — 61.
9. Гребенников А.О., Сеничкина О.А. Использование латентно-структурного анализа для оценивания и контроля уровня сформированности англоязычной профессионально-ориентированной коммуникативной компетенции студентов неязыковых факультетов // Дистанционное и виртуальное обучение. 2014. № 7 (85). С. 101 — 110.
10. Ильясов Ф. Н. Репрезентативность результатов опроса в маркетинговом исследовании // Социологические исследования. 2011. № 3. С. 112 — 116.

References

1. Ryabukhina Yu.V. Prepodavanie inostrannykh yazykov: ot obucheniya menedzherov k menedzhmentu obucheniya // *Nauchnyi zhurnal NIU ITMO. Seriya «Ekonomika i ekologicheskii menedzhment».* 2013. № 2.
2. Dmitrenko N.A. Gruppovaya forma obucheniya pri podgotovke menedzherov v sovremennom vuze // *Ekonomika i ekologicheskii menedzhment.* 2012. №2.
3. Rasch, G. Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition. — Chicago: The University of Chicago Press. — 1980.
4. Wright B. D., Masters G. N. Rating Scale Analysis: Rasch Measurement. — Chicago: Mesa Press. — 1982.
5. RUMM: Rasch Unidimensional Measurment Models for analysing assessment and attitude questionaire data – Rezhim dostupa: http://www.rummlab.com.au, svobodnyi.
6. Grebennikov A.O. Metod analiza Rasha v primenenii k testovym dannym. // Aktual'nye voprosy yazykovogo testirovaniya. — SPb.: Izd-vo SPbGU, — 2007 — S. 73 — 77.
7. Letova L.V. Model' Rasha kak instrument dlya ob"ektivnogo izmereniya urovnya uchebnykh dostizhenii po matematike // *Innovatsii v obrazovanii*. 2014. № 2. S. 139 — 148.
8. Letova L.V. Issledovanie kachestva testa edinogo gosudarstvennogo ekzamena po fizike s pomoshch'yu modeli Rasha // *Upravlenie obrazovaniem: teoriya i praktika*. 2013. № 3 (11). S. 52 — 61.
9. Grebennikov A.O., Senichkina O.A. Ispol'zovanie latentno-strukturnogo analiza dlya otsenivaniya i kontrolya urovnya sformirovannosti angloyazychnoi professional'no-orientirovannoi kommunikativnoi kompetentsii studentov neyazykovykh fakul'tetov // *Distantsionnoe i virtual'noe obuchenie.* 2014. № 7 (85). S. 101 — 110.
10. Il'yasov F. N. Reprezentativnost' rezul'tatov oprosa v marketingovom issledovanii // *Sotsiologicheskie issledovaniya.* 2011. № 3. S. 112 — 116.