

Анализ качества тестовых заданий статистическими методами

**Туракты Интымаков,
Жибек Касенова,
Айгуль Муйтунова**
Национальный центр тестирования
aigul.muitunova@mail.ru

В статье приведены результаты оценки качества тестовых заданий внешней оценки учебных достижений (далее ВОУД) по специальностям и дисциплинам всех направлений обучения.

Проанализированы коэффициенты корреляции тестовых заданий, использованных во время ВОУД 2014–2015 гг. Представлен анализ качества тестовых заданий ВОУД ВО с применением программного средства RUMM.

Статья предназначена для широкого круга специалистов в сфере образования, а также для лиц, интересующихся проблемами высшего образования в нашей стране.

Введение

Внешняя оценка учебных достижений (ВОУД), как один из видов независимого от организаций образования мониторинга за качеством обучения, проводится в соответствии с Государственной программой развития образования на 2011–2020 годы и статьей 55 Закона Республики Казахстан «Об образовании».

Задачи ВОУД:

- 1) осуществление мониторинга учебных достижений обучающихся;
- 2) оценка эффективности организации учебного процесса;
- 3) проведение сравнительного анализа качества образовательных услуг, предоставляемых организациями образования¹.

Результаты ВОУД используются для широкого информирования учащихся, студентов, родительской ответственности о качестве предоставления образовательных услуг в каждой организации образования. В результате организации образования, не обеспечивающие достаточный уровень подготовки, вынуждены принимать экстренные меры по исправлению ситуации².

Система ВОУД применяется для проверки знаний студентов выпускных курсов вузов, имеющих достаточный, более чем трёхлетний, опыт экзаменационного тестирования. Данные ВОУД позволяют вести систематическое наблюдение, оценивать и прогнозировать состояние и динамику изменения результатов учебных достижений обучающихся. Результаты ВОУД используются в анализах при рейтинговых исследованиях.

Таким образом, осуществление мониторинга образовательных услуг путем введения ВОУД стимулирует вузы на предоставление качественного образования, а также даёт возможность гражданам страны самостоятельно выбирать вуз для обучения с учётом качества его работы.

¹ Закон Республики Казахстан «Об Образовании», статья 55, п.4 Управление качеством образования, 87 с.

² *Интымаков Т.Ж.* Внешняя оценка в высшем образовании.— Информационно-методический журнал РК: Новости в образовании (Білімдегі жаңалықтар) №4 (40), 2012. С. 4–5.

О форме тестовых заданий ВОУД ВО

Главная задача вузов — сформировать у выпускников компетенции по применению полученных знаний в реальной жизни. Важную роль в этом играют умелая организация и стимулирование учебно-познавательной деятельности студентов посредством систем заданий в тестовой форме.

По мнению ведущих учёных, тестовые задания с выбором нескольких правильных ответов позволяют максимально полно и объективно оценить уровень теоретической и практической подготовленности обучаемых. А также, в отличие от традиционных (тестовых заданий с одним правильным ответом из пяти предложенных), тестовые задания нового формата позволяют радикально снизить угадываемость правильного ответа, проверить различные виды знаний, а также повысить качество педагогических измерений.

При проведении ВОУД используется закрытая форма с выбором нескольких правильных ответов.

Особенностью заданий с выбором нескольких правильных ответов является то, что испытуемому надо не только найти правильные ответы, но и определить полноту своего ответа.

В таких заданиях предполагается оценивание результатов тестирования следующим образом:

1) за выбор всех правильных ответов студенту присваивается 2 (два) балла.

2) за допущенную 1 (одну) ошибку один балл вычитается из двух.

3) за допущенные 2 (две) и более ошибки вычитается ещё один балл из двух; остаётся ноль баллов. Оценки становятся 2-1-0.

С целью подтверждения результатов исследований предыдущих лет об эффективности использования тестовых заданий нового формата в ходе внешней оценки учебных достижений студентов выпускных курсов вузов НЦТ проведён анализ качества тестовых заданий по результатам основного (2013, 2014 гг.) тес-

тирования. В процессе обработки результатов тестирования произведены: выявление коэффициента корреляции, применение программного средства RUMM, дистракторный анализ.

Коэффициент корреляции тестовых заданий

Коэффициент корреляции тестовых заданий — количественный показатель, характеризующий степень статистической связи ответов испытуемых на тестовые задания. Высокие значения, близкие к 1, этот коэффициент получает в том случае, если факт решения одного задания с высокой вероятностью сочетается с решением другого задания. Как правило, наличие такого высокого показателя характерно для заданий, относящихся к одной теме или требующих актуализации одного и того же умения (навыка, способности).

Корреляция задания с критерием (r_{xy}) является статистическим требованием к тестовым заданиям. Это распространённое в практике название представляет собой сокращённый вариант более правильного понятия — корреляция оценок, полученных испытуемыми в задании, с оценками, полученными ими же по какому-либо критерию, например, по сумме баллов.

Чем выше значение r , тем больше вероятность превращения задания в тестовой форме в тестовое задание, то есть быть включённым в тест. Особенно заметно эта вероятность повышается при $r > 0,4$.

Нулевая корреляция свидетельствует об отсутствии у задания системных свойств, присущих тесту. Такие задания, так же как и задания с отрицательными значениями r , x , y , устраняются из тестовых материалов, как не выдержавшие эмпирической проверки.

Интерпретация значений коэффициента корреляции приведена в табл. 1.

Таблица 1

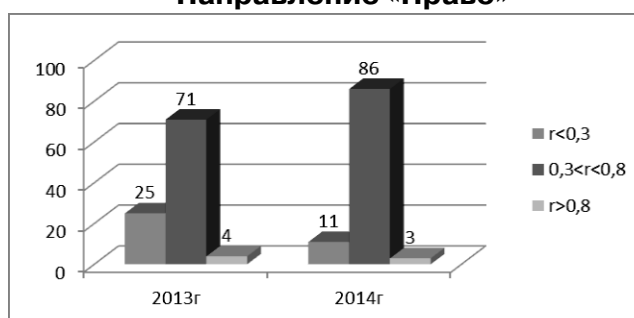
Интерпретация значений коэффициента корреляции

| R | Значение |
|-----------------|--------------------------|
| $r < 0,3$ | Низкая корреляция |
| $0,3 < r < 0,8$ | Хорошая корреляция |
| $r > 0,8$ | Очень хорошая корреляция |

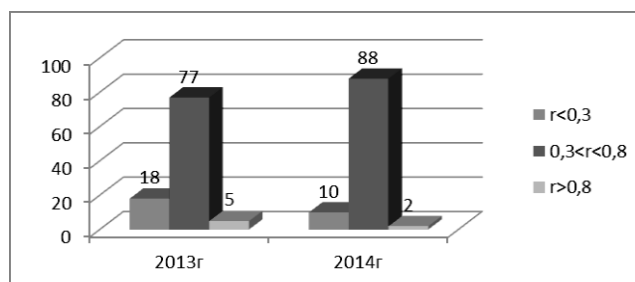
Рассмотрим анализ результата коэффициента корреляции тестовых заданий ВОУД ВО по дисциплинам направлений обучения.

В табл. 2 из 3 приведены показатели коэффициентов корреляции тестовых заданий по направлениям обучения.

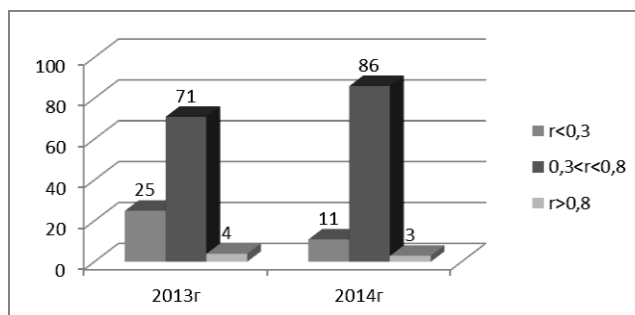
Динамика изменения коэффициента корреляции (2013–2014 гг.), баллы Направление «Право»



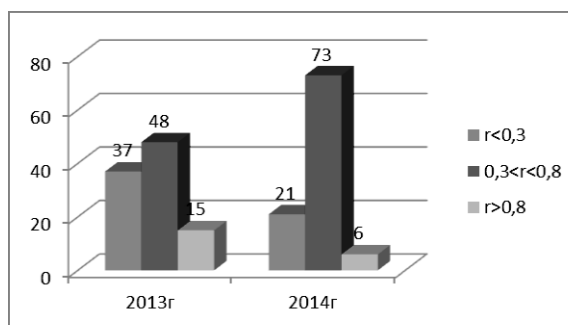
Динамика изменения коэффициента корреляции (2013–2014 гг.), баллы Направление «Социальные науки, экономика и бизнес»



Динамика изменения коэффициента корреляции (2013–2014 гг.), баллы Направление «Естественные науки»



**Динамика изменения коэффициента корреляции (2013–2014 гг.), баллы
Направление «Сельскохозяйственные науки»**



Выводы

Из множества известных методов расчёта применена формула наиболее часто применяемого метода расчёта коэффициента корреляции задания с суммой баллов. Ввиду громоздкости этой работы и во избежание ошибок вся работа выполнена на ЭВМ с помощью одного из множества имеющихся статистических пакетов Statistica, SPSS. Сводные результаты проведённого анализа отображены в таблицах в процентном соотношении.

Показатели свидетельствуют о существенном понижении количества тестовых заданий с низкой корреляцией ($r < 0,3$ и $r > 0,8$), увеличении количества тестовых заданий с хорошей корреляцией, уменьшению количества заданий с высокой корреляцией. Причиной увеличения количества тестовых заданий с хорошим показателем коэффициента корреляции могло быть проведение многократной экспертизы, апробационного тестирования. Тестовые задания с низкой корреляцией удалены с базы тестовых заданий.

Таким образом, база тестовых заданий 2014 г. в сравнении с 2013 г. значительно улучшилась.

Анализ качества тестовых заданий с применением программного обеспечения

RUMM

Для оценки качества набора тестовых заданий ВОУД ВО использовалась

модель Г. Раша. Выбор этой модели обусловлен следующим. Измерения на основе модели Раша позволяют получить объективные оценки латентных (скрытых) переменных. Измерения, полученные по модели Раша, являются объективными в том смысле, что они не зависят от того, кто измеряет, а также от набора тестовых заданий, который выступает в роли измерительного инструмента. Кроме того, измерения осуществляются на линейной интервальной шкале, что позволяет применять для их дальнейшей обработки и анализа широкий спектр статистических процедур.

С помощью программы измерения латентных переменных RUMM, разработанной под руководством проф. Дэвида Эндрича (Мердокский университет, Австралия), были обработаны результаты тестирования множества испытуемых.

Оценка качества набора тестовых заданий осуществлялась по критериям:

- определения экстремальных заданий;
- совместимости тестовых заданий;
- соответствия уровня трудности теста уровню подготовленности студентов.

Для анализа использовали задания по учебным дисциплинам с наибольшим числом тестируемых. Для применения программы необходимы объекты с большой выборкой с целью получения более достоверной картины. В качестве примера взяты дисциплины:

- 1) «Конституционное право» — вариант 441, студентов — 46, вузов — 30.

2) «Теория государства и права» — вариант 131, студентов — 67, вузов — 32.

Были проанализированы матрицы результатов тестирования 50 заданий двух вариантов данных дисциплин. В результате было выявлено, что экстремальные задания отсутствуют, т.е. нет заданий, которые решили бы все студенты, и нет заданий, которые не решили все студенты.

На основе критерия Хи-квадрат вычисляется совместимость разработанных тестовых заданий, т.е. образуют ли они измерительный инструмент. Чем меньше эмпирический уровень значимости статистики Хи-квадрат (по крайней мере меньше 0,05), тем в меньшей степени совместим набор тестовых заданий. В данном исследовании эмпирический уровень значимости составил:

• Конституционное право — 0,05;

• Теория государства и права — 0,721.

Полученное значение свидетельствует о высокой степени совместимости тестовых заданий. Это означает, что полученные результаты тестирования можно использовать для измерения уровня подготовленности студентов и для измерения уровня трудности заданий.

На рис. 1, 2 представлены распределения уровня подготовленности студентов (вверху) и оценка трудности заданий внизу. По оси абсцисс откладываются значения латентной переменной, по оси ординат — число испытуемых (вверху) и заданий (внизу). С левой стороны указаны абсолютные значения, а с правой — относительные величины.

Распределение трудностей заданий близко к равномерному закону распределения.

— Конституционное право РК:

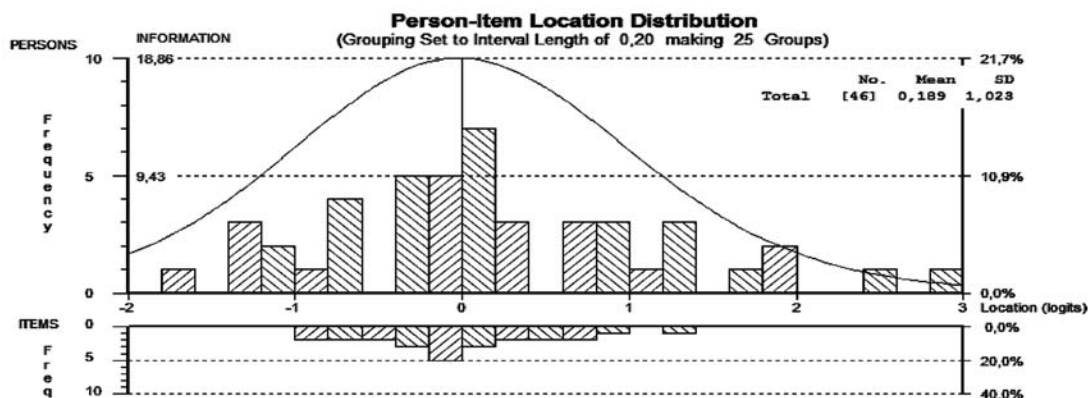


Рис. 1

Количество студентов 46. Среднеквадратичное отклонение 1,023. Уровень подготовленности тестируемых соответствует уровню трудности теста — 0,189. Коэффициент надёжности теста 0,983. Диапазон изменения уровня подготовленности составил от -2 до 3 логит. Длина интервала — 5 логит. Этот факт подтверждает диффе-

ренцирующую способность теста. Уровень трудности заданий изменяется от -1,0 до 1,4. Длина всего интервала составила 2,4 логит. Такой диапазон изменения трудности заданий не всегда обеспечивает измерение уровня знаний студентов. Поэтому данный тест требует небольшой корректировки заданий.

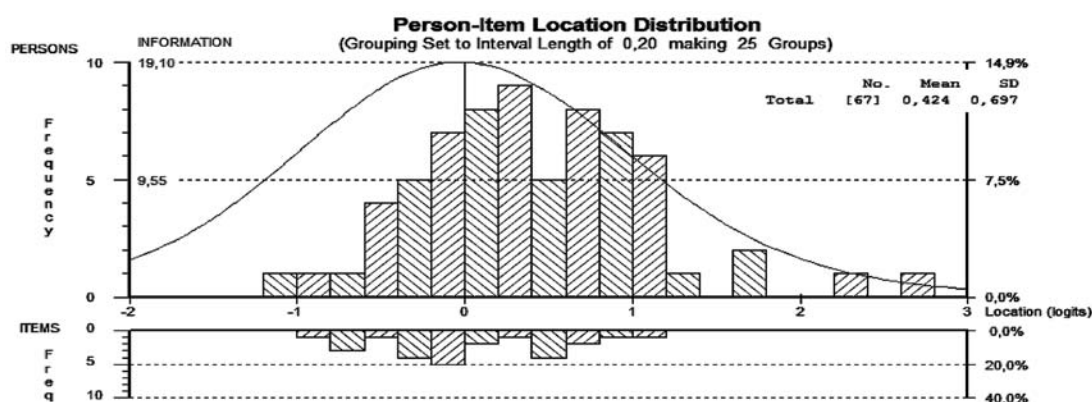


Рис. 2

— Теория государства и права:

Количество студентов 67. Средне-квадратичное отклонение 0,697. Уровень подготовленности тестируемых выше уровня трудности теста (0,424). Коэффициент надёжности тестовых результатов 0,756. Диапазон измерения уровня подготовленности составил от $-1,2$ до $2,8$ логитов. Длина интервала — 4 логита. Этот факт подтверждает дифференцирующую способность теста. Уровень трудности заданий изменяется от $-1,0$ до $1,2$ логита. Длина всего интервала составила 2,2 логита.

Выводы

Из рис. 2 вытекает, что в целом уровень качества теста годится для измерения уровня подготовленности основной массы студентов, но уровень трудности теста надо увеличивать. Тест имеет недостатки для измерения

хорошо и отлично подготовленных испытуемых. Для повышения дифференцирующей способности теста в него надо добавить трудные задания.

Заключение

В целом полученные результаты позволяют сделать вывод, что разработанный набор тестовых заданий является приемлемым средством контроля уровня подготовленности студентов.

Применение новой системы с использованием нескольких правильных ответов позволило существенно повысить качество проведения ВОУД. В свою очередь это способствует повышению эффективности управления системой высшего образования Казахстана. Анализ и интерпретация результатов ВОУД ВО способствует решению многих проблем повышения качества образования.