

ПРОВЕРКА КАЧЕСТВА БАЗЫ ТЕСТОВЫХ ЗАДАНИЙ

**Кали Абдиев,
Диана Абентаева,
Гульжан Примбетова,**
Национальный центр тестирования
Республики Казахстан

Целью данной статьи является исследование методов и результатов улучшения качества тестовых заданий, включаемых в базу заданий Национального центра тестирования. Это связано с новой политикой ежегодного обновления базы. Обновление осуществляется с опорой на результаты статистической обработки, проведённой по итогам Национального тестирования за 2009–2012 годы.

Статистическая обработка результатов выпускников школ в рамках национального исследования осуществлена тремя методами. Это расчёт классических коэффициентов корреляции Пирсона, коэффициентов надёжности результатов тестирования — альфа Кронбаха и проведение дистракторного анализа.

Проблемная ситуация

В Республике Казахстан тестирование абитуриентов во время вступительных экзаменов в вузы внедрено с 1999 года. На сегодняшний день метод тестирования используется в рамках девяти национальных и мониторинговых исследований, проводимых с участием Национального центра тестирования, на уровнях высшего и среднего образования. Содержание и форма тестовых заданий, используемых в данных исследованиях, отличаются друг от друга. Несмотря на эти различия способ формирования базы тестовых заданий для всех национальных и мониторинговых исследований один.

Формирование базы тестовых заданий для использования в рамках национальных и мониторинговых исследований в разных странах осуществляется по-разному. Выбор того или иного способа формирования базы зависит от особенностей системы образо-

вания данной страны, от целей тестирования и от других факторов.

Использование испытуемыми шпаргалок, подготовка заданий разработчиками в домашних условиях, недостаточное обновление базы тестовых заданий привели к искажению общей картины относительно уровня подготовленности выпускников школ.

Средний балл по всей стране по результатам ЕНТ-2010 и ЕНТ-2011 достиг 80–90 баллов, и обнаружился рост среднего балла ежегодно на 10 пунктов. Вместе с тем, с каждым годом увеличилось количество выпускников, выполнивших все тестовые задания по всем предметам, то есть выполнивших правильно 100% заданий. Все вышесказанное явилось доказательством существования серьёзных проблем и в вопросах организации экзамена и в вопросах формирования базы тестовых заданий.

Формирование базы тестовых заданий: анализ состояния

Анализ результатов тестирования выпускников школ во время Единого национального тестирования в 2009–2012 годы позволил сделать ряд выводов организационного и содержательного характера по пробле-

ме формирования базы тестовых заданий.

Вместе с оценкой качества знаний выпускников школ проводилась статистическая обработка результатов тестирования. Благодаря проведённому анализу были выявлены проблемы, связанные с формированием базы тестовых заданий, а также с организацией проведения тестирования школьников.

База тестовых заданий обновлялась в 2009 году на 3,7%. В 2010 году база не обновлялась вообще, в 2011 году база обновлялась на 3%.

При этом до 2012 года тестовые задания для ЕНТ разрабатывались по заказу разработчиками «на дому», то есть без выезда в Национальный центр тестирования. При таком положении дел невозможно было давать 100%-ную гарантию на то, что тестовые задания, разработанные в домашних условиях, не становились доступными для широкого круга заинтересованных в положительных результатах лиц (администрация школ, родители, учащиеся, учителя и т.д.).

Касательно заданий, использованных во время ЕНТ, до 2012 года ситуация складывалась следующим образом:

- задания, использованные в ЕНТ, одновременно публиковались в учебно-методических пособиях, в книжках для

пробного тестирования. Таким образом, вся база тестовых заданий, которая хранилась в Национальном центре тестирования, оказалась доступной для всех, что позволило разработать шпаргалки, позволяющие набрать учащимся на экзамене повышенные баллы без специальной подготовки к предметам.

Существующие способы формирования базы тестовых заданий

Первый способ. Разрабатываются отдельные тестовые задания по каждой теме учебных предметов, по которым сдаётся экзамен. Из разработанных тестовых заданий и результатов их применения создаётся база данных. Во время подготовки к экзамену из имеющихся в базе заданий создаются варианты тестов. Нередко тестовые задания отбираются путем случайной выборки согласно спецификации тестов. Количество вариантов тестов определяется в зависимости от количества тестируемых.

Второй способ. Разрабатываются целостные варианты тестов в определённом количестве, опять же согласно спецификации теста. После их использования на экзамене создаются параллельные варианты для следующего экзамена.

Параллельными называются варианты, которые имеют сходное предметное содержание в пределах укрупненной учебной единицы, равные средние арифметические, равные дисперсии и равные интеркорреляции. Ошибочные компоненты одного параллельного варианта не коррелируют с такими же компонентами другого варианта¹.

При этом тестовые задания для определённых тем имеют специально отведённое место в тесте. Варианты тестов, которые использовались во время экзаменов, в последующем публикуются в учебно-методических пособиях по подготовке к экзаменам, используются во время проведения пробного тестирования.

В Казахстане тестирование абитуриентов во время вступительных экзаменов в вузы внедрено с 1999 года. На сегодняшний день метод тестирования используется в рамках 9 национальных и мониторинговых исследований, проводимых Национальным центром тестирования на уровнях высшего и среднего образования. Содержание и форма заданий, используемых в данных исследованиях, отличаются друг от друга. Несмотря на эти различия способ формирования базы заданий для всех девяти национальных и мониторинговых

1

Аванесов В.С.
Композиция тестовых заданий. Учебная книга.
3-е изд., доп. М.: Центр тестирования, 2002.
240 с. С. 212.

исследований один. Это первый способ формирования базы тестовых заданий из двух указанных выше способов.

Выбору данного способа способствовало несколько факторов, именно отсутствие опыта Казахстана в проведении экзаменов путем тестирования до 1999 года и отношение общественности к тестовым заданиям как средствам измерения уровня подготовки учащихся.

Указанные две причины выбора способа формирования базы тестовых заданий путем разработки отдельных заданий и подбора тестовых заданий в варианты случайным образом, в соответствии со спецификацией теста взаимосвязаны между собой.

Шпаргалки

В Казахстане существует два типа шпаргалок: алфавитная и цифровая.

1. В алфавитной шпаргалке даётся условие тестового задания и правильный ответ. В шпаргалке тестовые задания расположены в алфавитном порядке.

2. В цифровой шпаргалке:

- для предметов естественно-гуманитарного цикла — указываются первые две буквы пер-

вого слова, затем в скобках указывается количество слов после него, затем первая буква последнего слова, и наконец, правильный ответ;

- для предметов естественно-математического направления — указываются все цифры, имеющиеся в тестовом задании по порядку, затем правильный ответ.

Два примера заданий по математике $16y^2 = 25y^2 - 9$

1. КОРНИ УРАВНЕНИЯ

1) $-1; 0$

2) $-1; -\frac{9}{16}; \frac{9}{16}; 1$

3) $-1; -\frac{3}{4}; \frac{3}{4}; 1$

4) $-1; 1$

5) $-\frac{3}{4}; \frac{3}{4}$

Правильный ответ: $(-1; -\frac{3}{4}; \frac{3}{4}; 1)$

В цифровой шпаргалке

указано: 1642529 $(-1; -\frac{3}{4}; \frac{3}{4}; 1)$

2. ВЫЧИСЛИТЕ: $16^{\log_{81} 3 + \log_2 \sqrt[4]{5}}$

1) 5

2) $4\sqrt{3}$

3) 6

4) $9\sqrt[4]{3}$

5) 7

Правильный ответ: 6

В цифровой шпаргалке указано: 16813243 (6)

Пути решения проблемы создания базы качественных тестовых заданий

На основе анализа сложившейся ситуации нами определены следующие пути решения указанной проблемы:

- обновление базы тестовых заданий через создание параллельных заданий;
- раздельное использование базы тестовых заданий (для учебно-методических пособий, пробного тестирования и ЕНТ);
- внесение изменений в организацию проведения ЕНТ путем увеличения количества представителей Министерства образования и науки Республики Казахстан, экзаменаторов, направляемых в регионы для проведения экзамена.

Подготовительная работа к проведению ЕНТ-2012 началась с начала 2011/2012 учебного года. Во-первых, были внесены изменения в нормативно-правовые акты по вопросам проведения ЕНТ, где указывалось, что в каждой аудитории, где будет проводиться тестирование, будут дежурить по 1 человеку, а также исключение дежурства местных работников. Этот шаг был предпринят для исключения возможности использования шпаргалок тестируемыми, то есть для ужесточения контроля.

Во-вторых, на 30% была обновлена база тестовых заданий путем создания параллельных заданий. Так же было принято решение разделить базу тестовых заданий на 3 блока, так чтобы в первом блоке хранились тестовые задания, которые будут использоваться на экзамене, во втором — задания, предназначенные для публикации в учебно-методических пособиях, в третьем блоке — задания, используемые в рамках пробного тестирования.

В-третьих, было принято решение разработку и экспертизу тестовых заданий проводить на территории Национального центра тестирования с приглашением разработчиков со всех регионов Казахстана.

Первые результаты

В результате предпринятых мер, перечисленных нами выше, средний балл по результатам ЕНТ-2012 снизился до 70,91%. Поскольку использование шпаргалок было предотвращено за счёт обновления базы тестовых заданий, а также увеличения количества представителей Министерства образования и науки в аудиториях, то данный показатель уровня подготовленности выпускников школ считается объективным.

Взаимосвязь между результатами тестирования и обновлением базы тестовых заданий выявлена также за счёт осуществления статистической обработки результатов ЕНТ 2009–2012 гг.

Статистическая обработка результатов тестирования в рамках ЕНТ за 2009–2012 годы даёт возможность проследить прямую связь между обновлением базы тестовых заданий и результатами выпускников школ. Статистическая обработка результатов осуществлена посредством расчётов классических коэффициентов корреляции Пирсона, коэффициента альфа Кронбаха и проведением дистракторного анализа. В данной статье приводятся результаты применения первых двух указанных методов.

Корреляция

Коэффициент корреляции рассчитывался по следующей формуле²:

$$r_{xy} = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}},$$

где SS_x — сумма квадратов отклонений баллов испытуемых от среднего арифметического балла в интересующем задании;

• SS_y — сумма квадратов отклонений тестовых баллов испытуемых от среднего арифметического балла по всему тесту;

• SP_{xy} — скорректированная на средние значения сумма попарных произведений x и y .

В табл. 1 даны показатели коэффициента корреляции тестовых заданий для школ по математике, с русским языком обучения (ЕНТ 2009–2012 гг.), в процентах. В ходе статистической обработки результатов тестирования были осуществлены расчёты коэффициентов корреляции тестовых заданий по всем предметам ЕНТ, но в рамках данной статьи рассмотрим результаты расчёта данного коэффициента по математике, так как этот предмет является одним из основных предметов ЕНТ.

Из табл. 1 видно, что в те годы, когда база тестовых заданий обновлялась (2009 г., 2012 г.), количество тестовых заданий с коэффициентом корреляции $r < 0,3$ выше, чем в годы, когда база обновлялась не более чем на 5%. Так же обстоят дела с тестовыми заданиями, у которых $0,8 < r < 0,3$ и $r > 0,8$. Наблюдается схожесть показателей коэффициентов корреляции в годы, когда база обновлялась не более чем на 5%.

Следовательно, существует прямая зависимость между показателями коэффициента корреляции тестовых заданий и процессом обновления базы тестовых заданий.

Рассмотрим расчёт коэффициента корреляции тестовых

Методология

Методология

2

Аванесов. В.С.
Композиция тестовых заданий. Учебная книга.
3-е изд., доп. М.: Центр тестирования, 2002.
240 с. С. 169.

Таблица 1

Значения коэффициентов корреляции тестовых заданий по математике для школ с русским языком обучения (ЕНТ 2009–2012 гг.)

Годы проведения ЕНТ	Показатели коэффициента корреляции (в процентах)		
	$r < 0,3$	$0,8 < r < 0,3$	$r > 0,8$
2009 год	12	87,6	0,1
2010 год	7,7	92,2	0,04
2011 год	7,9	92,1	0,02
2012 год	9,7	89,7	0,5

Таблица 2

Показатели коэффициентов корреляции тестовых заданий по математике для школ с казахским языком обучения, в процентах (ЕНТ 2009–2012 гг.)

Годы проведения ЕНТ	Показатели коэффициента корреляции (в процентах)		
	$r < 0,3$	$0,8 < r < 0,3$	$r > 0,8$
2009 год	13,12	86,8	0,08
2010 год	5,4	94,6	0
2011 год	7,92	92,08	0
2012 год	14,8	85,16	0,04

вых заданий, использованных для школ с казахским языком обучения. Показатели коэффициентов корреляции тестовых заданий по математике для школ с казахским языком обучения даны в табл. 2.

Ситуация с тестовыми заданиями, использованными во время ЕНТ 2009–2012 гг. для школ с казахским языком обучения обстоит следующим образом:

1) в 2009 и 2012 годах, когда база тестовых заданий обновлялась на 30%, показатели коэффициентов корреляций распределены таким образом, что около 0,04–0,08% тестовых заданий имеют коэффициент

корреляции больше 0,8 (высокая корреляция с суммой баллов). Значит, эти задания пригодные для использования.

2) в 2010 и 2011 годах, когда база тестовых заданий обновлялась на незначительное количество (3–3,8%), тестовых заданий с высокой корреляцией нет вообще!

Сравнительный анализ результатов расчёта коэффициентов корреляции по математике для школ с казахским и по математике для школ с русским языками обучения показал, что в 2009 и 2012 годах тестовые задания с коэффициентом корреляции меньше 0,3 находятся в диапазоне 85–89%, тогда как в

2010 и 2011 годах такие задания составляют 92–94% от общего числа задания.

Коэффициент альфа Кронбаха

Следующим показателем качества тестовых материалов, используемых на ЕНТ, является коэффициент альфа Кронбаха. Коэффициент альфа Кронбаха α показывает внутреннюю согласованность заданий, включённых в один тест. Расчёт коэффициента альфа Кронбаха производился по следующей формуле³:

$$\alpha_{st} = \frac{N \cdot \bar{r}}{1 + (N - 1) \bar{r}},$$

где N — количество заданий, а \bar{r} — средний коэффициент корреляции между заданиями теста.

Результаты расчёта данного коэффициента показаны в табл. 4.

Согласно данным табл. 4 среди вариантов тестов, использованных при тестирова-

Значения коэффициента альфа Кронбаха

αA	Значение
> 0,9	очень хорошее
> 0,8	хорошее
> 0,7	недостаточное
> 0,6	сомнительное
> 0,5	плохое
? 0,5	очень низкое

нии выпускников школ с русским языком обучения в 2009 году, когда база тестовых заданий обновлялась более чем на 20%, к достаточно хорошим и хорошим, очень хорошим тестам можно было отнести 87,8% и 99,6% тестов соответственно. Тогда как, в 2010 и 2011 годах, когда процент обновления базы тестовых заданий не достигал и 5%, эти показатели равны 100% и 100% соответственно.

Аналогичным способом рассчитан коэффициент альфа Кронбаха по результатам по математике для школ с казахским языком обучения (табл. 5).

Таблица 3

Методология

Таблица 4

Показатели коэффициента альфа Кронбаха по результатам ЕНТ 2009–2012 гг. (тесты по математике для школ с русским языком обучения)

Годы проведения ЕНТ	Показатели коэффициента альфа Кронбаха (количество вариантов тестов дано в %)					
	$\alpha ? 0,5$	$\alpha > 0,5$	$\alpha > 0,6$	$\alpha > 0,7$	$\alpha > 0,8$	$\alpha > 0,9$
2009 год	0	0	0,2	13,2	85,4	1,2
2010 год	0	0	0	6	93	1
2011 год	0	0	0	9,4	90,2	0,4
2012 год	0	0,2	0,2	5,4	91	3,2

3 http://ru.wikipedia.org/wiki/альфа_Кронбаха

Таблица 5

**Показатели коэффициента альфа Кронбаха по результатам
ЕНТ 2009–2012 гг. (тесты по математике для школ
с казахским языком обучения)**

Годы проведения ЕНТ	Показатели коэффициента альфа Кронбаха (количество вариантов тестов дано в %)					
	$\alpha \geq 0,5$	$\alpha > 0,5$	$\alpha > 0,6$	$\alpha > 0,7$	$\alpha > 0,8$	$\alpha > 0,9$
2009 год	0	0	0,8	22,5	75,1	1,6
2010 год	0	0	0	1,6	98,2	0,2
2011 год	0	0	0,3	11,5	88	0,2
2012 год	0	0,2	0,3	23,5	75,8	0,2

В 2009 году и 2012 годах к достаточно хорошим, хорошим и очень хорошим тестам можно было отнести 99,2% и 99,5% тестов соответственно. Тогда как, в 2010 и 2011 годах эти показатели равны 100% и 99,7% соответственно.

Согласно требованиям теории педагогических измерений показатели 2010 и 2011 годов говорят о том, что варианты тестов сформированы хорошо, тестовые задания достаточно хорошие. Но учитывая основной фактор, искажающий результаты тестируемых этих лет, при анализе качества тестов мы во внимание берем результаты 2009 и 2012 годов, когда за счёт обновления базы тестовых заданий на более 20% использование шпаргалок было исключено как проблема.

Таким образом, расчёт коэффициента альфа Кронбаха для определения качества тестов также как расчёт коэффициента корреляции для определения качества тестовых зада-

ний позволяют увидеть взаимосвязь между обновлением базы тестовых заданий и их качеством.

Обсуждение результатов и выводы

Проведённая работа по статистической обработке результатов тестирования выпускников школ в рамках ЕНТ 2009–2012 гг. позволила выявить следующие факторы, влияющие на качество тестовых заданий и тестов в целом:

- 1) процент обновления базы тестовых заданий;
- 2) способы хранения тестовых заданий в базе;

Первый фактор: базу тестовых заданий необходимо обновлять ежегодно, так как обновление базы позволит исключить как проблему, использование тестируемыми шпаргалок во время экзаменов. Благодаря чему появится возможность увидеть реальную карти-

ну об уровне подготовленности выпускников школ. Данный вывод был сделан на основе расчётов коэффициентов корреляции и альфа Кронбаха.

Второй фактор: необходимо хранить тестовые задания отдельно, при этом должна быть создана архивная база тестовых заданий, где будут храниться тестовые задания, предназначенные для публикации в учебно-методических пособиях, книжках для пробного тестирования. Раздельное использование тестовых заданий для различных мероприятий (публикация в пособиях для подготовки к экзамену, использова-

ние во время пробного тестирования, использование во время экзамена) позволит также исключить проблему, связанную с использованием шпаргалок выпускниками школ с целью искусственного увеличения количества баллов, набранных во время тестирования.

Кроме того, необходимо исключить фактор списывания на экзамене. Этого можно было добиться усилением контроля, поэтому был принято решение, что в каждой аудитории контроль будет осуществляться представителем Министерства, лицом, не заинтересованным в повышении результатов экзамена.

Методология

Методология