

**Прогнозирование успешности студентов первых курсов
высших учебных заведений Кыргызской Республики по
результатам Общереспубликанского теста 2003 года:
верификация валидности теста**

Дэн Е. Дэвидсон
Американские Советы по международному образованию: ACTR/ACCELS
Ddavidson@americancouncils.org



Прогнозирование успешности студентов первых курсов высших учебных заведений Кыргызской Республики по результатам Общереспубликанского теста 2003 года: верификация валидности теста

Содержание:

Краткий обзор	4
Почему необходимо верифицировать валидность Общереспубликанского теста?	6
Охваченные исследованием университеты	8
Индивидуальные оценки студентов как критерий валидности теста	9
Выбор предметов и оценка успеваемости	11
Трудности при анализе оценок	12
Заключения и направления дальнейшего изучения	12

Информация, представленная в таблицах:

1. Соотношение общего числа абитуриентов и получателей грантов, вошедших в выборку
2. Распределение вузов по популярности среди студентов с различными средними баллами
3. Средние баллы общего числа студентов и получателей грантов, вошедших в выборку
4. Средние баллы по математике в разрезе областей и успеваемости в вузе
5. Распределение средних баллов по специальностям в вузах



Приложения:

1. Соотношение между взвешенными баллами по математике и оценками по математике (тест 2002 года)
2. Соотношение между взвешенными баллами по математике и оценками по математике (тест 2003 года)
3. Распределение среднего балла по математике (2003 г.) по успеваемости и полу студентов
4. Соотношение между средним баллом за словесно-логический раздел и успеваемостью студентов в вузе
5. Распределение среднего балла по грамматике родного языка (2003 г.) по успеваемости и полу студентов
6. Соотношение между баллами за предметные тесты (2003 г.) и успеваемостью студентов в вузе
7. Распределение средневзвешенного балла за тест по английскому языку по языкам тестирования
8. Распределение средневзвешенного балла за тест по английскому языку по областям
9. Гендерные особенности распределения средних баллов за тест 2003 года: общие баллы и баллы по отдельным разделам
10. Распределение средневзвешенных баллов по разделам «Практическая грамматика», «Аналогии и дополнение предложений» и «Чтение и понимание» по языкам тестирования
11. Распределение общего средневзвешенного балла по областям Кыргызской республики
12. Распределение средневзвешенного балла за раздел «Математика» и за весь тест в целом по категориям школ
13. Распределение средневзвешенных баллов по разделам «Практическая грамматика», «Аналогии и дополнение предложений» и «Чтение и понимание» по категориям школ
14. Распределение среднего балла за раздел «Математика» по специальностям в вузе
15. Распределение среднего балла за раздел «Чтение и понимание» по специальностям в вузе

Краткий обзор



Общереспубликанское тестирование в Кыргызской Республике выступает в настоящее время в качестве основного механизма отбора при зачислении в государственные вузы страны. По результатам этого тестирования каждый год отбираются 5 085 человек (15%) из общего числа абитуриентов, которые получают полный государственный грант на обучение по выбранным ими специальностям в государственных вузах КР. В 2003 году 200 абитуриентов (3,9%), набравших самые высокие баллы на Общереспубликанском тестировании, получили так называемые «президентские» стипендии, а также право на обучение в любом вузе и на любой специальности по своему выбору. Общереспубликанское тестирование проводится по заказу Министерства образования КР Центром оценки в образовании и методов обучения (ЦООМО) – независимой тестовой организацией, учрежденной специально для этой цели при технической поддержке Американских Советов по международному образованию и финансировании Агентства США по международному развитию.

Общереспубликанское тестирование в Кыргызской Республике – это первый и пока единственный в странах СНГ действующий на общенациональном уровне вид вступительных испытаний, ориентированных исключительно на способности абитуриентов. Оно заменило предыдущий способ отбора, который был основан в основном на оценке объема заученного материала и давал возможность для процветания коррупции.

Тест измеряет общие навыки рассуждения и критического мышления и включает в себя несколько разделов: математический, словесно-логический (проверяющие математическое мышление и способность решать логические задачи) и грамматический (проверяющий уровень практического владения родным – кыргызским, русским или узбекским – языком). Валидность теста (т.е. его способность верно выявлять наиболее хорошо подготовленных абитуриентов, независимо от их социального и финансового статуса, национальности, пола, места проживания и т.п.) является важным фактором для достижения общественной и политической поддержки данного проекта. Верификация валидности теста проводится для оценки его возможностей, эффективности выбранного дизайна, достоверности результатов и т.д.



В настоящей работе представлена информация о валидности Общереспубликанского теста, а также предварительные результаты анализа взаимосвязанности тестовых баллов абитуриентов 2002 и 2003 гг. с их академическими успехами за первый учебный семестр в вузе. Выборка 2003 года основана на анализе оценок 1 641 студента по 90 различным специальностям в девяти государственных вузах за первый семестр 2003-2004 учебного года. Для сравнения приведена меньшая выборка (307 человек) студентов, сдававших тест в 2002 году, и их оценки за первый семестр первого курса.

Результаты исследования показывают, что словесно-логический и математический компоненты Общереспубликанского теста достаточно валидны для прогнозирования успешности первокурсников, и уровень валидности соответствует требованиям тестирования высокого значения, действующим в странах Запада. Необходимо отметить, что обычно коэффициенты валидности не достигают статистического значения по всем показателям. Выборки студентов по некоторым специальностям слишком малы для выведения статистических заключений. В некоторых случаях отклонения показателей валидности объясняются ошибками или ограничениями в выборке, которые нередки в подобных исследованиях в университетах. Некоторые показатели, например, сравнение университетов по отдельным специальностям, должны рассматриваться в свете явных различий между сравниваемыми величинами: кыргызские университеты, охваченные исследованием, отличаются друг от друга стандартами отбора студентов, практикой их оценивания и т.д. В будущем, после еще нескольких лет проведения тестирования, собранная информация позволит изучить и проанализировать те области, в которых в данном исследовании не представляется возможным сделать какие-либо выводы в связи с недостаточным количеством данных. Тем не менее, это первое эмпирическое исследование показывает, что в целом Общереспубликанское тестирование проводится на соответствующем уровне для всех категорий абитуриентов: представителей всех трех основных языковых групп, сельских местностей и городов, обоих полов.

Почему необходимо верифицировать валидность Общереспубликанского теста?

Тестирование, используемое для распределения бюджетных мест в высших учебных заведениях, должно являться эффективным инструментом выявления наиболее способных абитуриентов на основе



справедливой и непредвзятой оценки их знаний и навыков. Общереспубликанский тест в целом соответствует лучшим принципам проведения тестов высокого значения: его содержание не опирается на какие-либо определенные учебные программы или учебники; он ориентирован исключительно на измерение навыков аналитического и критического мышления, которые свидетельствуют об уровне подготовки к обучению в вузе.¹ Тест не является инструментом оценивания достижений учащихся и их успешности в усвоении школьного материала: это функция выпускных испытаний, проводимых самими школами. Общереспубликанский же тест является так называемым прогностическим тестом, т.е. его цель заключается в выявлении из общего числа абитуриентов тех, кто сможет успешно учиться в вузе. Лучшие абитуриенты получают полные образовательные гранты и имеют возможность выбрать любые специальности в любом университете; те, кто доказал свои способности к обучению, но не добился выдающихся результатов в тесте, имеют возможность поступить в вуз на основе конкурса соответственно количеству набранных ими баллов. К числу последних можно отнести абитуриентов из отдаленных сельских районов и абитуриентов, имеющих права на социальные льготы.

В Таблице 1 показано соотношение общего числа абитуриентов и получателей грантов, вошедших в выборку данного исследования.

¹ *Стандарты успешности в высшем образовании*, Ассоциация Американских Университетов, Нью-Йорк, 2004 г.



A Comparison of Population and Scholarship Sample Demographics		
2003 Kyrgyz National Exam		
American Councils for International Education: ACTR/ACCELS		
	Exam Population (N=35,246)	Scholarship Sample (N=1,641)
Gender		
Male	38%	30%
Female	62%	70%
Test Language		
Kyrgyz	71%	67%
Russian	23%	32%
Uzbek	6%	1%
High School Region		
Batken	9%	4%
Bishkek (Capital)	14%	18%
Chui	15%	15%
Issyk-Kul	12%	19%
Jalalabad	13%	16%
Naryn	9%	15%
Osh	24%	11%
Talas	5%	3%
High School Location		
Bishkek (Capital)	13%	17%
Oblast Center	10%	15%
Small City	9%	7%
Village	69%	61%



Основной характеристикой прогностических тестов является их прогностическая валидность. Понятие валидности тестов сегодня очень широко распространено во всем мире, особенно в сферах профессиональной сертификации, лицензирования, приема на работу и т.д. В образовании верификация валидности тестов необходима для каждого теста, определяющего процессы зачисления, формирования учебных групп, а также институциональную оценку.² Хотя существует очень много возможных критериев, по которым можно оценивать валидность прогностического теста для поступления в вузы, тем не менее наиболее широко используемым из них является уровень академической успешности студентов первого курса относительно набранных ими на тесте баллов.³ Понятие «академическая успешность» обычно определяется как способность студента адаптироваться к новой учебной обстановке, справляться с большим объемом работы и успешно усваивать требуемый вузом объем материала. Американские университеты используют данные по валидности для сравнения школьных оценок, набранных на тестах баллов и университетских показателей в процессе мониторинга общих

² Совет по высшему образованию, к примеру, включает в себя Службу оценки успеваемости студентов-первокурсников (ACES), которая бесплатно проводит анализ того, насколько верно тесты SAT, школьные оценки и другие виды показателей успешности прогнозируют дальнейшую успешность студента в американском колледже. Только за последние 30 лет в США было проведено свыше 4000 исследований валидности тестов SAT-I и SAT-II (предметных тестов). Создание в Кыргызстане подобной службы, возможно, на базе Центра оценки в образовании и методов обучения, могло бы внести значительный вклад в развитие системы образования в стране.

³ 1. Л. Ремист, С. Льюис, Л. МакКемли-Дженкинс, «Использование теста SAT-II для оценки достижений и прогнозирования академической успешности в колледже применительно к различным социальным группам», исследовательский отчет № 2001-5 Совета высшего образования, Нью-Йорк.



тенденций зачисления и выделения основных групп абитуриентов. Так как университеты стремятся к тому, чтобы были зачислены представители всех социальных групп, доступ к высшему образованию для студентов из мало представленных сегментов населения (по половому, этническому, языковому, расовому, экономическому и социальному признакам) является одной из основных проблем при зачислении. Данные по валидности тестов помогают разработчикам стратегии образования и университетам год от года совершенствовать процедуру зачисления и более эффективно использовать результаты тестов для анализа академической деятельности каждого абитуриента и его подготовленности к обучению на уровне высшей школы.

Общереспубликанское тестирование в настоящее время выступает в качестве единственного критерия отбора при распределении примерно 5 085 правительственных грантов на обучение, а также постепенно становится одним из критериев отбора во всех высших учебных заведениях Кыргызской республики. В 2003 году в тесте принимали участие примерно 35 245 выпускников общеобразовательных школ; примерно 22 000 из них поступили в государственные или частные вузов.

Охваченные исследованием университеты

Настоящим исследованием в 2003-2004 учебном году были охвачены 1 641 студент 90 различных специальностей девяти государственных вузов КР из 27. В выборку вузов вошли разные по популярности учебные заведения (см. Таблицу 2).



Таблица 2.

University Selectivity Based on Scholarship Sample Means			
American Councils for International Education: ACTR/ACCELS			
Total National Exam Score		Mean	S.D.
Most Selective Universities	Kyrgyz-Russian Slavonic University	194	26
	Manas-Kyrgyz Turkish University	172	11
	Kyrgyz National University	156	19
	Bishkek State Institute of Economy & Commerce	153	28
	Bishkek Humanities University	152	21
Least Selective Universities	Jalalabad State University	142	15
	Issyk-Kul State University	140	21
	Osh Technological University	138	16
	Naryn State University	135	13

Индивидуальные оценки студентов как критерий валидности теста

В странах Евразии и Восточной Европы студентам первого курса обычно предлагается по 6-9 предметов за семестр. Эти предметы являются общеобразовательными или подготовительными к специализированным предметам. Обычно лишь 1-3 предмета оцениваются через экзамен по пятибалльной системе (5 – отлично и 2 – неудовлетворительно), успешность по остальным предметам отмечается как «зачтено» или «не зачтено». На практике система еще более проста: отметка «не зачтено» или 2 ставится очень редко и означает, что студент полностью пропустил курс занятий по данному предмету.

Настоящее исследование охватило выборочные группы первокурсников 2002 и 2003 г.г., которые проходили Общереспубликанское тестирование и поступили по его результатам на бюджетные отделения. В исследовании принимали участие только те вузы, которые изъявили согласие сотрудничать с АКСЕЛС и ЦОМО в проведении мониторинга успеваемости. Необходимо помнить,



что классификация университетов на основе среднего балла по тесту – это новая практика для Кыргызстана, как, впрочем, и для Евразии в целом. В выборку исследования входили не все студенты первых курсов, а только обладатели правительственных грантов. Классификация вузов может изменяться в зависимости от того, какие студенты были приняты разными вузами на контрактные отделения (обычно это студенты с более низкими баллами). Общественное наблюдение за процессом зачисления в вузы нужно поощрять по многим причинам, и Общереспубликанское тестирование в Кыргызской Республике предоставляет очень важную эмпирическую базу для выведения тенденций и закономерностей этого процесса с течением времени. Средняя величина оценок успеваемости студентов первого курса не подсчитывается университетами в Кыргызстане, хотя ведение такой статистики, несомненно, помогло бы в будущем проводить исследования валидности тестов в самой республике.

Таблица 3.

A Comparison of 2003 Population and Scholarship Sample Test Means				
American Councils for International Education: ACTR/ACCELS				
	Mean	Minimum	Maximum	Standard Deviation
Scaled 2003 Total Score				
Total Population	114	27	233	27
Scholarship Sample	153	62	224	25
Scaled 2003 Math Score				
Total Population	50	22	120	15
Scholarship Sample	67	22	120	19
Scaled 2003 Analytical Score				
Total Population	50	11	112	15
Scholarship Sample	65	17	112	17
Scaled 2003 Reading Score				
Total Population	50	9	102	15
Scholarship Sample	69	25	102	12
Scaled 2003 Grammar Score				
Total Population	50	6	101	15
Scholarship Sample	66	15	98	12
Scaled 2003 Biology Subject Test				
Total Population	50	13	142	15
Scholarship Sample	62	27	131	17
Scaled 2003 Chemistry Subject Test				
Total Population	50	6	160	15
Scholarship Sample	59	21	136	17
Scaled 2003 English Subject Test				
Total Population	75	34	148	23
Scholarship Sample	90	39	148	24
Scaled 2003 German Subject Test				
Total Population	76	44	150	23
Scholarship Sample	96	54	136	25



Таблица 4.

Spring 2003 Mean Math Exam Score by Grade and Home Region			
(Data for pass/fail students not included.)			
American Councils for International Education: ACTR/ACCELS			
Course Name	1st Semester Grades		
	Grade: 5	Grade: 4	Grade: 3
Algebra (N=118)	60.5	55.5	n/a
Bishkek (Capital)	n/a	n/a	n/a
Oblast Center (Large City)	68.0	55.1	n/a
Small City	58.2	54.5	n/a
Village	59.3	55.6	n/a
Descriptive Geometry (N=90)	61.0	55.6	39.0
Bishkek (Capital)	n/a	46.0	n/a
Oblast Center (Large City)	59.2	54.1	n/a
Small City	66.0	51.7	n/a
Village	61.8	56.4	39.0
Modern Concepts of Natural Science (N=191)	73.3	65.7	63.9
Bishkek (Capital)	80.4	76.1	90.0
Oblast Center (Large City)	73.9	62.8	56.0
Small City	71.6	77.4	59.5
Village	70.6	61.0	59.9
Economics (N=83)	88.0	64.7	n/a
Bishkek (Capital)	102.0	99.3	n/a
Oblast Center (Large City)	77.0	63.7	n/a
Small City	92.0	54.0	n/a
Village	66.9	59.1	n/a

The pink cells are the ones in which the mean is higher than expected.

Выбор предметов и оценка успеваемости

Выбор предметов студентами в Кыргызской Республике несколько ограничен, так как для каждой специальности предусмотрен установленный ряд предметов, обязательных для прохождения. В отличие от американских университетов, кыргызские вузы предоставляют студентам очень мало свободы в выборе предметов по кредит-часам. Студенты зачисляются на определенную специальность; для каждой специальности есть уже утвержденный порядок прохождения предметов, а также требования по каждому предмету. Как и везде в мире, социальные науки, точные науки и математика больше привлекают юношей, чем девушек; часто система оценки в таких дисциплинах более строгая, чем в гуманитарных предметах.



Трудности при анализе оценок

Давняя проблема анализа оценок учащихся заключается в том, что анализу подвергаются оценки разных учителей и преподавателей совершенно разных дисциплин и в разных учебных заведениях. В свете этой проблемы необходимо принимать во внимание некоторые моменты, характерные для Кыргызстана в частности. Во-первых, в Кыргызстане существует относительно большое количество академических специальностей, где студенты не могут выбирать себе предметы сами. Во-вторых, в любой специальности есть установленный уровень способностей студентов, как показано в настоящем исследовании показателями стандартного отклонения средних баллов по Общереспубликанскому тестированию. И, наконец, существуют постоянные и значительные ограничения в системе оценок (всего 5 баллов). Поэтому представляется довольно сложным установить верное соотношение между баллами, набранными на тестировании, и оценками, полученными в университете. В тех специальностях, которые позволяли установить такое соотношение, эти показатели были исследованы и соотнесены стандартными методами.

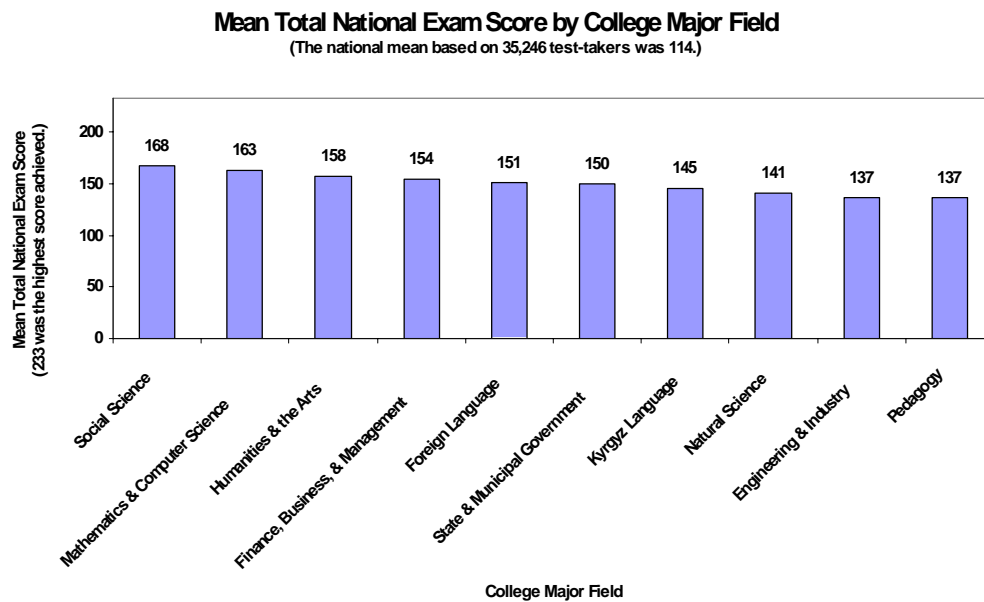
Заключения и области для дальнейшего изучения

Словесно-логический и математический компоненты Общереспубликанского теста достаточно валидны для прогнозирования успешности первокурсников, и уровень валидности соответствует требованиям тестирования высокого значения, действующим в странах Запада. Общая академическая успеваемость на первом курсе прогнозируется общими баллами по Общереспубликанскому тесту на уровне соотношения $.101^{**}$ ($p = .01$ и $N = 1620$). Соотношение с оценками по предметам было несколько выше для девушек ($.130^{**}$ $p = .01$ и $N = 1140$) и для студентов средних по популярности вузов ($.187^{**}$ $p = .01$ и $N = 737$). Для сравнения можно добавить, что соотношение между оценками по математике за первый семестр и средневзвешенным баллом по разделу «Математика» в тесте 2002 года было равно $.212$ ($p = .01$ и $N = 311$). Полностью коэффициенты соотношения баллов теста 2002 года и оценок по основным предметам приведены в приложении к данному отчету. Важно отметить, что не во всех специальностях коэффициенты валидности достигают статистической значимости. К примеру, на некоторых специальностях обучается слишком малое количество студентов для того, чтобы их показатели валидности теста имели какое-либо значение для настоящего исследования. Коэффициенты



валидности для теста по немецкому языку, хотя и высокие (.711), рассчитаны на слишком малом количестве показателей, чтобы делать какие-либо заключения; так же на данном этапе затруднительно сделать заключения о тесте по химии. Некоторые показатели, например, сравнение университетов по отдельным специальностям, должны рассматриваться в свете явных различий между сравниваемыми величинами: кыргызские университеты, охваченные исследованием, отличаются друг от друга стандартами отбора студентов, практикой их оценивания и т.д. В будущем, после еще нескольких лет проведения тестирования, собранная информация позволит изучить и проанализировать те области, в которых в данном исследовании не представляется возможным сделать какие-либо выводы в связи с недостаточным количеством данных. Выводы, сделанные в настоящем исследовании, должны в дальнейшем подвергаться повторному изучению со стороны университетов по мере сбора и анализа дополнительной информации по валидности.

Table 5.





Распределение тестовых баллов по специальностям представляет также интерес как показатель места, занимаемого каждой специальностью на «академическом рынке». В 2003-2004 учебном году наибольшее количество абитуриентов с высокими баллами поступали на такие специальности, как информатика, математика и экономика, что является свидетельством возрождения рыночного спроса на точные науки и инженерию, которые в советское время считались главными и элитными специальностями (см. Таблицу выше). В этом же контексте необходимо отметить и популярность социологических наук среди абитуриентов.

В целом можно сделать вывод, что Общереспубликанское тестирование проводится на соответствующем требованиям уровне для всех категорий абитуриентов: представителей всех трех основных языковых групп, сельских местностей и городов, обоих полов.