

Прогностическая валидность абитуриентских тестов и школьного аттестата

Дятлова Ксения Дмитриевна, директор Учебного центра НГСХА, доцент, кандидат биологических наук, e-mail: jaro@uic.nnov.ru

Михалёва Татьяна Григорьевна, руководитель отдела Центра тестирования Минобразования РФ, кандидат физико-математических наук, e-mail: tata@rustest.ru

Хлебников Владимир Алексеевич, директор Центра тестирования Минобразования РФ, доцент, кандидат технических наук, e-mail: tata@rustest.ru

Цель любых вступительных испытаний, в том числе и с использованием абитуриентских тестов, — отобрать наиболее способных и подготовленных к освоению основных образовательных программ граждан. Социальный заказ требует, чтобы абитуриентский тест выявлял не только знания по конкретному предмету, но и способность абитуриента к дальнейшему обучению по избранной специальности. Следовательно, при использовании тестирования для отбора абитуриентов в вуз необходимо правильно выбрать инструмент тестирования: предметные педагогические тесты или (и?) психологические тесты способностей.

Прежде всего необходимо чётко установить различия между психологическим и педагогическим тестом. В основу разграничения могут быть положены два фактора. Во-первых, это предмет измерения: психологические тесты измеряют некие психические свойства человека (способности — aptitude, особенности темперамента, характера, особенности познавательной сферы), тогда как с помощью педагогических тестов измеряется степень усвоения знаний, умений и навыков, которыми овладел испытуемый, то есть результаты обучения, степень усвоения содержания учебного материала, некие личные приращения [7, 10]. Таким образом, педагогические тесты предстают как частный случай тестов достижения (achievement), измеряющих влияние на респондента прошлого опыта, причём в качестве последнего выступает изучавшийся испытуемым учебный теоретический или практический курс. Второе отличие психологических и педагогических тестов в их назначении: первые служат для предсказания уровня последующего выполнения определённой функции или деятельности, вторые — представляют конечную оценку состояния индивидуума по завершении обучения, главное значение в них придаётся тому, что человек способен делать в настоящий момент [1].

Однако нужно признать, что, применяя тесты способностей и достижений, невозможно провести между ними жёсткую границу. Любой когнитивный тест, независимо от его традиционного названия, обеспечивает выборочную проверку того, что индивидуум знает на момент тестирования, и измеряет уровень развития одной или нескольких способностей (abilities). Ни один тест не показывает, как или почему конкретный человек достиг такого уровня. Но тот же результат выполнения теста позволяет предсказать, как поведёт себя данный человек в других, нетестовых ситуациях и к тому же по прошествии какого-то времени.

Мы полагаем, что абитуриентские предметные тесты должны интегрировать тесты достижений и психологические тесты способностей. В этом случае абитуриентские тесты смогут выполнять прогностическую функцию.

Дж. Равен (младший) вообще ставит под сомнение целесообразность использования тестов учебных достижений как валидную и надёжную оценку качества подготовки учащихся и полагает, что существующие тесты достижений не только не позволяют распознать способности учеников к обучению, но, более того, активно подавляют их развитие [11]. Однако независимо от суждений Дж. Равена и у нас в России неоднократно провозглашалось, что тесты (особенно состоящие из закрытых тестовых заданий) имеют свою нишу в образовательном процессе — они эффективны при оценке измерения достижения испытуемым наиболее простых уровней усвоения содержания обучения [12]. На наш взгляд, эта точка зрения и преуменьшает возможности тестового контроля [3].

Конечно, ни в одной современной дидактической системе тесты достижений не

выступают в качестве единственного средства контроля. А. Анастаси и С. Урбина пишут, что «тесты достижений наподобие SAT и АСТ никогда не предназначались на роль заменителя школьных отметок при прогнозировании достижений во время обучения в колледже. Высокие оценки в средней школе могут предсказывать оценки в колледже не хуже или даже чуть лучше, чем большинство тестов. Однако при объединении тестовых показателей с оценками за среднюю школу предсказание будущих успехов в освоении программ колледжа улучшается» [1. С. 530]. Очевидно, для определения соотношения тестовых и нетестовых форм контроля знаний нужны специальные исследования, критерием сравнения в которых может выступать прогностическая валидность методов контроля, понимаемая как эффективность прогноза учащихся в будущем. Прогностическая валидность теста устанавливается как корреляция между результатами тестирования и некоторым внешним критерием (внешней переменной), характеризующей измеряемое свойство в будущем. Для отбора абитуриентов наиболее существенным внешним критерием является успешность их обучения в вузе.

Изучение прогностической валидности абитуриентских тестов предпринималось в последнее время в России неоднократно. В частности, было показано, что результаты выполнения абитуриентских тестов централизованного тестирования* (далее — ЦТ) хорошо коррелируют с успеваемостью в вузе по итогам курсовых экзаменов, а результаты школьных аттестаций плохо согласуются с последующими достижениями учащихся [6, 13]. Казалось бы, вопрос о прогностической валидности абитуриентских тестов может быть решён положительно, но все вышеперечисленные работы имели серьёзные методические недостатки:

* Централизованное тестирование (ЦТ) — это проводимое Центром тестирования Минобразования РФ с 1991 г. ежегодно в апреле месяце одновременно по всей России тестирование выпускников одиннадцатых и девярых классов [14]. Особенности этого тестирования: добровольность участия, использование стандартизированных тестов, процедур проведения тестирования, проверки и выдачи результатов. Это самая массовая процедура независимой оценки знаний учащихся: за последние 4 года в Центре тестирования было обработано более 5 миллионов тестов.

— малые группы (12–60 чел.);

— учёт результатов только первой сессии, которая не может репрезентативно отражать успешность обучения (стресс новизны);

— учёт успеваемости только тех студентов, которые прошли в вуз по результатам ЦТ, что вызывает асимметричность выборки — учитываются только лучшие участники тестирования, находящиеся на гистограмме распределения результатов тестирования вправо от медианы.

Эти недостатки заставляют предпринять исследование прогностической валидности абитуриентских тестов на больших выборках студентов разных специальностей, учитывая их успеваемость в ходе, как минимум, двух сессий. Основное отличие предпринимаемой работы от аналогов — включение в выборку студентов, которые хотя и участвовали в ЦТ в апреле, не обязательно использовали его результаты в качестве вступительных испытаний, а прошли в вуз по итогам экзаменов или собеседований (коммерческие группы).

Нулевой гипотезой послужили два предположения:

1) прогностическая валидность абитуриентских предметных тестов не ниже таковой для среднего балла аттестата об окончании полной средней школы;

2) прогностическая валидность объединённых данных — аттестата и предметного абитуриентского тестирования выше, чем у любой составляющей.

Экспериментальной группой исследования послужили 310 студентов первого курса (приём 2002 г.), сдававших химию и биологию в форме централизованного тестирования в качестве вступительных экзаменов в вуз на шесть «биологических» факультетов Нижегородской государственной сельскохозяйственной академии (НГСХА) (см. табл. 1).

Таблица 1

Факультет	Всего зачислено,	Прошло ЦТ, чел.	Средний балл (M + т) аттестата		
			1-й сессии	2-й сессии	1-го курса

	чел.					
Технология						
переработки	89	66	4,55±0,05	3,88±0,06	3,72±0,08	3,80±0,07
Агрохимия	51	39	4,43±0,05	3,79±0,07	7,78±0,10	3,79±0,08
Ветеринария	99	58	4,29±0,04	4,13±0,06	3,76±0,08	3,94±0,08
Зоотехния	98	63	4,25±0,05	3,75±0,06	3,67±0,08	3,71±0,07
Агрономия	57	32	4,19±0,08	3,57±0,07	3,61±0,09	3,59±0,07
Лесное хозяйство	92	52	4,15±0,03	3,15±0,08	3,49±0,10	3,42±0,07
	486	310				

Прошедшие тестирование первокурсники составили 63,7% (от 56,1% до 76,5% на разные факультеты) от зачисленных на 1-й курс. При этом не все из них использовали итоги тестирования в качестве вступительных испытаний в вуз — около половины получили по ЦТ недостаточные оценки и пересдавали предметы в виде традиционного экзамена или собеседования (для медалистов и поступающих на коммерческое отделение). Тем не менее их данные представлены в исследуемой группе.

Для всех сдававших тестирование студентов учитывался их первичный тестовый балл (птб — процент верно выполненных заданий) и тестовый балл* (тб) по 100-балльной шкале, подсчитывался средний балл школьного аттестата, учитывались оценки аттестата по профильным предметам (биология, химия), итоги 1-й и 2-й сессий по ведомостям деканатов (от 3 до 5 экзаменов в сессию), средний балл успеваемости за 1-й курс и средний балл успеваемости на 1-м курсе по профильным предметам (2–3 предмета) (см. табл. 2).

* Окончательное численное выражение на определённой шкале индивидуальной оценки уровня подготовленности тестируемого, полученное на основании стандартизированной обработки результатов выполнения теста [8].

Таблица 2. Основные изучавшиеся показатели — пример исходных таблиц

Агрохимия	биология		химия		био. атт.	химия атт.	ср. атт.	1-я сессия			2-я сессия			ср. за 1-й курс		
	ЦТ птб, %	тб	ЦТ птб, %	тб				исто-рия	фи-зика	неор. хи-мия	ср.	ма-тем.	бо та-ника		ср.	орг. хи-мия
Базанкова	70,9	63	46,7	47	4	4	4,17	4	3	2	3	4	4	4	4	3,5
Берлова	52,7	50			5	5	4,82	4	4	4	4	3	4	3	3,33	3,67
Вафина			50	50	5	5	4,94	5	3	3	3,67	3	4	3	3,33	3,5
Веснина	38,2	42	26,7	34	4	4	4,06	4	3	3	3,33	3	5	3	3,67	3,5
Волгутов	67,3	67	70	64	5	5	4,89	4	4	4	4	3	3	5	3,67	3,83

Абитуриенты, сдававшие один предмет на разные специальности, были объединены в одну группу. Таким образом, из всей совокупности респондентов были сформированы две предметные группы: биология и химия.

Для доказательства нулевой гипотезы результаты обрабатывались с применением параметрических методов биометрии [5]. Подсчёт характеристик вариационных рядов включал нахождение среднего арифметического и ошибки репрезентативности. Связь между рядами показателей оценивалась с помощью коэффициента линейной корреляции Пирсона (r), разность между коэффициентами корреляции — по методу Z (зет). Достоверность данных оценивалась на уровне значимости $p < 5\%$.

В НГСХА биологию в качестве вступительного экзамена сдавали 643 абитуриента, из них поступили 303 человека.

В централизованном тестировании по химии участвовало 518 человек, из них 265 — стали студентами. Сопоставление результатов тестирования со средними данными по России свидетельствует, что абитуриенты НГСХА знают биологию несколько лучше, но химию хуже, чем средний российский абитуриент (см. табл. 3).

Таблица 3. Первичный тестовый балл по результатам централизованного тестирования в

России и НГСХА в 2002 г., %

Предметы	Россия	НГСХА
Биология	47,1	48,5
Химия	50,8	46,3

Анализ кросс-корреляций тестового балла и первичного тестового балла по биологии со средней оценкой аттестата, аттестационной отметкой по биологии, средней успеваемостью на 1-м курсе и успеваемостью по предметам биологического профиля показал, что прогностическая валидность школьного аттестата, тестового балла и первичного тестового балла по биологии в отношении успешности учёбы практически одинакова — коэффициенты данных показателей с успеваемостью на 1-м курсе были высоко достоверны, имели среднюю силу и не различались по критерию Z (табл. 4).

Таблица 4. Матрица кросс-корреляций тестового балла (ЦТтб), первичного тестового балла (ЦТптб) по биологии, средней оценки аттестата (ср. атт.), аттестационной отметки по биологии (био. атт.), средней успеваемости на 1-м курсе (ср. курс) и успеваемости по предметам биологического профиля (био курс) для 303 студентов биологических факультетов ($r_{st} = 0,11$, $p < 0,05$; $r_{st} = 0,15$, $p < 0,01$)

	ЦТ(тб)	ЦТ(птб)	ср. атт.	биоатт.	ср. курс	био курс
ЦТ(тб)		0,9699	0,3751	0,3768	0,5404	0,443
ЦТ(птб)			0,3748	0,3519	0,5287	0,4208
ср. атт.				0,7068	0,5473	0,4328
биоатт.					0,4638	0,396
ср. курс						0,7956

Эти показатели так же успешно, хотя и с меньшей силой, предсказывали успеваемость по предметам биологического цикла, а аттестационная отметка по биологии обладала наименьшей прогностической способностью. Весьма неприятным фактом является низкая корреляция аттестационных отметок по биологии и среднего балла аттестата с баллами ЦТ. Иными словами, **абитуриенты на тестировании не подтверждают свои аттестационные отметки**. Так, из 31 обладателя золотой и серебряной медалей из числа поступивших только 3 человека получили отметку «5» по централизованному тестированию. Это подтверждает данные Г.А. Гончаровой, которая в Политехническом колледже Саратовского государственного технического университета выявила, что традиционные испытания (аттестат за 9-й класс) не отражают объективную картину знаний учеников. Когда школьник попадает в условия независимой оценки знаний, результаты испытаний более объективны [2]. Возможно, что дело не в субъективности аттестата, иначе его прогностическая валидность в отношении успешности учёбы была бы минимальна, а в том, что абитуриентские тесты централизованного тестирования предъявляют к участникам более высокие требования, чем обычные экзамены средней школы.

Весьма интересно было объединить данные аттестата и предметного абитуриентского тестирования. Поскольку

А. Анастаси и С. Урбина не дают алгоритма этого объединения [1], мы просчитали для каждого студента **объединённый индекс успеха** (ОИУ): тестовый балл ЦТ по предмету плюс сумма оценок по всем предметам школьного аттестата. Прогностическая валидность ОИУ действительно выше, чем таковая для любой его составляющей — коэффициент корреляции с успеваемостью по итогам 1-го курса составляет 0,647, то есть связь весьма сильная.

Прогностическая валидность тестового балла и первичного тестового балла по химии в отношении успешности учёбы в вузе достоверно (по критерию Z) выше, чем таковая для школьного аттестата (см. табл. 5).

Таблица 5. Матрица кросс-корреляций тестового балла (ЦТтб), первичного тестового балла (ЦТптб) по химии, средней оценки аттестата (ср. атт.), аттестационной отметки по химии

(хим. атт.), средней успеваемости на 1-м курсе (ср. курс) и успеваемости по предметам химического профиля (хим. курс) для 265 студентов биологических факультетов ($r_{ст} = 0,13$, $p < 0,05$; $r_{ст} = 0,16$, $p < 0,01$)

	ЦТ(тб)	ЦТ (птб)	ср. атт.	хим. атт.	ср. курс	хим. курс
ЦТ(тб)		0,9675	0,4042	0,4788	0,6137	0,5496
ЦТ(птб)			0,4072	0,4756	0,6093	0,5377
ср. атт.				0,7026	0,5546	0,463
хим. атт.					0,5142	0,4763
ср. курс						0,8408

Эти показатели также успешно, хотя и с меньшей силой, предсказывали успеваемость по предметам химического цикла, а аттестационная отметка по химии обладала наименьшей прогностической способностью. Также отмечена невысокая, хотя и большая по величине, чем для биологии, корреляция аттестационных отметок по химии и среднего балла аттестата с баллами ЦТ, то есть абитуриенты и на тестировании по химии не подтверждают свои аттестационные отметки. Среди 24 медалистов только 4 человека получили на ЦТ по химии отличные баллы.

Прогностическая валидность объединённого индекса успеха (ОИУ) также выше, чем таковая для любой его составляющей, — коэффициент корреляции ОИУ с успеваемостью по итогам первого курса составляет 0,695, то есть связь является сильной.

Заключение

Обобщая всё вышеизложенное, можно сделать следующие выводы:

1. Прогностическая валидность абитуриентского централизованного тестирования по поводу успешности обучения студентов в НГСХА выше, чем у школьного аттестата для теста по химии, и равна ей для теста по биологии.
2. Выявлена слабая связь первичного тестового балла и тестового балла централизованного тестирования со школьной отметкой по данному предмету, выставленной в аттестате об окончании полной средней школы.
3. Объединение баллов тестирования и оценок школьного аттестата (объединённый индекс успеха) повышает эффективность прогноза успеваемости в вузе.

Литература

1. *Анастаси А., Урбина С.* Психологическое тестирование. СПб.: Питер, 2003.
2. *Гончарова Г.А.* Сравнительный анализ традиционных и тестовых форм контроля знаний // Развитие системы тестирования в России. Тезисы докладов III Всероссийской научно-методической конференции / Под ред. Л.С. Гребнева. М.: Центр тестирования МО РФ, 2001. С. 193.
3. *Дятлова К.Д., Гапонова С.А., Михалёва Т.Г.* Связь успешности выполнения тестов с показателями умственного развития абитуриентов // Стандарты и мониторинг в образовании. 2004. № 3.
4. *Кытманов А.М., Василенко Е.П., Лейнартас Е.К., Лукин В.Н.* Исследование прогностической валидности теста по математике // Вопросы тестирования в образовании. 2002. № 4. С. 128–134.
5. *Лакин Г.Ф.* Биометрия: Учебное пособие. М.: Высшая школа, 1980.
6. *Мальцев А.В.* и др. Сравнение различных процедур отбора абитуриентов и их прогностическая валидность // Развитие тестовых технологий в России. М.: Центр тестирования МО РФ, 2002. С. 194–195.

7. *Майоров А.Н.* Теория и практика создания тестов для системы образования. М.: Народное образование, 2000.
8. *Нейман Ю.М., Хлебников В.А.* Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей, 2000.
9. Об утверждении порядка приёма в государственные образовательные учреждения высшего профессионального образования (высшие учебные заведения) Российской Федерации, учреждённые федеральными органами исполнительной власти. Приказ Министерства образования Российской Федерации № 50 от 14 января 2003 г.
10. *Переверзев В.Ю.* Критериально-ориентированное педагогическое тестирование: Учеб. пособие. М.: Логос, 2003.
11. *Равен Дж.* Педагогическое тестирование: проблемы, заблуждения, перспективы. М.: Когнито-центр, 1999.
12. *Соколов В.М.* Роль и место тестов достижений в диагностике качества образования/ Вестник Нижегородского университета им. Н.И. Лобачевского. Серия «Инновации в образовании». Вып. 1. Н. Новгород: Изд-во ННГУ, 2000. С. 166–168.
13. *Турпищева М.С., Дербенёв Н.А.* Анализ результатов тестирования выборочных групп для оценки объективности тестовых методов контроля // Развитие тестовых технологий в России. М.: Центр тестирования МОРФ, 2002. С. 214.
14. *Хлебников В.А., Михалёва Т.Г.* Централизованное тестирование в России: необходимость, возможность, проблемы // Школьные технологии. 1999. № 1–2.