

Закрытые тестовые задания по биологии

Дятлова Ксения Дмитриевна — доцент Нижегородской государственной сельскохозяйственной академии, директор Учебного центра, кандидат биологических наук;

Михалева Татьяна Григорьевна — руководитель научно-методического отдела Центра тестирования Министерства образования РФ, кандидат физико-математических наук.

Анализ статистических характеристик тестовых заданий (трудность, дискриминативность, точечно-бисериальная корреляция, анализ дистракторов, невыполнение заданий), полученных при апробации тестов, позволяет выделить удачные и неудачные задания и определить причину неудачи.

Тестовое задание (ТЗ) — это минимальная содержательная составляющая педагогического теста в виде проверочного задания специфической формы. Тестовые задания закрытого типа предусматривают готовые варианты ответа на поставленный вопрос. Чаще всего при тестовом контроле знаний используются закрытые ТЗ с множественным выбором ответов (ТЗ множественного выбора), предполагающие определённый набор ответов. Один из предлагаемых ответов правильный, остальные — похожи на правильный, но не являются таковыми. Обычно варианты неправильных ответов называют дистракторами [1]. К закрытым ТЗ множественного выбора предъявляются следующие требования: логическая форма высказывания, правильность формы, однозначность правильного ответа, краткость, наличие определённого места для ответов, правильность расположения элементов задания, идентичность правил оценки ответов, одинаковость инструкций для всех испытуемых, адекватность инструкции форме и содержанию задания и т.д.[1,6]. Структура такого закрытого тестового задания имеет следующий вид:

Инструкция: Выберите верный, по вашему мнению, ответ; обведите кружком букву, соответствующую варианту правильного ответа (в бланке ответов поставьте крестик, который соответствует букве вашего ответа)

Форма задания:

Вопрос (Утверждение в форме незаконченного предложения)

Варианты ответа:

- а. Вариант ответа 1
- б. Вариант ответа 2
- в. Вариант ответа 3
- г. Вариант ответа 4

Это основной вид заданий, применяемый в педагогических тестах. Эти задания, во-первых, высокотехнологичны (контроль знаний легко автоматизируется); во-вторых, предоставляют широкую возможность создавать унифицированные, параллельные по содержанию и сложности варианты тестов, в-третьих, их можно применить к любой дисциплине.

Каждое ТЗ имеет качественные и количественные характеристики. К первым относятся: элемент содержания, тип знания-умения и уровень усвоения, контролируемые этим заданием, а также форма и вид ТЗ в соответствии с общепринятой их классификацией [1,4]. Характеристики тестовых заданий, включённых в конкретный педагогический тест, зависят от цели тестирования и задаются в спецификации теста. Соответствие реальных характеристик ТЗ заданным в спецификации определяется путём экспертизы. Количественные характеристики тестовых заданий определяются эмпирическим путём.

При составлении любого теста первоначальный вариант, априорно соответствующий спецификации, но подлежащий апробации и последующей коррекции, называется предтестом (предварительным или «пилотным» тестом). Его элементы — предтестовые задания (ПТЗ). ПТЗ — проверочное задание, содержание, логическая структура и форма представления которого удовлетворяют ряду специфических требований. После экспертизы и эмпирической проверки ПТЗ, удовлетворяющие предъявленным требованиям, превращаются в ТЗ. ПТЗ, имеющие по результатам апробации неудовлетворительные статистические характери-

стики, корректируются по форме, содержанию (либо по тому и другому) или удаляются из теста.

Если тест предназначен, чтобы сравнивать учащихся по уровню подготовки (дифференциации уровня учебных достижений), то интерпретация результатов статистической обработки проводится с позиции нормативно-ориентированного подхода. При таком подходе анализ статистических результатов выполнения заданий должен дать ответ на вопрос, соответствует или нет то или иное задание цели тестирования. Такой ответ можно получить, анализируя систему статистических показателей (норм), отражающих результаты выполнения заданий теста представительной выборкой учащихся.

Если педагогический тест используется для ранжирования тестируемых по уровню их подготовки и отбора лучших, то такой тест, являясь по сути нормативно-ориентированным, не требует разработки норм [4] и, следовательно, репрезентативной выборки при апробации. Именно с такой целью проводится абитуриентское централизованное тестирование. Это добровольное испытание, и определить репрезентативную выборку для апробации нереально. В связи с этим апробация проводилась, чтобы определить некачественные задания с позиции общих требований к статистическим показателям ТЗ нормативно-ориентированных тестов. В апробации участвовали учащиеся 11-х классов с разным уровнем подготовки.

По классической модели интерпретации результатов, к числу наиболее важных количественных показателей ТЗ множественного выбора относятся **трудность задания, дифференцирующая сила, коэффициенты точноно-бисериальной корреляции, правдоподобность дистракторов** [6].

В представленной работе проанализировано качество закрытых ТЗ множественного выбора по биологии на основе статистического подхода. Материалом послужили данные о выполнении учащимися тестов по биологии в ходе апробации пилотных тестов абитуриентского централизованного тестирования 2002 г. (вариант 1: число участников — 115 чел.; вариант 2 — 91 чел.) и абитуриентского централизованного тестирования 2001 г. (вар. 1 — 1633 чел. и вар. 5 — 1560 чел. — регионы России) [2]. Тесты включали 45 закрытых ТЗ множественного выбора, относящихся ко всем разделам биологии.

Трудность задания

Трудность задания определяется процентом учащихся, верно выполнивших задание. Наиболее эффективным считают задание с параметром трудности 50%. С помощью такого задания в наибольшей степени позволяет разделять учащихся на тех, кто владеет определённым знанием или умением, и тех, кто им не владеет. Если бы все задания теста имели трудность, близкую к 50%, то распределение баллов за выполнение теста имело бы максимальный разброс и приближалось к нормальному, а дифференцирующая способность теста была бы наивысшей. Подобный отбор заданий обеспечивает лучшую дифференциацию, чем в случае широкого разброса уровней трудности [6]. На практике сложно подобрать все задания, имеющие такую трудность, особенно если при отборе заданий необходимо обеспечить содержательную валидность теста. Поэтому для теста, цель которого разделить учащихся по уровню их подготовки, наилучшим считается отбор заданий с трудностью в интервале от 40 до 60%.

Чтобы поддержать мотивацию к работе, в тест иногда включают задания, которые хорошо выполняются большинством тестируемых. Задания, которые выполняет небольшой процент тестируемых, также могут быть включены в окончательный вариант теста, чтобы выявить более подготовленных абитуриентов.

Нормативно-ориентированный тест обычно содержит три группы заданий:

- ТЗ **малой** трудности (более 70% правильных ответов) — 20 % всех ТЗ;
- ТЗ **средней** трудности (25–69% правильных ответов) — 70% всех ТЗ;
- ТЗ **большой** трудности (менее 25% правильных ответов) — около 10 % ТЗ.

В спецификации анализируемых тестов ожидаемый уровень трудности декларировался

согласно этим критериям нормативно-ориентированного теста. После апробации был выявлен реальный уровень для каждого ТЗ. Общая картина распределения ТЗ по трудности представлена в табл. 1.

Таблица 1

Тест	Тестовые задания (%)		
	малой трудности	средней трудности	большой трудности
Пилотный, Вариант 1	13,3	64,5	22,2
Пилотный, Вариант 2	13,3	75,6	11,1
Абитуриентский, Вариант 1	8,9	80	11,1
Абитуриентский, Вариант 5	8,9	82,2	8,9

Все исследованные тесты отклонялись от заданных в спецификации параметров трудности, причём число трудных заданий было близким к заявленному, а простых было явно недостаточно. Во всех тестах преобладали ТЗ средней сложности.

При анализе заданий следует учитывать, что значение трудности зависит от основных характеристик выборки тестируемых, т.е. может меняться при изменении размера и состава выборки (уровня подготовленности испытуемых).

Статистические результаты апробации заданий свидетельствуют, что параметр трудности задания не главный критерий для отбора задания в нормативно-ориентированный тест. Более важный показатель для теста данного типа — дифференцирующая способность задания.

Дифференцирующая способность задания

Дифференцирующая способность (дифференцирующая сила, дискриминативность, индекс дискриминативности) задания рассчитывается как разность двух значений: трудности задания для сильной группы тестируемых, составляющих 27% от их общего числа, и трудности задания для слабой группы тестируемых (также 27%).

$$D = p_{\text{сильн.}} - p_{\text{слаб.}}$$

Этот показатель может быть выражен в процентах или в десятичных дробях. Чем выше дифференцирующая способность задания, тем чётче оно разделяет учащихся по уровню подготовки, а значит, тест, цель которого — разделить учащихся по уровню их знаний и умений, более валидный и надёжный.

Таблица 2

Критерии оценки дифференцирующей способности

Дифференцирующая способность (%)	Комментарии	Оценка
41–100	Очень хорошая	Отличное задание для оценки различия в подготовке тестируемых
31–40	Хорошая	Хорошее задание для дальнейшего использования
21–30	Средняя	Желательна дальнейшая доработка задания
11–20	Низкая	Критическое значение
–10*–10	Очень низкая	Задание не следует использовать

* Отрицательное значение дифференцирующей способности может указывать на ошибки в заданиях, например, возможна ошибка в правильном ответе.

В нормативно-ориентированном тесте дифференцирующая способность большинства ТЗ должна превышать 30% (0,3). Однако для небольшого числа лёгких заданий (трудность выше 70%) допускается значение дифференцирующей способности 20% (0,2).

Анализ полученных в ходе статистической обработки данных показал, что дискриминативность ТЗ абитуриентского тестирования была значительно выше, чем ТЗ пилотных тестов. Возможно, различия объясняются разными размерами выборки протестированных учащихся, а также низкой мотивацией последних при апробации пилотных тестов.

Во всех случаях **низкая дифференцирующая способность** ТЗ рассматриваемых тестов может быть обусловлена несколькими причинами:

Во-первых, ТЗ может быть плохо сформулировано и учащиеся его не понимают.

*Пример 1 (1633чел.)**. Дифференцирующая способность задания 3,5%

* В скобках указано количество учащихся, выполнивших данное ТЗ, правильный ответ выделен курсивом.

В лесах средней полосы России преобладают:

а) *ель европейская, сосна обыкновенная, берёза бородавчатая**;

* Здесь и далее правильный ответ выделен курсивом.

б) клён остролистный, липа сердцелистная, сосна обыкновенная;

в) дуб черешчатый, рябина обыкновенная, берёза бородавчатая;

г) клён остролистный, дуб черешчатый, ель европейская.

Неудачными следует признать выражения «преобладают» (в ТЗ не рекомендуется применять обороты типа «главным образом», «преимущественно», «всегда», «никогда» [3, 6] и «средняя полоса России» в силу неопределённости понятий).

Во-вторых, ТЗ может не содержать правильного ответа, что, естественно, вызывает недоумение и, как следствие, угадывание:

Пример 2 (115 чел.). Дифференцирующая способность задания 10,2%.

Пояс верхних конечностей у земноводных:

а) отсутствует;

б) включает 2 лопатки и 2 ключицы;

в) включает 2 лопатки, 2 ключицы, 2 вороньи кости, 2 грудины;

г) включает 2 лопатки, 2 ключицы, 2 вороньи кости и грудину.

ТЗ содержит типичную ошибку, репродуцируемую во многих учебных пособиях (грудина не входит в состав плечевого пояса конечностей).

В-третьих, ТЗ могут содержать **два верных ответа**:

Пример 3 (115 чел.). Дифференцирующая способность задания 2,7%.

Тело грибов:

а) разделено на органы (шляпка и ножка);

б) *представлено слоевищем;*

в) *представлено мицелием;*

г) представлено микоризой.

Термин «слоевище» (ответ б) в отношении грибов применяется редко, но этот ответ правильный, так же, как и предусмотренный автором ответ «мицелий». 53,9% учащихся выбрали не предусмотренный автором, но верный ответ (б), 21,7% — верный и предполагаемый автором (в).

В-четвёртых, ТЗ могут быть **слишком сложны** для данного контингента учащихся — трудность менее 25%. При этом учащимся может быть неизвестен содержательный элемент ТЗ — требуемые темы были не пройдены или не усвоены в надлежащем объёме — или не доступен требуемый когнитивный уровень. В случае большой трудности ТЗ сильная и слабая группы действуют методом угадывания или слабая группа угадывает (вероятность правильного ответа в задании с четырьмя альтернативами — 25%), а сильная пытается ответить на вопрос, но ошибается. В последнем случае дискриминативность даже может иметь отрицательные значения (см.: *Пример 5*).

Пример 4 (1560 чел.). Дифференцирующая способность задания 0,6%.

Корневищами в природе размножаются:

а) брусника, черника, лапчатка;

б) лапчатка, ландыш, пырей;

в) *черника, брусника, ландыш;*

г) лапчатка, пырей, брусника.

Это ТЗ на мелкие фактуальные знания, выходящие за рамки школьного учебника или не сфокусировавшие на себе внимания учащихся. Трудность задания — 22,5%. В сильной группе с ТЗ справились 21,9% учащихся, в слабой — 21,3% (процент выполнения в обеих группах близок к 25 — угадыванию из четырёх альтернатив, то есть все учащиеся работали «методом тыка»).

Пример 5 (91 чел.). Дифференцирующая способность задания 12,3%.

У белой планарии выделение происходит через:

- а) рот;
- б) сократительные вакуоли;
- в) протонефридии;
- г) метанефридии.

Трудность этого задания — 12,0%. Правильный ответ требует знаний, изложенных в Пособиях для поступающих в вузы, но не в школьном учебнике. В сильной группе с ТЗ справились 6,9% учащихся, в слабой — 19,2 (процент выполнения в слабой группе близок к 25).

Пример 6 (115 чел.). Дифференцирующая способность задания 7,2%.

Черепаша и тритон принадлежат к одному типу,

- а) но разным подтипам;
- б) одному подтипу, но разным классам;
- в) подтипу и классу, но разным отрядам;
- г) подтипу, классу и отряду.

Это ТЗ на классификационные знания и умение рассуждать. Нетрадиционная постановка вопроса позволяет отнести ТЗ к заданиям третьего (творческого) уровня усвоения учебного материала. Трудность задания — 30,4%. В сильной группе с ТЗ справились 37,9% учащихся, в слабой — 30,8%, у обеих групп вероятность угадывания близка к 33,3%, какая получается при угадывании из трёх альтернатив — последний дистрактор большинством отвергался как мало вероятный (его выбирали лишь 8,7%).

В-пятых, ТЗ могут быть **слишком просты** для данного контингента учащихся — трудность ТЗ близка к 80%. В этом случае и слабые, и сильные учащиеся отвечают одинаково успешно и разница между группами минимальна.

Пример 7 (115 чел.). Дифференцирующая способность задания 8,1%.

К биологическим факторам антропогенеза НЕ относится:

- а) труд;
- б) естественный отбор;
- в) борьба за существование;
- г) наследственная изменчивость.

Трудность этого задания — 78,3%. В сильной группе с ТЗ справились 78,4% учащихся, в слабой — 70,1%.

Точечно-бисериальная корреляция (ТБК)

Точечно-бисериальная корреляция позволяет ответить на вопрос, как связано выполнение конкретного задания с выполнением всего теста. Числовые значения точечно-бисериального коэффициента (ТБК) могут меняться от 1 до -1. (Формула расчёта ТБК приводится в [6]). **Положительная корреляция для верного ответа** показывает, что учащиеся, хорошо справившиеся с тестом и получившие достаточно высокий тестовый балл, правильно выполняют это задание, т.е. выбирают правильный ответ. **Отрицательная или нулевая корреляция для верного ответа** отражает случайный характер ответов учащихся, наличие каких-либо проблем в усвоении проверяемого материала либо некорректную формулировку ТЗ.

Можно установить следующий критерий отбора заданий по этому показателю: отбираются задания со значением точечно-бисериального коэффициента для правильного ответа, превышающим критическое значение 0,20. Хорошим значением точечно-бисериального коэффициента считается интервал 0,3–0,5. Интервал 0,2–0,3 считается критическим. При анализе заданий обращается внимание также на **точечно-бисериальные коэффициенты для дистракторов**. Их значения должны быть отрицательными и превышающими по модулю 0,1–0,2. Положительное или близкое к нулю значение точечно-бисериального коэффициента для неверного ответа говорит о наличии проблем с данным вариантом ответа.

Пример 8 (115 чел.). ТБК для верного ответа –0,0; для дистракторов –0,0; 0,1; –0,1.

В селекции животных только для пчёл применяют:

- а) полиплоидию;
- б) гетерозис;
- в) искусственный мутагенез;
- г) движущий отбор.

Это ТЗ на системные знания (требует глубокого понимания нескольких тем общей биологии) и умение рассуждать, явно превосходит возможности исследуемого контингента. Все учащиеся действовали методом свободного угадывания.

Пример 9 (115 чел.). ТБК для верного ответа –0,1; для дистракторов –0,0; 0,0; –0,1.

Правило единообразия гибридов первого поколения справедливо:

- а) только при моногибридном скрещивании и полном доминировании;
- б) при ди- и полигибридном скрещивании и полном доминировании;
- в) только при неполном доминировании;
- г) всегда.

Это ТЗ требует знания довольно сложного раздела, громоздко по формулировке. Варианты ответа не равнозначны по форме и смыслу (а и б оперируют двумя признаками, в — одним, г — неизвестным числом признаков). Все учащиеся плохо понимали задание и действовали методом угадывания.

Пример 10 (1560 чел.). ТБК для верного ответа 0,0; для дистракторов 0,3; –0,2; –0,2.

Шейный отдел, состоящий из 7 позвонков, имеют млекопитающие, за исключением:

- а) утконосов и ехидн;
- б) кенгуру и опоссумов;
- в) ежей и кротов;
- г) ленивцев и ламантинов.

Те же тип и уровень, что и в *Примере 4*. Все учащиеся действовали методом свободного угадывания.

Анализ распределения ответов учащихся по предложенным вариантам ответа

Анализ распределения выбора ответа из предложенных вариантов позволяет оценить, как «работает» каждый из них. Анализ проводился путём сравнения процента учащихся, выбравших каждый предложенный ответ. В хороших заданиях верный ответ по этому показателю имеет чётко выраженный максимум, а неверные ответы учащихся распределяются между всеми предложенными дистракторами приблизительно одинаково. «Хороший» неверный ответ может быть выбран большим числом слабых учащихся. Если неверный ответ выбирают сильные учащиеся, то его рекомендуется проанализировать и заменить. Возможно, что он также верный или частично верный [6].

Ответ, который выбирают **менее 5%** тестируемых, представляется явно неверным для всей выборки испытуемых, и его рекомендуется заменить на более привлекательный («более верный ответ») для слабой группы тестируемых.

Заметим, что наличие неудачных дистракторов (ответов, которые выбирают менее 5% тестируемых) может не только не уменьшать дифференцирующую силу, но и увеличивать её за счёт лёгкости выбора из 2–3 оставшихся альтернатив для сильных учеников (см. Примеры 14, 15). При этом необходимо отдавать себе отчёт, что ТЗ становится менее сложным по сравнению с другими.

Пример 11 (91 чел.). Неудачные дистракторы (в и г) выбирает 1,1% и 2,2% учащихся; дифференцирующая способность задания 23,1%

В световой фазе фотосинтеза происходит превращение энергии:

- а) световой в механическую;
- б) световой в химическую;

- в) химической в химическую;
- г) химической в механическую.

Формулировка вопроса содержит подсказку «световой», поэтому выбор делается всего из двух вариантов (а и б), где есть эта подсказка. Дистракторы, не содержащие подсказки, отвергаются.

Пример 12 (91 чел.). Неудачный дистрактор (г) выбирает 0%; дифференцирующая способность задания 12,1%.

У рыб кровеносная система:

- а) имеет один круг кровообращения, сердце двухкамерное;
- б) имеет два круга кровообращения, сердце трёхкамерное;
- в) имеет два круга кровообращения, сердце двухкамерное;
- г) такая же, как у ланцетника.

Дистрактор (г) вопиюще неудачный — он привлекает к себе внимание совершенно иной, чем остальные ответы, формулировкой и явно не является правдоподобным по содержанию.

Пример 14 (1560 чел.) Неудачные дистракторы (в и г) выбирают 4 и 5%; дифференцирующая способность задания 48,7%.

Слюна в организме человека участвует в:

- а) смачивании пищи и расщеплении углеводов;
- б) смачивании пищи и гуморальной регуляции пищеварения;
- в) гуморальной регуляции пищеварения и расщеплении углеводов;
- г) выработке антител и синтезе пищеварительных ферментов.

Неудачные дистракторы отвергаются, исходя из их малого правдоподобия. Сильная группа отвечает верно (90,6%), слабая свободно выбирает из двух оставшихся альтернатив (41,9%).

Пример 15 (1560 чел.). Неудачный дистрактор (б) выбирает 1%; дифференцирующая способность задания 54,3%.

Одной из основных причин истощения озонового слоя в атмосфере является:

- а) накопление фреонов;
- б) увеличение содержания кислорода;
- в) увеличение содержания углекислого газа;
- г) вулканы, пыльные бури, лесные пожары.

Очевидно, большинству учащихся известно, что озон — аллотропная форма кислорода. Сильная группа отвечает верно (81,7%), слабая свободно выбирает из трёх оставшихся вариантов (27,3%).

Для очень простых заданий (трудность более 75%) хорошие дистракторы могут иметь малый процент выбора, поскольку на их долю достаётся не более 25%, но они должны распределяться равномерно (Пример 16).

Пример 16 (1560 чел.) Трудность ТЗ 83,5%; % выбора дистракторов: 6, 5, 4; дифференцирующая способность задания 30,5%.

Использование живых организмов и биологических процессов в производстве продуктов питания — это:

- а) клеточная инженерия;
- б) генная инженерия;
- в) мутагенез;
- г) биотехнология.

Сочетание статистических показателей

Опыт показал, что неудачные ТЗ при апробации обнаруживают «плохие» статистические показатели сразу по двум или трём параметрам: индексу дискриминативности, точечно-бисериальным коэффициентам и распределению ответов (Табл. 3 — Примеры 17, 18). В этом

случае выбор ответов носит случайный характер, что обычно отражается на дискриминативности.

Таблица 3

Статистические показатели неудачных тестовых заданий по биологии

№ примера	Распределение ответов, %				ТБК				Трудность, %		индекс дискриминатив. (%)
	а. альт.	б. альт.	в. альт.	г. альт.	а. альт.	б. альт.	в. альт.	г. альт.	Слабая группа	Сильная группа	
17	4,3	35,7	54,8	2,6	0	0,1	0	0	54,1	45,9	-8,1
18	30	18	33	18	0,2	-0,1	-0,2	0,1	16,6	25,8	9,2
19	20,9	56,5	19,1	0,9	-0,2	0,4	-0,2	0	35,1	78,4	43,2
20	4,3	40,9	47	7,8	-0,2	0,2	0	-0,2	27	51,4	24,3

Пример 17 (115 чел.)

Для всех хордовых характерно наличие в течение всей жизни или только в зародышевой стадии:

- а) сердца;
- б) позвоночника;
- в) жаберных щелей в стенках глотки;
- г) черепа.

Правильный ответ на это ТЗ требует знаний, изложенных в Пособиях для поступающих в вузы, но не в школьном учебнике. ТЗ плохо сформулировано — смысл ускользает. Все учащиеся действовали методом свободного угадывания из двух альтернатив: дистракторы а и г не рассматривались как маловероятные.

Пример 18 (1633 чел.)

Цветок с простым околоцветником имеет:

- а) только венчик;
- б) только чашечку;
- в) венчик и чашечку;
- г) одинаковые листочки.

ТЗ сформулировано некорректно: альтернативы (а) и (б) могут считаться правильными ответами: «венчиковидный простой околоцветник» или «чашечковидный простой околоцветник».

Результирующим показателем, на который следует опираться в первую очередь, может служить индекс дискриминативности: если он мал, ТЗ должно отвергаться при тестировании аналогичного контингента учащихся. В то же время высокая дифференцирующая способность ещё не гарантирует высокого качества ТЗ. Возможно, она достигнута за счёт уменьшения числа альтернатив из-за плохо подобранных дистракторов (*Примеры 19, 20*).

Пример 19 (92 чел.)

Биологическими полимерами в клетке являются:

- а) только белки;
- б) белки и нуклеиновые кислоты;
- в) белки и АТФ;
- г) нуклеиновые кислоты и глюкоза.

Дистрактор (г) явно неправдоподобен, даже самые слабые ученики понимают это — он отвергается всеми (0,9% выбора). 78,4 % сильных учеников верно отвечают на вопрос. 35,1% слабых отвечают верно, угадывая ответ из трёх вариантов.

Пример 20 (115 чел.)

Трёхраздельная конечность впервые в процессе эволюции появилась у:

- а) рыб;
- б) амфибий;
- в) рептилий;
- г) птиц.

Дистрактор (а) выбирают 4,3% всех тестируемых, дистрактор (г) — 7,8%. Очевидно, сильные ученики отвергают эти неправдоподобные дистракторы и угадывают из двух оставшихся альтернатив — 51,4% правильных ответов. Слабые ученики почти не отвергают

неудачные дистракторы и угадывают из 4 вариантов — 27% верных ответов.

Невыполнение задания

При анализе задания рекомендуется обратить внимание на число (%) тестируемых, которые по каким-либо причинам не выполнили задание: пропустили его или не успели выполнить. Если это число превышает 5%, то важно понять причину невыполнения (не знают правильный ответ; задание пропущено в связи с нежеланием выполнять его, например, более трудоёмкое задание со свободным ответом; или задание расположено в конце теста и не хватило времени на его выполнение) [6]. В рассматриваемых тестах отсутствовали ТЗ, выполнять которые отказалось более 5% учащихся, однако 4,3% отказались выполнять стоящее в конце пилотного теста трудное и неудовлетворительное по всем статистическим показателям ТЗ (см. *Пример 8*).

Удачные тестовые задания

Удачными, достойными получить статус ТЗ и войти в банк тестовых заданий могут считаться задания, имеющие все статистические показатели, удовлетворяющие требованиям: трудность 40–60%; дискриминативность больше 30%; точно-бисериальный коэффициент правильного ответа 0,3–0,5 и дистракторов –0,1–0,2; относительная равномерность выбора дистракторов; выполнение большинством испытуемых. Заметим, что параметр трудности может быть иным, исходя из целей теста.

В рассмотренных абитуриентских и пилотных тестах большое количество ТЗ можно считать удачными.

Пример 21 (115 чел.)

Реакция среды щелочная в:

- а) ротовой полости и желудке;
- б) желудке и двенадцатипёрстной кишке;
- в) ротовой полости и двенадцатипёрстной кишке;
- г) желудке.

Пример 22 (1633 чел.)

Из почечной лоханки вторичная моча непосредственно поступает в:

- а) мочеточник;
- б) собирательную почечную трубочку;
- в) извитой почечный каналец;
- г) мочеиспускательный канал.

Пример 23 (115 чел.)

Найдите соответствие:

Компоненты биоценоза:	Организмы:
1. Продуценты	А. Травы
2. Консументы	Б. Зайцы
3. Редуценты	В. Бактерии и грибы-сапротрофы (сапрофиты)
	Г. Деревья
	Д. Кустарники

- а) 1А, Г, Д; 2Б; 3В;
- б) 1А, Б; 2В, Г; 3В;
- в) 1Б; 2А, Г, Д; 3В;
- г) 1Б; 2А, В, Г; 3Б.

Пример 24 (1633 чел.)

Эпителий альвеол лёгких формируется из:

- а) эктодермы;
- б) энтодермы;
- в) эпидермы;
- г) мезодермы.

Пример 25 (115 чел.)

Более высокая жизнеспособность и продуктивность гибридов первого поколения — это проявление:

- а) точечных мутаций;
- б) гетерозиса;
- в) полиплоидии;
- г) модификации.

Пример 26 (1560 чел.)

В состав нефрона НЕ входит.

- а) почечная пирамида;
- б) почечный каналец;
- в) капсула Шумлянского-Боумена;
- г) капиллярный (мальпигиев) клубочек.

Заметим, что указанные здесь удачные, с нашей точки зрения, ТЗ составлены в соответствии с принципами конструирования ТЗ В.С. Аванесова [1], лаконичны по форме и относятся (за исключением *Примера 24*) к фактуальным знаниям первого уровня трудности (репродукция).

Таблица 4

Статистические показатели неудачных тестовых заданий по биологии

№ примера	Распределение ответов, %				ТБК				Трудность, %	индекс дискриминатив. (%)
	а. альт.	б. альт.	в. альт.	г. альт.	а. альт.	б. альт.	в. альт.	г. альт.		
21	23,9	20,7	34,8	19,6	-0,2	-0,2	0,5	-0,2	34,8	61,3
22	54	20	11	14	0,4	-0,1	-0,1	-0,3	53,6	49,8
23	62	15,2	12	9,8	0,4	-0,2	-0,1	-0,2	62	48,1
24	16	45	20	18	-0,1	0,4	-0,3	-0,1	44,8	41,4
25	33	15,7	15,7	34,8	-0,2	0,5	-0,1	-0,1	15,7	35,1
26	37	13	26	22	0,5	-0,1	-0,3	-0,1	36,8	59,5

Заключение

Предтестовые задания пилотных тестов, неудовлетворительные по одному или нескольким статистическим показателям, кардинально переработаны или исключены из тестов. Неудачные ТЗ абитуриентского тестирования не должны включаться в формирующийся банк тестовых заданий по предмету. При разработке ТЗ авторы должны учитывать отмеченные недостатки. Удовлетворительные по статистическим (количественным) и «откалиброванные» по качественным (элемент содержания, форма, вид и уровень контролируемых знаний и умений) тестовые задания могут служить основой для формирования банка ТЗ и компоновки тестов любого назначения [5,3]. Следующий этап анализа материалов, полученных в ходе статистической обработки результатов тестирования — определение разделов содержания предмета, которые вызывают затруднения учащихся, чтобы выявить причины устойчивых ошибок.

Литература

1. *Аванесов С.В.* Композиция тестовых заданий. Учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов. М.: Адепт, 1998.
2. Биология: Тесты для 11-го кл. Варианты и ответы централизованного тестирования. М.: Центр тестирования МО РФ, 2001.
3. *Ван Крикен Р., Баккер С.* Подготовка к проведению экзаменов. Руководство для организации и разработки централизованных экзаменов// Стандарты и мониторинг в образовании. 1998, № 3.
4. *Майоров А.Н.* Теория и практика создания тестов для системы образования. М.: Народное образование, 2000.
5. *Михалёва Т.Г., Хлебников В.С.* Методические принципы формирования единого банка

тестовых заданий/ Тез. докл. Второй Всероссийской конференции «Развитие системы тестирования в России» 23–24 ноября. Часть 3. М.: Прометей, 2000.

6. *Чельщикова М.Б.* Теория и практика конструирования педагогических тестов. Учебное пособие. М.: Исследовательский центр проблем качества подготовки специалистов, 2001.