

**В.Ю. БОДРЯКОВ,  
Н.Г. ФОМИНА**  
*Уральский государственный  
педагогический университет*

## **Простая вероятностно-статистическая модель количественной оценки уровня знаний учащихся**

*Описывается простая вероятностно-статистическая модель количественной оценки уровня знаний учащихся при проведении как письменного (тестирование), так и устного экзаменов. Официальная итоговая оценка в общепринятой системе «2» — «3» — «4» — «5» определяется путем перевода набранной учащимся суммы баллов в собственно оценку по одной из предложенных схем.*

### **Традиционный номенклатурный подход к оценке знаний**

Не вызывает сомнений актуальность комплексной проблемы адекватной количественной оценки качества образовательного процесса в любом образовательном учреждении (ОУ), будь то школа, колледж, вуз. Такая оценка связана с последующей обработкой и интерпретацией результатов наблюдений, выработкой плана необходимых корректирующих мероприятий и отслеживания их эффективности [1—4]. Отметим, что изложенная последовательность операций составляет основу мониторинга качества учебного процесса. Эффективность его проведения основывается на использовании современного аппарата математической статистики [5, 6] для обработки информации на всех этапах мониторинга [7—10].

Особую важность приобретает самостоятельная проблема адекватной количественной оценки знаний учащихся (студентов) как в течение семестра (текущий контроль), так и при итоговом контроле, например в ходе экзаменационных сессий. Если в текущем контроле могут при-

меняться самые разнообразные, в том числе и неформальные его виды, то официальная оценка знаний студентов во время экзаменационной сессии имеет детально регламентированный формальный характер. Знания учащихся по экзаменуемой дисциплине должны быть оценены одной из четырех оценок: «2» (неудовлетворительно), «3» (удовлетворительно), «4» (хорошо), «5» (отлично). При таком предельно ограниченном наборе значений признака (рангов) без применения достаточно тонких статистических технологий просто не обойтись. Например, при оценке статистической значимости изменений, произошедших в результате тех или иных педагогических воздействий.

Традиционно применяемый ответственными специалистами по качеству образования анализ с заключениями вида «стало лучше», «стало хуже» по сравнению с «аналогичным периодом прошлого года» вряд ли может быть признан удовлетворительным в смысле содержательного количественного анализа качества учебного процесса.

Ряд авторов считают, что «... оценивание знаний учащихся с помощью знаков «5», «4», «3», «2» и т.п. представляет собой не количественный, а качественный (описательный) способ

оценивания знаний»<sup>1</sup>. Придерживается того же мнения А.И. Орлов, говоря об оценивании учащихся. Он отмечает, что «вряд ли кто-либо будет утверждать, что знания отличника равны сумме знаний двоечника и троечника (хотя  $5 = 3 + 2$ ), хорошист соответствует двум двоечникам ( $2 + 2 = 4$ ), а между отличником и троечником такая же разница, как между хорошистом и двоечником ( $5 - 3 = 4 - 2$ ). Поэтому очевидно, что для анализа подобного рода качественных данных необходима не всем известная арифметика, а другая теория...»<sup>2</sup>.

По сути, при таком традиционном подходе (назовем его номенклатурным) экзаменатор, выступающий в роли эксперта, непосредственно относит экзаменуемого к качественной номенклатурной категории «отличников», «хорошистов», «троечников» и «двоечников», что и символизируется оценками «5» — «4» — «3» — «2». Помимо общего субъективизма такого подхода, к его недостаткам можно отнести то, что часто эти категории так прочно «закрепляются» за конкретными учащимися, что переход из одной категории в другую почти невозможен. «Троечнику» готовы поставить удовлетворительную оценку, а отличнику — отличную, хотя бы просто полистав зачетную книжку (взглянув в аттестационный лист), даже если первый объективно показывает уверенно хорошие знания по экзаменуемой дисциплине, а отличник едва готов к экзамену. При такой ситуации неудивительно, например, что ежегодно за последние пять лет 60—62% выпускников-медалистов страны не подтверждают свои отличные знания при поступлении в вузы [4]. Несомненно, что одной из причин недостаточной объективности аттестации знаний учащихся является неколичественный характер оценивания этих знаний.

Знания учащихся можно (и нужно!) оценивать по объективной количественной шкале, и лишь затем набранные ими баллы согласно заранее определенной процедуре переводить (округлять) в привычные «5» — «4» — «3» — «2», т.е. «отлично» — «хорошо» — «удовлетворительно» — «неудовлетворительно».

<sup>1</sup> Поташник М.М. Требования к современному уроку. М.: Центр педагогического образования, 2008. — С. 209.

<sup>2</sup> Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — С. 26.

## Функции традиционной номенклатурной оценки знаний

Заметим, что традиционная номенклатурная оценка знаний учащихся не слишком хорошо выполняет такие основные функции, как:

- стимулирующая;
- воспитательная;
- ориентирующая;
- организационная;
- контролирующая;
- диагностическая.

Считается, что *стимулирующая функция* оценки является наиболее важной, так как побуждает личность к изменениям, оказывает воздействие на ее дальнейшую деятельность. В связи с этим педагогическая оценка выполняет воспитательную функцию. Между тем не секрет, что в силу большой субъективности в номенклатурном подходе превалирует не столько оценка собственно знаний учащегося, сколько категоризация его личности. Поэтому такая оценка зачастую ни к чему не побуждает учащихся, а подчас играет и прямо антистимулирующую и антивоспитательную роль.

*Ориентирующая функция* проявляется, когда оценка выступает как определение уровня достижений обучаемых. Хотя в номенклатурном подходе говорить о таком уровне можно лишь условно в силу его качественного характера к оцениванию знаний.

Суть *организационной функции* оценки состоит в том, что экзаменатор (проверяющий) руководствуется установленными (квалификационными) требованиями, предъявляемыми к уровню образованности (подготовки). Качественный характер номенклатурного подхода к оцениванию знаний не позволяет принять объективные количественные критерии достижения учащимся соответствующего уровня подготовки по дисциплине.

*Контролирующая функция* обеспечивает целевое назначение педагогической оценки — управление качеством образования.

*Диагностическая функция* педагогической оценки характеризует степень (уровень) освоения обучаемым образовательной программы на

конкретный момент и выявляет причины его незнания. Однако диагностика причин проходит не слишком эффективно, так как номенклатурная оценка имеет субъективный и качественный характер.

Таким образом, традиционная номенклатурная оценка знаний не может являться объективной, а сама процедура проверки — прозрачной. Поэтому мы предлагаем перейти к количественному оцениванию знаний учащихся на основе построенной вероятностно-статической модели с последующей интерпретацией полученных результатов.

## Количественный подход к оценке знаний

Простая (в смысле общедоступности ее реализации в учебном процессе) вероятностно-статистическая модель (ВСМ) предполагает, что полученные количественные предварительные оценки знаний суть обычные числа — элементы некоторого ограниченного множества. Их можно складывать и вычитать по обычным правилам арифметики. Количественный подход, основанный на ВСМ, по нашему мнению, лишен недостатков, присущих традиционному номенклатурному подходу к оцениванию знаний учащихся, а сформированная итоговая оценка может более эффективно выполнять присущие ей педагогические функции.

Для определенности будем говорить об оценке знаний студента в ходе экзаменационной сессии, хотя предложенная модель применима к любому тестированию знаний учащихся.

Ранее на примере достаточно объемного статистического анализа экзаменационных результатов студентов математического факультета УрГПУ нами было показано, что эмпирические распределения относительных частот оценок по профильным математическим дисциплинам удовлетворительно описываются нормальными частотами, даваемыми нормальным законом распределения  $N(a; \sigma^2; x) = N(3,5; 1,0; k)$  [7, 10]:

$$N(a; \sigma^2; k) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(k-a)^2}{2\sigma^2}\right\} \Delta k, \quad (1)$$

где  $k = 0, 1, 2, 3, 4, 5$  — количество баллов, набираемых студентами при ответах на пять оцениваемых блоков эк-

заменяемой дисциплины;  $\Delta k = 1$  — «ширина» оценочных интервалов. Это нормальное распределение оценок, устойчиво реализуемое для типичного экзамена, принимаемого опытными педагогами-экспертами, мы назвали *опорным, или базовым*.

При базовом распределении на долю (относительная частота) неудовлетворительных оценок (количество набранных баллов  $k = 0, 1, 2$ ) приходится примерно 15%, на долю удовлетворительных и хороших оценок приходится по 35% ( $k = 3, 4$ ), на долю отличных оценок приходится 13%.

Сопоставление полученных модельных вероятностных распределений оценок с нормальными кривыми вида (1) позволило предположить, что нормальное распределение оценок является частным отражением общей нормальности распределений учащихся по уровню интеллекта IQ и другим способностям, отмечаемых многими авторами [1, 2, 4]. Это хорошо видно из рис. 1, на котором приведены эмпирические распределения уровней интеллекта IQ подростков по Векслеру [4] и сессионных экзаменационных оценок по профильным математическим дисциплинам студентов—математиков УрГПУ. Тесная коррелированность приведенных распределений очевидна.

## Описание вероятностно-статистической модели оценки знаний

Примем следующую ВСМ оценки знаний студента экзаменатором. Будем считать, что весь экзаменуемый материал разбит на некоторое количество (обычно несколько десятков) примерно равных по сложности вопросов и содержит достаточно представительную базу (желательно не менее нескольких сотен) практических заданий (задач). Для простоты будем считать:

1. Задачи имеют приблизительно равный с теоретическими вопросами уровень сложности.
2. Теоретические вопросы заранее сообщаются студентам для подготовки.
3. База задач также доступна студентам, например, через компьютерную сеть ОУ.

Согласно модели, каждый экзаменуемый получает  $n$  примерно одинаковых по сложности

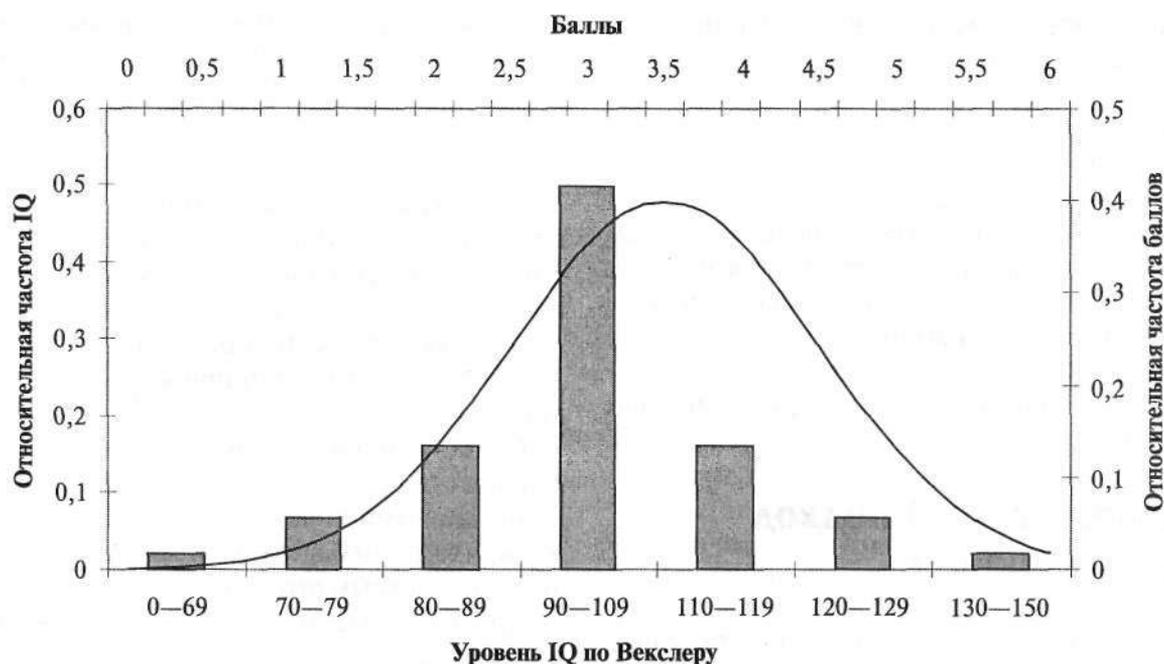


Рис. 1. Частотные распределения уровней интеллекта IQ подростков (гистограмма) и оценок знаний учащихся при типичном экзамене (сплошная кривая).

теоретических и практических заданиях по оцениваемой дисциплине, каждое из которых решает с вероятностью  $p$  (или не решает с вероятностью  $q = 1 - p$ ). Задания должны соответствовать требованиям ГОС ВПО. Фактически экзаменационное испытание проходит по схеме Бернулли [5, 6]. Для технического удобства (хотя это и не обязательно) выберем  $n = 10$ . При десяти экзаменационных заданиях достаточно полно и широко можно охватить весь учебный материал, что повысит объективность оценки знаний учащегося. Набранная экзаменуемым итоговая сумма баллов  $k = 0, 1, \dots, 9, 10$  подчиняется биномиальному закону распределения. Вероятность результата с набранными  $k$  баллами определяется законом биномиального распределения:

$$B_n(k) = C_n^k \cdot p^k \cdot q^{n-k} \quad (2)$$

Итак, экзаменационное испытание начинается с выдачи студентам  $n = 10$  экзаменационных заданий в виде подготовленных экзаменационных комплектов. Эти десять заданий должны включать теоретические и практические задания из различных частей учебного курса. Процедура формирования экзаменационных комплектов может быть организована как в традиционном

виде (уже сформированный набор заданий), так и со случайным, но контролируемым выбором заданий из компьютерной базы заданий персонально для каждого экзаменуемого с применением переносного персонального компьютера. Инновационный подход к реорганизации традиционного экзамена в вузе с применением переносного персонального компьютера подробно описан нами ранее [8]. Преимуществом «компьютерного» подхода к организации экзамена является более широкий охват курса, чем это возможно при ограниченном наборе заранее подготовленных билетов, большая объективность, а при необходимости, и «протоколизация» экзаменационного процесса. Экзамен можно проводить непосредственно в тестовой форме. Однако мы рассмотрим нетестовый вариант экзамена.

Экзамен может быть письменным. Тогда экзаменационные комплекты одновременно получают, начинают работать и сдают все студенты экзаменуемой группы. Экзамен может быть устным. Тогда экзаменационный процесс протекает в традиционной последовательности. После завершения подготовки студентов к ответу, начинается проверка сданных письменных экзаменационных работ или заслушивание устных ответов с одновременным подсчетом на-

бранных баллов. Появление той или иной оценки зависит от вероятности набора баллов, которые распределены по биномиальному закону (табл. 1).

Таблица 1

**Вероятность набора баллов по степени усвоения учебного материала**

Сумма баллов	Вероятность набора баллов по группам $B_n(k)^*$				
	Слабые знания		Средние знания	Хорошие знания	Сильные знания
$k$	$p = 0,5$ $q = 0,5$	$p = 0,6$ $q = 0,4$	$p = 0,7$ $q = 0,3$	$p = 0,8$ $q = 0,2$	$p = 0,9$ $q = 0,1$
0	0,0010	0,0001	0	0	0
1	0,0098	0,0016	0,0001	0	0
2	0,0439	0,0106	0,0014	0,0001	0
3	0,1172	0,0425	0,0090	0,0008	0
4	0,2051	0,1115	0,0368	0,0055	0,0001
5	0,2461	0,2007	0,1029	0,0264	0,0015
6	0,2051	0,2508	0,2001	0,0881	0,0112
7	0,1172	0,2150	0,2668	0,2013	0,0574
8	0,0439	0,1209	0,2335	0,3020	0,1937
9	0,0098	0,0403	0,1211	0,2684	0,3874
10	0,0010	0,0060	0,0282	0,1074	0,3487

\*Биномиальные вероятности  $B_n(k)$  набора экзаменуемым  $k$  баллов (вероятности правильного ответа для «слабых» студентов —  $p \leq 0,6$ , для «средних» —  $p = 0,7$ , для «хороших» —  $p = 0,8$  и «сильных» —  $p \geq 0,9$ ).

## Схемы перевода набранной суммы баллов в оценку

Официальная итоговая оценка в общепринятой системе «2» — «3» — «4» — «5» формируется путем перевода набранной суммы баллов  $k$  в собственно оценку по одной из двух предлагаемых схем.

По одной из схем перевод набранной экзаменуемым суммы баллов  $k$ , выраженной в процентах от максимально возможной суммы баллов  $k_{\max} = n$ , в традиционную систему оценок «2» — «3» — «4» — «5» осуществляется с помощью таблицы 2. Применение «табличной» схемы просто, наглядно и делает процесс оценивания знаний предельно открытым и объективным. Целесооб-

разно по завершении экзамена проводить собеседование с учащимися для обсуждения ошибок.

Таблица 2

**Таблица перевода набранной студентом на экзамене суммы баллов в оценку**

$k = k/k_{\max}, \%$	Оценка
$k \leq 40\%$	2 (неудовлетворительно)
$40\% < k \leq 67\%$	3 (удовлетворительно)
$67\% < k \leq 87\%$	4 (хорошо)
$87\% < k$	5 (отлично)

По второй схеме процесс формирования итоговой официальной оценки несколько сложнее и проходит в два этапа. На первом этапе формируется предварительная оценка («предоценка») путем простого деления набранной суммы баллов  $k$  пополам. В том случае, когда полусумма набранных баллов позволяет однозначно выставить оценку, т.е. для  $k = \{0; 1; 2; 3; 4; 6; 8; 10\}$ , она выставляется и в принципе для таких студентов экзамен закончен, если не проводится дополнительное собеседование по его итогам. Конкретно при  $k \leq 4$  оценкой будет «неудовлетворительно»; при  $k = 6$  — «удовлетворительно» и т.д.

В том случае, когда полусумма набранных баллов полуцелая и не позволяет однозначно выставить оценку, т.е. для  $k = \{5; 7; 9\}$ , то для принятия окончательного решения об оценке знаний экзаменатор дает студенту дополнительный вопрос (задание). В случае правильного ответа на дополнительный вопрос итоговая оценка округляется в сторону увеличения, при неправильном — в сторону уменьшения. Например, студент набрал  $k = 9$  баллов, что дает предоценку  $k/2 = 4,5$ , лежащую в интервале между «хорошо» и «отлично». Для объективного приведения оценки к общепринятой градации экзаменатор в ходе индивидуального устного собеседования задает студенту еще один вопрос и в зависимости от правильности или неправильности ответа округляет оценку в сторону «отлично» или «хорошо». Преимуществом такого подхода является предоставление студенту с неточно определенной оценкой уровня знаний дополнительной возможности улучшить свои показатели. Помимо демократичности, такой подход, как нам представляется, несет в себе еще и боль-

шую воспитательную функцию, что особенно важно при обучении студентов, особенно будущих педагогов.

Заметим, что конкретные формы уточнения итоговой оценки при полуцелой предоценке могут быть различными и выбираются экзаменатором с учетом педагогической ситуации. Например, дополнительно заданный вопрос может быть более сложным, чем экзаменационные задания (с меньшей вероятностью правильного ответа  $p$ ), а может быть, напротив, более простым. Первый вариант рекомендуется для хороших и сильных студентов ( $p \geq 0,8$ ), второй — для слабых ( $p \leq 0,6$ ).

На рис. 2 приведены модельные частотные распределения оценок знаний учащихся с различной степенью подготовленности. Преобразование набранных студентами баллов в итоговую оценку осуществлялось согласно табл. 1 (белые кружки) и по схеме с предоценкой (черные кружки). Данные для различных вероятностей  $p$  смещены друг относительно друга на единицу по вертикальной оси. Поэтому для определения фактических относительных частот целая часть чисел оцифровки по вертикальной оси должна быть отброшена. Блоки данных снизу вверх соответствуют вероятностям  $p = 0,5; 0,6; 0,7; 0,8; 0,9$  соответственно. При определении итоговых оценок по схеме с предоценкой предполагалось, что дополнительный уточняющий вопрос имеет тот же уровень сложности, что и экзаменационные задания. Из рис. 2 видно, что табличная форма преобразования набранных экзаменуемых баллов в итоговую оценку является несколько «мягче», чем схема с предоценкой, особенно для слабых студентов: меньше доля неудовлетворительных оценок и больше доля удовлетворительных и хороших оценок.

Проанализируем полученные результаты. Как видно из рис. 2, предложенная модель оценивания знаний с фиксированным уровнем сложности экзаменационных заданий хорошо подходит для средних и хороших студентов, дающих правильные ответы с вероятностями  $0,6 \leq p \leq 0,8$ . При оценке знаний слабых студентов ( $p = 0,5$ ) в том смысле, что доля двоек и троек превышает 80%; а при оценивании знаний сильных студентов ( $p \geq 0,9$ ) почти все оценки составляют «4» (около 30%) или «5» (около 70%). Это означает, что организованный подобным образом модельный экзамен с фиксированными заданиями, как правило, по степени

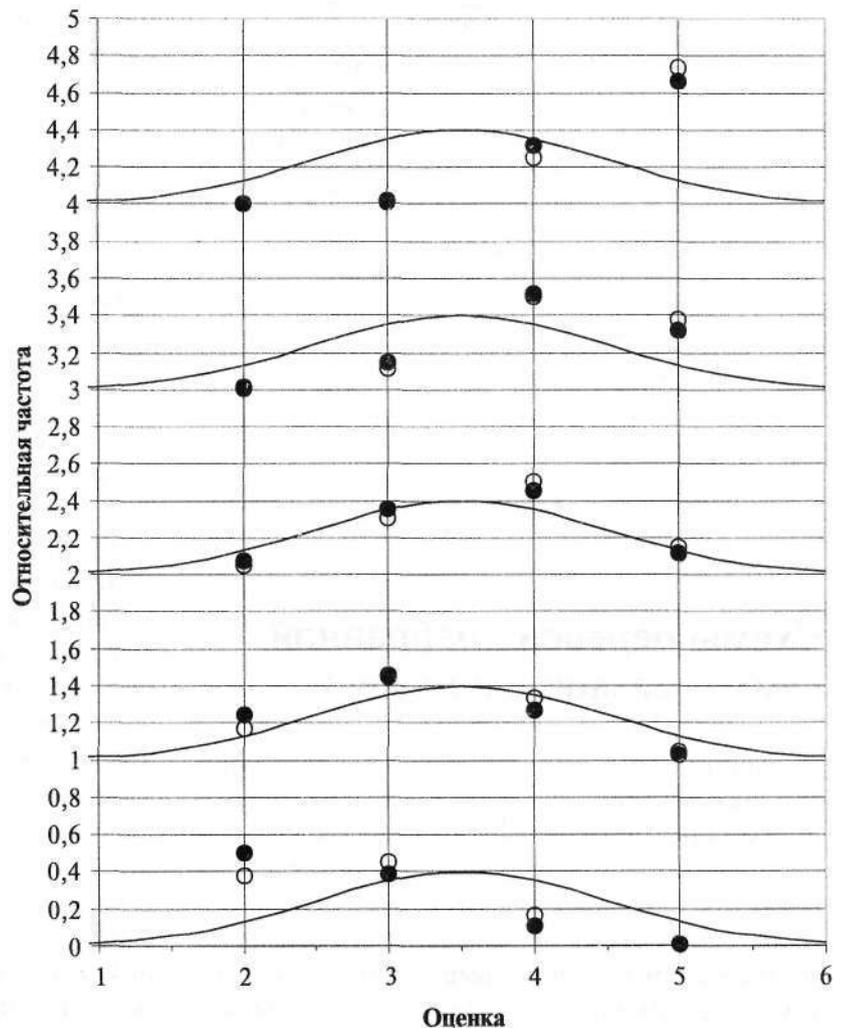


Рис. 2. Частотные распределения оценок знаний учащихся. Символы «кружки» — вероятности правильных ответов, полученные по табличной схеме (белые кружки), с использованием предоценки — черные; сплошные линии — нормальные кривые  $N(3,5; 1,0; x)$

сложности задания ориентирован на контроль знаний средних и хороших студентов, но является чрезмерно трудным для слабых студентов и недостаточно напряженным для сильных студентов. Прогнозируя такую ситуацию в течение учебного периода, преподавателю с учетом дифференцированного подхода следует заранее разработать более простые индивидуальные задания для слабых студентов и, наоборот, более сложные задания для сильных студентов так, чтобы вероятности правильных ответов на них составляли в среднем  $0,6 < p < 0,8$ . В результате экзамен будет одинаково напряженным и развивающим мероприятием для всех студентов.

Следует также добавить, что рассмотренная схема с достаточно большим числом проверочных заданий (в нашем случае  $n = 10$ ) наиболее комфортна для письменной (тестовой) формы проведения экзамена. Для случая с меньшим числом заданий, как это бывает на устном экзамене, модель также вполне применима. В этом случае доля набранных студентом за правильные ответы баллов (от максимально возможного) преобразуется в традиционную систему оценок «2» — «3» — «4» — «5» с помощью таблицы 2.

## Заключение

Подведем основные итоги работы.

Разработана удобная вероятностно-статистическая модель подхода к количественному оцениванию знаний учащихся и интерпретации наблюдаемых результатов.

Результаты применения модели проиллюстрированы расчетами итогов сдачи экзаменов для студентов  $cp = 0,5; 0,6; 0,7; 0,8; 0,9$ . Даны рекомендации по применению модели для студентов с различными уровнями  $p$ .

Разработанная модель универсальна, проста в реализации и делает процесс количественного оценивания знаний учащегося объективным, открытым, наглядным и убедительным, а полученная совместно с учащимся оценка в полной мере будет служить своим педагогическим целям.

## Литература

1. Шишов С.Е., Кальней В.А. Мониторинг качества образования в школе. — М.: Педагогическое общество России, 1999. — 78 с.

2. Кальней В.А., Шишов С.Е. Технология мониторинга качества образования в системе «Учитель — ученик». — М.: Педагогическое общество России, 1999. — 86 с.

3. Медведенко Н.В., Рубцова С.Ю. Оценка и ее взаимосвязь с контролем, измерением и диагностикой в управлении качеством образования // Стандарты и мониторинг в образовании. — 2008. — №2. — С. 19-22.

4. Поташник М.М. Требования к современному уроку. — М.: Центр педагогического образования, 2008. — 272 с.

5. Гмурман В.Е. Теория вероятностей и математическая статистика. — М.: Высшее образование, 2006. — 479 с.

6. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.

7. Бодряков В.Ю., Торопов А.П., Фомина Н.Г. Статистические методы исследования и оценки результатов экзаменационной сессии в вузе / Математика в образовании. Сб. науч. статей. Вып. 4. — Чебоксары: ЧувГУ, 2008. — С. 18-21.

8. Бодряков В.Ю., Фомина Н.Г. Инновационный подход к реорганизации традиционного экзамена в вузе с применением переносного персонального компьютера / Математический вестник педвузов и университетов Волго-Вятского региона. Сб. науч. статей. Вып. 10. — Киров: ВятГУ, 2008. — С. 136—139.

9. Бодряков В.Ю., Торопов А.П., Фомина Н.Г. Оценка остаточных знаний по математическому анализу студентов математического факультета педагогического университета / Актуальные вопросы преподавания математики и информатики. Сб. науч. трудов 3-й Межд. заочной науч.-практ. конф. — Биробиджан: ДГСГА, 2008. — С. 11-15.

10. Бодряков В.Ю., Торопов А.П., Фомина Н.Г. Динамика успеваемости студентов-математиков при обучении в УрГПУ в период поступления 2000 — выпуска 2005 г. / Новые технологии в образовании. — 2008. — №3. — С. 71-75.