

**УДК 37.02**

**Технология тестирования в процессе изучения информатики**

Горбунова Т.Н.

Национальный исследовательский Московский государственный  
строительный университет, Московский политехнический университет  
Москва, Россия

**Аннотация**

Предметом исследования является построение методики изучения Информатики в ВУЗе, способствующей наилучшему усвоению материала. Автор обращает внимание на важность получения обратной связи об уровне усвоения знаний в процессе обучения. В качестве такого средства рассматривается тестирование. Объектом исследования является моделирование системы тестов для защиты лабораторных работ. Особое внимание уделяется надежности смоделированных тестов, их дифференцирующей способности для оценки знаний. Особое внимание автор обращает внимание на обучающий аспект тестов, который предполагает возможность выбора студентом способа решения предложенных тестовых задач.

В качестве методологии построения тестов выбраны открытые и закрытые тестовые задания, предлагаемые студентам и оцениваемые в двоичной системе. В работе предложен алгоритм проверки гипотезы о нормальном распределении с использованием критерия Шапиро-Уилки. Также приведены используемые шкалы оценивания.

Основными выводами проведенного исследования являются: во-первых, возможность применения тестов для оценивания понимания студентами освоенного материала, во-вторых, определения тем имеющих большую и меньшую степень корреляции между собой и влияющих на получение устойчивых знаний. Главным результатом является разработанная технология проведения тестирования с автоматизацией разработки вариантов и обработки результатов. Предложенный алгоритм тестов может быть включен в общую методику изучения дисциплины.

## **Ключевые слова**

тестирование, нормальное распределение, критерий Шапиро-Уилки, шкала оценивания, критерий усвоения, открытый тест, закрытый тест, корреляция, статистическая обработка, балл

## **Testing technology in the process of studying computer science**

**Announce:** The subject of research is the construction of a technique of studying computer science, promotes the best mastering of a material. The author draws attention to the importance of obtaining feedback on the level of assimilation of knowledge in the learning process. The test is regarded as such a tool. The object of research is the simulation test system for the protection of laboratory work. Particular attention is paid to the reliability of simulated tests, differentiating their ability to assess knowledge. Special attention the author pays attention to the training aspect of the test, which suggests the possibility of the student's choice proposed ways to solve test problems.

As the methodology of the tests chosen open and closed test tasks offered to students and evaluated in the binary system. An algorithm testing the hypothesis of normal distribution using the Shapiro-Wilk criterion. And given the scale of assessment used.

The main conclusions of the study are as follows: firstly, the possibility of applying the test for evaluating the understanding of the students mastered the material, and secondly, the identification of topics that have a greater and lesser degree of correlation between themselves and affecting the acquisition of sustainable knowledge. The main result is the developed technology of testing with the automation of the development of options and the processing of results. The proposed algorithm of tests can be included in the general methodology of studying the discipline.

**KeyWords:** testing; normal distribution; Shapiro-Wilk criterion; evaluation scale; the criterion of assimilation; open test; closed test; correlation; statistical processing; reliability rating

## **Введение**

Весь учебный процесс является сложной системой. И для согласованности ее функционирования важным фактором является возможность получения и расшифровки обратной связи, которая необходима всем участникам учебного процесса. Преподаватель, получая необходимую информацию об уровне

достижений студентов, имеет возможность скорректировать учебный процесс для его оптимизации. Обучаемый также сможет осуществлять самоконтроль и самодиагностику своего обучения. Подобная информация также важна и для различного уровня управленческих органов.

В наши дни существует достаточно много форм для организации такой обратной связи, в том числе и в виде разнообразных тестов, которые при выполнении целого комплекса мероприятий могут служить относительно объективным инструментом педагогической диагностики, позволяя организовать эффективную систему обратной связи. [1-5]

Педагогическое тестирование служит не только целям мониторинга. Как отмечает В.Аванесов [6], одной из функций педагогического тестирования является обучающая функция, которая наиболее ярко проявляет себя в программированном обучении.

Целью данного исследования была разработка технологии комплекса мероприятий для проведения тестирования и включения его в учебный процесс.

### **Подготовка и проведение теста**

В МГСУ уже не первый год на кафедре Прикладной математики проводится тестирование студентов по математике. Положительный опыт использования разработанной методики, описанной в работах [7,8], позволил распространить его и на Информатику.

Из всего процесса изучения данной дисциплины был выбран семестр, в котором происходит изучение основ программирования в среде MATLAB, введение в численные методы с использованием знаний линейной алгебры. Поэтому данный семестр можно рассматривать как применение изученных математических моделей в Информатике, а сам этот модуль можно рассматривать как междисциплинарным.

Данный курс Информатики состоит из лекционного курса и 7 практических работ. Курс завершается дифференцированным зачетом. Темы, вынесенные на изучение:

1. решение СЛАУ методом Гаусса;
2. решение СЛАУ итерационными методами, такими как методом простой итерации и методом Зейделя;
3. вычисление обратной матрицы методом Гаусса;
4. вычисление собственных значений и собственных векторов матрицы с изучением степенного метода;
5. численное интегрирование методами средних, трапеций и Симпсона;
6. решение нелинейного уравнения методами половинного деления и методом касательных;
7. метод наименьших квадратов для построения оптимальной прямой.

Выполнение каждой практической работы состоит из трех этапов:

1. выполнения ручного счета для изучения функционирования математической модели и получения тестового результата;
2. реализации программы на компьютере, в данном случае в среде MATLAB и сравнения результатов с ручным счетом;
3. защита работы.

И именно последний этап защиты первых четырех работ проводилась в тестовом виде.

### **Моделирование тестов**

Вся технология моделирования тестов должна определяться поставленным целям. В данном случае в качестве такой цели формулировалась задача оценить понимание студентом математических моделей и численных методов рассмотренного класса задач. Для достижения этой цели при моделировании тестов были использованы следующие технологии.

### **Принципы формирования содержания заданий**

1. Согласно теории тестирования, задания должны быть сформулированы в лаконичной форме исключающей неоднозначное трактование. На тему каждой работы было отведено по 4 тестовых заданий. Формулировка всех заданий была выполнена в единой логической форме высказывания в виде утвердительного предложения. Согласно В.С.Аванесову, это средство

способствует упорядочению и эффективной организации содержания теста [9].

2. В группу тестов для каждой работы входил теоретический вопрос закрытого типа, когда студент выбирает правильный или несколько правильных ответов из предложенных вариантов. Также содержалось задание на вычисление также закрытого типа, и два задания на вычисления открытого типа, в которых испытуемый должен был решить задачу и записать полученный результат. Таким образом, количество заданий открытого и закрытого типа распределялось поровну.

Бесспорно, что задания открытого типа исключают возможность угадывания, но также налагают более жесткие требования к самой формулировке задания, исключая его двусмысленность.

3. Используя закрытую и открытую форму задания, удается выстроить задания с возрастающей сложностью и, тем самым, повысить и их обучаемую способность.

В качестве примера ниже приведены четыре тестовых задания по теме итерационные методы решения СЛАУ.

1. Укажите системы с диагональным преобладанием

$$\bullet \begin{cases} 3x - 2y + z = 1 \\ 3x - 11y + 7z = 2 \\ x + 2y - 3z = 3 \end{cases} \bullet \begin{cases} 2x_1 + 3x_2 - x_3 = 9 \\ x_1 - 2x_2 + x_3 = 3 \\ x_1 + 2x_3 = 2 \end{cases} \bullet \begin{cases} 2x_1 + 10x_2 - 3x_3 = 38 \\ -3x_1 - 12x_2 + 13x_3 = -82 \\ x_1 + 3x_2 - 5x_3 = 27 \end{cases} \bullet \begin{cases} 4x_1 + x_2 - 3x_3 = -12 \\ -x_1 - 12x_2 + 11x_3 = -28 \\ x_1 + 3x_2 - 5x_3 = 51 \end{cases}$$

2. Решить систему уравнений 
$$\begin{cases} 5x_1 + 2x_2 - x_3 = -1 \\ 3x_1 + 10x_2 + x_3 = 9 \\ x_1 + x_2 - 5x_3 = -2 \end{cases}$$
 методом простой итерации. Выполнить

1 шаг. В качестве начального приближения выбрать  $x_1^0 = 0$ ,  $x_2^0 = 0$ ,  $x_3^0 = 0$ . В ответе указать  $x_2^1$ .

3. Дана система линейных уравнений 
$$\begin{cases} x_1 + 5x_2 = 1 \\ 2x_1 + 2x_2 = 3 \end{cases}$$
. Для сходимости метода простой

итерации ее надо записать в виде

$$\bullet \begin{cases} x_2 = (1 - x_1)/5 \\ x_1 = (3 - 2x_2)/2 \end{cases}$$

- $\begin{cases} x_1 = 1 - 5x_2 \\ x_2 = (3 - 2x_1)/2 \end{cases}$
- $\begin{cases} x_1 = 2x_1 + 5x_2 - 1 \\ x_2 = 2x_1 + 3x_2 - 3 \end{cases}$
- $\begin{cases} x_2 = x_1 + 6x_2 - 1 \\ x_1 = 3x_1 + 2x_2 - 3 \end{cases}$

4. Используя метод простой итерации (3 итерации) определить 1-й столбец обратной матрицы к матрице А:  $A = \begin{bmatrix} 6 & -15 \\ 15 & 8 \end{bmatrix}$

Последнее задание позволяет проверить понимание не только итерационного процесса, но и понимание метода нахождения обратной матрицы. Таким образом, позволяет студенту выполнить действия анализа и синтеза для получения решения задачи, что соответствует уже верхним уровням таксономии Блума [10].

4. Применение принципа фасетности позволил на каждое из 16 видов заданий составить по 30 однотипных задач. Сами варианты теста формировались случайным образом по оригинальной методике, описанной в работе [7,8]. "Авторская программа в виде макроса на языке Visual Basic генерирует 30 параллельных индивидуальных вариантов тестов, распределяя задания между испытуемыми случайным образом, что обеспечивает уникальность набора заданий для каждого студента во всех учебных группах".

5. Студенты выполняют тесты в письменном виде. Подобная форма может вызвать некоторое недоумение и требует дополнительного разъяснения. Безусловно, сейчас в период всеобщей компьютеризации и информатизации всех сфер жизнедеятельности общества использование бумажного варианта для тестирования, требует пояснения.

Тестирование в наше время используется повсеместно. Программных средств для проведения автоматизированного тестирования существует

достаточно много. Об этом достаточно подробно написано в работах [10, 11, 12, 13, 14].

При проведении тестирования по математике подобный способ был оправдан отсутствием компьютеров в учебных аудиториях. Занятия по Информатике проходят в компьютеризированных помещениях.

Главное достоинство подобного способа, по мнению авторов - это методическая организация самого процесса обучения. Тестирование - часть этого процесса. Защита практических работ должна показать понимание методов решения определенного класса задач. Все вычислительные задания были сформулированы таким образом, что их можно решить как вручную [15], так и на компьютере в системе MATLAB или при помощи других программных средств, например Excel. Поэтому именно бумажный вариант теста позволял студенту сделать выбор в пользу того или иного способа решения, записав способ решения.

Конечно, недостатки подобной формы теста также очевидны. Это и дополнительные затраты на бумагу, картриджи. Проверка осуществляется преподавателем вручную с использованием файла ответов, который генерируется программой одновременно с индивидуальными вариантами заданий. Тем самым проверка проходит достаточно быстро.

И еще одно несомненное достоинства подобной формы - это возможность индивидуального разбора теста в присутствии студента «по горячим следам», а также обобщенного анализа для всей группы.

6. Тестирование было ограничено 60 минутами.

### **Тестирование**

Тестирование было проведено в трех группах первого курса и двух группах второго курса Московского государственного строительного университета.

Результаты тестирования были статистически обработаны с помощью табличного процессора Microsoft Excel [16, 17]. Каждая задача оценивалась по двухбалльной шкале: 0 – задача решена неверно, 1 – задача решена верно.

Согласно классической теории тестов результат тестирования испытуемого определяется тестовым баллом – суммой набранных баллов при решении тестовых заданий.

### Статистическая обработка тестов

Для каждой группы были вычислены основные характеристики распределения тестовых баллов:  $\bar{X}$ , дисперсия  $s_x^2$ , стандартное отклонение  $s_x$  и коэффициент асимметрии  $A$ . Результаты представлены в таблице 1.

Таблица 1

	курс	Кол-во студентов	$\bar{X}$ математическое ожидание	$s_x^2$	$s_x$	$A$
Группа А1	1	25	8,08	7,4	2,7	0,35
Группа В1	1	18	8,3	7,0	2,7	-0,08
Группа С1	1	17	9,9	7,1	2,7	-0,01
Группа А2	2	20	9,5	3,7	1,9	0,05
Группа В2	2	13	10,2	3,3	1,8	-0,07

Результаты тестирования оказались предсказуемыми: лучший результат показали студенты второго курса.

### Проверка гипотезы о нормальном распределении тестовых баллов

Известно, что для нормативно-ориентированного теста, предназначенного для ранжирования испытуемых по уровню знаний с использованием методов стандартизации, кривая распределения тестовых баллов испытуемых должна быть симметрична и близка к гауссовой кривой.

Для проверки соответствия распределения баллов нормальному существует достаточно много способов. Выбор был сделан в пользу технологии, изложенной в работе [18]. Предварительный этап состоит из визуального. Используя возможности того же Excel, можно построить гистограммы распределения индивидуальных баллов для групп испытуемых в зависимости от количества решенных задач. И, если визуально распределение близко к нормальному, сделать предварительный вывод. В качестве начальной стадии анализа была выполнена эта процедура, все гистограммы показали нужное распределение.

Следующий этап состоял в анализе статистических характеристик, таких как: медиана, мода, ошибка среднего и др. Все эти характеристики могут быть получены в Excel Данные - Анализ данных - Описательная статистика. И анализ этих характеристик также показал, что распределение баллов близко к нормальному.

Эти результаты не приводятся в этой статье, т.к. окончательное решение было принято после выполнения критерия Шапиро-Уилка. Данный критерий применяется как раз для небольшой выборки  $<50$ .

Алгоритм расчета критерия Шапиро-Уилка в Excel:

1. Данные измерений сортируют от минимума к максимуму.

$$X_1 < X_2 < \dots < X_n$$

2. Вычисляют среднее значение выборки  $\bar{X}$  и квадрат отклонений от среднего

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

3. Вычисляют коэффициент  $b$  по формуле:

$$b = \sum_{i=1}^k a_i \cdot (X_{n-k+1} - X_k), \quad \text{где } k = \begin{cases} n/2, & \text{если } n \text{ четного} \\ (n-1)/2, & \text{если } n \text{ нечетное} \end{cases}$$

$a_i$  - известные коэффициенты, которые находятся либо по специальным таблицам [19], либо могут быть вычислены программно по соответствующим соотношениям [20].

4. Находят эмпирическое значение критерия по формуле

$$W_{эмп} = \frac{b^2}{S^2}$$

5. Сопоставляют полученное значение критерия  $W_{эмп}$  с табличным значением, которое также можно взять из ГОСТа [19]. Либо вычислить программно.

Если  $W_{эмп} \gg W_{табл}$ , то гипотеза о соответствии полученных результатов нормальному распределению не отвергается.

В таблице 2 приведены расчетные и табличные значения статистики критерия Шапиро-Уилка при уровне значимости  $\alpha=0,05$ .

Таблица 2

Группа	$W_{эмп}$	$W_{табл}$
A1	0,97764	0,918
B1	0,9141	0,897
C1	0,92651	0,892
A2	0,9647	0,905
B2	0,90173	0,866

В результате проведенного исследования был сделан вывод, что распределение тестовых баллов во всех группах является нормальным.

### **Дифференцирующая способность теста**

Для оценки дифференцирующей способности теста, т.е. возможности разделять при помощи набранного тестового балла учащихся с различным уровнем подготовки, используется отношение математического ожидания к стандартному отклонению. Дифференцирующая способность теста считается удовлетворительной, если указанное отношение около 3 или больше [6].

Таблица 3

	$\bar{X} / s_x$
Группа A1	2,97

Группа В1	3,12
Группа С1	3,71
Группа А2	4,89
Группа В2	5,58

Данные в таблице 3 указывают на достаточно высокую дифференцирующую способность теста. По приведенным данным, видно распределение исследуемых групп по двум: группы первого курса и группы второго курса. Поэтому дальнейшее оценивание - выставление оценки по полученным данным проводились для каждой из этих групп.

### **Шкалы оценивания**

В таблицах 4 и 5 в столбцах 5 и 6 приведены соответствия значений 16-бальной системы с использованием нормализации в четырехбальную систему (от 2 до 5).

Согласно поставленной цели, а именно, проведение защиты четырех работ в тестовом варианте, при оценивании каждая тема рассматривается отдельно. Результатом каждой темы является балл той же четырехбальной шкалы. Соответственно 1 и 2 столбец. Тема считается освоенной и соответственно защищенной, если решены три или более задач. С учетом этого положения результат по всему тесту корректировался, т.е. результирующая оценка вычислялась с учетом сданных тем. Данные приведены соответственно в 3 и 4 столбцах таблиц.

Таблица 4

### **Шкалы оценивания: первый курс**

Шкала оценивания по исходным баллам.		Корректировка шкалы с учетом засчитанных тем		Шкала оценивания по нормальному распределению	
Исходные сырые баллы	Оценка по тесту	Баллы	Оценка по тесту	Баллы	Оценка
1	2	3	4	5	6
0-5	2	0-7	2	0-6	2

6-9	3	7-9	<b>3</b>	7-10	<b>3</b>
10-13	4	10-13	<b>4</b>	11-13	<b>4</b>
14-16	5	14-16	<b>5</b>	14-16	<b>5</b>

Таблица 5

### Шкалы оценивания: второй курс

Шкала оценивания по исходным баллам.		Корректировка шкалы с учетом засчитанных тем		Шкала оценивания по нормальному распределению	
Исходные сырые баллы	Оценка по тесту	Баллы	Оценка по тесту	Баллы	Оценка
1	2	3	4	5	6
0-6	2	0-8	<b>2</b>	0-6	<b>2</b>
7-9	3	8-10	<b>3</b>	7-10	<b>3</b>
10-13	4	11-13	<b>4</b>	11-13	<b>4</b>
14-16	5	14-16	<b>5</b>	14-16	<b>5</b>

По приведенным результатам видно, что для второго курса диапазон низких баллов смещается с большую сторону. Также сужается диапазон хорошей оценки, что согласуется с рекомендациями, изложенными в работе [10 с.151, 21].

### Исследование критерия усвоения темы

Критерием усвоения  $j$ -й темы  $q_j$  будем считать отношение числа испытуемых решивших три и более задач из  $j$ -й темы, к общему числу испытуемых. Соответственно критерием характеризующим не усвоения  $j$ -й темы назовем число равное  $p_j = 1 - q_j$  - отношение числа менее трех задач из темы к общему числу испытуемых.

Данные критерии всех 4 тем определялись по результатам тестирования отдельно в каждой группе испытуемых. Среднее значение  $\bar{p}$  простоты

заданий в группе равно отношению среднего числа выполненных заданий  $\bar{X}$  к объему теста  $n = 20$ ; среднее значение сложности  $\bar{q} = 1 - \bar{p}$ . В исследуемом тесте

Таблица 6

Критерии усвоения тем

группа	1 тема	2 тема	3 тема	4 тема
A1	0,64	0,36	0,28	0,19
B1	0,75	0,52	0,43	0,24
C1	0,8	0,6	0,55	0,22
A2	0,75	0,65	0,6	0,25
B2	0,62	0,77	0,70	0,31

По результатам таблицы 6 видно, что темы 1- 3 усвоены на достаточном уровне. Результаты усвоения темы 4 ниже. Анализ учебного материала показывает, что первые три темы взаимосвязаны между собой, тем и обуславливается высокая их усвояемость.

Средняя корреляция тем между собою подтверждает этот вывод.

Тема 1	Тема 2	Тема 3	Тема 4
0,34	0,27	0,31	0,15

Средняя корреляция результатов по заданию и индивидуальными баллами составляет:

Тема 1	Тема 2	Тема 3	Тема 4
0,92	0,89	0,72	0,48

## Выводы

В результате данного исследования было сделано несколько выводов. Первое. Проведенное тестирование студентов показало о недостаточности проработки четвертой темы при существующем подходе и необходимости

разработки дополнительных заданий с целью организации более тесного взаимодействия с темами курса.

Главным результатом проведенной работы является выработка технологии проведения тестирования для защиты работ по Информатике, которая представлена в виде алгоритма. Его этапы включают: формирование заданий, проведение тестирования, статистическая обработка с проверкой гипотезы о нормальном распределении, получения шкал оценок и анализ усвояемости материала.

Все этапы, кроме формирования самих заданий, осуществляются в автоматизированном виде с помощью макросов Excel. В качестве дальнейшего развития предполагается полностью компьютеризировать весь процесс без потери его методологических достоинств для распространения на дистанционное обучение.

### Литература

1. Fisher W.P., Jr. The Central Theoretical Problem of the Social Sciences // Rasch Measurement Transactions. 2014. Vol. 28:2. P. 1464-1466.
2. Kenneth D.R. Measuring Liberal/Conservative Voting Tendencies among U.S. Senators. // Rasch Measurement Transactions. 2012. Vol. 26:2. P. 1366-1367.
3. Reeve B.B. Item response theory modeling in heart outcomes measurement // Expert Review of Pharmacoeconomics and Outcomes Research. 2003 3(2). P. 131-145.
4. Bartram D. The development of standards for the use of psychological tests in occupational settings: The competence approach. The Psychologist, 1995,5. P. 219-223.
5. Baayen R.H., Davidson D.J., Bates D.M. Mixed-effects modeling with crossed random effects for subjects and items // Journal of Memory and Language. 2008. Vol. 59, is. 4. P. 390–412.
6. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе, 1989. –М., МИСИС. –168 с.
7. Сафина Г.Л., Осипов Ю.В., Керимова Д.Х., Красовская И.А. Полуавтоматическая система тестирования по математике // Открытое и дистанционное образование. 2015. № 2 (58). с. 56-62.
8. Осипов Ю.В., Сафина Г.Л., Ветухновский Ф.Я. Моделирование тестов по математике // Открытое и дистанционное образование. 2016. № 3 (63). с. 69-77
9. Аванесов В.С. Форма тестовых заданий. М.: Центр тестирования, 2005. 156 с.
10. Ким В.С. Тестирование учебных достижений. Уссурийск: УГПИ, 2007. 214 с.

11. Артищева Е. К. Об инструментарии педагогической диагностики в учебном процессе вуза // Вестник Российского государственного университета им. И. Канта. Калининград: Изд-во РГУ им. И. Канта. 2007. Вып. 4. С. 27-34
12. Майоров А.Н. Теория и практика создания тестов для системы образования. М.: «Интеллект-центр», 2001. 296 с
13. Кузьмина Л.И., Осипов Ю.В. О тестировании студентов по «школьной» математике // Качество. Инновации. Образование. 2014. № 2. с. 9-13.
14. Кузьмина Л.И., Осипов Ю.В. Новые технологии преподавания и "старые" дисциплины // Качество. Инновации. Образование. 2013. № 12. с. 3-7.
15. Горбунова Т.Н., Чалмов И.Д., Черных С.И. Ручной и автоматизированный подход при изучении информационных технологий. // Наука, образование, общество: актуальные вопросы и перспективы развития: сб. науч. тр. М., 2015. С. 22-23.
16. Гюнтер Штайнер. Visual Basic 6.0 для приложений. – М.: Лаборатория Базовых Знаний, 2000. – 832 с.
17. Горбунова Т.Н., Журавлева Т.Ю. Автоматизированный лабораторный практикум по информатике Освоение работы в MS Excel 2007. [Электронный ресурс]// Монография / Вузское образование, ЭБС «IPRbooks», — Саратов, 2014. 77 с. —Режим доступа: <http://www.iprbookshop.ru/20699> (дата обращения:06.12.2016).
18. Попов О.А. Психологическая статистика [Электронный ресурс]: Режим доступа: [http:// www.psystat.at.ru](http://www.psystat.at.ru).
19. ГОСТ Р ИСО 5479-2002 Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения [Электронный ресурс]: Режим доступа: <http://gostexpert.ru/gost/gost-5479-2002>
20. Заляжных В. В. Статистические расчёты при планировании и обработке результатов испытаний [Текст] : учебное пособие / В. В. Заляжных ; Северный (Арктический) федеральный ун-т им. М. В. Ломоносова. - Архангельск : САФУ, 2014. - 82с.
21. Дубас В. Об оценивании знаний при программированном контроле //Физика в школе, 1990, №3. -С 83.