

Система контролю знань до дистанційного курсу “Теорія ймовірностей і математична статистика”

Сучасне тестування являє собою комплекс стандартизованих методів вимірювання параметрів, через які визначають рівень підготовки людини і відповідність освітнім стандартам у конкретній галузі знань [1, 2]. При цьому широко використовуються математичні методи планування й опрацювання результатів тестування, а також сучасні технології опрацювання даних. Об’єктивний контроль знань, вмій і навичок вдається здійснити при критеріально-орієнтованій інтерпретації тестування. Критеріально-орієнтоване тестування призначене не тільки для оцінювання рівня знань, а й для визначення рівня індивідуальних досягнень відносно певного критерію на підставі аналізу змісту завдань. Тому, враховуючи індивідуалізацію навчання, конструювання критеріально-орієнтованих тестів є одним із провідних та найактуальніших напрямків розвитку теорії тестів.

В сучасних навчальних системах потрібно знати складність кожного завдання тесту. Така складність апробується і визначається емпірично [3]. Так виникає одне із головних питань теорії тестів – питання побудови ефективного тесту.

Історично виділяють два основні підходи до створення тестів. Перший з них набув широкого розвитку в рамках класичної теорії тестів. Згідно з ними, рівень знань учасників тестування оцінюється за допомогою індивідуальних балів. Бал обчислюють як алгебраїчну суму оцінок виконання кожного завдання тесту. Класична теорія тестів ґрунтується на статистичних методах аналізу результатів тестування.

Складність завдань можна визначати двома способами [6]:

- на основі оцінки передбачуваного числа і характеру розумових операцій, необхідних для вдалого виконання завдань;
- на основі емпіричної перевірки завдань, з підрахунком частки неправильних відповідей.

У класичній теорії тестів багато років розглядалися тільки емпіричні показники складності. У сучасних теоріях навчальних тестів, які використовуються в дистанційному навчанні, більше уваги приділяється характеру розумової діяльності у процесі виконання тестових завдань різних форм.

Наступною вимогою до тестових завдань є варіація балів.

Якщо на деяке завдання правильно відповідають всі студенти, то таке завдання стає не тестовим. Учасники тестування відповідають на нього однаково: між ними немає варіації. Нетестовим вважається також завдання, на яке немає жодної правильної відповіді. Варіація при цьому теж дорівнює нулеві. Нульова варіація означає практичну необхідність викидання завдання із тесту.

Штучність деяких припущень класичної теорії тестів і деякі її практичні недоліки помітно вплинули на розвиток критичних тенденцій. Цьому в першу чергу сприяли сумніви в об’єктивності емпіричних оцінок складності завдань тесту. А саме: виникло питання про правомірність традиційного оцінювання складності завдань за допомогою частки правильних чи неправильних відповідей.

При традиційному підході до зміни рівня складності завдань на різних вибірках відповідей студентів з різними рівнями підготовки залишається відкритим питання про об’єктивність значень параметра складності завдань тесту [1]. Спроба введення вагових коефіцієнтів, що відображають вагу завдання в індивідуальному балі студента, суттєво не виправляє такі недоліки. Значення цих коефіцієнтів можна, в свою чергу, поставити під сумнів. Деякі з них визначаються суб’єктивно, на основі думки педагога про складність завдання. Оцінки решти з них базуються на емпіричних даних тестування і, відповідно, залежать від рівня знань опитаних студентів.

Другий підхід до створення тестів та опрацювання результатів тестування використовуються в так званій сучасній теорії тестування, що набула широкого розвитку в 1960 - 1980 роках в багатьох країнах [7].

Item Response Theory (IRT) – математична теорія параметричного оцінювання тестових завдань і результатів тестування. Відповідно до цієї теорії встановлено, що між результатом виконання, що спостерігається, і рівнем знань учасників тестування є деяка залежність, яку можна виразити за допомогою деякої функції.

До найбільш вагомих переваг IRT відносять:

- стійкі об’єктивні оцінки параметра складності завдань, що не залежать від властивостей вибірки відповідей студентів, які виконують тест;

- вимірювання значень параметрів знань студентів і завдань тесту в одній і тій самій шкалі, що дозволяє поставити у відповідність рівень знань кожного учасника тестування із рівнем складності кожного завдання тесту;

- можливість оцінити ефективність різних за рівнем складності завдань для вимірювання рівня знань студента.

Аналіз сучасних методів тестування та практичних особливостей проведення тестового контролю, а також загальних вимог до комп'ютерних навчальних систем та практичного досвіду їх використання дозволяє визначити загальні вимоги, яким повинні задовольняти комп'ютерні тестові програми. Дотримання цих вимог визначає основні та додаткові можливості використання середовищ тестування.

До складу основних можливостей належать такі:

(Далі кожен можливість покажемо на прикладі її використання при вивченні курсу "Теорія ймовірностей і математична статистика")

- проводити тестовий контроль на базі тестових завдань закритого та відкритого типів, тобто з наданням (відображенням для вибору) та без надання варіантних відповідей на питання тестових завдань, оскільки тестові завдання інших типів (на відповідність, на встановлення правильної послідовності) можуть бути перетворені в одне або кілька тестових завдань закритого та відкритого типів;

При роботі з системою Moodle, яка вибрана для розробки курсу, можна додавати всі перераховані типи тестових завдань. В курсі "Теорія ймовірностей і математична статистика" переважно використовуються завдання типу: відповідь "Так/Ні", "Множинний вибір" та "Числовий".

- використовувати кілька бібліотек тестових завдань при проведенні одного тесту з послідовним виконанням тестових завдань з кожної бібліотеки, що дозволить розмістити в окремих бібліотеках тестові завдання з різних тем;

- вибрати тестові завдання з бібліотек тестових завдань в заданій кількості та відобразити варіантні відповіді в тестових завданнях закритого типу у випадковому порядку.

Для курсу "Теорія ймовірностей і математична статистика" тестові завдання до кожної окремої теми містяться в окремій бібліотеці "Категорії". Для переходу до бібліотеки тестових завдань потрібно у блоці "Управління" (рис. 1) перейти за посиланням "Банк питань" і відкрити вкладку "Категорії" (рис. 2). Щоразу для створення підсумкового тесту модуля використовуються щоразу два тестові завдання, які випадковим чином вибираються з кожної підкатегорії даного модуля (рис. 3).

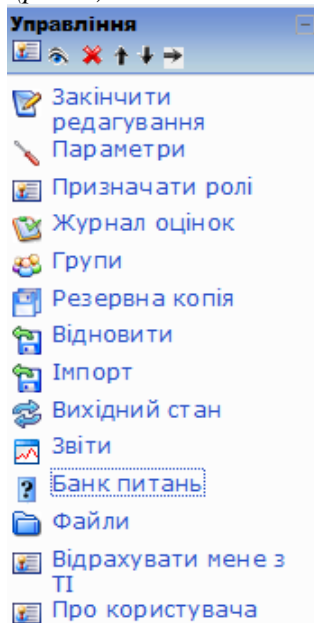


Рис. 1



Рис. 2

Редагування: Тест

Інформація Результати Перегляд Редагувати

Тест Банк питань Категорії Імпорт Експорт

Питання в цьому тесті

Коротке означення питання	Тип	Оцінка	Дія
Випадкове питання (Test 01)	?	0.5	⏪ ⏩
Випадкове питання (Test 01)	?	0.5	⏪ ⏩
Випадкове питання (Test 02)	?	0.5	⏪ ⏩
Випадкове питання (Test 02)	?	0.5	⏪ ⏩
Випадкове питання (Test 03)	?	0.5	⏪ ⏩
Випадкове питання (Test 03)	?	0.5	⏪ ⏩
Випадкове питання (Test 04)	?	0.5	⏪ ⏩
Випадкове питання (Test 04)	?	0.5	⏪ ⏩
Випадкове питання (Test 05)	?	0.5	⏪ ⏩
Випадкове питання (Test 05)	?	0.5	⏪ ⏩
Випадкове питання (Test 06)	?	0.5	⏪ ⏩
Випадкове питання (Test 06)	?	0.5	⏪ ⏩

Банк питань

Категорія Модуль 3

Відображати питання, що знаходяться в підкатегоріях
 Також показувати сховані питання
 Показати формулювання запитання у списку питань

Створити питання Обрати...

Сторінка: 1 2 3 4 5 6 7 8 9 10 11 12 (Далі)

Сортувати за типом, назвою

Дія	Коротке означення питання
⏪ ⏩ ⏴ ⏵	2. Якщо дійсна функція, визначена на просторі елементарних подій коли множина розв'язків рівняння (де – відоме, а – невідоме) є поді
⏪ ⏩ ⏴ ⏵	22. Математичне сподівання довільного випадкового процесу є
⏪ ⏩ ⏴ ⏵	24. Кореляційна функція
⏪ ⏩ ⏴ ⏵	4. Розподіл ймовірностей на множині значень випадкової величини коли
⏪ ⏩ ⏴ ⏵	Якщо розподіл ймовірностей на множині значень випадкової величини параметрами m та σ , то її центральні моменти k -го порядку


Рис. 3

– оцінювати правильну відповідь або будь-яку відповідь на кожне тестове завдання певною кількістю балів та підраховувати загальну кількість набраних балів і відповідну оцінку за заданою шкалою оцінок;

Кожне запитання можна оцінити числом, яке відображає відносну складність завдання і записується у колонку “Оцінка”. Якщо запитання вибираються довільним чином, то доцільно вибирати однакові значення оцінки тестового завдання. На рис. 3 оцінка кожного завдання дорівнює 0,5.

– відводити або не відводити певний час на виконання всього тесту або кожного тестового завдання окремо, що дозволить виконувати тест в різних режимах обмеження часу.

Проходження тестів не передбачає контролю за користуванням додатковими матеріалами з боку викладача, тому доцільно обмежити час на виконання тесту, щоб при пошуку відповіді на завдання користувач не встигав проходити тестування у повному обсязі. Налаштувати обмеження в часі можна або під час створення тесту, або звернувшись до пункту “Оновити”

 і встановити відповідні параметри (рис. 4).

При роботі з системою Moodle є можливість обмежити проходження тесту за датою. Але при дистанційному вивченні курсу “Теорія ймовірностей і математична статистика” ця можливість не використовується, оскільки вона є доцільною для стимулювання вчасного виконання завдань, коли дистанційний курс використовується для підтримки традиційної форми навчання, або коли термін проходження курсу обмежений в часі (рис. 4).

Вибір часу

Почати тестування Відключити

Закінчити тестування Відключити

Обмеження в часі (хвилини) Включити

Час між першою та другою спробами

Час між пізнішими спробами

Показати

Питань на одній сторінці

Випадковий порядок питань

Випадковий порядок відповідей

Спроби

Дозволено спроб

Кожна спроба ґрунтується на попередній

Навчальний режим

Оцінки

Метод оцінювання

Нараховувати штрафи

Десяткове значення в оцінках

Опції перегляду

Безпосередньо після спроби	Пізніше, але тільки поки тест відкритий	Після того, як тест буде закритий
<input checked="" type="checkbox"/> Відповіді студента	<input checked="" type="checkbox"/> Відповіді студента	<input checked="" type="checkbox"/> Відповіді студента
<input type="checkbox"/> Правильні відповіді	<input type="checkbox"/> Правильні відповіді	<input checked="" type="checkbox"/> Правильні відповіді
<input type="checkbox"/> Коментар	<input checked="" type="checkbox"/> Коментар	<input type="checkbox"/> Коментар
<input type="checkbox"/> Основний коментар	<input type="checkbox"/> Основний коментар	<input checked="" type="checkbox"/> Основний коментар
<input checked="" type="checkbox"/> Бали	<input checked="" type="checkbox"/> Бали	<input checked="" type="checkbox"/> Бали
<input type="checkbox"/> Розширений відгук	<input type="checkbox"/> Розширений відгук	<input checked="" type="checkbox"/> Розширений відгук

Рис. 4

До складу додаткових належать можливості:

– задавати кількість можливих разів виконання кожного тесту для обмеження спроб повторного виконання тесту однією особою.

В залежності від призначення тесту (тренувальний чи контролюючий) можна добирати кількість дозволених спроб. Наприклад для тесту на контроль знань доцільно обмежити кількість спроб двома, а метод оцінювання обрати – “Середня оцінка” (рис. 4). При таких параметрах тесту навіть при невеликому банку тестових завдань оцінка досить точно відповідає рівню знань студента. Прикладом тренувального тесту може бути тест з параметрами: кількість спроб – “Необмежена”, метод оцінювання – “Перша оцінка”. При такому виборі параметрів оцінки подальших спроб не будуть враховуватись у загальному заліку.

– зберігати результати тестування для подальшого їх використання, відображати та вилучати результати тестування при необхідності;

Для перегляду всіх результатів проходження обраного тесту треба перейти за посиланням відповідної назви тесту і відкрити вкладку “Результати” (рис. 5).

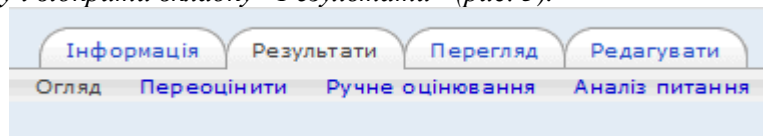


Рис. 5

У вікні браузера з'явиться таблиця (рис. 6)

– аналізувати відповіді на завдання тесту кожного окремого студента або групи.

На відміну від класичної теорії тестів, де індивідуальний бал розглядається як стає число, в IRT параметр трактується як деяка змінна. Початкове значення параметра отримується безпосередньо на основі емпіричних даних тестування. Змінний характер вимірюваної величини вказує на можливість послідовного наближення до об'єктивних оцінок параметрів за допомогою ітераційних методів.

Спроби:139										
Всі спроби впливають на фінальну оцінку.										
Ім'я : Усі А Б В Г Г Д Е Є З Ж І Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ю Я										
Прізвище : Усі А Б В Г Г Д Е Є З Ж І Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ю Я										
Сторінка: 1 2 (Далі)										
Прізвище / Ім'я	Тест початий	Завершено	Затрачений час	Оцінка/5	#1	#2	#3	#4	#5	#5
	19 вересня 2009, 19:27	19 вересня 2009, 19:30	3 хв 3 сек	3.5	0.5/0.5	0.5/0.5	0/0.5	0.5/0.5	0.5/0.5	0/0.5
	19 вересня 2009, 19:33	19 вересня 2009, 19:36	2 хв 46 сек	4.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5
	21 вересня 2009, 19:24	21 вересня 2009, 19:26	2 хв 48 сек	3.5	0/0.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5
	21 вересня 2009, 18:17	21 вересня 2009, 18:20	3 хв	2.5	0.5/0.5	0/0.5	0/0.5	0.5/0.5	0.5/0.5	0.5/0.5
	21 вересня 2009, 18:20	21 вересня 2009, 18:20	зараз	0	0/0.5	0/0.5	0/0.5	0/0.5	0/0.5	0/0.5
	21 вересня 2009, 18:03	21 вересня 2009, 18:06	3 хв 4 сек	3.5	0/0.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5	0.5/0.5

Рис. 6

Взаємний аналіз двох множин значень вимірюваної величини породжує результати виконання тесту. Елементи першої множини – це значення параметра, за яким визначається рівень знань учасників тестування θ_i , де $i=1, \dots, n$. Другу множину утворюють значення параметра δ_j , де $j=1, \dots, m$, параметри що відповідають рівням складності m завдань тесту.

На практиці ставиться задача: за відповідями студентів на завдання тесту оцінити значення параметрів θ і δ [8]. Для її розв'язання потрібно відповісти на два питання:

1. Як вибрати співвідношення між θ і δ ?

2. Як правильно вибрати математичну модель, тобто таку модель, в якій пов'язуються емпіричні результати тестування та параметри θ і δ ?

В рамках IRT датським математиком Джорджем Рашем у 1957 році була запропонована модель контролю знань, яку часто називають простою логістичною моделлю. Модель Раша спирається на поняття „складність завдання" та „рівень підготовки студентів". Так, одне завдання вважається складнішим, ніж друге, якщо ймовірність правильної відповіді на перше завдання менша, ніж на друге, незалежно від того, хто його виконує.

Таким чином, оцінка складності тестових завдань не залежить від вибірки учасників тестування. Крім того, модель Раша характеризується найменшим числом параметрів: один параметр рівня знань для всіх випробованих та тільки один параметр складності для всіх завдань. Дж. Раш запропонував ввести співвідношення між θ і δ у вигляді різниці $\theta - \delta$, вважаючи, що параметри θ і δ оцінюються в одній шкалі.

У такій математичній моделі параметри θ і δ виражаються як показники, задані в одній шкалі логітів. Введення однієї шкали для елементів двох множин θ і δ дозволяє ввести взаємозв'язок між змінними у вигляді різниці $\theta - \delta$, коректно порівняти результати тестування студентів, отримані за допомогою різних тестів, оцінити рівень складності завдань незалежно від рівня підготовки груп студентів.

Можна розглядати умовну ймовірність правильного виконання j -ого завдання із рівнем δ_j різними студентами. Тут незалежною змінною є θ , а δ_j – параметр, що визначає складність j -ого завдання:

$$P_j(x_{ij} = 1 | \delta_j) = \varphi(\theta - \delta_j), j=1, \dots, m.$$

В теорії IRT функцію $\varphi(\theta)$ називають „Item response function" (IRF). Спеціальну назву має графік такої функції – характеристична крива j -ого завдання (ICC). При виборі вигляду функції P_j враховують результати як емпіричного, так і математичного характеру. Припустивши, що на множинах значень параметрів змінних θ і δ розподіли ймовірностей нормальні, маємо дві такі функції. Одна з них позначається $\Psi(x)$ – деяка логістична функція, інша $\Phi(x)$ – інтегральна функція

нормованого нормального розподілу. Оскільки для одних і тих самих значень x ординати точок графіків функцій $\Phi(x)$ і $\Psi(1.7x)$ відрізняються достатньо мало, а саме

$$|\Phi(x) - \Psi(1.7x)| < 0.01$$

то на практиці перевагу віддають функції $\Psi(1.7x)$, адже у неї значно простіше аналітичне задання, зручне для оцінювання δ [8].

Кількість параметрів у такому аналітичному заданні функції поділяє сімейства IRF на класи. Серед логістичних функцій розглянемо однопараметричну модель Дж. Раша:

$$P_j(\theta) = \frac{e^{1.7(\theta - \delta_j)}}{1 + e^{1.7(\theta - \delta_j)}}$$

Розглянемо як за допомогою вбудованих засобів Moodle можна проаналізувати запитання тесту.

Для цього переходимо за посиланням “Аналіз питання” (рис. 5).

У вікні браузера відкриється таблиця аналізу питань (рис. 7).

Кількість відповідей. У чисельнику записане число, що показує, скільки разів студенти давали правильну відповідь на це запитання, а у знаменнику – кількість спроб відповісти на дане запитання.

% відповідей показує, у скількох відсотках спроб обиралась саме ця відповідь.

% правильних відповідей показує відсоткове співвідношення суми балів, набраних за відповіді на це запитання, до тієї суми балів, яку вони могли б набрати якби завжди відповідали на це запитання правильно. Цей показник іноді називають індексом простоти запитання.

відносна оцінка варіанту	Кільк. відп.	% відповідей	% правильних відповідей	Станд. відгук	Дискр. індекс	Дискр. коэф.
(1,00)	13/14	(93%)	93%	0,267	0,90	0,00
(0,00)	1/14	(7%)				
(1,00)	13/15	(87%)	87%	0,352	0,90	0,50
(0,00)	1/15	(7%)				
(1,00)	9/14	(64%)	64%	0,497	0,80	0,67
(0,00)	5/14	(36%)				
(1,00)	11/18	(61%)	61%	0,502	0,82	0,52
(0,00)	7/18	(39%)				
(1,00)	7/11	(64%)	64%	0,505	1,00	0,62

Рис. 7

Стандартний відгук (відхилення). За допомогою цього показника можна оцінити, наскільки відрізняються між собою відповіді різних студентів. Якби всі студенти, відповідаючи на це запитання, набрали однакову кількість балів, то стандартне відхилення дорівнювало б нулеві. Стандартне відхилення обчислюється як корінь квадратний із суми квадратів відхилень оцінок від їх середнього значення, розділеної на кількість цих оцінок.

Індекс дискримінації – це показник, за допомогою якого можна проаналізувати якість тестового запитання, точніше, як за допомогою нього відрізнити сильних студентів від слабких. Для обчислення індексів дискримінації за результатами відповідей на всі запитання тесту студенти умовно розподіляються на три групи: сильні, середні і слабкі. Вираз для індекса дискримінації i -того запитання має вигляд: $DI_i = \frac{X_{сильн(i)} - X_{слаб(i)}}{n}$, де

$X_{сильн(i)}$ – сума відношень набраних балів до максимальних при відповіді на i -те запитання третиною студентів, які отримали найвищі бали за тест взагалі; $X_{слаб(i)}$ – аналогічна сума відносних балів, отриманих третиною слабких студентів за результатами тесту, n – загальна кількість студентів.

Індекс дискримінації може набувати значень від -1 до +1. Наприклад, +1 означає, що на дане запитання всі сильні студенти відповіли правильно, а всі слабкі неправильно, 0 – і сильні і слабкі студенти відповіли правильно однаково кількість разів. Якщо питання має від'ємний індекс дискримінації, то воно, як правило, містить помилку і його слід вилучити з тесту.

Коефіцієнт дискримінації – це показник в якому співставляються відповіді сильних і слабких студентів. Коефіцієнт дискримінації є Коефіцієнт кореляції між сумою балів, набраних за це запитання, і в тесті взагалі: $\frac{\sum(XY)}{n \cdot s_x \cdot s_y}$, де $\sum(XY)$ – сума добутків відхилень набраних балів для даного запитання і тесту взагалі; n – кількість відповідей на дане запитання; s_x – стандартне відхилення набраних балів за це запитання; s_y – стандартне відхилення набраних балів за тест взагалі. Цей показник також набуває значень з відрізка [-1;1]. Додатні значення вказують на те, що на це запитання краще відповіли сильні студенти, від'ємне – на нього краще відповіли слабкі. Коефіцієнт дискримінації дає точніші результати, ніж індекс дискримінації, тому що для його обчислення беруться до уваги результати тестування всієї групи без умовного поділу на три групи.

Також можна вказувати, які спроби проходження тесту вважати показовими і включити їх у аналіз тесту, вказавши відповідні параметри. Іноді студенти можуть просто проглядати тест, не намагаючись правильно відповідати, і в результаті отримують низькі бали. Щоб відкинути такі спроби можна задати відсотковий поріг включення оцінки в аналіз (рис. 8).

Для подальшого аналізу результати можна подати у форматі Excel, або у текстовому форматі.

Рис. 8

Описані методи аналізу результатів тестування дозволяють провести найпростіші та необхідні процедури опрацювання результатів тестування знань і визначити методи оцінювання якості тесту. Розглянутий підхід до конструювання тестів, згідно з яким рівень знань учасників тестування оцінюється за допомогою їх індивідуальних балів, а складність завдань – за допомогою частки правильних та неправильних відповідей на них, показав необхідність використовувати нові методи конструювання тестів, представлені в так званій сучасній теорії тестування на основі математичної теорії параметричного оцінювання тестових завдань.

Список літератури

1. Аванесов В.С. Теория и методика педагогических измерений (материалы публикаций). – М.: ЦТ и МКО УГТУ-УПИ, 2005. – 98 с.
2. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.
3. Люсин Д.В. Основы разработки и применения критериально-ориентированных педагогических тестов. – М.: Исследовательский центр, 1993. – 51 с.
4. Moodle.org: open-source community-based tools for learning URL: <http://www.moodle.org/>
5. Анисимов А.М. Работа в системе дистанционного обучения Moodle. Учебное пособие. 2-е изд. испр. и дополн.– Харьков, ХНАГХ, 2009. – 292 стр.
6. Аванесов В.С. Научные проблемы тестового контроля знаний. – М.: Учебный центр при ИЦПКПС, 1994. – 136 с.
7. Чельшкова М.Б. Адаптивное тестирование в образовании (теория, методология, технология). – М.: Исследовательский центр проблем качества подготовки специалистов, 2001. – 165 с.

8. Янченко С.И. Математическая модель оценки результатов тестирования // Тезисы докладов Всероссийской конференции “Развитие системы тестирования в России”. – Москва, 2000.