

МЕТОДИЧНИЙ ПІДХІД ДО ВРАХУВАННЯ РІЗНИХ ТИПІВ ШКАЛ ВИМІРЮВАННЯ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

Застело Ольга В'ячеславівна,

старший викладач кафедри іншомовної підготовки, girzov@ukr.net.

Анотація. Розглядаються питання, пов'язані з впливом типу обраної шкали вимірювання результатів комплексного тестування на інтегральні оцінки рівня сформованості іншомовної комунікативної компетентності слухачів у навчальному закладі.

Ключові слова: інтегральна оцінка, результати комплексного тестування, вимірювання, типи шкал.



Концепція безперервної освіти в умовах динамічного життя суспільства є однією з провідних ідей суспільства ХХІ століття. Ще більш актуальною стає ця ідея на етапі інтенсивного розвитку міжнародних відносин з урахуванням важливої ролі знання іноземної мови. Тому питання формування високого рівня іншомовної комунікативної компетентності у слухачів вищих навчальних закладів — фахівців у різних галузях міжнародної діяльності — набуває все більшої актуальності. Отже, якість сучасної вищої освіти в країні має відповідати вимогам сьогодення. Проблемою управління якістю освіти й підготовки фахівців займався Б. К. Коломієць [3], Н. О. Селезньова [8], А. І. Субетто [10] та інші науковці. Управління якістю освіти можливе завдяки контролю рівня сформованості певної професійної компетентності майбутніх фахівців за допомогою тестування з використанням сучасних інформаційних технологій.

Тестування набуло широкого застосування в освітньому процесі середніх і вищих навчальних закладів, але існують труднощі інтерпретації тестових результатів. Важливість коректного визначення оцінки обумовлена тим, що оцінка є могутнім педагогічним інструментом, за допомогою якого педагог достатньо ефективно може впливати на навчальний процес.

В освітній практиці дослідники мають перетворювати тестові бали у звичні оцінки. Попри те, що таке переведення даних знижує дисперсію тестових результатів і погіршує диференціувальну функцію тесту, на практиці часто тестові бали мають бути представлені у звичайній чотирибальній шкалі ВНЗ. Така шкала позначок, зазвичай, критикується, однак вона має цілу низку переваг, що пояснює її широке використання. Головною перевагою чотирибальної системи є її простота, яка зумовлена обмеженою можливістю людини як вимірювального інструмента. Викладачеві досить легко визначити градації обсягу знань у межах 4–7 рівнів. А якщо ввести, наприклад, 20-бальну шкалу оцінок, то викладачу буде практично неможливо відрізнити 19 балів від 20 балів.

Результати комплексного тестування — інтегральна оцінка рівня сформованості іншомовної комунікативної компетентності (ІКК) слухачів — значною мірою відображає якість навчання. Саме тому питання

визначення типу шкали вимірювання результатів тестування потребує ретельного вивчення.

Аналіз останніх досліджень і публікацій. У висвітленні проблеми впливу обраного типу шкали вимірювання результатів тестування на оцінку слухачів зробили внесок В. Дубас [1], В. С. Ким [2], О. Г. Молибог [4], Ю. М. Нейман [5], О. І. Орлов [6]. Проте вони пропонують різні методики вибору типу шкали вимірювання результатів тестування, за якими отримані інтегральні оцінки слухачів часто не співпадають.

Метою цієї статті є виклад певних результатів дослідження щодо мінімізації впливу різних типів шкал вимірювання результатів комплексного тестування на виставлення інтегральної оцінки рівня сформованості іншомовної комунікативної компетентності слухачів ВНЗ.

Шкала вимірювання — це встановлення типу відповідності між значенням змінних і результатів вимірювань. Стивенс С. [9] розглядав чотири шкали вимірювання.

1. Шкала найменувань — найпростіша шкала вимірювання. Вона відображає ті відношення, завдяки яким об'єкти групуються в окремі неперетинні класи. У шкалі найменувань не можна визначити, чи є об'єкт більший або менший ніж інший, наскільки одиниць чи у скільки разів, можлива лише одна операція — класифікація.

2. Порядкова (рангова) шкала — це відображення відношень порядку. Єдино можливі відношення між об'єктами вимірювання — це більше/менше, краще/гірше. Виконавши вимірювання в порядковій шкалі, не можна дізнатися, наскільки одиниць чи у скільки разів відрізняються об'єкти.

3. Інтервальна шкала — це відображення відношень, зазначених не лише для шкал найменування і порядку (рангу), але й різниці між об'єктами. За інтервальною шкалою можна визначити, наскільки одиниць один об'єкт більший чи менший іншого.

4. Шкала відношень — це відображення відношень того, у скільки разів один об'єкт більший чи менший іншого.

Отже, будь-яке вимірювання виконується за допомогою шкали вимірювання. Результати вимірювання можуть бути різної якості, тобто їх можна відобразити за допомогою однієї з чотирьох шкал вимірювання. Кожна шкала обмежує застосування певних методів обробки результатів тестування з іноземної мови.

Перші два типи шкал (найменувань і порядку) належать до «слабких» шкал, оскільки присвоєння числових показників об'єктам пов'язано з мінімальними обмеженнями щодо якостей, які відображаються за допомогою шкали. Замість числових систем у цих шкалах можуть бути використані словесні системи, наприклад, під час оцінювання рівня сформованості ІКК слухачів у ВНЗ (порядкова шкала: *незадовільно, задовільно, добре, відмінно*). Ці два типи шкал залишаються інваріантними лише за загальних математичних перетворень групової структури шкал і передбачають тільки прості статистики.

Два інші типи шкал (інтервалів і відношень) належать до «сильних». Вони залишаються інваріантними за загальних лінійних перетворень (шкала інтервалів) і за перетворень подібності (шкала відношень). Відповідно, вони передбачають застосування більш складних статистик (середнє арифметичне, дисперсія, лінійна кореляція, середнє геометричне, коефіцієнт варіації).

Наразі шкала — це інструмент (прийнята система правил) оцінювання (вимірювання) будь-яких об'єктів чи явищ. У першу чергу становлять інтерес способи вимірювання, які дозволяють розташувати об'єкти на порядковій та інтервальній шкалах, тому що саме такий тип оцінювання найчастіше використовується під час проведення експертного оцінювання, зокрема у навчальному процесі.

Головною метою вимірювання у педагогіці є отримання кількісного еквіваленту ступеня наявності певної ознаки на інтервальній шкалі.

Під час математичної обробки результатів тестування з іноземної мови, перетворення результатів в оцінки за тією чи іншою процедурою, обчислення середньої оцінки, треба враховувати, що позначки визначені на порядковій шкалі. Зокрема, не можна використовувати як середню оцінку середнє арифметичне, воно застосовується для інтервальних шкал.

Складовою процедури переведення тестових балів в оцінки є таблиця відповідності певного діапазону тестових балів оцінкам або певний математичний вираз, який дозволяє визначити оцінку [7]. Різні автори пропонують різні таблиці відповідності.

Першими методами шкалювання, які почали застосовувати для вимірювання результатів тестування, були прості лінійні методи.

Розглянемо одну з таких таблиць відповідності та адекватний метод шкалювання. Наприклад, у праці В. Дубас [1] пропонується така таблиця (табл. 1).

Таблиця 1

«2»	«3»	«4»	«5»
$0,1 \leq V < 0,4$	$0,4 \leq V < 0,7$	$0,7 \leq V < 0,9$	$0,9 \leq V = 1$

V — відносний обсяг знань.

За цією таблицею В. Дубас пропонує номограму (рис. 1), яка дозволяє швидко здійснювати процедуру лінійної трансформації тестових балів в оцінки, де N — максимальна кількість балів, X — індивідуальний бал слухача.

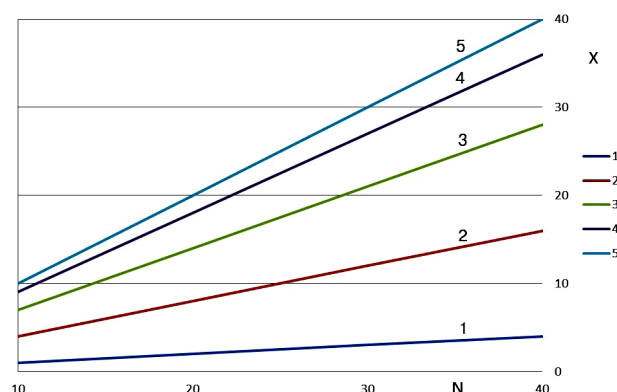


Рис. 1. Номограма для визначення оцінки

Зважаючи на широке застосування обчислювальної техніки в навчальному процесі, такі методи розрахунку виявляються досить слушними.

З часом почали широко використовувати нелінійні шкали. Одним із найзручніших способів переведення балів у оцінки є узагальнювальна функція бажаності Харрінгтона [11]. Автор реалізував ідею переведення позначок у безмірну шкалу бажаності, яка належить до психофізичних шкал. Вона виникла в результаті спостережень за реальними діями тестологів. Узагальнювальна функція бажаності Харрінгтона задана рівнянням такого вигляду:

$$d_i = \exp[-\exp(-y_i)], \quad (1)$$

де \exp — прийняте значення експоненти;

d_i — окрема бажаність, тобто значення i -го окремого показника, переведене в безрозмірну шкалу бажаності Харрінгтона, $0 < d_i < 1$;

y_i — значення i -го показника в безрозмірному вигляді, $y_i > 0$.

Стандартні оцінки на шкалі бажаності наведено у табл. 2.

Графіку бажаності Харрінгтона (рис. 2) притаманні такі якості, як безперервність, монотонність

Таблиця 2

Шкала бажаності Харрінгтона з проміжними значеннями

Бажаність / Лінгвістичні оцінки	Бальні оцінки	Позначка на шкалі бажаності
Відмінно/Дуже добре	5	1.00 – 0.80
Добре	4	0.80 – 0.63
Задовільно	3	0.63 – 0.37
Незадовільно/Погано	2	0.37 – 0.20
Дуже погано	1	0.20 – 0.00

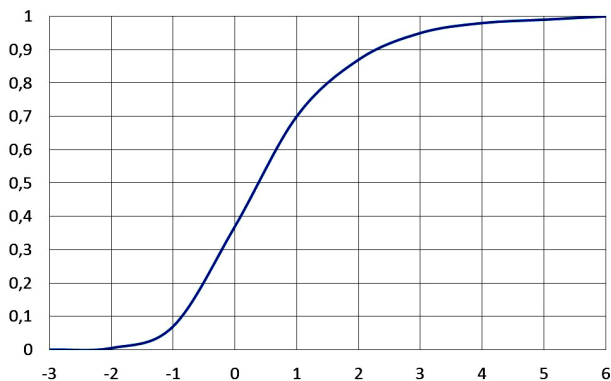


Рис. 2. Графік бажаності Харрінгтона

та гладкість. Утім, ця крива добре відображає те, що в областях бажаності, близьких до 0 та 1, чуттєвість її значно нижча, ніж у середній області.

Вітчизняні автори також використовують нелінійні типи шкал вимірювання. У своїй праці Молибог О. Г. [4] пропонує наближене *нелінійне* співвідношення такого вигляду:

$$Y = 3,3 \lg(1/(1-V)), \quad (2)$$

де Y — оцінка в балах ($Y=2, 3, 4, 5$);

V — обсяг знань матеріалу в долях від 1,

Криву, що відповідає вищезазначеній залежності, подано на рис. 3.

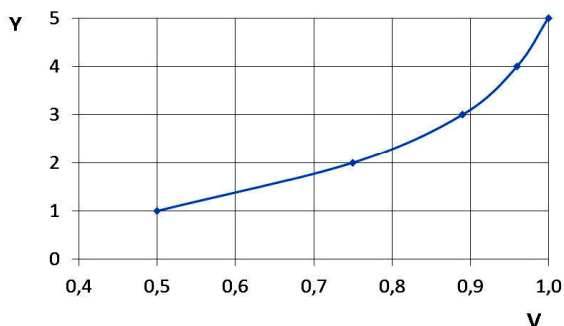


Рис. 3. Крива для визначення оцінки за надлінійною залежністю

Як можна бачити, присутні нелінійна залежність величини оцінки від відносного обсягу знань, а також високі вимоги до рівня сформованості ІКК слухачів. Зокрема, оцінці «3» відповідає значення $V=0,75$, що забагато.

За іншими таблицями тестування для переведення тестових балів у оцінки значення $V=0,75$ відповідає оцінці «5». Тобто, такі таблиці трансформації тестових балів у оцінки є суб'єктивними й відображають багато прихованих факторів. Простежимо закономірності розроблення процедури переведення тестових балів у оцінки. Відомо, що табульовані функції можна визначити простими рівняннями. Тому проаналізуємо не таблиці відповідності, а поведінку функцій, що описують ці таблиці. Надалі передбачається, що оцінка y пов'язана з індивідуальним балом X слухача нелінійною залежністю такого вигляду:

$$y = aX^n + b, \quad (3)$$

де a, b, n — коефіцієнти, які треба визначити. Серед цих коефіцієнтів розглянемо коефіцієнт n , що

відповідає типу залежності, криві для визначення оцінки за яким відображено на рис. 4.

При $n=1$ отримуємо лінійну залежність B .

Крива A відповідає випадку $n < 1$ і характеризує сублінійну залежність.

Крива C відповідає випадку $n > 1$ і характеризує надлінійну залежність.

У випадку лінійної залежності ($n=1$) спостерігається пряма пропорційна залежність між оцінками й індивідуальними балами слухача. Це — найпростіша залежність, але на практиці вона рідко застосовується.

Значення $n < 1$ (сублінійна залежність) показує, що дослідник у першу чергу цікавиться підвищеною диференціальною здатністю використаної шкали в області низьких оцінок. Надлінійні залежності, де $n > 1$, застосовуються, коли прагнуть підвищити диференціовальну здатність шкали вимірювання в області високих оцінок. Отже, сублінійні залежності слід застосовувати для тестів, які містять завдання підвищеної складності. Тоді переважна кількість слухачів отримуватиме відносно низькі індивідуальні бали. Відповідно, для тестів, які містять легкі завдання, слід застосовувати надлінійні залежності.

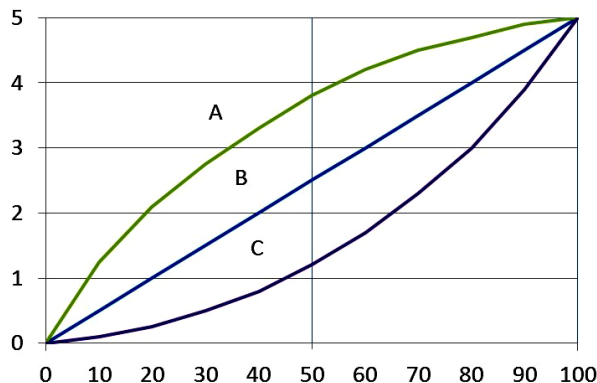


Рис. 4. Криві для визначення оцінки за сублінійною, лінійною і надлінійною залежностями

Отже, під час розроблення процедури переведення тестових балів у оцінки слід використовувати сублінійні або надлінійні залежності, зважаючи на характеристики тесту і слухачів.

Експериментальне порівняльне оцінювання цих методів у процесі дослідження для різної кількості об'єктів підтвердило цей висновок.

З огляду на викладене, підвищити стійкість результату можна за рахунок отримання інтегральної оцінки за результатами розв'язання задачі вибору всіма можливими методами з наявних, за умови, якщо такий спосіб оброблення результатів відомий. Вибір найбільш раціональної шкали вимірювання результатів тестування з іноземної мови потребує додаткових досліджень.

Розглянемо за допомогою спеціалізованого програмного засобу VERDICT, як різні типи шкал вимірювання впливають на інтегральну оцінку рівня сформованості ІКК слухачів. Спеціалізований програмний засіб VERDICT у цьому експерименті застосовується в режимі ітераційної процедури обчислення інтегральної оцінки з використанням 9 різних типів згорток. Блок-схему методу наведено на рис. 5.

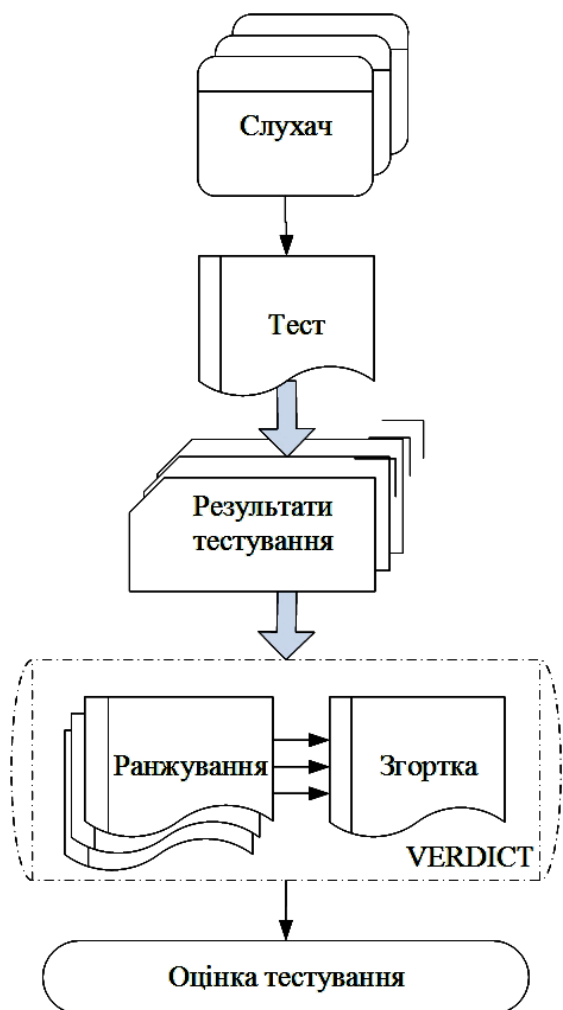


Рис. 5. Блок-схема методу застосування різних методів шкалювання результатів тестування

Програма VERDICT враховує 10 різних типів шкал вимірювання (табл. 3), що підвищує достовірність результатів тестування.

На рис. 6, 7, 8 зображено бали, які отримали слухачі за різними типами шкал (логарифмічною, відносно середнього, ненормованою), і результати інтегрального оцінювання рівня сформованості ІКК слухачів, обчислені за допомогою програми VERDICT.

Файл	Настройка	Модель	Выбор	Классификация	Опции
Модель качества					
1	1	1	1	1	1
stud_2	0.68	0.79	0.50	0.58	0.43
stud_3	0.00	0.50	0.79	1.00	0.68
stud_4	1.00	1.00	0.79	0.00	0.43
stud_5	0.00	0.50	0.79	1.00	0.68
stud_6	0.68	0.50	0.00	0.00	0.43
stud_7	0.00	0.50	0.79	1.00	1.00
stud_8	1.00	1.00	0.79	0.00	0.43
stud_9	0.00	0.00	0.50	0.00	0.68
stud_10	0.68	0.79	1.00	1.00	0.68

№ з/п	Типи шкал вимірювання програми VERDICT
1	Ненормована шкала
2	Пропорційна шкала
3	Пропорційно-ступенева шкала
4	Логарифмічна шкала
5	Логарифмічно-ступенева шкала
6	Шкала відносно середнього
7	Шкала відносно екстремуму
8	Шкала до одиниці
9	Рівнева шкала
10	Вибіркова шкала

Як можна бачити на рис. 6, за логарифмічною шкалою вимірювання найвища інтегральна оцінка у слухача №7. На рис. 7 відображено, що найвищий результат тестування з використанням шкали вимірювання відносно середнього у слухача №2. На рис. 8 продемонстровано, що за ненормованою шкалою вимірювання найвищий результат тестування у слухача №8.

Аналіз наведених результатів свідчить, що інтегральна оцінка рівня сформованості ІКК слухачів значною мірою залежить від застосованої шкали вимірювання результатів тестування.

Проведене дослідження дає підстави для висновків про те, що в теперішній час для оцінки результатів комплексного тестування переважно використовується пропорційний тип шкали. Використання сублінійних і надлінійних шкал вимірювання певною мірою розширює методи шкалювання результатів.

Для підвищення точності оцінювання під час комплексного тестування слухачів пропонується використовувати різні типи шкал вимірювання. Вибір певного методу шкалювання має відповідати виду тестування, що може бути напрямом подальших наукових досліджень. Для вимірювання результатів тестування пропонується використовувати розгорнутий перелік 10 типів шкал програмного засобу VERDICT, що дозволяє більш точно врахувати специфіку

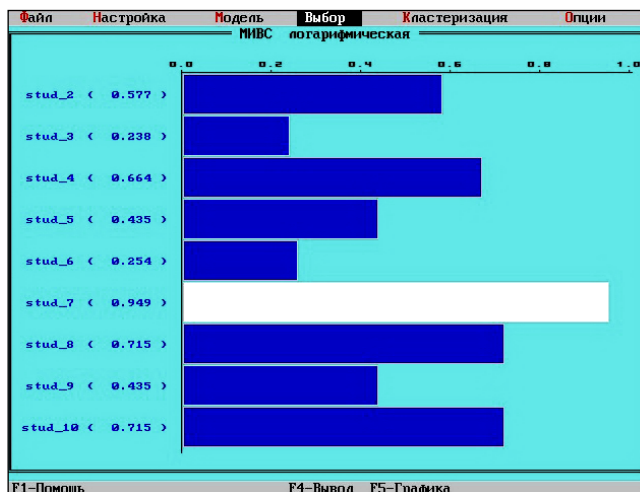


Рис. 6. Результат тестування, обчислений за логарифмічною шкалою вимірювання з використанням програми VERDICT

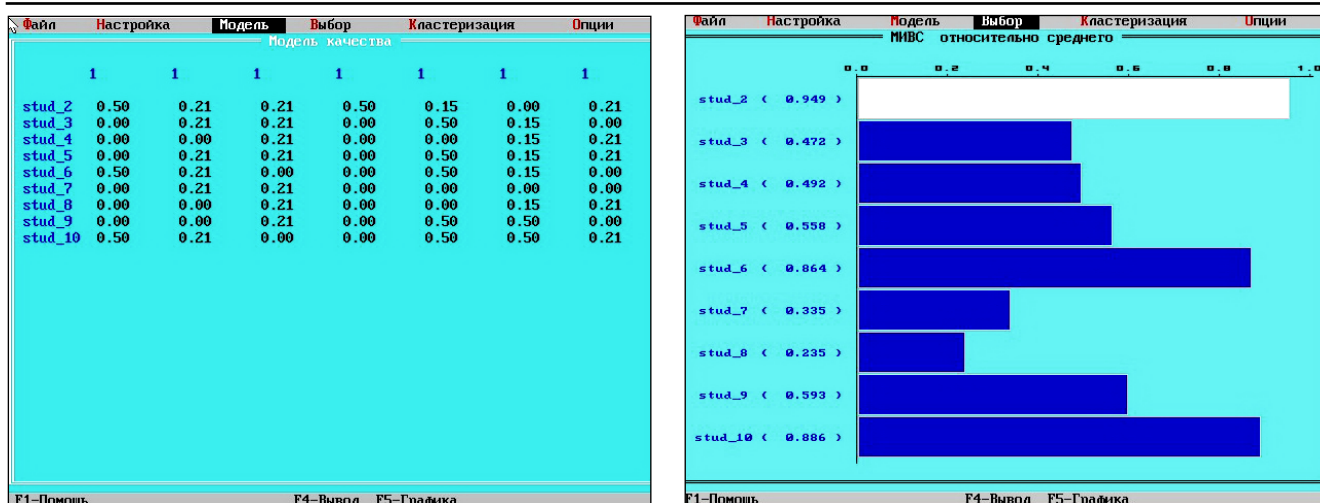


Рис. 7. Результат тестування, обчислений за шкалою вимірювання відносно середнього з використанням програми VERDICT

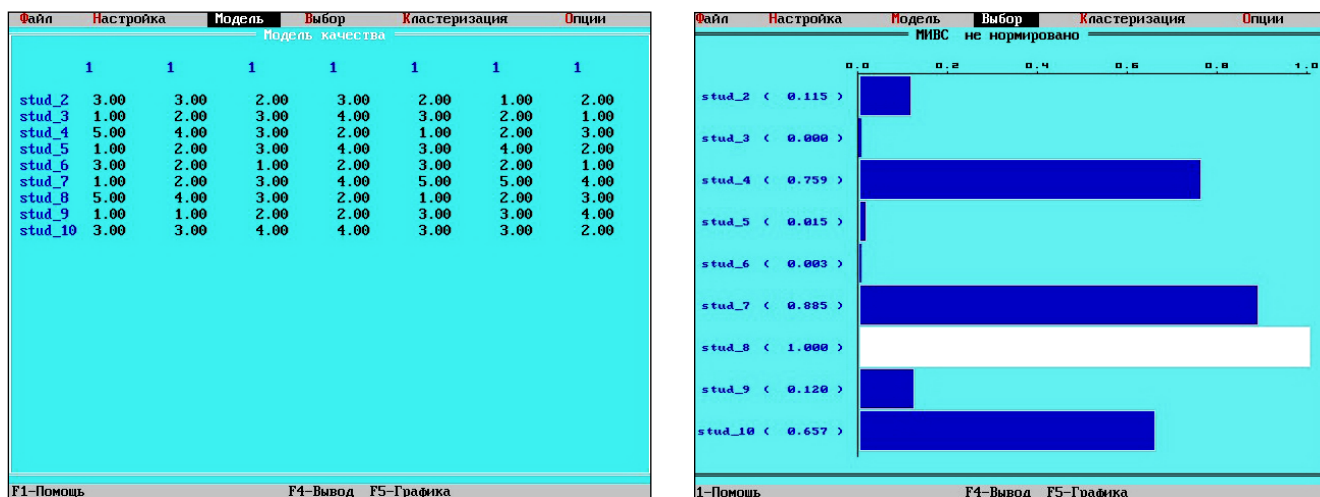


Рис. 8. Результат тестування, обчислений за ненормованою шкалою вимірювання з використанням програми VERDICT

різних тестів. Особливістю цієї програми також є можливість вимірювати результати різних видів тестів за різними типами шкал як у сукупності, так і окремо, залежно від доцільності певного типу шкали.

* * *

Застело О. В. Методический подход к учету разных типов шкал измерения результатов тестирования

Аннотация. Рассматриваются вопросы, связанные с влиянием типа выбранной шкалы измерения результатов комплексного тестирования на интегральные оценки уровня сформированности иноязычной коммуникативной компетентности слушателей в учебном заведении.

Ключевые слова: интегральная оценка, результаты комплексного тестирования, измерения, типы шкал.

* * *

Zastelo O.V. Technical approach to the account of different measurement scales of the testing results.

Abstract. This article studies the issues connected with the influence of a certain type of measurement scale of the complex testing results on the integral esteems of students' foreign language competency level.

Key words: integral esteem, results of complex testing, measurements, scale types.

Література

1. Дубас В. Об оценивании знаний при программированном контроле / В. Дубас // Физика в школе. — 1990. — №3. — С. 83.

2. Ким В. С. Тестирование учебных достижений : монография / В. С. Ким. — Уссурийск, 2007. — 214 с.
 3. Коломиец Б. К. Управление качеством образования: инвариантные аспекты / Б. К. Коломиец. — М. : Исследовательский центр проблем качества подготовки специалистов, Московская государственная технологическая академия, 2003. — 60 с.
 4. Молибог А. Г. Технические средства обучения и их применение / А. Г. Молибог, А.И. Тарнопольский. — Минск, 1985. — 208 с.
 5. Нейман Ю. М. Об оценивании уровня подготовленности учащихся по результатам единого государственного экзамена / Ю. М. Нейман. — М. : Poligraph, 2002. — 30 с.
 6. Орлов А. И. Теория измерений и педагогическая диагностика / А. И. Орлов // Педагогическая информатика. — 2004. — №1. — С. 22–31.
 7. Рекомендации по переводу тестового балла централизованного тестирования (вузовского) в пятибалльную шкалу оценок в 2005 году [Электронный ресурс]. — Режим доступа : <http://www.rustest.ru/test/scale100in5.php>.
 8. Селезнева Н. А. Качество высшего образования как объект системного исследования / Н. А. Селезнева. — М. : Исследовательский центр проблем качества подготовки специалистов, 2004. — 95 с.
 9. Стивенс С. С. Экспериментальная психология / под ред. С. С. Стивенса. — М., 1963. — 1038 с.
 10. Субетто А. И. Оценочные средства и технологии аттестации качества подготовки специалистов в вузах: методология, методика, практика / А. И. Субетто. — СПб.; М. : Исследовательский центр проблем качества подготовки специалистов, 2004. — 68 с.
 11. Harrington E. C. The desirable function / E. C. Harrington. — 1965. — №21, Vol. 10. — P. 124–131.