

ТЕСТУВАННЯ ЗІ ШКІЛЬНОГО КУРСУ ІНФОРМАТИКИ: ІНТЕРПРЕТАЦІЯ РЕЗУЛЬТАТІВ

Статтю присвячено висвітленню проблеми тестування зі шкільного курсу інформатики та інтерпретації його результатів.

The article is dedicated to the coverage of testing in school course of informatics problem and its results interpretation.

Сучасні тенденції у вітчизняній системі освіти вимагають перегляду технології і методики педагогічної діагностики школярів з метою отримання об'єктивних оцінок їх навчальних досягнень та діагностики навчального процесу в цілому. Найбільш передовою технологією у цьому сенсі є тестування. За рубежом дослідження різних аспектів педагогічного тестування проводяться ще з початку минулого століття і пов'язані з такими іменами як Ф. Лорд, Дж. Раш, Дж. Стэнли и Дж. Глас, Б. Райт та інші. Створені спеціальні установи і лабораторії, які займаються дослідженням і створенням педагогічних тестів. В Україні дослідженнями у цьому напрямку займаються І.Е. Буллах, Л.М. Гриневич, С.А.Раков, О.І.Ляшенко та ін. Вже є результати експериментального зовнішнього тестування школярів. Але тестування є дуже складною задачею. Одним із досить важливих питань при цьому є коректна інтерпретація результатів.

У Кримському інженерно-педагогічному університеті впродовж двох років проводився педагогічний експеримент по використанню тематичних тестів як засобу педагогічної діагностики школярів з інформатики. Одним із важливих завдань цього експерименту є отримання об'єктивних оцінок, які можна порівняти і зробити висновки про індивідуальні навчальні досягнення учнів. В даному випадку розглядається оцінка навчальних досягнень з різних розділів шкільного курсу інформатики. Метою цієї статті є розгляд способів вираження тестових балів за допомогою різних шкал.

Якщо розглядати тестові завдання з множинним вибором (тобто закриті), які використовуються в більшій частині зарубіжних тестів і були використані, зокрема, нами при проведенні педагогічного експерименту, то правильну відповідь учня можна оцінювати в 1 бал, а неправильну – в 0 балів. Таким чином, бал учня є кількістю правильних відповідей (сирий бал). Проте такий сирий бал сам по собі ніякої істотної інформації не несе, оскільки він не враховує складність тестових завдань. Наприклад, при проведенні нами тематичних тестів один і той же учень отримував 20 балів з розділу “Алгоритмізація і програмування” і 20 балів з розділу “Електронні таблиці”. Це зовсім не свідчить про те, що цей учень в однаковій мірі володіє програмуванням і прийомами роботи з електронними таблицями MS Excel. Таким чином, необхідна похідна шкала, яка дозволить врахувати складність завдань даного тесту і порівняти результати декількох учнів.

Відомо, що розрізняють чотири типи шкал: номінальна, порядкова, інтервальна і відносна [1: 26]. Номінальна шкала дозволяє віднести людей або предмети за якою-небудь ознакою до певної категорії. Проте не можна сказати, що одна категорія вища або краща за іншу. Порядкова ж шкала дозволяє не тільки розділити на категорії, але і розставити їх у певному порядку. Але про ці категорії не можна сказати наскільки одна з них відрізняється від іншої, тобто інтервали між різними сусідніми категоріями можуть не співпадати [2]. Інтервальна шкала, включаючи властивості перших двох, припускає ще і рівні інтервали між сусідніми категоріями, але не має абсолютного нуля. Нарешті, відносна шкала має нульову точку. Іншими словами, вона дозволяє, крім всього зазначеного, зробити висновок про те, у скільки разів одне значення вище за інше.

Наприклад, прийняті у вузах оцінки можна віднести до порядкової шкали: “незадовільно”, “задовільно”, “добре”, “відмінно”. Не можна сказати, що інтервал між

оцінками “незадовільно” і “задовільно” такий самий, що й інтервал між оцінками “добре” і “відмінно”, проте оцінка “незадовільно” безумовно є найменшою по порядку.

Якщо говорити про педагогічні тести, то тут до порядкової шкали відноситься шкала процентілей, яка є однією з найпоширеніших. Зокрема, ця шкала використовувалася при проведенні Центром тестових технологій зовнішнього тестування в Україні [3].

Бал за цією шкалою виражає відсоток учнів в даній групі, сирий бал яких менший, ніж у даного учня. Наприклад, бал одного з учнів, отриманий за тематичний тест “Інформація і інформаційні процеси”, складає 65 процентілей. Це означає, що 65% учнів в його групі отримали менший сирий бал, а 35% учнів справилися з даним тестом краще за нього. Така шкала дозволяє розставити учнів у групі по порядку, залежно від рівня їх навчальних досягнень на даний момент.

Перетворення сирих балів в процентілі проводиться таким чином [4]:

$$P = \frac{fr + 0.5fr_i}{n} \cdot 100\%,$$

де fr – накопичена частота всіх балів, менших i -го;

fr_i – частота i -го бала;

n – кількість учасників тестування.

Наприклад, на малюнку 1 представлені результати тематичного тестування з розділу шкільної інформатики “Електронні таблиці”, представлені у вигляді сирих балів (r) і в процентілях (P).

A		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
B →	Ex001	0	1	1	1	1	0	1	0	0	0	1	1
	Ex002	1	1	1	1	0	1	1	1	0	1	1	1
	Ex005	1	1	1	1	1	1	1	1	0	1	1	1
	Ex006	0	0	0	0	1	1	0	0	0	0	0	0
	Ex007	1	0	0	0	0	0	0	1	0	0	0	0
	Ex009	0	0	0	0	0	0	0	1	0	0	0	0
	Ex010	0	1	0	0	1	0	1	1	0	1	0	0
	Ex011	0	1	0	0	1	0	0	0	0	0	0	0
	Ex012	0	1	0	0	0	1	0	1	0	1	0	0
	Ex004	0	0	0	0	0	1	1	0	0	0	0	0
	Ex012	0	0	0	1	0	0	0	1	0	1	1	0
	Ex015	0	0	0	0	0	1	0	0	0	0	0	0
	Ex020	0	0	0	0	0	1	1	0	1	0	0	0
	Ex028	0	0	0	0	0	0	0	0	0	0	0	0
	Ex031	0	0	0	1	0	1	1	1	0	0	1	1
	Ex034	0	0	0	0	0	0	0	0	0	1	0	0
C →	r	7	10	8	7	7	12	11	14	5	10	10	7
D →	P	33	63	40	33	33	80	75	94	14	63	63	33

Мал.1. Результати тестування, виражені у шкалі процентілей.

A – номер учня

B – шифр завдання

C – сирі бали учнів

D – бали, виражені в процентілях

Якщо порівнювати результат за шкалою процентілей окремого учня з результатами контрольної групи з цього ж тесту, то можна судити про його індивідуальну успішність. При проведенні ряду тестів протягом року (або більшого періоду) можна судити про здібності до навчання окремого учня, про його досягнення за цей період. Наприклад, в навчальних закладах Великобританії використовують таке поняття як додаткова вартість, яку отримують

при порівнянні результатів тесту учнів конкретної школи з середнім результатом інших шкіл країни [5; 6]. При цьому можна, наприклад, порівняти, наскільки рівень навчальних досягнень учнів даної школи вище або нижче, ніж середній рівень по країні. Проте бали такого роду мають свої обмеження, оскільки дають інформацію в межах однієї школи або міста, або області і так далі залежно від контрольної групи. Складно створити таку контрольну групу в масштабах держави і тим більше робити висновки про те, що учень однієї школи навчився за даний період більшому, ніж учень іншої школи. Цілий ряд чинників впливає на успішність учнів, зокрема з дисципліни “Інформатика”: провінційна школа або столична, оснащеність комп'ютерною і комунікаційною технікою, кваліфікація вчителя і т.д.

У зв'язку зі специфікою порядкової шкали у школярів може виникнути помилкове розуміння свого результату. Наприклад, в [4] вказується така особливість процентілей, що при порівнянні результатів на різних вибірках різниця сирих балів в одиницю може викликати різну величину різниці в процентілях. Такі особливості шкали процентілей, і взагалі порядкової шкали, необхідно враховувати як педагогічним працівникам, які використовують результати тестування, так і учням.

При статистичному аналізі результатів педагогічного тесту досить часто використовуються стандартизовані бали, які виражаються в інтервальній шкалі. Тобто, якщо бал одного учасника тестування складає 100, іншого – 150, а третього – 200, то інтервал між результатами першого і другого рівний інтервалу між результатами другого і третього. Тобто, можна сказати, що другий випробовуваний підготовлений краще першого на стільки ж, наскільки третій – краще другого. Проте якщо учень отримує 0 балів, це ще не означає, що він взагалі нічого не знає. На основі різниці між балами учнів (якщо вона перевищує стандартну помилку вимірювання) можна судити лише про відмінності в рівні їх навчальних досягнень.

Для перетворення сирих балів у шкалу стандартизованих балів обчислюється середнє значення сирого балу і стандартне відхилення розподілу сирих балів, після чого результуючий бал обчислюється за формулою:

$$z = \frac{r_i - \bar{r}}{SD},$$

де r_i – сирий бал i -го учня;

\bar{r} – середнє значення сирого балу;

SD – стандартне відхилення розподілу сирих балів.

На малюнку 2 представлені результати згаданого вище тестування, виражені у шкалі стандартизованих балів.

У [4] вказуються наступні властивості шкали: абсолютне значення похідного балу відображає відхилення сирого балу від середнього значення вибірки. Негативні значення відповідають сирим балам, які опинилися менше середнього значення, а позитивні – сирим балам, які перевищують його. Оскільки шкала є інтервальною, то з отриманими балами можна проводити арифметичні операції.

З погляду учня або вчителя бали в стандартизованій шкалі мають наступну перевагу перед сирими балами. Припустимо, учні пройшли три тематичні тести з інформатики: “Операційна система”, “Електронні таблиці”, “Текстовий редактор”. Учень А в першому тесті набрав 26 балів, у другому – 13 балів, а в третьому – 9. Виникає питання: Як порівняти ці результати? Який можна зробити висновок? Грунтуючись лише на сирих балах, легко помилитися. Наприклад, чи можна сказати, що учень А засвоїв розділ “Операційна система” краще, ніж розділ “Електронні таблиці”? Для цього необхідна загальна шкала, яка дозволила б порівняти результати всіх трьох тестів для одного учня, а також результати учнів між собою. Якщо перевести сирі бали в стандартизовану z -шкалу, то отримуємо наступну картину для учня А:

“Операційна система” – -0,22 балів

“Електронні таблиці” – 1,17 балів

“Текстовий редактор” – -1,03 балів

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Ex001	0	1	1	1	1	0	1	0	0	0	1	1
Ex002	1	1	1	1	0	1	1	1	0	1	1	1
Ex005	1	1	1	1	1	1	1	1	0	1	1	1
Ex006	0	0	0	0	1	1	0	0	0	0	0	0
Ex007	1	0	0	0	0	0	0	1	0	0	0	0
Ex009	0	0	0	0	0	0	0	1	0	0	0	0
Ex010	0	1	0	0	1	0	1	1	0	1	0	0
Ex011	0	1	0	0	1	0	0	0	0	0	0	0
Ex012	0	1	0	0	0	1	0	1	0	1	0	0
Ex004	0	0	0	0	0	1	1	0	0	0	0	0
Ex012	0	0	0	1	0	0	0	1	0	1	1	0
Ex015	0	0	0	0	0	1	0	0	0	0	0	0
Ex020	0	0	0	0	0	1	1	0	1	0	0	0
Ex028	0	0	0	0	0	0	0	0	0	0	0	0
Ex031	0	0	0	1	0	1	1	1	0	0	1	1
Ex034	0	0	0	0	0	0	0	0	0	1	0	0
г	7	10	8	7	7	12	11	14	5	10	10	7
P	33	63	40	33	33	80	75	94	14	63	63	33
A → Z	-0,57	0,30	-0,28	-0,57	-0,57	0,88	0,59	1,46	-1,15	0,30	0,30	-0,57
B → $\bar{r}_{\text{ср}}$	9,0		$\bar{r}'_{\text{ср}}$	0								
C → SD	3		SD'	1								

Мал. 2. Результати тестування, виражені у шкалі стандартизованих балів

A – стандартизовані бали

B – середнє значення сирого балу

C – стандартне відхилення розподілу сирих балів

D – середнє значення стандартизованого балу

E – стандартне відхилення розподілу стандартизованих балів

Тепер можна сказати, що середній результат учень А отримав за тест “Операційна система”, оскільки його бал близький до 0, тобто до середнього балу вибірки. Найкращий результат він показав з розділу “Електронні таблиці”, оскільки його бал вище середнього. Найгірше учень А склав тест з розділу “Текстовий редактор”.

Проте з однієї і тієї ж дисципліни або розділу дисципліни можуть бути різні тести, тобто різні набори тестових завдань. Один тест може виявитися легше за інший. У спеціалізованому Інтернет-виданні “Rasch Measurement Transactions” указується, що оцінки, отримані об’єктивними вимірюваннями, повинні залишатися постійними і незмінними (в межах допустимої помилки) для всіх груп учасників тестування, при використанні будь-якого інструментарію і незалежно від того, хто проводив ці вимірювання [7]. Стандартизована шкала дозволяє порівняти результати випробовуваних, отримані на різних тестах. В ході педагогічного експерименту, проведеного автором статті, учням було запропоновано два тести з розділу шкільного курсу інформатики “Інформація і інформаційні процеси”. Назвемо їх умовно тест А і тест Б. Тест А виявився легшим за другий тест. У такому разі 13 балів, отриманих учасником тестування при проходженні тесту А не можна дорівняти 13 балам, які були отримані при проходженні тесту Б. Переходячи до стандартизованої шкали, стає видно: -0,11 балів відповідає 13 сирим балам у складнішому тесті А і 19 сирим балам у легшому тесті Б (мал.3). Тобто, щоб отримати однаковий результат, учень повинен в тесті А виконати правильно на 6 завдань більше, ніж у більш складному тесті Б.

A →	30	29	28	28	28	28	26	25	25	20	19	19
B →	2,03	1,84	1,64	1,64	1,64	1,64	1,25	1,06	1,06	0,08	-0,11	-0,11
	середнее											
										19,58		
										SD		
										5,123	Тест А	
A →	25	22	20	23	24	25	8	10	8	13	10	7
B →	1,86	1,37	1,04	1,53	1,69	1,86	-0,93	-0,60	-0,93	-0,11	-0,60	-1,09
	середнее											
										13,66		
										SD		
										6,106	Тест Б	

Мал. 3. Результати тесту А і тесту Б.

A – сирі бали

B – стандартизовані бали

Бали, виражені в стандартизованій шкалі, піддають подальшим перетворенням з тим, щоб позбавитися від негативних значень і десяткових чисел, які складні для сприйняття учнями і педагогами, що не володіють математичним апаратом теорії педагогічних вимірювань. Нові бали отримують наступним лінійним перетворенням:

$$z' = k_1 * z + k_2,$$

де z – стандартизований бал, а k_1 і k_2 – коефіцієнти.

Наприклад, для переведення в шкалу College Entrance Examination Board (СЕЕВ), розроблену в США, використовують коефіцієнти k_1 і k_2 , які складають відповідно 100 і 500. Таким чином, бал -0,57 у стандартизованій шкалі відповідає балу 443 за шкалою СЕЕВ. Ще одна поширена в педагогічних вимірюваннях шкала – Т-шкала, розроблена в 1920 році. В ній стандартне відхилення дорівнює 10, а середнє значення сирого балу беруть рівним 50.

Коефіцієнти 100 і 500, відповідно стандартне відхилення і середній бал, використовуються в таких відомих тестах як SAT и GRE (Graduate Entrance Examination), які проводяться кожного року організацією ETS (Educational Testing Service).

Таким чином, тестування, що регулярно проводиться, дозволить виявити здатність окремих учнів до навчання і забезпечить індивідуалізований підхід до учнів. Використання тестів у школі як засобу педагогічної діагностики дозволить отримати об'єктивну оцінку успішності учнів, а також об'єктивну оцінку навчального процесу. Вибір тієї або іншої шкали залежить від мети тестування. Використовуючи шкалу процентілей, можна виявити, наприклад, найбільш слабо підготовлених учнів, порівняти знання деякого питання своїми учнями і учнями іншої школи. Провівши тематичні тестування серед учнів (або студентів), можна перетворити бали в стандартизовану шкалу, яка дозволить порівняти рівень навчальних досягнень кожного учасника тестування з різних розділів дисципліни, а також порівняти їх між собою навіть при тому, що вимірювання проводилися на різних наборах тестових завдань. При використанні стандартизованої шкали можна визначити коефіцієнт перетворення сирого балу з тим, щоб розташовувати на одній шкалі тести з різними наборами завдань, з різними параметрами складності (за умови, що ці тести дозволяють вимірювати одні і ті ж знання).

Подальше дослідження надасть змогу розробити методику тематичного тестування школярів з інформатики, яка зокрема передбачатиме порядок шкалування тестових балів та їх інтерпретацію.

ЛИТЕРАТУРА:

1. Ингенкамп К. Педагогическая диагностика. Пер. с нем. – М.: Педагогика, 1991. – 240с.
2. Encyclopedia of educational technology. – <http://coe.sdsu.edu/eet>.

3. Гриневич Л. Зовнішнє тестування як засіб підвищення якості освіти в Україні в умовах децентралізації управління освітою// Шлях освіти. – 2004. – №4(34). – С. 24-28.
4. Rodríguez M. Norming and Norm-referenced Test Scores. – <http://ericae.net>.
5. Value added. – <http://www.teachernet.gov.uk>.
6. How value added works. – <http://news.bbc.co.uk/1/hi/education/3401219.stm>.
7. Definition of Objective Measurement. – <http://www.rasch.org>.