

Проблема становлення національної системи педагогічних вимірювань

В.С. Аванесов, доктор педагогічних наук, професор, головний редактор журналу «Педагогічні вимірювання»

(За матеріалами доповіді на II Міжнародній конференції «Вимірювання навчальних досягнень учнів і студентів: гуманістичні, методологічні, технологічні аспекти» (м. Харків, 13—14 грудня 2007 року))

Сьогодні в Україні є широкі можливості і велика потреба в обговоренні проблеми становлення Національної системи педагогічних вимірювань. Проте не можна не помітити, що водночас дискутується ідея проведення в Україні єдиного державного іспиту. Навряд чи варто пояснювати, що іспит є нетехнологічною формою оцінювання знань, яка має стародавню історію, що налічує понад чотири тисячі років. Ефективне вимірювання на підставі іспитів є неможливим, і застосовувати їх для такої цілі безглуздо. Для одержання якісних вимірювань використовують справжні тести. Мета даної доповіді — дати визначення ключових понять: *іспити, тестування, педагогічні вимірювання* і навести аргументи на користь створення в Україні системи тестування та її подальшої трансформації в Національну систему педагогічних вимірювань.

Вихідні визначення

Іспит — це традиційна форма педагогічного контролю рівня підготовленості. Характеризується високими витратами часу, суб'єктивізмом, нетехнологічністю, оцінками невідомого рівня надійності й валідності. В масових процесах дуже витратний метод так званої парної педагогіки, і нерідко є корупційним методом, якщо використовується для прийому у ВНЗ. Тому не застосовується в країнах, які мають розвинені системи освіти.

Тест визначається як система паралельних завдань рівномірно зростаючої складності, яка дозволяє оцінити структуру і якісно виміряти рівень підготовленості випробовуваних. Особливо ефективний для застосування у масових процесах завдяки короткочасності та економічній ефективності. Добре проведений тест виключає можливість виникнення корупції, що у сукупності з критеріями об'єктивності та справедливості пояснює широке застосування тестів в розвинених країнах, які мають також і найнижчі рівні корупції. Педагогічне тестування визначається переважно як практика застосування тестів для оцінки рівня і структури підготовленості учнів та студентів. Приблизно до середини 50-х років XX століття наука про розробку тестів називалася частіше «Теорією тестів», ніж теорією педагогічних вимірювань. І лише з початку XXI століття як наукова основа педагогічних вимірювань майже повсюдно затвердилася не «теорія тестів», а «теорія педагогічних вимірювань». На жаль, обидві ці теорії виявилися переважно статистичними і математичними, а не педагогічними теоріями. Про те,

що ми маємо справу саме з математико-статистичними теоріями, було чітко сказано в класичній роботі кінця шістдесятих років минулого століття. Система педагогічного тестування — це організація збору і обробки даних за допомогою спеціального застосування тестів на множині випробовуваних. Ця система може виявитися науково обґрунтованою, не обґрунтованою і навіть антинауковою. Науковою основою системи тестування повинна стати теорія педагогічних вимірювань. Звідси легко висувати, що система тестування може бути науково обґрунтованою лише настільки, наскільки центри тестування спираються на цю теорію. Зараз вони спираються на неї недостатньою мірою.

Педагогічні вимірювання — це наукова теорія, що сформувалася на стику педагогіки, психології, статистики, математики, логіки і філософії. Водночас *педагогічне вимірювання* є процесом, націленим на створення наукових методів отримання об'єктивних оцінок поточної і підсумкової підготовленості учнів та студентів. F.M.Lord і M.Novick визначають вимірювання як таке присвоєння чисел, яке правильно відображає розташування випробовуваних на числовій шкалі в залежності від вираженості у них вимірюваної якості.

У наш час істотна ознака педагогічних вимірювань — це педагогічно обґрунтований відбір змісту тесту, точний підбір форми завдань для того або іншого змісту, система основних педагогічних понять цієї теорії, отримання даних про випробовуваних і завдання на інтервальній шкалі. Останнє досягається використанням відповідних алгоритмів трансформації початкових результатів практичного тестування за допомогою комп'ютерних програм типу RUMM 2020, Winsteps та ін. У такому тлумаченні початкові результати тестування вже не можуть розглядатися як педагогічні вимірювання; вони виявляються лише першим етапом збору даних для подальшого логарифмічного шкалювання. Можна сказати так: тестування — це незавершене педагогічне вимірювання. У новітній літературі тільки шкальовані результати тестування визнаються повноцінними педагогічними вимірюваннями. Багаторічне відставання педагогічних вимірювань від потреб практики тестування і фактична відсутність учених у штаті більшості центрів тестування на теренах СНД перетворюють практику тестування на зону постійного ризику отримання ненадійних і невалідних результатів. Цей висновок легко підтверджується незалежною експертизою роботи будь-якого центру тестування. Вивести тестування із зони перманентного ризику отримання помилкових оцінок особи може тільки стабільна опора на науку — теорію педагогічних вимірювань. Але для цього в центрах тестування потрібні наукові підрозділи, яких практично немає. Служби ETS у США та Європі є винятком, який підтверджує загальний стан тестування, що зазвичай існує поза необхідним йому науковим забезпеченням. Усі існуючі у країнах колишнього СРСР центри тестування — практичні організації, створені на незначних бюджетах освітніх відомств. Вони, можна сказати, приречені на виробництво неякісних, псевдотестових оцінок, якщо не одержать статусу центрів педагогічних вимірювань, а також ресурсів і мотивів для залучення до співпраці справжніх учених.

Наукова основа системи педагогічного тестування

Проблему ускладнює стан самої теорії педагогічних вимірювань. Усі відомі в цій теорії методи обґрунтування якості тестових результатів незмінно виявлялися статистичними і математичними. Так склалося історично, таким був ідеал науковості тестових результатів. Проте спроби побудови власне педагогічної теорії педагогічних вимірювань, пошуки інших ідеалів науковості не припинялися ні в минулому, ні у наш час. Відомі спроби створення педагогічної теорії педагогічних вимірювань у США. У Росії проблемою створення педагогічної теорії педагогічних вимірювань займався автор даної доповіді. Ідея необхідності переходу від тестування до системи педагогічних вимірювань виникла як результат порівняльного дослідження питань методології педагогічних вимірювань і тестування, наукового аналізу практики централізованого тестування, введення російського Єдиного державного іспиту (ЄДІ) та особистого досвіду взаємодії з практиками — працівниками центрів тестування. Останні частіше прагнуть виконувати вказівки зверху, ніж рекомендації вчених. До того ж багато практиків вважають себе вченими не меншою мірою, ніж інші. Такого роду менталітет характерний і для деяких урядовців міністерств, що мають той чи інший науковий ступінь. Як педагогічна основа теорії педагогічних вимірювань автором доповіді пропонуються:

- концепція п'яти етапів тестового педагогічного процесу і система основних понять педагогічних вимірювань;
- формулювання цілей педагогічних вимірювань;
- результати дослідження змісту, форм і принципів композиції тестових завдань;
- математичні моделі та аксіоматика теорії педагогічних вимірювань.

Цілі національних систем

Україна за мету становлення Національної системи тестування ставить боротьбу з корупцією. За наявними даними, за рівнем корупції сфера освіти поступається лише міліції та митниці і стоїть на рівні податкової служби та судів. Шлях боротьби з таким станом речей - стовідсоткове тестування в усіх навчальних закладах. Україна зараз на шляху до цієї мети: у 2006 році тестування пройшли 44 тисячі абітурієнтів, в 2007 — 116 тисяч.

Президент Росії В. Путін на перше місце ставить мету забезпечення рівного доступу молоді до якісної освіти. Боротьба з корупцією — на другому місці. Важливо відзначити, що як засіб досягнення обох цілей він називає не тестування, а Єдиний державний іспит. У результаті в Росії виконується саме ця політична воля. Замість якісного тестування проводиться неякісний і витратний Єдиний

державний іспит. Треба пояснити, що в ЄДІ немає тестів. Там є так звані контрольні-вимірювальні матеріали (КВМ). У першій і другій частині цих КВМ використовуються зовні тестоподібні завдання невизначеної якості, а в третій частині містяться екзаменаційні завдання поза межної труднощі, за виконання яких суб'єктивним, комісійним методом виставляються підвищені бали. За даними автора цієї доповіді, по багатьох територіях РФ ці бали корелюють негативно з балами за першу і другу частини ЄДІ. Інтерпретувати це можна як масові факти цілеспрямованих спотворень результатів єдиного державного методу. Використання КІМ-ЄДІ замість тестування і методів педагогічних вимірювань є прямим шляхом в освітній глухий кут і до зростання корупції у сфері освіти. В результаті Росія виявилася на 143 місці за рівнем некорупційності. За даними ООН, у Росії щорічна загальна сума хабарів за вступ до ВНЗ складає 520 мільйонів доларів. Уведення ЄДІ породило нові витки корупції, тепер уже навколо нього самого.

Президент Казахстану Н. Назарбаєв поставив завдання входження його країни до числа 50-ти найрозвиненіших країн світу. Багато що вже зроблено, але гальмують системи освіти й охорони здоров'я. У країні створене і функціонує Єдине національне тестування (ЄНТ), здатне видати підсумкові результати в день проведення тестування, причому за 3—5 тестами відразу. Цьому успіхові ЄНТ сприяє, зокрема, придбання високошвидкісних оптосканів, розроблених в ETS — тестовій службі США. У казахських колег, як і скрізь у країнах СНД, потребує істотного поліпшення якості тестів, а база тестових завдань взагалі вимагає принципового оновлення через появу організованих груп професійних репетиторів, які уважно відстежують зміст завдань щороку.

Методи

Ідеального методу для використання в національних системах оцінювання підготовленості випускників шкіл і абітурієнтів ВНЗ різної професійної спрямованості немає.

Автор цієї доповіді неодноразово відзначав, що в масовому процесі атестація випускників шкіл — це переважно питання оцінки освітнього мінімуму, а питання прийому до ВНЗ — зовсім інша проблема — проблема професійного відбору, яка вимагає тестів із високою розрізнявальною здатністю на окремих інтервалах континууму підготовленості. Відбір залежить від квоти прийому. Відповідно, з погляду психометрики, теорії педагогічних вимірювань та теорії професійного відбору якісна і соціально відповідальна робота з прийому до ВНЗ вимагає принципово іншого підходу. Одним набором завдань ці дві різні задачі не вирішуються. На сьогодні склалися чотири основні методи, використовувані для становлення національних та міжнародних систем оцінок рівня підготовленості учнів і студентів. Найпоширеніший — метод тестування. Він застосовується у США, країнах Європи, Казахстані та в країнах Південно-Східної Азії. Актуальність

тестового методу пояснюється деякими його перевагами перед іншими методами. П'ять основних переваг такі:

1. Висока наукова обґрунтованість справжнього тесту, яка дозволяє одержувати об'єктивні оцінки рівня підготовленості випробовуваних.
2. Технологічність тестових методів.
3. Точність вимірювань, що об'єктивно фіксується і відкрито перевіряється незалежними експертами.
4. Наявність однакових, для всіх користувачів, правил проведення педагогічного контролю й адекватної інтерпретації тестових результатів.
5. Сполучуваність тестової технології з іншими сучасними освітніми технологіями. Таким чином, можна стверджувати, що переваги тестового методу — це швидкість і точність вимірювання, масовість і технологічність, об'єктивність і справедливість у сенсі однаковості змісту й умов контролю.

Але тестовий метод має і деякі негативні наслідки — стосовно розвитку мови учнів та студентів, негативного впливу на зміст навчального процесу. Він також має соціально-психологічні вади, особливо у випадках, якщо тестова оцінка стає вирішальною для формування життєвої і професійної траєкторії розвитку молоді [23]. При становленні національних систем тестування про ці та багато інших недоліків тестових методів треба знати й постійно відшукувати форми мінімізації негативних наслідків застосування тестів. Другий метод — традиційні іспити або завдання, нерідко олімпіадного рівня. Вони застосовуються — більшою чи меншою мірою — у програмах міжнародного дослідження PISA, в Росії, на Тайвані [24] та в інших країнах Південно-Східної Азії. Переваги цього методу — позитивний вплив на розвиток мови учнів, можливості поглибленої перевірки знань за вибраними темами, на вибірках випробовуваних. Краще забезпечується зв'язок між змістом навчання та змістом іспиту. Але для масового процесу вимірювання генеральної сукупності випробовуваних ці методи практично не придатні з огляду на їх підвищену трудомісткість і високу вартість. Третій метод — так звані контрольні-вимірювальні матеріали (КВМ). Застосовуються виключно в рамках проведення російського ЄДІ. Якість цих матеріалів є головною таємницею Міністерства освіти і науки РФ. Проте, за наявними фрагментарними даними, якість цих матеріалів нижча за будь-який припустимий рівень. Можливо, що саме цим пояснюється відсутність виразних публікацій про якість КІМ. Головна метрична вада КІМ — це спроба оцінити одним набором завдань рівень підготовленості ординарних випускників шкіл і одночасно рівень підготовленості абітурієнтів престижних ВНЗ. Теоретично ясно, що точність вимірювання одним і тим же тестом рівня підготовленості таких різних контингентів може бути тільки низькою. До того ж множина цілей, поставлених перед КІМ ЄДІ, виявилася незрозумілою навіть запрошеному міжнародному експерту.

Статус систем

Незважаючи на згадані недоліки, найбільш відповідним методом для становлення Національної системи об'єктивної (зовнішньої) оцінки є все-таки педагогічний тест. Можна із задоволенням відзначити, що стратегічний вибір на користь створення Української національної незалежної (зовнішньої) системи тестування зроблено правильно. Разом із тим, саме зараз час обговорювати можливість рухатися далі, у бік становлення Української системи педагогічних вимірювань. В Україні, США, Казахстані та в багатьох інших країнах прийнято статус Національної системи тестування. У західному тлумаченні поняття «національна» означає, що тестування є суспільним і професійним, але не державним. У визначенні статусу Національного тестування важливо зрозуміти співвідношення ролі держави й суспільства. Цивілізованій державі не слід займатися тестуванням знань кожної дитини або студента. Як відомо, в США центральний уряд не має права проводити тестування в тому або іншому штаті. І це правильно, оскільки стримує державні органи перед спокусою розширення владних повноважень та перед корупцією. Але держава там стоїть на сторожі у разі порушення прав громадянина у вигляді необ'єктивно виставленої оцінки, одержаної на основі неякісного методу. У справжньому Національному тестуванні є й інша важлива складова. Це професіоналізм тих, кого суспільство залучає до проведення Національного тестування. Саме професіоналізм забезпечив істотний успіх уже згадуваної недержавної служби тестування в США — Educational Testing Service (ETS).

Висновок

Якість вимірювань — головна невирішена проблема у роботі нинішніх центрів тестування в Росії та інших країнах колишнього СРСР. Основна причина — відсутність можливості та мотивів запрошення учених до роботи в центрах тестування. Для стабільного і якісного розвитку Національної (суспільно-професійної) системи тестування треба знайти шляхи для поступового переведення існуючих центрів тестування у статус незалежних центрів Національної системи педагогічних вимірювань. Це дасть змогу посилити науковий потенціал центрів, помітно підвищити якість педагогічних вимірювань. Створення якісної системи педагогічних вимірювань — наукоємна проблема. Без участі вчених багатьох спеціальностей ця система буде знаходитись у стані постійного ризику роботи з оцінками дуже низької якості. Слід якомога швидше переходити від розмов про створення системи іспитів до діяльності зі створення системи педагогічного тестування і далі — до становлення Національної системи педагогічних вимірювань.

(за матеріалами журналу «Вісник ТІМО» №1/2008, с. 38-41)