МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Таврійський державний агротехнологічний університет

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
з розробки і застосуванню тестів для контролю знань студентів

Розглянуто і схвалено
методичною радою ТДАТУ
протокол № 6
від 26.04.2012

Перезатверджено
методичною радою ТДАТУ
Протокол № 4
від 17.03.2015

МЕЛІТОПОЛЬ 2015

Розробили: проректор з НПР доц. Ломейко О.П.
в.о.зав.кафедри ТС АПК Смєлов А.О.

Рецензенти: к.т.н., доц. кафедри ОП та БЖ Малюта С.І.

ЗАТВЕРДЖЕНО:
Наказ ректора університету
від 14.04.2015 року № 107-ОД

ЗМІСТ

1 Загальні поняття про тести…….....4

2 Вимоги до тестів...……5

3 Принципи тестування ...……..……6

4 Порядок розробки тестів...…….....8

4.1 Завдання закритої форми...…….…11

4.2 Завдання відкритої форми...……...14

4.3. Завдання на встановлення відповідності...…….....15

4.4 Завдання на встановлення правильної послідовності дій……15

5 Надійність тестів...……...……..18

6 Валідність тестів...……..……20

6.1 Методи валідізації тестів...……....……24

6.2 Оцінка якості тестів...…….…....25

1 ЗАГАЛЬНІ ПОНЯТТЯ ПРО ТЕСТИ

Процес навчання розглядається як загальна цілеспрямована система діяльності викладача і студентів, що озброює останніх знаннями, уміннями і навичками, а також формує певний світогляд. Головна роль при цьому належить викладачу. Він повинен вміло відбирати необхідний матеріал, організовувати його засвоєння, закріплення і навчати застосовувати їх на практиці.

Перевірка й оцінка знань, умінь і навичок - одна з головних задач педагогічної роботи.

Урахування успішності повинно бути індивідуальним, плановим і систематичним.

Перевірка й оцінка знань повинні показати, наскільки студенти оволоділи навичками й умінням самостійної роботи, встановити обсяг і якість залишкових знань кожного індивідуально.

Особливо велике значення має поточний облік знань. За допомогою такого обліку викладач може своєчасно допомогти студенту заповнити прогалини в його знаннях і в процесі навчання виправити недоліки, які мають місце.

Результат поточного обліку знань спонукає студентів до регулярної самостійної роботи з дисциплінами, що вивчаються, а це потім виражається в стійкості набутих знань.

У різноманітних системах освіти існує дуже багато видів і методів навчання, відповідно з ними існують різноманітні форми і методи контролю.

Необхідність застосування тестових форм контролю викликана рядом причин.

По-перше, дуже багато часу витрачається викладачами на всі види традиційного контролю. Із збільшенням частини самостійної, творчої діяльності студентів буде зростати і розширюватися спектр контролюючої діяльності викладача. Таким чином, це потребує нових додаткових витрат часу. При наявності тестового контролю час на контроль зменшується до мінімуму. Тестування також дозволяє організувати ефективний самоконтроль студентів.

Тестування може бути ефективним і дати об'єктивну інформацію про рівень знань лише за умови, якщо викладач твердо володіє методикою їхньої розробки.

По-друге, за тестовою системою контролю виключається суб'єктивізм викладача при оцінці знань студента.

Педагогічний тест являє собою сукупність завдань зростаючої складності, що дозволяє надійно і валідно (адекватно) оцінити знання й інші характеристики особистості.

Іншими словами, під тестами слід розуміти набір формалізованих задач для самостійного рішення в межах відведеного часу, результати яких співвідносять із визначеною системою оцінок.

Існує багато форм тестів. Вони можуть бути у вигляді питань, задач, проблем, виконання яких може оцінюватися однозначно (так, ні), або за визначеною шкалою оцінок, якщо передбачені різнорівневі відповіді.

Виходячи із різноманітних стилів викладання різних дисциплін, а також великої кількості технологій навчання, то поряд з цим існує достатньо велика кількість форм і видів тестів.

Проте, враховуючи все це необхідно твердо знати основні принципи й ознаки, на яких базується їхня розробка. У противному випадку, та задача, яка поставлена перед тим або іншим тестом, не буде вирішена. Більш того, можна завдати шкоди навчальному процесу, отримавши ненадійну інформацію.

Повний набір тестів з дисципліни, що охоплює зміст навчання і забезпечує надійний контроль знань, умінь, навичок усіх рівнів складності називають банком тестів.

Перед тим, як приступити до розробки тестів, необхідно мати деякий мінімум знань і навичок по їх складанню.

Запропоновані приклади тестових завдань допоможуть викладачам системно підійти до складання тестів.

2 ВИМОГИ ДО ТЕСТІВ

Головні вимоги до тестів такі: простота, однозначність, визначеність, надійність, валідність.

Тест повинний мати системоутворюючі властивості:

1) приналежність до предметної області;

2) головний структурний принцип системоутворення полягає в тому, що кожне завдання відрізняється своєю складністю;

3) стислість;

4) ясність і чіткість;

5) коректність форми;

6) кумулятивність ефекту (шкала Гутмана), яка визначає, що студент, відповівший вірно на чергове завдання, з великою ймовірністю відповів

позитивно й на попередні;

7) час - усі тести виконуються в обмежений час;

8) диференційна властивість завдання - розділити слабких і сильних студентів.

Надійність тесту - це стійка стабільність результатів тестування того самого студента. Це перевіряється дослідницьким шляхом. Якщо розбіг в оцінках незначний, то такий тест називають надійним.

Валідність тесту - це його спроможність забезпечити контроль наперед визначеної групи якостей знань на визначеному рівні навчання. Іншими словами, якщо даний тест може бути виконаний позитивно тільки особою, що має відповідний набір фахових знань, умінь, навичок, котрі характерні для цього рівня навчання, то такий тест - валідний як функціонально, так і за змістом.

Простота теста полягає в тому, що він має чіткі на мові предмета, дисципліни формулювання задач.

Визначеність тесту – це така його якість, що дозволяє студенту надійно зрозуміти ту діяльність, яку він повинний виконувати. Визначеність тестів також перевіряється дослідницьким шляхом. Однозначність тестів повинна виключити різне тлумачення вірних і невірних відповідей експертами.

3 ПРИНЦИПИ ТЕСТУВАННЯ

Основними цілями контролю знань студентів є:

- визначити ступінь досягнень студентів у навчанні;
- виявити ставлення студентів до навчання;
- виявити, який матеріал повинний бути винесеним для самостійної роботи або для організації додаткових занять;
- одержати інформацію, необхідну для управління процесом навчання, для удосконалення методик викладання і вдосконалення форм самостійної роботи студентів.

Для досягнення зазначених цілей при проведенні контролю знань необхідно виходити з таких принципів:

1. Повнота контролю, тобто охоплення контролем усіх без винятку студентів для того, щоб забезпечити надійний зворотній зв'язок між викладачами і рівнем засвоєного студентами матеріалу.

2. Плановість контролю, де враховується обсяг, зміст, рівень, витрати часу і терміни проведення контролю з даної дисципліни.

3. Достовірність результатів контролю, що важливо як у навчальних, так і у виховних цілях. Це стимулює зацікавленість до навчання, підвищення особистої відповідальності студента.

4. Об'єктивність контролю. Цей принцип передбачає максимальне зниження суб'єктивного впливу викладача на результат контролю.

5. Диференційованість контролю, яка залежить від рівня знань. Ця вимога передбачає обов'язковість забезпечення глибини контролю відповідно для тих цілей, що поставлені при вивченні навчального матеріалу.

Щодо кількості рівнів знань існує декілька класифікацій. Найбільше поширена чотирьохрівнева система оволодіння навчальним матеріалом.

Перший рівень - це просте розпізнавання. Характер діяльності студента полягає в пошуку потрібного серед знайомого матеріалу.

Другий рівень - відновлення, або репродуктивний. На цьому рівні виконуються заучені дії, вирішуються задачі типові з застосуванням відомих формул, законів і т. ін.

Третій рівень - суб'єктивно творчий, пошуковий.

Студент застосовує відомі правила, формули в нових умовах або нові закони в стандартних умовах.

Четвертий рівень - об'єктивно творчий, конструкторсько-дослідницький. Знання цього рівня дають змогу продукувати нову інформацію, формувати нові підходи в рішенні відомих задач і дослідницький характер діяльності.

Відповідно до названих рівнів знань повинні відповідати рівні контролю, щодо яких розробляється зміст тестових завдань у банку тестів. При цьому, якщо навчальний процес передбачає зростання складності завдань із часом вивчення предмета або теми, то цьому повинний відповідати ріст складності тестів із цього предмета, або теми.

Таким чином, для перевірки якості засвоєння матеріалу на першому рівні повинні використовуватися тести на розпізнавання, можливість відрізнити співвідношення, класифікацію матеріалів, засвоєння понять і т.п.

Тести другого рівня використовуються для перевірки ступеня оволодіння репродуктивної (заданої в стандартних умовах) діяльності. Як правило, тести другого рівня базуються на рішенні типових задач, відомих формул, конструкцій.

Тести третього рівня - це нетипові задачі, що вирішуються в типових стандартних умовах, або типові - для нестандартних умов. Знаючи структуру, специфіку й зміст предмета, викладачу з досвідом такий тип тесту скласти

неважко. В силу того, що цей тип тестів специфічний і зрозуміти тонкощі змісту може тільки певне коло спеціалістів, то приклад такого тесту опускається.

Найбільш складним для викладача є побудова тестів четвертого рівня, тому що вони дозволяють контролювати таку творчу діяльність, у процесі якої повинна бути отримана об'єктивно нова інформація, або виявлена спроможність до дослідницької діяльності. У таких тестах використовують матеріали досліджень, оригінальні рішення студентів.

В тестах такого рівня можуть бути використані проблемні ситуації, підходи до рішення проблем з елементами теорії ймовірності.

Спеціалісти застерігають, що при розробці тестів усіх рівнів складності слід уникати формулювань питань, котрі взяті безпосередньо з навчальної літератури (підручників і навчально-методичних посібників).

4 ПОРЯДОК РОЗРОБКИ ТЕСТІВ

Розробка тестів - складний, трудомісткий процес і виконується в такій послідовності:

1) виявляються понятійні індикатори;

2) розробляється план тесту;

3) складаються і підбираються тестові завдання;

4) проводиться математична обробка результатів тестування.

Спочатку дається відповідь на питання - що таке знання предмета у вигляді достатньо загальних визначень, але намічаючи перехід до окремих понять, що у сукупності розкривають зміст вихідного поняття.

Далі окремі поняття розкриваються через сукупність так званих емпіричних індикаторів, що є завданнями тесту.

Узяті разом відповіді на завдання повинні свідчити про знання предмета в цілому.

Далі розробляється план тесту. Число завдань, що виділяються на кожний розділ, залежить від його вагомості у дисципліні, що вивчається.

Загальна кількість – 30-60 коротких завдань різного рівня складності.

Після розробки плана тесту починається найвідповідальніший етап - упорядкування і вибір завдань. Ця робота доручається самим досвідченим викладачам.

Завдань повинно бути в два - три рази більше необхідного, тому що багато їх відсівається.

Завдання повинні мати коректну форму.

Форма - це принцип упорядкування елементів тестового завдання. Розрізняють завдання:

1) закритої форми - пропонуються варіанти відповідей;
2) відкритої форми - відповідь довільна;
3) завдання на встановлення відповідності;
4) завдання на встановлення правильної послідовності дій.

Загальні правила при розробці завдань:

1) кожна форма потребує своєї інструкції;
2) вибір форми залежить від предмета і часу тестування;
3) використовується природна структура мови;
4) завдання повинно бути стислим, однозначним і належати до предметної області.

Саме форма й зміст є головними компонентами процесу створення тестових завдань. Ціль процесу- створення таких завдань, які можна було б включити в тест і використовувати як у традиційно організованому навчальному процесі, так і в автоматизованих системах контролю знань.

У процесі створення до завдань у тестовій формі пред'являються наступні вимоги:

- логічна форма висловлення;
- правильність форми;
- стислість;
- наявність певного місця для відповідей;
- правильність розташування елементів завдання;
- однаковість правил оцінки відповідей;
- однаковість інструкції для всіх випробуваних;
- адекватність інструкції формі й змісту завдання.

Логічна форма висловлення є універсальною формою чіткого вираження думки людиною, здатним виявити знання по досліджуваній навчальній дисципліні. Ця форма в багатьох випадках заміняє питання. Навчальні питання багатослівні й породжують відповіді, повні й неповні, правильні й неправильні, різні за формою, змісту й за структурою, внаслідок чого оцінка таких відповідей вимагає обов'язкової участі викладача й супроводжується деякої часток суб'єктивізму у той час як технологічная методика тестування припускає чітку й швидку дифференцируемость відповідей. Логічна перевага завдання в тестовій формі полягає в можливості його природнього перетворення, після відповіді студента, у форму дійсного або неправильного висловлення.

Правильна форма завдань - це засіб упорядкування й ефективної організації змісту тесту. Організація припускає аналіз змісту навчальної дисципліни, класифікацію навчального матеріалу, установлення меж тематичних і меж предметних зв'язків, укрупнення дидактичних одиниць, вистава цих одиниць через елементи композиції завдань. Форма завдань правильна, якщо вона дозволяє точно виразити зміст, зрозуміла для всіх випробуваних, виключає можливість появи помилкових відповідей по формальних ознаках.

Стислість завдань у тестовій формі випливає з переваг логічної форми висловлення, Тому завдання завжди коротше завдань і питань. Стислість забезпечується ретельним добором слів, символів, графіків, що дозволяють мінімумом засобів домагатися ясності змісту змісту завдання. Виключаються повтори, малозрозумілі, рідко вживані, а також невідомі для символи, що вчаться, іноземні слова, що утрудняють сприйняття змісту. Добре, коли завдання містять не більш одного придаткової пропозиції. Для досягнення стислості завдання краще запитувати про що-небудь одному.

Наявність певного місця для відповідей є одним із зовнішніх ознак завдання в тестовій формі. У завданнях з вибором відповідей - це код (цифри або букви) необхідного відповіді, із числа прикладених до кожного завдання. У завданнях відкритої форми відповідь пишеться замість прочерку, следуемого за текстом завдання. У завданнях на встановленні відповідності відповіді або пишуться в спеціально відведеній для цього рядку нижче тексту кожного завдання, або фіксується за допомогою так званої "миші" при комп'ютернім тестуванні. І, нарешті, у завданнях на встановленні правильної послідовності випробуваний ставить ранги в спеціально відведеному для цього місці.

Правильність розташування елементів завдання є вимогою, що допомагають випробуваним не витрачати час на визначення місця для відповідей і швидше зафіксувати свій розв'язок.

Однаковість правил оцінки відповідей у рамках прийнятої форми є важливим засобом організації тестування. Жодному випробуваному не дається ніяких переваг перед іншим, усе відповідають на ті самі завдання, усім дається однаковий час. Правила оцінки визначаються заздалегідь і абсолютно однаково застосовуються до всіх випробуваних. Зрозуміло, що все це робиться для зменшення помилок виміру, викликаних порушеннями стандартних умов тестування.

Відповідь на завдання являє собою коротке судження, зв'язане по змісту й за формою зі змістом завдання. Критерії правильності відповідей заздалегідь визначаються автором завдання.

Адекватність інструкції формі й змісту завдання означає взаємна відповідність перерахованих компонентів, необхідне для виконання завдання своєї функції. Умова адекватності дозволяє довести до свідомості випробуваних усі вимоги, закладені в змісті завдання. Невідповідність викликає помилку розуміння змісту завдання, і відповідно, приводить до помилкових відповідей там, де таким не місце.

Оволодіння формою є необхідним, але недостатньою умовою створення повноцінних тестів. Форма надає завданням лише структурну цілісність і визначеність, зовнішню організованість.

Завдання в тестовій формі тільки зовні схожі на тестові завдання, а це недостатньо для їхнього включення в тест. Потрібна ще перевірка так званих тестоутворюючих властивостей завдань.

4.1 Завдання закритої форми

В завданнях закритої форми даються готові відповіді на вибір. Завдання такого типу можуть бути ***одновибіркові*** з яких студент повинен вибрати одну вірну відповідь та ***багатовибіркові*** з яких студент вибирає ті і стільки, які він вважає вірними.

При використанні тестових завдань з множинним вибором студенти не знають точне число правильних відповідей, що робить такі завдання складнішими. Крім підвищення труднощів і технологічності, завдання даного варіанта дозволяють перевірити знання повніше, глибше й точніше. У таких завданнях випробувані можуть одержати від нуля до n-балів (n – кількість вірних відповідей в завданні), що підвищує варіацію даних і, як наслідок, точність виміру.

У таких завданнях, поряд з однією правильною відповіддю, є декілька неправильних, але правдоподібних. Останні називають дистракторами (від англ. to distract - відволікати); Число дистракторов може коливатися, звичайно, від одного до 5.

ІНСТРУКЦІЯ:

Позначте номер правильної відповіді

МІЦНІСТЬ ТВЕРДОГО СПЛАВУ ІЗ ЗБІЛЬШЕННЯМ ВМІСТУ КОБАЛЬТА

1. зменшується (принцип
2. збільшується альтернативності)

При зовнішній простоті і ймовірності угадування однієї відповіді рівної 1/2, можливість угадування *п'ятьох відповідей із десятьох завдань* дорівнює 0. 246, тобто достатньо велика.

Розглянемо завдання з чотирма альтернативами:

ОСНОВНИМ ПРОДУКТОМ ДОМЕННОЇ ПЕЧІ Є

1. сталь
2. чавун (принцип
3. мідні сплави класифікаційності)
4. алюмінієві сплави

В цьому разі така можливість відгадування дорівнює тільки 0.058 !!!

Крім цього можна проводити кореляцію на здогад за формулою:

$$X_{кор} = X - \frac{E}{K-1},$$

де Хкор - скоригований бал;

Х - тестовий бал;

Е - кількість помилкових відповідей;

К - кількість відповідей у завданні.

Приклад:

Якщо Х=20; Е=10; К=5; то

$$X_{кор} = 20 - \frac{10}{5-1} = 17{,}5$$

Якщо Хкор - негативний, то студент або упевнений в "антизнаннях", або нездогадливий і не знає матеріалу.

Основна трудність у розробці таких завдань - у виборі правдоподібних, але все-таки неправильних відповідей. Явно правильні відповіді студенти розрізняють відразу.

Застосування ЕОМ полегшує роботу з тестування такими завданнями. При цьому з'являється можливість провести частину опитування на ЕОМ, а потім викладач опитує по найбільш важким питанням, або по тим розділам, в яких була допущена найбільша кількість помилок.

ЕОМ дозволяє організувати самоконтроль студентів і видачу завдань випадковим способом. Студентам подобаються такі завдання, тому що разом із неправильними відповідями дається і правильний при цьому складається уява, що самому легше вибрати правильну відповідь, ніж її сформулювати.

При застосуванні тестів на паперовому носії необхідно часто змінювати завдання, тому що студенти швидко розкодовують ключ до тесту. Цього можна уникнути, якщо застосовувати, в тому числі і на ЕОМ, фасетні завдання, коли для утворення нового завдання в ньому змінюється одне, два слова. Наприклад:

У МАРКУВАННІ СТАЛІ БУКВОЮ { Ю / Б / Г } ПОЗНАЧЕНИЙ

1. бор
2. алюміній бор, берилій - псевдонатяк
3. берилій
4. ніобій
5. марганець

Завдання закритої форми застосовуються для перевірки міцності і упорядкування структури знань. Такі завдання можна використовувати при усіх видах контролю (у підсумковому і заключному - обмежено).

Нижче поданий приклад псевдофасети.

ЗА 10 ДНІВ ДО { ОТЕЛА / ОПОРОСУ / ОКОТУ } ВИКЛЮЧАЮТЬ З РАЦІОНУ

1. сіно
2. солому завдання змінюється, а відповідь
3. силос таж сама

У ПОЛОВИНЧАТОМУ ЧАВУНІ ВУГЛЕЦЬ ЗНАХОДИТЬСЯ У ВИГЛЯДІ

1. графіту
2. графіту, цементиту (принцип
3. графіту, цементиту і суміші графіту з залізом кумулятивності)

ІЗ ЗБІЛЬШЕННЯМ ПОДАЧІ УСАДКА СТРУЖКИ

1. збільшується
2. зменшується (подвійна
3. спочатку збільшується, потім зменшується (альтернатива)
4. спочатку зменшується, потім збільшується

У ІЗОТЕРМІЧНОМУ ПРОЦЕСІ З ІДЕАЛЬНИМ ГАЗОМ ЗАЛИШАЮТЬСЯ ПОСТІЙНИМИ

1. ентальпія й ентропія
2. ентропія і внутрішня енергія (принцип
3. внутрішня енергія і ентальпія сполучення)

У тесті можна комбінувати завдання різноманітних принципів.

4.2 Завдання відкритої форми

В завданнях цієї форми необхідно дописати правильну відповідь – звичайно одне слово, або один знак.

Основні вимоги до завдань відкритої форми такі:

- у кожному завданні такої форми повинно бути, як правило, тільки одне доповнення;
- доповнювати необхідно тільки найважливіше;
- доповнення краще ставити наприкінці речення;
- чітке питання - чітка й однозначна відповідь, і про це необхідно зазначити в інструкції;
- починати з ключового слова;
- при використанні паперового носія для відповіді відводиться необхідне місце.

Точність виміру відкритих завдань вище, ніж закритих.

Принцип єдності форм: із закритої форми можна зробити відкриту і навпаки.

Завдання відкритої форми можна використовувати при всіх видах контролю й особливо в підсумковому.

ІНСТРУКЦІЯ:
Доповнити
ШВИДКІСТЬ РІЗАННЯ ПРИ ТОЧІННІ ВИМІРЮЄТЬСЯ У _____

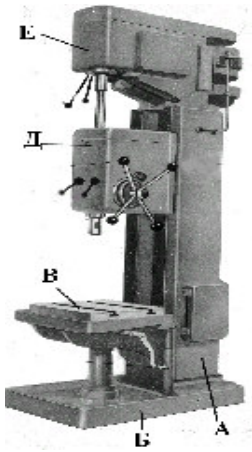
4.3 Завдання на установлення відповідності

Такі завдання состоять із елементів двох стовпців. Завдання такої форми використовуються для тематичного контролю і модульного навчання. У підсумкових тестах використовувати не рекомендується через громіздкість.

Тестові завдання "на відповідність" дозволяють виявити наявність у студентів знань зв'язків між об'єктами та їх функціями, термінами і позначеннями і таке інше.

ІНСТРУКЦІЯ:
Встановити відповідність

ВУЗЛИ ***НАЗВИ***



1. Колона
2. Траверса
3. Поперечина
4. Коробка швидкостей
5. Консоль
6. Стіл
7. Шпиндельна бабка
8. Електродвигун
9. Основа

ВІДПОВІДІ: А , Б , В , Д , Е .

В пропущених позиціях строки відповідей студенти вписують цифру правильної відповіді з правого стовпця.

Якщо в запитаннях використовуються цифри, то у відповідях – букви і навпаки.

4.4 Завдання на встановлення правильної послідовності дій

У завданнях такої форми необхідно цифрами у відповідних місцях зазначити правильну послідовність, наприклад, дій, виконання технологічного процесу складання агрегату і т. ін.

Тестові завдання зазначеної форми доцільно використовувати при контролі знань нормативної діяльності, процесу роботи машин та механізмів, послідовності виконання технічних і технологічних операцій, процедурної діяльності.

При розробці тестових завдань на встановлення правильної послідовності необхідно попередньо відпрацювати процедуру діяльності, тобто скласти однозначний покроковий опис (алгоритм, інструкцію, правила і т.ін.) виконання крок за кроком дій щодо вирішення будь-якого конкретного завдання з будь-якого класу завдань визначеного типу

ІНСТРУКЦІЯ:
Встановити правильну послідовність

РОЗРАХУНОК ОСНОВНОГО ТЕХНОЛОГІЧНОГО ЧАСУ

	- L
	- $n_ф$
	- V
	- $n_д$
	- t
	- T_o
	- $V_д$
	- $S_т$
	- $V_ф$
	- $S_ф$

У даному завданні в прямокутниках проставляється цифрами, у порядку зростання, правильна послідовність розрахунку основного часу.

Для комплексних завдань використовуються ланцюгові і ситуаційні тести, у яких можуть бути тестові завдання різноманітної форми.

Наприклад:

Пораненому в { стегно / передпліччя / стопу } накладений джгут.

При знятті джгута в медпункті полку через { одну / дві / три } години

кровотеча продовжувалася. Судина, що кровоточить, { видима / не видима }

1. У цій ситуації припинення кровотечі провадиться із застосуванням

1. джгута
2. затискача на судину в рані
3. тампонади рани
4. тугої пов'язки

Поранений транспортується

3. У

1. польовий шпиталь
2. тил

4. У черговості

1. першій
2. другій
3. третій

5. У положенні

1. сидячи
2. стоячи
3. лежачи

6. Використовується транспорт

1. гужовий
2. санітарний автомобіль
3. літак

і так далі. У ланцюгових тестах відповідь на наступне завдання залежить від правильності відповіді на попереднє.

Наприклад:

Виявлено п'ять ящиків із різноманітними невідомими будматеріалами, але відомо, що в них або крейда, або негашене вапно, або гашене вапно, або цемент, або гіпс.

У Вашому розпорядженні є кислота і вода.

При приливанні кислоти до однієї з пробірок виділився газ.

Тоді в даній пробірці, а:

1. Значить і в ящику знаходиться _____

2. Формула цієї речовини _____

До зразків із чотирьох ящиків, що залишилися, долили воду. У пробірці з одним із зразків відбулося сильне розігрівання,

3. Значить це _____

4. Формула цієї речовини_____

і так далі.

5 НАДІЙНІСТЬ ТЕСТІВ

Науково обґрунтований тест – тест, який заснований на методах, що відповідають установленим стандартам надійності і валідності. У вимогах перевірки тесту на надійність і валідність реалізується важлива ідея методологічного характеру - до істинного знання веде лише істинний метод. Таким чином, якість педагогічної інформації залежить від якості використовуваного для цього інструментарія.

Традиційно виділяються два основних критерії якості тестів. Перший із них пов'язаний із поняттям точності вимірів і відомий, головним чином, у вигляді поняття надійності тесту.

Існує декілька практичних способів визначення надійності тесту. Найбільш бездоганний із статистичної точки зору метод визначення надійності - це корелювання двох паралельних тестів, створених для виміру однієї і тієї ж властивості. Проте цей метод застосовується надзвичайно рідко або через звичайну відсутність паралельних тестів, або невідповідності наявних

варіантів суворим вимогам паралельності. Введене в останні роки комп'ютерне пред'явлення завдань із значеннями параметрів, що змінюються у випадковому порядку, знову відродило інтерес до створення рівнобіжних тестів з оцінкою їхньої надійності за допомогою корелювання. Чим вищий коефіцієнт кореляції між двома такими тестами, тим вище надійність теста. Для кореляції використовується формула розрахунку класичного коефіцієнта кореляції Пірсона.

Інтуїтивно найбільш зрозумілий і простий спосіб визначення надійності теста - це дворазове, щонайменше, використання того самого тесту у тій же самій групі студентів. Результати обох опитувань (якщо вони не анонімні) корелюються.

Метод повторного тестування має як переваги так і недоліки. Переваги полягають у порівняльній простоті його використання, ясності основних посилок, що лежать у визначенні надійності, простоті розрахунків. До недоліків можна віднести невизначеність у виборі інтервалу часу між першим і другим опитуванннями.

Цей інтервал може коливатися від декількох хвилин до декількох днів, місяців і навіть років. Природно, що при цьому по-різному виявляють себе такі фактори, як запам'ятовування або, навпаки, забування (у різні інтервали часу), вплив досвіду, отриманого в першому опитуванні на другий, вплив спілкування студентів між собою після першого опитування, придбання досвіду і т. ін.

Результати застосування методу повторного тестування з наступним корелюванням можна і потрібно контролювати іншими методами.

Наступний метод дозволяє оцінити надійність тесту при однократному тестуванні. Для цього весь тест поділяють на дві рівні частини. Частіше усього при цьому окремо підраховуються бали випробуваних у парних завданнях і в парних. Отримані суми балів корелюються.

Розглянутий вище метод визначення надійності тесту заснований на припущеннях паралельності і еквівалентності, які не завжди можуть бути витримані.

Визначення надійності тесту за допомогою середнього коефіцієнта кореляції потребує проведення ряду розрахунків, що не завжди зручно.

Крім перерахованих, розроблені й інші методи оцінки надійності тестів, засновані на застосуванні методів дисперсійного і факторного аналізу.

Надійність - необхідний, але не достатній критерій якості тесту. Іншим необхідним критерієм служить, валідність.

6 ВАЛІДНІСТЬ ТЕСТІВ

Якість педагогічного контролю у вузі залежить не тільки від надійності методів, які використовуються, але і від їхньої валідності. Проте на відміну від надійності, визначення якої зводиться до вибору однієї з поданих тут форм, обгрунтування валідності тесту являє собою задачу методологічного характеру. Як і обгрунтування будь-якої діяльності, процес валідізації починається з уточнення цілі і конкретних задач педагогічного контролю. Якщо ставиться мета

перевірити знання студентів з якої дисципліни і при цьому однаково, яким методом це можна буде зробити, то легко зрозуміти, що ця мета може бути досягнута за допомогою використання заліків, іспитів, курсових і дипломних робіт і тощо.

Ці й інші методи нерівноцінні з погляду об'єктивності і якості оцінки. Тому питання про валідність легко переводиться в прагматичну площину оцінки порівняльної придатності того або іншого методу для досягнення поставленої цілі.

Валідність будь-якого використовуваного тесту можна спробувати оцінити з позицій нормативного типу мислення, при якому будь-який результат повинний бути співвіднесений із нормою, і, уже виходячи, із неї, оцінюватися, як гарний або поганий. Якщо немає норми, то важко сказати, що означає наприклад, отриманий піддослідним М в тесті бал 35- добре, це або погано? Якщо ж відомо, що нормою є 25 балів, то картина проясняється: М знаходиться, мабуть, у числі найкращих. Норма - дуже відносне поняття.

Тест може бути валідним, якщо крім інших вимог середні результати відповідають більшій частині студентів вузу, а самі дані розподіляються по нормальному закону. Якщо ця умова не виконується, то тест вважається невалідним із точки зору відповідності стандартам розподілу. Саме звідси виникає прагнення творців тест домогтися нормального розподілу за рахунок варіювання числа легких і важких завдань.

Якщо в тесті немає достатнього числа легких і важких завдань, то виникає питання про його збалансованість по трудності, суть якої коротко можна звести до того, що звичайно в тесті повинно бути побільше завдань середньої трудності і декілька менше відверто легких або важких завдань. У процесі створення теста міра трудності регулярно перевіряється на випадковій вибірці з того контингенту, для якого тест призначається. У збалансованому тесті автори звичайно легко домагаються нормальності розподілу. Подальше

удосконалювання йде по шляху заміни ряду завдань, відповіді на який порушують нормальність розподілу.

Трудність завдань впливає на надійність і валідність. Якщо тест дуже важкий, то студенти частіше змушені догадуватися, яка відповідь правильна. Але чим частіше вони удаються до догадки, тим більше розподіл результатів теста наближається до випадкового розподілу. Тому придатність теста для оцінки всієї маси студентів буде тим нижче, чим складніше тест. Такий же вплив на надійність - але по іншій причині робить легкий тест, у якому студентам, навпаки, не треба здогадуватися, їхні відповіді стійкі, але майже немає розходжень між піддослідними.

Валідність теста нерідко залежить від часу, відведеного на тестовий контроль. Ще до його створення вирішується питання - яким за часом буде тест. Якщо на тестування приділяється обмежений час, то це помітно впливає на результати. Тому заздалегідь визначається, чи буде час тестування обмежений або ні? В залежності від відповіді на це питання застосовуються різноманітні формули розрахунку надійності теста, по-різному розташовуються завдання, встановлюються інші норми. Це положення стосується, в основному, педагогічних тестів і менше - до психологічного.

Якщо час тестування обмежено, то кращих успіхів домагаються ті студенти, у кого швидше протікають розумові процеси. Але чи більше вони знають від тих, хто встиг відповісти на менше число питань - це потребує спеціального аналізу. В чистому виді тест з обмеженням часу являє собою метод контролю, в якому жодному не удасться встигнути відповісти на всі питання. У тестах без обмеження часу практично кожному дається можливість відповісти на всі завдання. .Для зручності класифікації в практиці прийнято, якщо 95% студентів встигають спробувати відповісти на кожне завдання, то такий тест відносять до контролю без обмеження часу; у противному випадку його можна класифікувати як тест з обмеженням часу.

Оскільки обмеження виражене може бути самим найрізноманітнішим, пошук оптимального обмеження часу на практиці зводиться до декількох спроб дати той самий тест у тій же групі. Якщо виявиться, що немає тренда - стійкого поліпшення результатів від спроби до спроби, то за оптимум береться такий час, при якому виявляється найбільша варіація в результатах.

Валідність тесту істотно знижується через зниження його розрізняючої спроможності. Якщо десять чоловік в групі одержують «відмінно», така оцінка не дозволяє розрізняти, хто з цих десятьох краще, а хто гірше знає предмет.

Розрізняючи спроможність теста тим вище, чим менше однакових оцінок студенти по ньому одержують. Отже, чим більше варіація результатів,

тим більш чутлива шкала до індивідуальних розходжень. Тому підвищенню спроможності розрізняючій теста (РСТ) у стадії його створення приділяється велика увага. При цьому застосовуються декілька методів:

1. Регулювання за часом тестування; чим більше стандартне відхилення результатів, тим розрізняюча спроможність тесту.

2. Правильне комплектування груп. Якщо відібрана група осіб, приблизно з однаковими спроможностями, то РСТ різко погіршується. Знижується надійність і особливо різко - валідність. Це питання виникає при подальшому доборі з групи раніше відібраних осіб, і вирішується він застосуванням формул для корекції валідності.

3. Оптимальний добір завдань. У принципі РСТ, а разом із нею і надійність теста, зростають із збільшенням частки завдань середньої трудності в тесті. Проте в тесті обов'язково повинна бути деяка частина легких і важких запитань, точна кількість яких залежить від конкретних обставин;

4. Точність вимірів. Якщо, наприклад, час реакції вимірювати у випробуваних із точністю до однієї десятої, сотої, тисячної і т.д. секунди, то ми одержимо різноманітну РСТ. Процес підвищення точності має розумні межі і пов'язаний, як правило, із придбанням дорогої апаратури. Варто уникати надточних вимірів, що штучно підвищують розрізняючи спроможність теста, проте логічно невиправдану.

Валідність тесту пов'язана, крім іншого, із поняттями «Гомогенний і гетерогенний тест». Якщо тест створений із метою перевірки знань по одній навчальній дисципліні і всі питання теста пов'язані саме з нею, то такий тест рахується гомогенним, а значить і валідним для цієї конкретної цілі. Наприклад, тест з алгебри, що не містить завдань з інших розділів математики, можна назвати гомогенним.

Але гомогенність - поняття відносне, тому що в курс алгебри входить багато розділів - рішення нерівностей, рішення лінійних рівнянь і ін. Тому в більш чистому вигляді гомогенний тест являє собою тест для вивчення знань по якомусь окремому розділу програми.

Для комплексної оцінки знань студентів може бути створений тест, що складається з питань по декількох дисциплінах. Це - приклад гетерогенного теста, що складається з групи гомогенних тестів. Відповідно такий тест є валідним саме для комплексної оцінки. Він не придатний, як надмірний, для оцінки знань по окремій дисципліні. У таких випадках гетерогенний тест розділяється і використовується тільки та його частина (шкала), що безпосередньо відповідає поставленій цілі.

Валідність тесту залежить і від так називаної довжини його. Під довжиною теста слід розуміти кількість завдань, що входять в тест. Існують тести дуже короткі, що складаються з 7-15 завдань, і дуже довгі, що складаються з більш ніж п'ятисот завдань.

Якщо тест дуже довгий, то погіршується мотивація й увага у студентів, а це знижує надійність і валідність. Практика показує, що якщо тестування займає більше ніж півтори години, то при цьому виникають організаційні проблеми. Студенти неохоче погоджуються відповідати на питання тесту. З іншого боку, із погляду теорії, чим довше тест, тим він надійніше. Виникаюче протиріччя між теорією і практикою вирішується поступом в ту або іншу сторону, в залежності від конкретного випадку.

Валідність теста залежить ще і від розташування завдань у тесті. Існує різноманітна практика розташування завдань:

1. За ступенем зростання трудності. Таке розташування характерне в основному для гомогенних тестів. Для гетерогенних тестів зберігання цього принципу виражається в так називаній «спіральній» формі розташування завдань. Наприклад, якщо є гетерогенний тест по чотирьох дисциплінах, то перші чотири завдання тест складають самі легкі питання по всіх цих дисциплінах; наступні чотири – дещо трудніші, але по всім чотирьох дисциплінах і так далі один за іншим, як би по спіралі зростаючої трудності.

2. У випадковому порядку. Цей спосіб розташування завдань широко застосовується в психологічних тестах.

Виняток звичайно робиться для першого й останнього питань, що можуть відбиратися спеціально, із розумінь більшої привабливості, зручності, інтересу і т.д.:

3. У спеціальному порядку, відповідно до якоїсь теорії, набуття спеціальних навичок, концентрації уваги й ін.

4. У порядку, що поєднує спеціальний і випадковий добір.. Звичайно це робиться в гетерогенних тестах. Наприклад, у тесті питання 15, 25, 35, 45, 55, 65 і т.д. з однієї шкали, а скажімо, питання - 30, 40, 50, 60, 70 і т.д. - з іншої. У психологічних тестах на ці номера питання кожної шкали видаються у випадковому порядку, у педагогічних - у порядку зростання трудності. Така побудова гетерогенного теста зручна для запам'ятовування номерів завдань при обробці результатів.

6.1 Методи валідізації тестів

Існують декілька підходів до валідізації тестів, що різняться в залежності від використовуваних критеріїв. У педагогічній практиці найбільше поширення в останні роки одержали такі тести, валідність яких не потрібно доказувати емпірично: у таких тестах критерієм їхньої придатності є сам зміст теста, схвалений досвідченими викладачами-експертами. При цьому у викладача повинна бути впевненість у тому, що:

- завдання тесту відповідають програмі;
- завдання тесту охоплюють не один якийсь розділ, а всю програму курсу;
- висока імовірність того, що студент, що успішно відповів на завдання теста, знає предмет у відповідності з отриманою оцінкою.

Перераховані три пункти об'єднуються загальною ідеєю - чи містить тест завдання, придатні для оцінки знань з конкретної навчальній дисципліні? Якщо в результаті статистичної перевірки виявляється, що відповіді на питання тесту цілком дозволяють обгрунтовано судити про знання студентів, то вважається, що тест містить валідні питання; він валідний по змісту. Вимога валідності по змісту подається до кожного питання теста, мірилом валідності є коефіцієнт кореляції відповідей завданням і критеріям.

При створенні теста в якості критерію звичайно беруться оцінки, що виставляються студентам групою викладачів-експертів без тестів. Результати студентів із питань теста і по оцінках експертів корелюються. Висока узгодженість оцінок по тесту й в експертів вказує і на високу валідність. Звичайно педагогічні тести мають достатню змістовну валідність, якщо до їхнього упорядкування залучалися досвідчені педагоги.

Водночас, при практичній валідізації виникає ряд труднощів, пов'язаних із недостатньою узгодженістю думок експертів, великою обчислювальною роботою, недостатньо суворим контролем для запобігання можливості списування і т.д. Надалі, коли відібрані валідні питання, у якості критерію можна брати сумарний індивідуальний бал студентів по всіх питаннях. Завдання, які знову вводяться в тест можна корелювати із сумарним критерієм.

Прогностична валідність тесту багато в чому залежить не тільки від надійності самого теста, але і від надійності критерію. У практиці не завжди вдається підібрати високонадійний критерій. Тому що тести підбираються за критерієм, то часто виникає питання, якою валідністю володів би тест, якби

надійність критерію була максимальною, рівною одиниці? У таких випадках застосовується корекція на ненадійність.

Використання тестів дозволяє об'єктивно оцінити знання, швидко одержати інформацію про результати навчання й оцінити ефективність роботи кожного педагога, підвищити ефективність самостійної роботи за рахунок створення умов для самоперевірки знань.

6.2 Оцінка якості тестів

Тест як інструмент дослідження може в залежності від різноманітних умов і факторів показувати різну надійність. Звідси випливає, і це треба підкреслити, що немає раз і назавжди встановлених показників надійності і валідності теста. У кожному окремому дослідженні рекомендується перевіряти якість тесту і лише на цій основі робити висновки про достовірність даних.

До показників надійності, як і валідності, пред'являються певні вимоги:

Розмір кореляції	Надійність	Валідність коефіцієнта
0,90-0,99	Відмінна	Відмінна
0,850,89	Дуже добра	Відмінна
0,80-0,84	Гарна	Відмінна
0,75-0.79	Задовільна	Відмінна
0.70-0,74	Малозадовільна	Гарна
0,60-0,69	Сумнівна	Гарна
0,50-0,59	Незадовільна	Гарна
0,40-0.49	Зовсім незадовільна	Задовільна
0,30-0,39	те ж	Малозадовільна
0,20-0,29	- * -	те ж
0,10-0,19	- *-	- *-
0,00-0,09	- * -	Незадовільна

Зовсім мало тестів, що мають валідність вище 0,80. З іншого боку, у практиці широко застосовуються тести, що мають низьку валідність.

Тест може мати відмінну надійність і в той же час бути зовсім незадовільним по валідності. Інтерпретувати цю обставину можна приблизно так: тест являє собою досить точний інструмент, але цей інструмент непридатним для вивчення того, що їм намагаються вивчати. З іншого боку, якщо тест валідний, то одне це вже припускає достатню надійність тесту, тому що він не може бути валідним, не будучи надійним.