

РЪКОВОДСТВО ЗА СЪСТАВЯНЕ НА ТЕСТОВЕ*

Фернандо Картрайт
 Джери Мусио¹⁾

Резюме. Образованието се превърна в основен фактор, определящ социалното и икономическото развитие на индивидите и нациите. Затова реформата на образователната система е обикновено сред ключовите национални приоритети на правителствата. Националните програми за външно оценяване на учениците са основният инструмент за измерване на качеството на образованието и за текущо наблюдение на ефекта от образователните реформи. Това налага да бъде гарантирано, че разработените тестове и изпитни програми са на най-високо професионално равнище.

Целта на настоящото ръководство е да подпомогне всеки екип, отговорен за разработването и провеждането на тестове, както и при анализа и обявяването на резултатите. Текстът е фокусиран върху практическото приложение на теорията за разработване на тестове въз основа на опит от провеждането на различни програми за оценяване в Канада и в други държави.

Главните теми, които са разгледани, са следните: планиране на оценяването, разработване на тест, създаване на тестови задачи, съставяне на тестове и провеждане на пилотни тестове, анализ на тестовете и тестовите задачи и интерпретиране на тестовите резултати.

Keywords: education, assessment, item development, test design, testing

Създаване на тестови задачи

Задачи и тестове

Задачата е основната структурна единица на оценяването. В този смисъл е възможно да се състави тест, съдържащ една-единствена задача. Всички задачи в един тест трябва да отговарят на единни общи принципи, даващи насока за коректното провеждане на оценяването.

Всички задачи трябва:

- да измерват конкретната образователна цел, с която са обвързани в спецификацията на теста;
- да не измерват образователна цел, с която не са обвързани;
- да не предизвикват отговори, които биха могли да бъдат повлияни от различия в произхода на учениците или условията, при които се провежда тестът.

* Продължение от книжка 1, 2013 на сп. „Стратегии на образователната и научната политика“

В допълнение разработването на задачите трябва да минимизира разходите по провеждането на теста, оценяването, анализа и докладването на резултатите, постигнати от учениците.

Видове задачи

Съществуват два основни вида задачи – със свободен (отворен) отговор и с избираем отговор. Първите изискват учениците да отговорят със собствени думи на задачата. Вторите предлагат кратък списък с възможни отговори, от който учениците трябва да изберат онзи, който най-много се доближава до тяхната представа за верен отговор.

Следната таблица обобщава нещата, които трябва да се знаят за двата типа задачи:

	Със свободен отговор	С избираем отговор
Примери	Кратък отговор (дума, цифра) Есе Решение на математическа задача	Избор от списък Вярно/невярно Откриване на сходство или еквивалентност
Предимства	Лесни за изготвяне Предоставят богата информация Една задача може да оценява няколко дяла от съдържанието на учебния план	Лесни за провеждане Лесни за оценяване Опростен механизъм за разработване на задачата (често основан на статистическия анализ) Относително независещи от произхода и социално-икономическата среда на учениците
Недостатъци	Трудни за администриране Възможно занижаване на валидността (силно зависещи от произхода и социално-икономическата среда на учениците и от контекста) Скъпи за оценяване Трудни за обработване (въвеждане на резултати)	Трудни за разработване, особено при оценяване на по-високи познавателни нива Необходими са много задачи, за да се покрият спецификациите на теста Възможен риск от изтичане на информация и нарушаване секретността на теста
Най-подходящи за оценяване на:	Способността за комплексно мислене и аргументиране (например възприемане на писмен текст, тълкуване на доказателства) Способността за извършване на действия и възпроизвеждане на процедури (например математически изчисления)	Познаването на факти (например исторически дати и цифри, научни наименования) Преценка на алтернативи (например вземане на аргументирани решения, тълкуване на данни)

Изисквания	Засилена роля на екипа за експертна оценка на теста в процеса на одобряване на задачите и особено при разработването на инструкциите за оценяване Голям екип от обучени оценители на тестови задачи Достатъчен ресурс (финансов и времеви) за периода от провеждането до докладването на резултатите	Секретно съхранение на одобрените тестови форми и задачи Съставяне на банка със задачи и форматиране на самите задачи (Незадължително) Използване на стандартизирани листове за отговори и пишещи средства при електронно оценяване на тестове
------------	--	--

Най-общо, разходите за мащабните оценявания могат да се намалят, ако по-голямата част от образователните цели се измерват чрез задачи с избираем отговор. Задачите със свободен отговор трябва да се използват единствено за оценяване на образователни цели, които не биха могли да бъдат измерени с малко на брой задачи с избираем отговор.

Структура на тестовите задачи

Независимо от това, към кой вид принадлежат, тестовите задачи се състоят от едни и същи 3 компонента:

- Условие
- Въпрос
- Отговор

Условие

Условието (наричано още и основа на задачата) предоставя информацията, необходима за коректен отговор на въпроса в задачата. Целта на условието е да постави всички ученици при равни условия по отношение на информацията, чието познаване няма да бъде измервано от задачата. Включването на минимална информация в условието понижава риска отговорът да бъде повлиян. Предоставянето на твърде много информация понижава валидността на тестовата задача.

Информацията в условието на задачата може да се представи в следните формати:

- Изображения
- Пасажи писмен текст
- Схеми
- Графики
- Списък с термини (при задачите, изискващи намиране на съответствие)
- Подробни инструкции

Разходите за разработване на тестови задачи могат да се занижат чрез разработване на групи от еднотипни, хомогенни задачи. В групата отделните задачи използват едно и също условие, като могат да оценяват различни образователни цели и познавателни нива. Някои групи се състоят само от неделимими задачи (например такива за намиране на съответствия), докато в други задачите могат да се използват и самостоятелно.

Валидността на теста се увеличава, когато задачите от една група са относително независими една от друга. Независимостта на задачите се постига, когато информацията, дадена в условието, или знаенето на верния отговор на една задача не увеличава вероятността да се отгатне отговорът на друга задача.

Въпрос

Въпросът на тестовата задача очаква определено поведение от ученика. Въпросът може да бъде експлицитен или имплицитен.

Експлицитният въпрос ясно заявява какво точно се очаква от учениците (например „Коя от възможностите по-долу е най-подходяща?“ или „Кое събитие предизвиква Първата световна война?“).

Имплицитният въпрос предоставя на учениците непълна поредица или конструкция, като подсказва на ученика да попълни липсващите звена. Примери за това са изречения с липсващи думи или числови редици.

Като цяло, валидността на задачите нараства, когато въпросът е експлицитен. Ако се използват задачи с имплицитни въпроси, трябва да се предоставят достатъчно примери и инструкции, за да могат учениците да разберат условието на всяка задача.

Отговор

Разработването на частта с отговорите в една задача обикновено отнема най-много време. Тази част трябва да съдържа цялата необходима информация за преобразуване на дадените от учениците отговори в оценка.

Форматът на отговора на задачата може да бъде свободен или избираем, но информацията в частта за оценяване трябва да бъде достатъчно точна и ясна, така че един и същ отговор да доведе до една и съща оценка, независимо от изпитвания ученик, условията на провеждане на теста и вида на оценяването. При задачите с избираем отговор това се постига с ключа на верните отговори. Ключът за оценяване обвързва всеки възможен отговор с оценка с определено цифрово изражение (брой точки).

При задачите със свободен отговор са възможни твърде много варианти за отговор, затова информацията за оценяване се придружава от критерии за оценяване, които предоставят инструкции и принципи за поставяне на оценки. Оценителите трябва да бъдат подходящо обучени, за да използват критериите за оценяване ефективно.

Процес за разработване на задачите

Процесът за разработване на задачите трябва да стартира, след като екипът, който ръководи оценяването, приключи със спецификацията на теста.

Възможно е подготвянето на задачи за определена област да става без или преди изработването на спецификацията на теста. Въпреки това без тази таблица авторите на задачите рискуват да съставят твърде много задачи, които да не отговарят на целта на оценяването, и да не успеят да обхванат частта от учебното съдържание, което тестът цели да оцени.

Използването на спецификацията на теста като насока при разработването на задачи увеличава валидността на финалния тест и намалява разходите за процеса по съставяне на теста.

При изготвена спецификация на теста процесът по разработване на тестовите задачи трябва да следва следните стъпки:

1. Избор на автори.
2. Разработване на съдържателна рамка на теста.
3. Разработване на въпроси.
4. Разработване на отговори.
5. Форматиране на задачите.
6. Преглед и одобряване на задачите.
7. Апробиране на задачите чрез пилотен тест.
8. Анализ и преработване на определени характеристики на задачите.
9. Създаване на банка със задачи или допълване на съществуваща такава.

Ако задачите, съставени в края на процеса, не отговарят изцяло на изискванията на спецификацията, е възможно да се наложи целият процес да бъде повторен. За да се намалят разходите за създаване на качествени тестови задачи, преди апробирането трябва да са подготвени поне 2 пъти повече задачи, отколкото са необходими за крайния вариант на теста. По-конкретно за дяловете от учебното съдържание, които са трудни за оценяване, трябва да се предвидят повече тестови задачи в сравнение с по-ясни и разбираеми дялове.

Избор на автори на тестови задачи

Авторите на тестови задачи трябва да познават както учебната програма, така и когнитивните процеси на учениците при работа върху задачи.

Като цяло, при селекцията на автори на задачи трябва приоритетно да се вземе под внимание познаването на учениците. Валидността на задачите зависи повече от способността на авторите да предизвикат желаното когнитивно поведение у учениците, изучавали материала, отколкото от знанието за познавателните процеси и различните подходи.

За да се постигне баланс между необходимостта да се познават учениците и съдържанието на учебната програма, най-добре е за автори на задачи да се подберат учители с голям преподавателски опит. Специалисти по съдържанието ще бъдат по-полезни като членове на експертната група, която ще проверява съответствието на задачите със съдържателната рамка на теста. Специалисти по съдържанието на учебната програма без опит в работата с ученици не бива да се използват като автори на задачи, тъй като познавателните процеси при експертите са различни от тези при учениците.

Ако тестът засяга чувствителни културни или демографски теми, авторите на задачите трябва да бъдат подбрани така, че да представят пропорционално демографския състав на учениците. Макар възможните пристрастия в задачите да могат да бъдат минимизирани при прегледа им от екипа, ръководещ оценяването, процесът на разработване на задачите би бил много по-ефективен, ако авторският екип успее да идентифицира пристрастията и предварителните нагласи на учениците.

В техническата документация на теста за всеки от авторите трябва да бъде записана следната информация:

- Образователен ценз.
- Опит (брой години като преподавател, практически опит по съответния предмет).
- Населено място и вид училище, в което преподава.

Необходимата демографска информация може да се променя в зависимост от културната чувствителност на съдържанието, предмет на оценяването. Някои области – например математическите изчисления, са относително независими от културни фактори, докато предмети, свързани с хуманитарните науки, са доста културно чувствителни. Като минимум демографската информация трябва да включва пол и възраст. В зависимост от предмета на оценяването може да бъде включена и информация относно етническа принадлежност, местоживее (градска или селска среда) и основен език.

Разработване на условията на задачите

Едно от изискванията за постигане на валидност на теста е съдържанието на оценяването да е съобразено с нивото на учениците и да покрива съответния дял от учебната програма. За придържане към практическата цел на теста често е добре процесът по съставянето му да започне със събирането на набор от съпътстващи помощни документи, свързани със съдържанието на съответния дял на учебната програма.

Целта на съпътстващата документация е да гарантира, че отговорите на всички ученици са върху задачи, съдържащи една и съща практическа информация. Най-общо, един тест трябва да съдържа цялата необходима информация, за да може да се отговори на всички въпроси, които се измерват съгласно съдържателната рамка на теста.

В **Пример 3.1.** са дадени двете задачи, които илюстрират два подхода за съставяне на условие. Примерите са взети от тест, който би трябвало да оценява познанията за храносмилателните ензими, но НЕ и за човешката анатомия.

Пример 3.1. Съставяне на условие на тестова задача

Добър пример	Лош пример
Условие: Следните храносмилателни ензими се произвеждат от различни жлези и се използват в различни части на храносмилателния тракт: Ензим – Произведен от – Място в организма амилаза – слюнчени жлези – устна кухина пепсин – стомашни жлези – стомах жлъчка – черен дроб – тънко черво трипсин – панкреас – тънко черво Въпрос: Коя част на тялото отговаря за разграждането на мазнините? Отговор: (0) А. уста (0) В. стомах (1) С. тънко черво	Условие: Следните храносмилателни ензими се произвеждат от различни жлези: Ензим – Произведен от амилаза – слюнчени жлези пепсин – стомашни жлези жлъчка – черен дроб трипсин – панкреас Въпрос: Коя част на тялото отговаря за химическото разграждане на мазнините? Отговор: (0) А. уста (0) В. стомах (1) С. тънко черво

Следните два примера използват различни формати за представянето на информацията в условията на задачи, включени в програма за оценяване от Британска Колумбия.

В **Пример 3.2.** е използвана политическа карикатура, за да се стимулира мисленето, свързано с важно историческо събитие и да се предостави база за оценяване на аналитичните способности на учениците. В Канада политическите карикатури често се използват при социологически проучвания и тестове по история, тъй като провокират мисленето за определено историческо събитие или за актуално събитие.

Пример 3.2. Задача, използваща изображение в условието

Използвайте следната карикатура, за да отговорите на въпрос № 40.



40. Какво е посланието на карикатурата, свързано с пътя към мира?

- A. Труден е.
- B. Почти е достигнат.
- C. Може да се постигне само посредством виетнамизация.
- D. Може да се постигне чрез американско производство.

Източник: Изпит по история за XII клас, Британска Колумбия, август 2006

Пример 3.3. Задача, използваща таблица в условието

Разгледайте таблицата и определете какъв е крайният баланс на 16 юни?

Справка за транзакциите по сметка 412-783-27					
Чек №	Дата	Описание	Дебит	Кредит	Баланс
					212,11
15	28 май	E-bay	54,70		
Депозит	30 май	заплата		260,50	
16	15 юни	Shell	80,25		
17	15 юни	Visa	60,53		
Депозит	16 юни	заплата		130,25	

Източник: Изпит по математика за X клас, Британска Колумбия, 2007–2008

В **Пример 3.3.** се използва обикновена банкова форма за проследяване на личните разходи с цел да се прецени способността на учениците да прилагат основни математически умения в своето ежедневие. В последните години министерствата на образованието из цяла Канада все по-често наблюдават върху връзката между това, което се преподава в училище, и знанията и уменията, необходими за реализацията на младите хора в обществото.

Друг инструмент за оценяване на способността да се прилагат знания и умения е използването на текст със специфично съдържание – например инструкции за безопасност при работа с химически торове в една ферма.

В Таблица 3.4. са посочени примерни формати за условия на задачи по различни предмети, които могат да се използват при разработването на задачи.

Таблица 3.4. Примерни формати за условия на задачи

Езикови умения:

- изображения
- поезия
- поредица от думи
- есета (или откъси от есета)
- разкази (или откъси от разкази)
- друга кратка проза (изречения, абзаци)

Социални науки (история, география, гражданско образование):

- карти
- исторически документи
- есета
- карикатури
- описателни текстове

Математика и природни науки:

- изображения
- работни, официални или технически формуляри, образци, модели
- кратки, свързани с предмета текстове
- карикатури
- информация за търговски продукти (цени, съдържание, инструкции и т.н.)

Съставянето на условията на задачите може да изисква сериозна инвестиция на време, необходимо за селекция, създаване, договаряне на авторски права и графичен дизайн. Ето защо този процес трябва да започне значително преди началото на основния процес по съставяне на задачите. Трябва да бъде предвидено и достатъчно време между съставянето на задачите и пилотното тестиране, за да може да се доразвият условията на задачите в съответствие с въпросите и отговорите.

Съставяне на въпроси към тестови задачи

Въпросът и частта с отговорите на една задача се съставят едновременно, тъй като са тясно свързани. Автори на задачи с малък опит в разработването на качествени и ефективни задачи в рамките на мащабни програми за оценяване трябва да бъдат обучени. Обучението трябва да включва следните стъпки:

1. Обсъждане на целта и целевата група на оценяването.

2. Представяне на спецификацията на теста и процеса на съставяне на теста.
3. Предоставяне на общи насоки за писане на въпроси (източници от учебната програма, съществуващи документи, подходящи за съставяне на условията на задачите, авторски права, продължителност на теста и време, необходимо за администрирането му).
4. Представяне на одобрените примерни формати на задачите (с избираем отговор, за съответствие, есе, с кратък отговор).
5. Представяне на примери с подходящи въпроси.
6. Представяне на неподходящи въпроси за съответната целева група и цел (твърде общи, недостъпни, неясни, подвеждащи). Съставяне на примерни въпроси и разглеждането им в светлината на общите принципи за съставяне на тест.
7. Задаване на теми и въпроси, които да бъдат взети под внимание от авторите (общите теми в задачите правят работата по теста по-добре съгласувана, а самият тест – по-последователен, а това минимизира нежелателното натоварване на учениците с твърде много отговори).
8. Преглеждане и одобряване или преработване на темите.
9. Създаване на авторски въпроси въз основа на одобрените теми.

Обученията трябва да бъдат групови. Ако е невъзможно всички автори да бъдат обучени едновременно, всяка отделна група трябва да включва автори с разнообразен опит и от различни населени места и видове училища. Разнообразният състав на групите предизвиква съгласуване на нагласите и нивото на информираност относно процеса за разработване на тестови задачи. Това своеобразно стандартизиране намалява вероятността от наличието на задачи, които могат да бъдат дискриминиращи към определена група ученици и също така повишава валидността на теста.

Най-добрият източник на информация за авторите са примерите с качественни и некачествени задачи от предходни оценявания или пилотни тестове. Ако няма такива задачи, проектозадачите трябва да бъдат преценени от групата обучители, съобразно общите принципи за съставяне на задачи, като е необходимо да се достигне до консенсус относно тяхното качество. Задачите с незадоволително качество изискват същото време и усилия, за да бъдат разгледани, както и добре написаните задачи. Ако авторите на задачи бъдат запознати с често срещаните грешки, които да избягват, и с моделите, които да следват, то стойността на преглеждането на проектозадачите и пилотния тест може да се намали.

Задачи с избираем отговор

Съставянето на задачи с избираем отговор се различава значително от съставянето на задачи със свободен отговор.

По същество задачите с избираем отговор са обикновено по-ясни и съдаването им е линейно. Тъй като верният отговор е предварително известен,

предизвикателството е да се измисли поредица от подвеждащи отговори (дистрактори). Те трябва да бъдат достатъчно правдоподобни, така че учениците, които не могат сами да достигнат до верен отговор, да изберат по-скоро подвеждащия, а не правилния отговор. Има няколко типа задачи, подходящи за мащабни или национални оценявания. Следната таблица излага в резюме техните характеристики:

Формат	Описание
Задачи с избор на един верен отговор измежду няколко възможни или алтернативен избор (един от два).	Въпрос с от два до пет възможни отговора. Този тип тестови задачи могат да се състоят от разширено условие, следвано от няколко въпроса. Отговарянето на всеки въпрос е базирано на информацията в общото условие.
Задачи с поредица от твърдения, които трябва да бъдат определени като верни или неверни.	Условие с от 3 до 30 твърдения (имплицитни въпроси). Всяко от твърденията трябва да се определи като вярно или невярно в зависимост от информацията в условието.

Разработването на задачи с избираем отговор трябва да отговаря на следните изисквания (адаптирани от Haladyna, Downing, & Rodriguez (2002), „Преглед на правилата за писане на задачи с избираем отговор за оценяване в час“, Приложно оценяване в образованието, 15(3), 309–334):

1. Всяка задача трябва да отразява конкретен учебен материал и да измерва специфичен мисловен процес, както е определено в спецификацията на теста.
2. Всяка задача трябва да се концентрира върху съществени придобити знания, като избягва несъществената част от учебната програма.
3. При оценяването на по-високите когнитивни равнища е добре да се използват непознати материали. За да се избегне неволното оценяване на способността на учениците просто да възпроизвеждат запаметена информация, е добре да се перифразира характерният език от учебниците.
4. Съдържанието на всяка задача трябва да е независимо от съдържанието на останалите тестови задачи.
5. При разработването на задачи с избираем отговор трябва да се избягва твърде конкретно и твърде общо съдържание.
6. Трябва да се избягват задачи с отговор, зависещ от личното мнение на ученика.
7. Трябва да се избягват задачи с подвеждащи елементи (подвеждащи въпроси, подвеждащи отговори или и двете).
8. Трябва да се използва прост език, разбираем за учениците.
9. Трябва да се избягват сложни задачи с избираем отговор, които предлагат припокриващи се отговори.

10. Възможните отговори на задачите е добре да се формират вертикално, а не хоризонтално.

11. Текстът на задачите трябва да премине редакция и коректура.

12. Трябва стриктно да се съблюдават граматиката, пунктуацията и правописът.

13. Текстът в задачата трябва да се сведе до възможния минимум.

14. Инструкциите в условието трябва да са много ясни.

15. Основната идея на задачата трябва да се съдържа в условието, а не в отговорите.

16. Трябва да се избягва многословието.

17. В условието трябва да се подбират положителни форми и да се избягват негативните изрази „НЕ“ и „С ИЗКЛЮЧЕНИЕ НА“. Ако се налага използването на негативни изрази, това трябва да става с особено внимание, като те трябва да се изпишат с главни букви и удебелен шрифт.

18. Всяка задача може да съдържа толкова възможности за отговор, колкото авторите сметнат за необходимо, но опитът показва, че четири отговора са достатъчни.

19. Само един от отговорите трябва да е верен.

20. Мястото на верния отговор е добре да се променя за всяка следваща задача.

21. Отговорите трябва да са подредени в логически ред, а ако съдържат цифри – според величината им.

22. Отговорите трябва да са независими и да не се припокриват.

23. Задачите трябва да имат хомогенна структура на съдържанието и граматиката.

24. Отговорите е добре да бъдат със сходна дължина.

25. Изразът „НИТО ЕДИН ОТ ПРЕДИШНИТЕ“ трябва да бъде използван с особено внимание.

26. Трябва да се избягва изразът „ВСИЧКИ ПРЕДИШНИ“.

27. Трябва да се използват положителни форми и да се избягват отрицанията „НЕ“ и „НЯМА“.

28. Трябва да се избягват насочващи фрази към верния отговор, като:

– специфични определения: винаги, никога, напълно и абсолютно;

– изрази и думи, напомнящи на тези от условието;

– граматически неправилни изрази, които насочват към верния отговор;

– твърде очевиден верен отговор;

– двойки или тройки еднакво недопустими отговори, които директно насочват към верния;

– откровено абсурдни и невероятни отговори.

29. Всички дистрактори трябва да се направят да звучат достоверно.

30. В дистракторите трябва да се избягват грешки, които учениците често правят.

31. Добре е да се използва хумор, само ако е съвместим с нагласата на преподавателя и общата атмосфера на учебния процес.

Информацията, придружаваща отговорите на задачи с избираем отговор, трябва да съдържа:

- възможните отговори;
- подреждането на отговорите;
- ключа с верните отговори;
- начина, по който липсващите отговори да бъдат третираны – като грешни или пропуснати. При повечето текущи тестове (национални оценявания) липсващите отговори се третират като пропуснати. При приемни изпити, конкурси и тестове за получаване на сертификат, липсващите отговори трябва да се третират като грешни. При тестове със задачи, подредени с нарастваща трудност, до определена задача трябва да се третират като липсващи, а към края на теста – като грешни. Обратното важи за текущи тестове, чиито резултати не се съобщават индивидуално на учениците.

Задачи със свободен отговор

Създаването на частта с отговорите за задачи със свободен отговор е далеч по-просто, в смисъл че не се губи време за писане на дистрактори. Същевременно от авторите се изисква повече информация за изработване на критериите и принципите, въз основа на които оценителите ще преценяват отговорите като верни или грешни и ще разграничават качеството на верните отговори при различните ученици. Критериите за оценяване на всяка задача със свободен отговор трябва да включват:

- ключ на верните отговори и критерии за оценяване;
- план за обучение на оценителите;
- протоколи за оценяване;
- примерни оценени задачи с реални отговори на ученици.

Ключ на верните отговори и критерии за оценяване

Аналитичните критерии определят отделните аспекти на постиженията на учениците и позволяват на оценителя да постави отделна оценка на всеки от тях. Общата оценка на задачата е обобщение (средно аритметично или сбор) на оценките на отделните аспекти. Холистичните критерии предоставят по-разширено описание на постигнатото от учениците, включително подробно описание на възможното качество на отговора. Това помага на оценителя прецизно да определи нивото на конкретния отговор. Оценителят трябва да намери описанието, което най-точно характеризира постигнатото от всеки ученик, и да постави съответната оценка. Холистичните критерии са по-полезни, ако са придружени с примери на оценени задачи. Аналитичните критерии предоставят по-добра информация за откриване на проблеми в преподаването.

План за обучение на оценителите

Обучителният план е методология за обучение на оценителите за поставяне на оценки на отговорите. Обучаемите имат нужда от внимателно разясняване на обучителните цели, достигането на които ще бъде измервано със съответната задача, както и подробно описание на факторите, които е възможно да присъстват в отговорите, но не следва да се вземат под внимание при поставяне на оценката (например, ако се оценява дали ученикът е разбрал смисъла, без да се оценява правописът му). Оценителите трябва независимо да поставят оценки на многобройни отговори. Когато техните оценки са в противоречие с тази на главния оценител (ако има такъв) или с общото мнение на останалите оценители, различията трябва да се разрешат. Обучението приключва, когато се достигне минималният праг на съгласуване на оценките. Оценители, които не успеят да стигнат до това ниво на съгласуваност, не трябва да бъдат допускани като оценители на реални отговори на ученици.

Протоколи за оценяване

При провеждането на теста определен процент от задачите (от тези със свободен отговор) трябва да бъдат оценени от двама независими оценители. Протоколите за оценяване трябва да съдържат и методология за разрешаване на случаи с твърде различни оценки на една и съща задача. В повечето случаи тези отговори се препращат на главния оценител (арбитър). Неговата преценка определя окончателния резултат.

Примерни оценени задачи

Примерните оценени задачи представляват набор от отговори с разнообразни характеристики, които недвусмислено отговарят на критериите за поставяне на определена оценка. Примерните оценки могат да бъдат изработени и по време на обучението на оценителите. В зависимост от сложността на изискванията към тестовите задачи всяка оценъчна категория трябва да бъде онагледена с от един до пет примера, които да служат за справка.

Форматиране на задачите

Следните фактори могат да окажат негативно влияние върху валидността на оценяването:

- Четивност – намалява при използване на твърде дребен или нечетлив шрифт.
- Претоварване на паметта – когато учениците трябва да правят справки на различни страници или в отдалечени пасажи на една и съща страница.
- Разконцентриране – поради значително количество ненужна или трудно обяснима информация, като усложнено номериране, многословни заглавия на отделните части или забележки под линия или в полето на страницата.

Всяка задача трябва да бъде форматирана съобразно установените указания преди нейното одобряване, за да се избегне влиянието на изброените по-горе фактори.

Одобряване на задачите

Всички задачи трябва да бъдат одобрени от специално сформирана експертна група (това може да бъде и екипът, ръководещ оценяването), преди те да бъдат включени в пилотния тест. Целта на процеса на одобряване е броят на задачите, които ще бъдат включени в пилотния тест, да бъде сведен до максималния брой, който е възможно да бъде адекватно анализиран въз основа на данните от пилотната тестова извадка. Ако след одобряването се окаже, че определен дял от учебното съдържание не е достатъчно застъпен, задачите трябва да се върнат на авторите за преработване или за разработване на нови задачи.

Като вземем под внимание колко скъпо и трудоемко е създаването и обучаването на екип от автори, по-ефикасно е да се напишат първоначално много повече от необходимия брой задачи. Това би спестило евентуално пренаписване или дописване на нови задачи. Обикновено 150% резерв е достатъчен, като това зависи от естеството на съответния дял от учебната програма (при трудни за оценяване области и умения трябва да се предвиди по-голям резерв от задачи). Този резерв позволява в процеса на редактиране и одобряване на задачите да бъде изключен достатъчно голям брой от тях. Така задачите, които ще бъдат включени в пилотния тест, трябва да надхвърлят с около 50% броя на задачите, които ще бъдат включени в същинския тест.

Експертната група, която одобрява задачите, има пет отговорности:

1. Да прегледа и коригира граматичните грешки и двусмислията. Освен установяването на очевидни грешки групата трябва да се погрижи задачите да не са обременени с твърде много текст, както и сложността на теста да е под нивото на оценяваните ученици. Ако например се оценяват ученици в VII клас по математика, граматическата сложност на задачите трябва да е под нивото, изисквано за VII клас. Текстовете в задачите по четене, които не са пряко свързани с измерване резултатите от обучението (например инструкциите), трябва да бъдат на достатъчно ясен и прост език.

2. Да установи и коригира грешките във формата. Всякакви визуални характеристики на задачата, които понижават четивността, разсейват или ненужно натоварват мисълта, трябва да бъдат отстранени. Ако учениците трябва да интерпретират една фигура, то тя трябва да има точно толкова визуални елемента, колкото е нужно, за да оцени адекватно учениците. Нерядко по време на одобряването много оригинални текстове в условията трябва да бъдат опростени, за да се премахнат ненужните елементи.

3. Да установи и коригира технически грешки, съгласно инструкциите за писане на задачи.

4. Да открие и елиминира културни, религиозни или влияния на пола. Експертната група трябва да има демографски разнообразен състав. Всеки от членовете ѝ трябва да се увери, че темата или изказът на задачите не би довел до затрудняване на учениците, които са с определени демографски характеристики.

5. Да потвърди съответствието на задачите с целта на теста. Експертната група трябва да отговори на всяка задача и да потвърди, че процесът, през който следва да премине ученикът, за да отговори на въпроса в задачата, е пряко свързан само с онези цели на обучението, с които задачата е обвързана съгласно спецификацията на теста. В идеалния случай в този процес трябва да участват и ученици, като те трябва да мислят на глас, докато отговарят на въпросите, за да може да бъде описан когнитивният процес, през който преминават.

Съставяне на тестове и провеждане на пилотни тестове

Фактори, които влияят върху ефективността на задачите

Основен принцип при съставянето на тестове е качествата на теста да не влияят върху способността на задачите да са ефективни съобразно първоначалния си замисъл.

Някои от факторите, които влияят върху ефективността на задачите, са:

- емоционални отговори на теста;
- умора на отговарящия;
- разконцентриране;
- взаимовръзка между задачите.

Емоционални отговори

Емоционалните отговори в теста добавят „шум“ към тестовите резултати, тъй като те са повлияни от изненада, разочарование, гняв, объркване, а не са отражение на знанията на отговарящия. Влиянието им може да бъде минимизирано чрез спазване на следните принципи:

- Представяне на ясни инструкции.
- Групиране на задачи с общи или сходни условия.
- Поставяне в началото на теста на няколко по-лесни задачи или вмъкване на лесна задача след някоя особено трудна.
- Групиране на задачите със сходен формат.
- Редуване на задачи с относително дълги условия с такива с кратки условия.

Умора на отговарящия

Умората се наблюдава, когато учениците са изтощени (психически или физически) от процеса на оценяване. Прагът на умора е различен при различните ученици, но е известно, че средно когнитивните способности започват да спадат след осемдесетата минута. Ако тестът е важен (конкурсен) за отделния ученик, ефектът от умората настъпва по-бавно, което позволява по-голяма продължителност на теста. Текущите и пилотните тестове обаче е възможно да дадат неточни резултати, ако се провеждат за повече от един час, освен ако не бъде предоставен някакъв стимул.

Следните подходи биха понижали ефекта от умората на отговарящите върху финалните резултати от теста:

- За важни (конкурсни/зрелостни) тестове най-трудните задачи се поместват към края на теста. Вероятността учениците, които не са успели да стигнат до тях, да се справят успешно, ако бяха стигнали, е много малка.

- Ако съдържанието на теста налага време за провеждане повече от 2 часа, задачите се разделят в два отделни свитъка. След като учениците попълнят и предадат първия свитък, имат кратка почивка, преди да продължат с втория свитък.

- При текущите тестове продължителността на провеждане не бива да надхвърля 90 минути. Обикновено се предоставят достатъчно сериозни аргументи (документация, административни правила), които ясно подчертават, че макар това да не е входен или финален тест, резултатите от него са важни за други цели. Например тестът може да не е важен за отделните ученици, но резултатите от него могат да повлияят върху структурирането на учебната програма или разпределянето на финансирането за образователните програми.

Разконцентриране

Всички страници от тестовата документация, върху които са разположени задачите, трябва да съдържат минимална друга информация, като номера на задачи и страници. Номерата на страниците и задачите трябва да са поредни (1, 2, 3,...), а не обвързани с номера на секцията (1.1, 1.2,..., 2.1, 2.2,...). Изключение от това правило се допуска при използване на повече от един тестов свитък. При този случай номерацията при всеки свитък започва от 1. За стимулиране на мисленето на учениците често се използват рисунки (карикатури). Те трябва да бъдат поместени в началото или в края на тестовия свитък и не бива да се поставят на една и съща страница с други тестови задачи.

Взаимовръзка между задачите

Макар задачите да са съставени като независими единици, неизбежно някои от тях ще съдържат подсказваща информация, която ще помогне на учениците да се справят по-добре с други задачи. Задачи, свързани по този начин, не трябва да се включват в един и същи тест.

Обикновено причините за тази взаимовръзка са три:

- Информация от условието или дистрактор от една задача описва или дава подробности за образователна цел, измервана от друга задача. Например в химията една задача може да описва атомната структура на 2 молекули и да очаква учениците да преценят коя е с по-голяма атомна маса, докато друга задача, която изисква учениците да изчислят масата на едно химическо съединение, е възможно да предоставя таблица с атомните маси на всички елементи. Тази връзка може да се елиминира, като информацията от втората задача, която съдържа отговора на първата, се премахне.

– Основното твърдение в една задача се подлага под въпрос в друга. Например в тест по литература една задача може да очаква учениците да идентифицират основния конфликт в едно произведение (например човек срещу собственото си его), докато втора задача може да иска от учениците да обяснят защо лирическият герой е в конфликт със себе си. В този случай една от задачите следва да отпадне.

– Особено при решаването на задачи по математика и природни науки е възможно междинната стойност от изчисленията от една задача да се окаже междинен или краен резултат от друга. Такива задачи трябва да се преработят, така че да няма две задачи с еднакъв път на изчисление в един и същи тест.

Условия за провеждане

Тестът трябва да съдържа подробно описание за условията, при които трябва да бъде проведен, за да даде валидни резултати. Условията за провеждане влияят върху представянето на учениците. Те не оказват влияние само ако всички ученици са поставени при еднакви условия.

Следните условия за провеждане трябва да бъдат стандартизирани:

1. Условия за работа: определяне на минималното разстояние между учениците, вида на столовете и работните маси, които трябва да се използват. Като цяло учениците не бива да са с лице един към друг по време на теста.

2. Температура: определяне на температурните граници, при които учениците са концентрирани.

3. Време на провеждане: всички ученици трябва да попълнят теста по едно и също време.

4. Помощ (или изключения) за ученици със специални образователни потребности: определяне на необходимите ресурси, за да бъдат резултатите им съпоставими с тези на останалите ученици.

5. Правила за квесторите: определяне на броя на учениците, които могат да бъдат в една и съща зала по време на теста, и възможностите за напускане на залата преди края на теста; определяне на необходимия брой квестори; инструктиране на квесторите.

6. Секретност на тестовите материали: определяне на процедури за осигуряване на поверителността на тестовите материали и гарантиране на неприкосновеността им преди провеждането на теста. Ясно описание на механизмите, чрез които се гарантира еднозначно идентифициране на учениците с тяхната анонимна изпитна работа. Описание на процедурите за транспортиране и съхранение на изпитните материали.

7. Събиране на данните и обработване на резултатите: при използване на електронни методи за обработване на резултатите учениците трябва да бъдат обучени да използват специални листове, които подлежат на сканиране. Ако се прилага ръчно оценяване, трябва да се изработи и използва дефиниция за допустим

отговор (например допустимо е ограждане на верния отговор, докато отмятане или отбелязване с „X“ е недопустимо).

8. Правила за използването на помощни средства: определяне на правила за използването на:

- учебници
- пособия за чертане
- уреди за измерване
- калкулатори
- компютри
- Интернет
- други справочни материали.

Ако използването на някои от гореизброените е позволено, трябва да се направи нужното всички ученици да разполагат с тях по време на теста. С оглед на сигурността не бива да се позволява на учениците да си разменят помагала. Трябва да се осигурят достатъчно помощни материали, така че всеки ученик да има свои.

Инструкциите за провеждане на теста са част от самия тест и трябва да бъдат разпространени заедно с него. За да могат различните места на провеждане на теста да са адекватно подготвени, е необходимо инструкциите за провеждане на теста да се разпространят достатъчно време преди самото му провеждане.

Провеждане на пилотни тестове

Целта на пилотните тестове е да се провери дали тестовите задачи постигат заложената в тях цел, а не способността на учениците да се справят. По тази причина дизайнът на един пилотен тест се различава от този на финалния тест. Резултатите от пилотния тест никога не бива да се използват за правене на изводи за постиженията на отделен ученик или група от ученици.

Избор на задачи за пилотен тест

В идеалния случай всички одобрени задачи би следвало да бъдат включени в пилотното изследване. Ако обаче ресурсите са ограничени, приоритет за включване се дава на тези задачи, чиито измерителни качества са неубедителни. Например, ако всички членове на експертната група са безрезервно съгласни, че една задача е отличен инструмент за измерване на определена образователна цел, необходимостта от нейното пилотно тестване е по-малка в сравнение с тази на задача, за която има съмнения, че е двусмислена или която измерва повече от една цел. Задачи, свързани с дялове от учебното съдържание, които в предходни тестове са се оказали трудни за измерване, също трябва да бъдат включени приоритетно. В резултат съдържанието на пилотния тест може доста да се различава от съдържанието на финалния тест, определен в спецификацията.

Забележка: Тази препоръка не означава, че в пилотния тест преобладаващата част от задачите трябва да бъдат с проблемни качества. Ако повече от 20%–30% от задачите в теста са с незадоволително качество, резултатите от пилотния тест няма да бъдат надеждни.

Селекция на участници в пилотен тест

В идеалния случай извадката трябва да представлява всички ученици от целевата популация. Структурата на извадката се определя от изводите и решенията, които трябва да се вземат въз основа на показаните резултати. Ако резултатите ще се използват за вземане на решения, свързани с отделните ученици, извадката трябва да отразява предварително дефинираните демографски групи (например момичета/момчета, градско/селско население), за да може да се постигнат достоверни резултати за всички.

Ако резултатите ще бъдат предмет на статистически анализ, тогава трябва да бъдат дефинирани демографски групи, които да представляват групите, сравнявани в анализа (например социално-икономически статус, етническа принадлежност, регион). Ако изводите ще бъдат базирани на взаимовръзката между характеристиките на тези групи, носителите на повече от една характеристика ще бъдат третирани като отделна група (например момчета от селски региони; момчета, живеещи в градска среда; момичета от селски региони; момичета, живеещи в градска среда, и т.н.). Групите, определени по този начин, определят логическата плоскост на пилотната извадка. Макар да не е възможно винаги извадката да се стратифицира по всички групи, все пак трябва да се направи максималното групите да бъдат равномерно представени.

Често финансовите ограничения и други практически съображения не позволяват да бъдат включени ученици от всички групи. В подобни случаи приоритет се дава на две категории ученици:

– Групи, които най-вероятно ще бъдат засегнати от решенията, основаващи се на резултатите от теста. Например, ако тестът цели да идентифицира недостатъци в усвояването на знания, приоритетно се включват групи, които се очаква да се представят слабо. Ако тестът цели да селектира най-добрите, то трябва да бъдат включени групите, които се очаква да се представят най-добре.

– Групи, по отношение на които ефективността на задачите е под съмнение. Например, ако експертната група единодушно реши, че повечето задачи са ефективни по отношение на повечето ученици, но не и за техните връстници от определен регион, трябва да бъдат включени ученици именно от него.

Размер на извадката

Макар по-голямата извадка да предоставя повече информация, финансовите ограничения често налагат пилотните извадки от ученици да са по-малки от необходимото за задълбочен статистически анализ. За да могат да се направят

изводи за тестовите задачи обаче, извадката трябва да включва поне по 50 ученици от всяка идентифицирана за целите на пилотния тест демографска група.

Дизайн на пилотния тест

Съставянето на пилотния тест трябва да следва общите принципи за съставяне на тест. Обикновено в пилотния тест са включени повече задачи в сравнение с финалния и мотивацията на учениците е относително по-ниска. Затова ефективният пилотен тест е добре да се състои от няколко тестови свитъка, като за попълването на всеки от тях да е предвиден 1 час. Задачите във всеки от свитъците трябва да бъдат подбрани в произволен ред.

Ако учебният материал, който предстои да бъде тестван, има три дяла, във всеки отделен свитък трябва да се включат пропорционално задачи от всеки дял. Поне 50% от задачите във всеки свитък трябва да присъстват поне в още един свитък. Например, ако в пробния тест има 100 задачи, а всеки от свитъците съдържа 50, трябва да се съставят минимум 3 свитъка.

Рискове при управлението на теста

Една от най-сериозните заплахи за осъществяването на едно оценяване е недобронамерената политическа воля. Ако политически важни групи, като представители на обединенията на учителите или учениците, или влиятелни представители на образователната общност загубят доверие в оценяването, те могат да повлияят върху степента на участие в оценяването или да се опитат да елиминират оценяването изцяло. Ако някоя от тези групи участва или има пряк интерес от процеса на провеждане на пилотния тест, те трябва да бъдат осведомени за целта и ограниченията на пилотния тест.

Резултатите от пилотния тест може да не отразят точно картината на успеваемостта сред учениците. В зависимост от приоритетното включване на определен тип задачи и групи ученици средният резултат може да е по-висок или по-нисък от очаквания за целевата популация на финалния тест. Резултатите могат да бъдат претеглени в зависимост от определените групи задачи и групи ученици, като така те ще представят пропорционално групите в тестовата спецификация или в целевата популация, съответно.

Много от задачите в пилотния тест могат да се окажат много по-не-ефективни от очакваното. Учениците използват различни познавателни процеси от преподавателите и експертите. По тази причина задача, която експертната група очаква да е ефективна, може да се окаже неефективна. Авторите на задачи, комисията по теста и експертната група трябва да са подготвени за тази възможност. Тъй като ефективността на задачите е приоритет на пилотното тестиране, влиятелните членове на експертните групи не бива да бъдат поставяни в позиция да защитават репутацията си, дискредитирайки теста.

Кога провеждането на пилотен тест е неподходящо

Провеждането на пилотни тестове за апробиране на задачите е основно изискване в процеса по организиране на изпити и оценявания. Съществуват обаче обстоятелства (например при приемни или конкурсни тестове, при които поверителността е особено важна), при които провеждането на пилотен тест без компрометиране на същинското оценяването би било трудно.

При подобни обстоятелства изборът на автори на задачи и процесът на писане и одобряване на тестовите задачи стават още по-важни. Колкото по-добре е организиран този процес, толкова по-малко проблеми ще бъдат идентифицирани при анализа на ефективността на задачите след провеждането на теста.

Както след провеждането на всеки друг тест, и в този случай, трябва да се направи анализ на задачите. Всички некоректни задачи трябва да бъдат премахнати от оценяването или преработени.

БЕЛЕЖКИ

1. Ръководството е разработено от Фернандо Картрайт и Джери Мусио за Център за контрол и оценка на качеството на училищното образование по проект, финансиран от Световна банка.

DEVELOPMENT OF TESTS, A MANUAL

Abstract. Education has become an important determinant of the social and economic progress of individuals and of nations, and it typically ranks as one of the top priorities established by national governments.

National student assessment programs play a central role in measuring the quality of education and monitoring the impacts of education reform. As such, assessment agencies have a responsibility to ensure that the development of tests and examinations meet the highest professional standards.

The purpose of this manual is to assist a staff responsible in the design and development of student tests, and the analysis and reporting of results. Our focus here is on the practical application of test theory based on our experiences in carrying out a variety of assessment programs in Canada and internationally.

The main themes are as follows: planning the assessment, test design, item development, test development & pilot testing, item and test analysis, interpreting test results.

**Fernando Cartwright
Jerry Mussio**