

Дизајнирање ефикасних питања вишеструког избора за оцењивање исхода учења

УДК 371.275

Јасмина Јовановска
jasmina.armenska@gmail.com

Педагошки факултет
„Св. Климент Охридски“
Универзитет
„Св. Тирило и Методије“
Скопље, Македонија

САЖЕТАК: Тестови вишеструког избора данас представљају популаран алат за оцењивање исхода учења јер су флексибилни, релативно једноставни за спровођење, а омогућавају и оцењивање садржаја великог обима за кратко време. Много важнија чињеница је да су ова питања свестрана и поуздана, што повећава њихову атрактивност. Осим тога, за потребе оцењивања испитивачи могу користити колекције питања које су већ припремљене за претходне курсеве или су доступне онлајн. С друге стране, многи аутори тврде да се питања вишеструког избора обично користе за оцену когнитивних способности нижих нивоа, дефинисаних Блумовом таксономијом. Међутим, свеобухватне анализе потврђују да ова врста алата за оцењивање омогућава и процену мишљења вишег реда.

КЉУЧНЕ РЕЧИ: тестови вишеструког избора, когнитивни нивои, индекс тежине задатка, индекс дискриминације задатка

Превод с енглеског:
Јелена Јаћимовић

РАД ПРИМЉЕН: 10. мај 2018.
РАД ПРИХВАЋЕН: 28. мај 2018.

1. Увод

Оцењивање је систематско прикупљање и анализирање информација ради побољшања учења (Stassen et al., 2001). Мерење учења путем оцењивања је важно јер пружа корисну повратну информацију и наставницима (испитивачима) и ученицима о томе у којој мери ученици испуњавају циљеве учења једног наставног предмета. Такође је

корисно наставницима за образлагање педагошких одлука у настави.¹ Најчешће коришћени традиционални алати за оцењивање су тестови вишеструког избора, тестови двочланог избора тачно/нетачно, тестови кратког одговора и есеји.² Brown and Knight тврде да коришћење више различитих алата побољшава поузданост оцењивања. Међутим, овај приступ је прилично изазован јер од испитивача захтева правилно тумачење резултата различитих инструмената оцењивања.

Данас наставници, школе, универзитети и организације за оцењивање обично користе тестове вишеструког избора јер представљају прави ефикасан начин за оцењивање исхода учења (детаљно објашњење о предностима, као и ограничењима питања вишеструког избора дато је у наредна два одељка). Свако *питање вишеструког избора*, познато као *ставка*, састоји се од датог проблема (познатог као *основа*) и списка понуђених одговора (познатих као *могућности*). Могућности обично садрже један *тачан одговор* (најбољи избор), као и неколико нетачних или мање значајних опција, познатих као *дистрактори* (Brame, 2013). Задатак ученика је да одаберу опцију која представља најбољи одговор за дати проблем. Сврха ових дистрактора је да се оним ученицима који нису постигли циљ који ставка мери учине као веродостојна решења датог проблема. Насупрот томе, ученицима који су постигли циљ, ови дистрактори морају да изгледају као решења која нису разумна.

Овај рад је организован на следећи начин. У овом делу дат је преглед предности и мана коришћења питања вишеструког избора у процесу оцењивања исхода учења. Одељак 2. посвећен је стандардним протоколима за повећање валидности питања вишеструког избора. На крају рада наглашен је потенцијал тестова вишеструког избора за процену способности мишљења вишег реда, укључујући и неке препоруке како направити такве тестове.

1.1 Предности коришћења питања вишеструког избора

Тестови вишеструког избора имају одређене предности, али и ограничења, као и било која друга врста тестова. Испитивачи морају да буду упознати са овим одликама како би могли ефикасно да користе питања вишеструког избора. У даљем тексту су представљене

¹ Michael R. Fisher, Jr., "Student Assessment for Teaching and Learning", Center for Teaching, Vanderbilt University (на вебу).

² Maryellen Weimer, "Advantages and Disadvantages of Different Types of Test Questions", Faculty Focus, (на вебу).

најважније предности коришћења тестова вишеструког избора као алата за оцењивање (Burton et al., 1991; Chan, 2009; Dikli, 2003; Towns, 2014).

Свестраност. Тестови вишеструког избора су применљиви у многим различитим предметним областима и могу се користити за оцењивање различитих нивоа исхода учења (од једноставног подсећања до много сложенијих нивоа, као што су примена, анализа и процена). Међутим, ови тестови се не могу применити у сваком тестирању јер ученици бирају из скупа потенцијалних одговора. На пример, они нису ефикасан начин за тестирање способности ученика да организују своје мисли или изразе своје креативне идеје (део 1.2).

Поузданост. Поузданост се дефинише као степен до кога тест доследно мери исходе учења. Одговарајући тестови вишеструког избора су поузданији у односу на тестове који садрже друге врсте питања. На пример, мање су подложни погађању него питања тачно/нетачно. Такође, њихово оцењивање је лакше разумети него оцењивање тестова са кратким одговорима јер нема потребе за разрешавањем делимичних или погрешно написаних одговора. Штавише, оцењивање путем тестова вишеструког избора је објективније у односу на оцењивање које укључује питања чији одговор треба дати у облику есеја. Недоследности испитивача могу утицати на оцењивање тестова у облику есеја, који нису отпорни ни на утицај фактора обмањивања и умешност писања, што може умањити њихову поузданост.

Валидност. Валидност је дефинисана као степен до ког тест мери циљане исходе учења. Будући да се на питања вишеструког избора обично одговара брже него на питања у форми есеја, тестови вишеструког избора могу да буду усмерени на релативно обимнији материјал наставног предмета, чиме се увећава пуноважност оцењивања (Васон, 2003).

Ефикасност. Коришћење тестова вишеструког избора веома је важно за испитиваче јер им они омогућавају једноставну и брзу процену. Ови тестови су посебно значајни испитивачима који држе више наставних предмета са великим бројем полазника. Тестови вишеструког избора омогућавају брже извештавање о резултатима ученика, чиме се испитивачу омогућава брз увид у њихова постигнућа и давање додатних разјашњења и упутстава пре завршетка наставе.

1.2 Ограничења коришћења питања вишеструког избора

Упркос наведеним предностима, оцењивање исхода учења путем тестова вишеструког избора често је критиковано. У даљем тексту овог пододелка налази се преглед ограничења коришћења тестова вишеструког избора (Burton et al., 1991; Chan, 2009; Dikli, 2003; Towns, 2014).

Свестраност. Одређени истраживачи истичу да тестови вишеструког избора пре оцењују способност ученика да памте, него да разумеју, примене и анализирају информације (Walsh and Seldomridge, 2006). Ипак, очигледно је да се ти тестови текође могу користити и за оцењивање мишљења вишег реда. Ово се може постићи укључивањем питања која су усмерена на више нивое спознаје, представљене у добро познатој таксономији Андерсона и Кратвола (прерађеној Блумовој таксономији) (Bloom, 1977; Anderson et al., 2001). Основа таквих врста питања представља проблем, који се може решити само анализом и применом одређених принципа области испитивања. Неопходност процене датих могућности такође може допринети овом процесу. Међутим, израда тестова вишеструког избора намењених оцењивању мишљења вишег реда захтева више вештина и способности од оних које су потребне за развијање питања која оцењују једноставно препознавање и памћење (Palmer and Devitt, 2007).

На слици 1 приказана је таксономија когнитивног домена Андерсона и Кратвола,³ који је подељен на 6 нивоа циљева. Слика 1 такође садржи и коментаре о подобности тестова вишеструког избора за оцењивање приказаних нивоа. Осим тога, у табели 1 дат је детаљан опис сложености и специфичности нивоа. На основу таксономије Андерсона и Кратвола, питања вишеструког избора нису одговарајућа само за проверу највишег нивоа спознаје (креирање). Разлог је то што процес креирања захтева од ученика да делове саставе на нов начин или да их споје у нешто ново и другачије, стварајући нови облик или производ. Процес креирања је најтежа ментална функција у представљеној таксономији.⁴

Поузданост. Питања вишеструког избора су мање подложна погађању него питања из тестова тачно/нетачно, али су у одређеној мери томе ипак

³ Workshop: Designing Effective Multiple-Choice Questions, Teaching and Learning Services, McGill (на вебу).

⁴ Leslie Owen Wilson, "Anderson and Krathwohl – Bloom's Taxonomy Revised", The Second Principle (на вебу).



Слика 1. Прикладност питања вишеструког избора за различите нивое спознаје Андерсонове и Кратволове таксономије

подложна. Погађање смањује поузданост питања вишеструког избора, али се са повећањем броја ставки позданост повећава.

Табела 2 приказује вероватноћу постизања 50% или више на тесту вишеструког избора слепим погађањем тачних одговора (P). Резултати су добијени помоћу биномне расподеле:⁵

$$P = \begin{cases} \sum_{k=\frac{n}{2}}^n \binom{n}{k} p^k (1-p)^{n-k}, & \text{за } n \text{ парно} \\ \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} p^k (1-p)^{n-k}, & \text{за } n \text{ непарно} \end{cases}$$

где је $\binom{n}{k} p^k (1-p)^{n-k}$ вероватноћа да се на питања одговори тачно , а $p = \frac{1}{4}$ је вероватноћа да се на појединачно питање да тачан одговор. Као што се може видети из табеле 2, вероватноћа P једнака је 0,0139, када се тест састоји од 20 питања вишеструког избора. Ако је укупан број студената једнак N , очекивано је да ће од њих постићи 50% или више

⁵ Ronald E. Walpole et al., *Probability & Statistics for Engineers & Scientists*, Prentice Hall (на вебу)

Табела 1. Опис 6 нивоа сложености у таксономији Андерсона и Кратвола

Ниво	Дефиниција
Памћење	Присећање информација
Разумевање	Идентификовање примера датог термина, концепта или принципа. Интерпретирање значења неке идеје, концепта или принципа.
Примена	Употреба информација, правила и процедура у конкретним ситуацијама.
Анализирање	Рашчлањивање информација како би се истражили обрасци и односи. Анализирање графикана, података који подржавају закључивање.
Вредновање	Образлагање одлуке или поступка.
Креирање	Стварање нових идеја или производа.

Табела 2. Вероватноћа резултата 50% или више слепим погађањем, у зависности од броја питања вишеструког избора

Број питања вишеструког избора у тесту са четири могућности (n), са само једним тачним одговором	Вероватноћа резултата 50% или више слепим погађањем (P)
5	0,1035
10	0,0781
20	0,0139

слепим погађањем. На пример, ако је 100 ученика тестирано са по 20 питања вишеструког избора (свако садржи четири могућности, а само је једна од њих тачан одговор), очекивано је да један ученик положи тест слепим погађањем.

Сложеност израде. Кључ за искоришћавање раније представљених предности питања вишеструког избора јесте израда добрих ставки вишеструког избора. Међутим, углавном је теже израдити добра питања вишеструког избора него друге типове питања, што захтева и више времена. Ово је посебно видљиво у процесу одређивања могућих дистрактора, што захтева одређене вештине. Ове способности, међутим, могу се повећати кроз учење, праксу и искуство.

2. Повећање валидности питања вишеструког избора применом стандардних протокола

Свестраност и поузданост су основне одлике питања вишеструког избора, али се њихова валидност не може унапред претпоставити, имајући у виду чињеницу да ученик или ученица могу да погоде тачан одговор чак и када им недостаје очекивано знање. Стога се, ради повећања валидности питања вишеструког избора, препоручује примена стандардних протокола пре и после тестирања (Ramakrishnan et al., 2017).

2.1 Стандардни протоколи за претходну валидацију питања вишеструког избора

Процес *претходне валидације*, који подразумева коришћење смерница и контролних листа, спречава појаву грешака приликом израде питања вишеструког избора. Најзначајније запажање је изградња теста вишеструког избора са независним ставкама. Ово ће онемогућити ученике да користе информације из једног питања како би дали одговор на друго, чиме би се умањила валидност теста. Најупечатљивије смернице које треба поштовати приликом развијања ефикасних ставки вишеструког избора представљене су у даљем тексту. Ради боље илустрације, смернице су допуњене примерима питања вишеструког избора из колекције текстова описаних у (Jovanovska, 2018). Сва питања су преузета из колекције отворених питања Државне матуре у Македонији и јавно су доступна на веб страници Државног испитног центра.⁶ Тачни одговори су истакнути.⁷

⁶ Државен испитен центар, Биро за развој на образование (на вебу)

⁷ У онлајн издању су истакнути црвеном бојом, у штампаном издању су дати малом капиталом.

Табела 3. Пример ставке са бесмисленом основном и њене побољшане формулације

Ставка са бесмисленом основном	Побољшана формулација ставке
Који од наведених израза је тачан:	Феноменологија је наука о?
А. феноменологија је наука о околностима у природи	А. околностима у природи
Б. феноменологија је наука о лепим стварима	Б. лепим стварима
В. ФЕНОМЕНОЛОГИЈА ЈЕ НАУКА О ФЕНОМЕНИМА У СВЕСНОСТИ	В. ФЕНОМЕНИМА У СВЕСНОСТИ
Г. феноменологија је наука о принципима добре воље	Г. принципима добре воље

Табела 4. Пример основе са небитним информацијама и њене побољшане формулације

Основа са небитним информацијама	Побољшана формулација основе
Када је Аристотел имао 18 година, отишао је у Атину како би учио на Платоновој Академији. Као најбољи ученик, сарадник и предавач, након тога постао је тотор:	Чији је учитељ био Аристотел?
А. АЛЕКСАНДРУ ВЕЛИКОМ	А. АЛЕКСАНДРА ВЕЛИКОГ
Б. Филипу Македонском	Б. Филипа Македонског
В. Сократу	В. Сократа
Г. Ел Фарабију	Г. Ел Фарабија

Изградња ефикасне основе. Следећи захтеви су кључни у процесу изградње ефикасне основе (Frey et al., 2003):

1. Основа би сама по себи требало да буде смислена и треба да представља јасно дефинисан проблем. Таква основа обезбеђује да је ставка усмерена на оцењивање исхода учења (табела 3).
2. Основа не би требало да садржи небитне информације јер може смањити поузданост и валидност резултата испитивања (табела 4) (Haladyna and Downing, 1989).
3. Основа би требало да буде исказана са негацијом само онда када то захтева значајан исход учења. Студије потврђују да ученици имају потешкоће у разумевању негативно формулисаних ставки (Michael Rodriguez, 1997). Ако постоји потреба за негативним изразом како би се оцениле специфичне вештине и знање (као што је у медицини), онда негација мора бити истакнута курзивом или великим словима (табела 5).
4. Основа би требало да буде питање или део реченице (Statman, 1988). Питање као основа је пожељније јер омогућава ученику да се усредсреди на одговарање на питање, а не на памћење дела реченице и њено завршавање сваком од понуђених опција. Студије су већ потврдиле да се когнитивно оптерећење повећава ако је основа конструисана са почетном или унутрашњом празнином (табела 6) (Brame, 2013).

Изградња ефикасних могућности. У процесу креирања могућности требало би задовољити следеће препоруке (Frey et al., 2003):

1. Све могућности би требало да буду вероватне. Невероватне опције не представљају функционалне дистракторе и не треба да се користе. Уобичајене грешке ученика представљају најбољи извор дистрактора (табела 7).
2. Могућности треба да буду наведене јасно и концизно и треба да буду међусобно искључиве. Ученици сматрају да ставке које садрже могућности чији се садржаји преклапају могу да умање веродостојност процене.
3. Могућности не би требало да пружају индиције на основу којих ће бити искључене из скупа потенцијално тачних одговора. У супротном, пажљивији ученици могу лако да открију тачан одговор. Стога је важно да понуђене опције буду сличне дужине, користе исти стил израза и граматику доследну основи.
4. Могућности „све горе наведено“ и „ништа од горенаведеног“ требало би избегавати приликом дизајнирања ставки вишеструког избора.

Ако се опција „све горе наведено“ користи као тачан одговор, ученик или ученица који могу да идентификују више од једне могућности као тачне, могу да изаберу тачан одговор чак иако нису сигурни у вези са другим понуђеним опцијама. Исти аргумент важи и када се опција „ништа од горе наведеног“ користи као тачан одговор. У оба случаја је могуће применити делимично знање како би тачно одговорили на ставку.

5. Општа претпоставка у процесу дизајнирања питања вишеструког избора је да је редослед могућности у потпуности небитан, све док су одговори насумично или међусобно једнако распоређени (Attali and Bar-Hillel, 2003). У том смислу, (Hohensinn and Baghaei, 2017) су испитали да ли тежина ставке зависи само од њене основе или на њу утиче позиција тачног одговора. Анализа је потврдила да позиција тачног одговора има веома мали утицај на тежину ставке вишеструког избора и да је уобичајена пракса насумичне расподеле тачних одговора валидна. (Haladyna et al., 2002) су представили таксономију смерница за креирање питања вишеструког избора. Део који се односи на позицију могућности наглашава да би оне требало да буду дате у неком логичном следу (попут алфабетског или нумеричког) како би се избегла наклоност ка одређеним позицијама.
6. Број могућности се може разликовати међу питањима вишеструког избора, све док су све понуђене опције веродостојне. Не постоје чврсти докази који потврђују значајне разлике у тежини ставке и поузданости резултата испитивања између питања која садрже два, три или четири дистрактора (Haladyna, 2004).

2.2 Стандардни протоколи за накнадну валидацију питања вишеструког избора

Накнадна валидација помаже да се идентификују питања вишеструког избора чија је валидност упитна, тако да се могу адекватно модификовати пре поновног коришћења или одбацивања. Анализа ставки је поступак накнадне валидације који карактерише свако питање вишеструког избора додељивањем нумеричких вредности, као што су: индекс тежине, индекс дискриминативности и анализа дистрактора. На основу стандардних прихватљивих ограничења за ове нумеричке вредности, питања вишеструког избора могу бити прихваћена, модификована и поново потврђена, или одбачена.

Табела 5. Пример основе са негацијом и њене побољшане формулације

Основа са негацијом	Побољшана формулација
Шта од наведеног НЕ чини један од стубова ислама?	Пет стубова ислама, као муслиманске верске и етичке дужности, укључују све наведено осим:
А. шехадет	А. шехадет
Б. намаз	Б. намаз
В. зекат	В. зекат
Г. ХАЛАХА	Г. ХАЛАХА

Табела 6. Пример основе са унутрашњом празнином и њена побољшана формулација

Основа са унутрашњом празнином	Побољшана формулација
Заједно са Холбахом, _____ је један од најзначајнијих француских енциклопедиста.	Ко је уз Холбаха један од најзначајнијих француских енциклопедиста?
А. ХЕЛВЕЦИЈУС	А. ХЕЛВЕЦИЈУС
Б. Ками	Б. Ками
В. Прудон	В. Прудон
Г. Дерида	Г. Дерида

Индекс тежине. Индекс тежине је један од најчешће коришћених статистичких показатеља за анализу ставки. То је мера пропорција оних ученика који су на ставку одговорили тачно, и стога се често назива *p*-вредношћу. Виша *p*-вредност указује на то да је већи проценат ученика правилно одговорио на ставку, те се стога ставка сматра једном од лакших. Индекс тежине се добија тако што се број ученика који су на ставку одговорили тачно подели са укупним бројем ученика који су решавали задатак, тако да се креће између 0,0 и 1,0 (Crocker and Algina, 1986). Табела 8 представља три различите категорије ставки у зависности од опсега коме вредност индекса тежине припада.

Табела 7. Пример невероватне могућности за дату основу

Невероватне могућности (Б и Г) Који филозоф је најутицајнији представник модерног интуиционизма?
А. Анри БЕРГСОН
Б. Алберт Ајнштајн
В. Норберт Винер
Г. Џон Кенеди

Табела 8. Категорије ставки у зависности од вредности индекса тежине (Wiersma and Jurs, 1990)

Опсег индекса тежине	Категорија ставке
$p \leq 0,30$	тешка
$0,30 < p \leq 0,70$	прихватљива
$p \geq 0,70$	лака

Индекс дискриминативности. Индекс дискриминативности ставке показује у којој мери ставка поседује својство да разликује ученике који су постигли исходе учења од оних који то нису. При израчунавању ове мере, анализирана је група ученика са најбољим постигнућем (горња група), заједно са групом ученика који су слабо урадили цео тест (доња група). Како би се осигурала стабилност, пожељно је да групе укључују већи број ученика. Такође је пожељно да ове групе буду разноврсније, како би способност разликовања била јаснија. Према (Wiersma and Jurs, 1990), коришћење 27% од укупног броја ученика у свакој групи, ове две функције води ка максимуму.

Индекс дискриминативности је дефинисан као:

$$D = \frac{U - L}{N}$$

где су U и L број ученика у горњој и доњој групи који су одговорили тачно на ставку, а N је број ученика у највећој од ове две групе. Wood (1960) наводи да када више ученика из доње него из горње групе одабере тачан одговор на ставку, тада ставка има негативну вредност. Негативна вредност указује на то да је ставка не само

бескорисна, већ и да смањује валидност теста. Табела 9 приказује четири различите категорије ставки у зависности од опсега коме вредност индекса дискриминативности припада.

Табела 9. Категорије ставке у зависности од вредности индекса дискриминативности (Ebel and Frisbie, 1986)

Опсег индекса дискриминативности	Категорија ставке
$D \geq 0,40$	веома добра
$0,30 \leq D < 0,40$	разумно добра (евентуално подложна побољшању)
$0,20 \leq D < 0,30$	маргинална (неопходна ревизија)
$D < 0,20$	слаба/лоша (неопходна већа ревизија или искључивање)

Анализа дистрактора. Две горе дефинисане мере не узимају у обзир карактеристике дистрактора ставки и начин на који они утичу на одлуку ученика да изабере једну од понуђених могућности. Анализа дистрактора се бави овим проблемима тако што испитује квалитет дистрактора као једног важног елемента квалитета ставке. Сваки дистрактор мора да буде вероватан и очигледно нетачан.

Један једноставан приступ анализе дистрактора укључује одређивање удела ученика који су изабрали сваку од понуђених опција. Ове пропорције могу бити посебно информативне. На пример, када је проценат ученика који су изабрали одређени дистрактор већа од процента ученика који су изабрали тачан одговор, онда би требало испитати ставку како би се утврдило да ли је тачан одговор погрешан. Анализа дистрактора такође може да открије невероватне дистракторе. На пример, ако ученици доследно не бирају одређени дистрактор, то може бити доказ за његову неподобност. Дистрактори које не бира 5% или више ученика сматрају се неефикасним и требало би их ревидирати или елиминисати (Linn and Gronlund, 2000).

2.3 Релевантна истраживања

Спроведено истраживање је потврдило да постоји значајан простор за унапређење квалитета многих тестова заснованих на питањима вишеструког избора. На основу анализе узорка од 60 питања вишеструког избора из области медицине, [Ramakrishnan et al. \(2017\)](#) су закључили да више од једне трећине свих дистрактора није било функционално тј. да нису били прихватљиви. Ови дистрактори би требало да буду модификовани или замењени и тестирани поново, све док се не испуне дефинисани критеријуми (постизање ефикасности дистрактора једнаке или веће од 5%). [Halikar et al. \(2016\)](#) су детаљно анализирали 20 питања вишеструког избора из исте области (медицина) и приметили да су сва питања имала барем по један нефункционалан дистрактор, док је 23% укупног броја дистрактора било нефункционално. Резултати су такође показали да је проценат прихватљивих питања, базиран на индексу тежине и индексу дискриминативности, био 35%, односно 50%. Због тога су аутори препоручили израду фонда валидних питања вишеструког избора, где ће свако питање бити удружено са својим индексним вредностима. Стога, испитивачи могу изабрати одговарајућа питања вишеструког избора из тог фонда за потребе одговарајућег тестирања. У свом истраживању су Дибатиста и сар. [Battista and Kurzawa \(2011\)](#) истакли да су испитивачима потребни курсеви обуке и подршке, како би били сигурни да су њихови тестови са ПВИ добро дизајнирани и да имају прихватљиву дискриминативну моћ. Процес стварања високо квалитетних тестова вишеструког избора је вештина која се може научити ([Jozefowicz et al., 2002](#)).

3. Разматрања везана за писање питања вишеструког избора која тестирају мишљење вишег реда

Упркос чињеници да је дизајнирање таксономија когнитивних вештина испрва спроведено како би се превазишле специфичности различитих домена, стручњаци се слажу око тога да су когнитивни процеси вишег реда природно специфични за одређени домен. [Anderson et al. \(2001\)](#) су потврдили да би свака велика област требало да има своју сопствену таксономију. Стручњаци су суочени са изазовом коришћења нивоа опште таксономије за потребе својих

специфичних области (Morrison and Free, 2001). Због тога је број основних правила за израду питања вишеструког избора за различите когнитивне нивое делимично ограничен. Ипак, идентификоване су неке стратегије које могу помоћи приликом дизајнирања питања вишеструког избора, које превазилазе пуко присећање. Следећи пасуси нуде неке препоруке које би могле олакшати овај процес.

Примена одређених глагола. Morrison and Free (2001) удружују одређене глаголе са различитим когнитивним процесима (табела 3.). Када је одређени глагол постављен у ставку, може да послужи као показатељ да је одговарајући когнитивни ниво оцењен. Међутим, ову стратегију би требало користити обазриво јер се неки глаголи могу поставити на више нивоа, а много тога зависи и од контекста ставке у који је глагол смештен. Ипак, овакво мапирање стручњацима који развијају ставке пружа објективну и транспарентну основу.

Табела 10. Примери глагола придружених различитим категоријама Блумове таксономије (Morrison and Free, 2001)

Знање	Разумевање	Примена	Анализа	Синтеза	Евалуација
идентификовати	описати	применити	анализирати	саставити	проценити
дефинисати	разликовати	проценити	категоризовати	изградити	оценити
знати	расправљати	класификовати	поредити	створити	евалуирати
набројати	објаснити	развијати	упоредити	осмислити	просудити
именовати	преформулисати	испитати	разликовати	формулисати	
препознати	поново исказати	решити	одредити	модификовати	
исказати	другачије формулисати	користити	истражити	планирати	

Употреба реалних сценарија. Један од најбољих начина за унапређење и процену мишљења вишег реда јесте коришћење питања која се заснивају на реалним ситуацијама, посебно оним које симулирају стварна радна искуства (Scully, 2017).

Анализа визуелних садржаја. Вештине критичког мишљења ученика се могу оценити на основу анализе или интерпретације информација визуелних садржаја, који су дати као саставни део основе питања или датих могућности. У многим случајевима ови визуелни садржаји, попут дијаграма и графикона, симулирају праве задатке различитих радних окружења (Burton et al., 1991).

Заклетв за разраду одговора. Мишљење вишег реда се такође може проценити ако се од ученика затражи да синтетишу оно што су научили. То значи да би одговори требало да садрже објашњења која их подржавају (Lord et al., 2009).

4. Закључни коментари

Оцењивање исхода учења је пресудна за унапређење образовања. Процес оцењивања ученика треба ускладити са наставним и образовним циљевима. Идентификовање одговарајућих стратегија за оцењивање напредовања ученика у оквиру појединачних програма подједнако је важно као и утврђивање наставног садржаја и метода њиховог презентовања. Питања вишеструког избора, као једна од најчешће коришћених стратегија за оцењивање, имају одређене предности и мане. Она су ефикасна, флексибилна, објективна, једноставна за спровођење и могу се користити за оцењивање садржаја наставног плана и програма великог обима. Међутим, развој доброг теста вишеструког избора за евалуацију постигнућа ученика је веома изазован циљ. Чак и када испитивачи прате низ смерница за израду фер и систематских тестова, различити фактори могу утицати на то како ће ученици доживети ставке теста. Како би се повећао квалитет питања вишеструког избора, веома је важно анализирати индексе тежине и дискриминативности ставки, што би онима који развијају тестове могло да помогне да оцењивање учине далеко значајнијим.

Честе су тврдње да су ставке вишеструког избора погодне само за оцењивање вештина мишљења нижег реда. Ипак, много прецизнија тврдња би била да се ставке вишеструког избора које мере сложене когнитивне процесе једноставно ређе изграђују. Поштујући одређене стратегије, могуће је конструисати ставке вишеструког избора које мере процесе као што су примена знања, анализа и процена.

Литература

- Anderson, Lorin W., David R. Krathwohl and Benjamin Samuel Bloom. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman, 2001.
- Attali, Yiagal and Maya Bar-Hillel. "Guess where: The position of correct answers in multiple-choice test items as a psychometric variable". *Journal of Educational Measurement* Vol. 40, no. 2 (2003): 109–128.

- Bacon, Don R. “Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context”. *Journal of Marketing Education* Vol. 25, no. 1 (2003).
- Battista, David Di and Laura Kurzawa. “Examination of the quality of multiple-choice items on classroom tests”. *The Canadian Journal for the Scholarship of Teaching and Learning* Vol. 2, no. 2 (2011). http://ir.lib.uwo.ca/cjsotl_rcacea/vol2/iss2/4
- Bloom, Benjamin S. *Taxonomy of Educational Objectives*. New York: David McKay Company Inc., 1977.
- Brame, Cynthia J. *Writing good multiple choice test questions*. Vanderbilt University Center for Teaching, 2013. <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>
- Brown, Sally and Peter Knight. *Assessing Learners in Higher Education*. London : Routledge Falmer, 1998.
- Burton, Steven J., Richard R. Sudweeks, Paul F. Merrill and Bud Wood. *How to prepare better multiple-choice test items: Guidelines for University Faculty*. Brigham Young University Testing Services and The Department of Instructional Science, 1991.
- C., Chan. *Assessment: Multiple Choice Questions. Assessment Resources*. HKU, University of Hong Kong, 2009. http://ar.cetl.hku.hk/am_mcq.htm
- Crocker, Linda and James Algina. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston, 1986.
- Dikli, Semire. “Assessment at a distance: Traditional vs. alternative assessments”. *The Turkish Online Journal of Educational Technology* Vol. 2 (2003).
- Ebel, Robert L. and David A. Frisbie, *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- Frey, Bruce B., Stephanie Petersen, Lisa M. Edwards, Jennifer Teramoto Pedrotti and Vicki Peyton. “Toward a consensus list of item-writing rules”, 2003.
- Haladyna, Thomas M. *Developing and validating multiple-choice test items (3rd ed.)*. Lawrence Erlbaum Associates, 2004.
- Haladyna, Thomas M. and Steven M. Downing. “Validity of a taxonomy of multiple-choice item-writing rules”. *Applied Measurement in Education* Vol. 2, no. 1 (1989): 51–78.
- Haladyna, Thomas M., Steven M. Downing and Michael C. Rodriguez. “A review of multiple-choice item-writing guidelines for classroom assessment”. *Applied Measurement in Education* Vol. 15, no. 3 (2002): 309–334.

- Halikar, Swapnagandha S., Veerendra Godbole and Saurabh Chaudhari. "Item Analysis to Assess Quality of MCQs". *Medical Science* Vol. 6, no. 3 (2016): 123–125.
- Hohensinn, Christine and Purya Baghaei. "Does the position of response options in multiple-choice tests matter?". *Psicologica* Vol. 38 (2017): 93–109.
- Jovanovska, Jasmina. "Multiple-choice question answering system for Macedonian and English test-collections". Докторска дисертација, Ss. Cyril and Methodius University, Faculty of Computer Science and Engineering, Skopje, Republic of Macedonia, 2018.
- Jozefowicz, Ralph F., Bruce M. Koeppen, Susan Case, Robert Galbraith, David Swanson et al. "The quality of in-house medical examinations", *Academic Medicine* Vol. 77, no. 2 (2002): 156–161.
- Linn, Robert R. and Norman E. Gronlund. *Measurement and assessment in teaching (8th ed.)*. Prentice Hall, Upper Saddle River, NJ, 2000.
- Lord, Thomas R., Donald P. French and Linda W. Crow. *Guide to Assessment*. National Science Teachers Association, 2009. <http://static.nsta.org/files/PB231Xweb.pdf>
- Michael Rodriguez, K. "The art and science of item-writing: A meta-analysis of multiple-choice item format effects", 1997.
- Morrison, Suzan and Kathleen Walsh Free. "Writing multiple-choice items that promote and measure critical thinking". *Journal of Nursing Education* Vol. 40, no. 1 (2001): 17–24.
- Palmer, Edward J. and Peter G. Devitt. "Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple-choice questions?" *BMC Medical Education* Vol. 7 (2007): 49–55.
- Ramakrishnan, Minu, Aditya B. Sathe and Vinayak A. "Item analysis: A tool to increase MCQ validity". *Indian Journal of Basic and Applied Medical Research* Vol. 6, no. 3 (2017): 67–71. <http://ijbamr.com/pdf/June201767-71.pdf.pdf>
- Scully, Darina. "Constructing multiple-choice items to measure higher-order thinking". *Practical Assessment, Research and Evaluation* Vol. 22, no. 4 (2017). <http://pareonline.net/getvn.asp?v=22&n=4>
- Stassen, Marta L., Kathryn Doherty and Mya Poe. *Course-based review and assessment: Methods for understanding student learning*. Office of Academic Planning & Assessment, University of Massachusetts Amherst, 2001.
- Statman, Stella. "Ask a clear question and get a clear answer: An enquiry into the question/answer and the sentence completion formats of multiple-choice items". *System* Vol. 16, no. 3 (1988): 367–376.

Towns, Marcy H. “Guide To Developing High-Quality, Reliable, and Valid Multiple-Choice Assessments”. *Journal of Chemical Education* Vol. 91 (2014): 1426–1431.

Walsh, Catherin and Lisa Seldomridge. “Critical thinking: Back to square two”. *Journal of Nursing Education* Vol. 45, no. 6 (2006): 212–219.

Wiersma, William and Stephen G. Jurs. *Educational measurement and testing* (2nd ed.). Boston, MA: Allyn and Bacon, 1990.

Wood, Dorothy Adkins. *Test construction: Development and interpretation of achievement tests*. Columbus, OH: Charles E. Merrill Books, Inc, 1960.