

**Abstrakt:** Je možné spravodlivé hodnotenie študentov len pomocou zatvorených testov multiple choice? Koľko musí mať test otázok a každá otázka volieb aby bolo možné odhaliť náhodné uhádnutie? Okrem toho, že test musí byť dobre zostavený (s prítlačivými distraktormi) máme možnosť zaviesť v Moodle **direktívne** alebo **kooperatívne** metódy vyhodnocovania (bodovania). Direktívne metódy môžu napríklad spočívať v komplikovaných otázkach s viacerými správnymi voľbami (MC-MR) a negatívnym skórovaním. Za kooperatívnu metódu považujeme napríklad Confidence-Based Marking. Neexistuje optimálny univerzálny systém hodnotenia vhodný pre všetky okolnosti testovania. Ako najdôležitejšie odporúčame vybrať metódu podľa účelu hodnotenia, tak aby boli využité direktívne alebo kooperatívne možnosti odradenia od hádania správnych odpovedí. Určite sa prihovárame za kombinovanie rôznych typov otázok a aj využitie otázok typu MC-MR s negatívnym skórovaním.

**Kľúčová slova:** Vyhodnocovanie testov, kvízov

**Abstract:** Is there a fair way to evaluate students by multiple choice (MC) quizzes? How many questions and options have to be included. Sure that the quiz must be well prepared (with attractive distractor) but there is an opportunity in Moodle to use prescriptive or cooperative assessment methods of scoring. Prescriptive methods may for example be to use many hard question with multiple responses and negative scoring systems. An example of the cooperative approach is Confidence-Based Marking. There is no optimal universal scoring system suitable for all circumstances of testing. We recommend to choose the method according to the purpose of evaluation, to use prescriptive or cooperative method to discourage guessing.

**Keywords:** MOODLE, quiz, multiple choice, random guess, formula scoring, negative scoring, Confidence-Based Marking

## 1 Úvod

Didaktické testovanie je jednou z podstatných súčastí vzdelávania pomocou e-learningu. Využitie počítačových systémov na tvorbu a vyhodnotenie najmä objektívne skórovateľných úloh ponúka širokú paletu testov. Cieľom tohto príspevku je zhrnúť odporúčania pre zostavenie a ohodnotenie testov zostavených predovšetkým z otázok s jednou a viacerými správnymi voľbami, ktoré sú pri testovaní v e-learningu s Moodle najpoužívanejšie.

## Formy didaktických úloh a testov

Profesor Ivan Turek v učebnici Základy pedagogického výskumu (Turek 2002<sup>[1]</sup>) rozdeľuje formy didaktických úloh na otvorené úlohy (so širokou odpoveďou: neštrukturalizované/štrukturalizované, so stručnou odpoveďou: produkčné, doplňovacie) a zatvorené úlohy (dichotomické, s výberom odpovede (polytomické), priradovacie, usporiadacie). Vyhodnotenie otvorených úloh so širokou odpoveďou, pri ktorých probant (žiak / študent / skúšaný a pod.) tvorí (píše, kreslí, počíta,..) je vo veľkej miere závislé na hodnotiteľovi. Najčastejšie je ním priamo examinátor / skúšajúci. Vo zvyšných typoch úloh probant pri otvorených úlohách píše a pri zatvorených úlohách vyberá správnu odpoveď z niekoľkých ponúkaných možností/odpovedí/volieb. Preto sa posledné menované úlohy nazývajú **objektívne skórovateľné**. Štulajterová (2002)<sup>[2]</sup> upozorňuje na to, že slovo test je prevzaté z angličtiny: „V slovenčine chápajú mnohí slová test a testovať zúžene, najčastejšie ako skúšku písomnú, ktorú možno objektívne ohodnotiť. V tomto zmysle je zaužívané spojenie didaktický test, ktorý je obyčajne objektívnou písomnou skúškou vedomostí“. Rovnako, teda na označenie objektívne skórovateľných úloh, používame slovo test aj my v ďalšom texte, pričom máme na mysli didaktický test.

### 1.1 Výhody a nevýhody otvorených a zatvorených úloh a použitie otázok s výberom odpovede

Burjan<sup>[7]</sup> upozorňuje na niektoré nevýhody testov s uzavretými úlohami v školskej praxi: „znevýhodňujú nepozorných žiakov aj žiakov hlbavých, ktorí pomaly rozmýšľajú. (Naopak sú zasa omnoho vhodnejšie pre komunikatívne slabších žiakov a žiakov, ktorí pomaly písomne formulujú svoje myšlienky alebo pomaly píšú) a ľahšie sa pri nich opisuje.“

Všetci autori sa zhodujú na základnej výhode zatvorených úloh a tou je jednoduchosť spracovania a vyhodnotenia a teda v konečnom dôsledku vyššia efektivita, najmä pri potrebe testovania veľkého počtu probantov a samozrejme pri zachovaní základných zásad správneho zostavenia a prezentovania testov.

Zatvorené úlohy podľa DePalma<sup>[3]</sup>(1990) nielenže vyhovujú viac niektorým typom probantom ale vo všeobecnosti, tým, že odstraňujú subjektívnu úlohu hodnotiteľa napríklad hodnotením rukopisu alebo zrozumiteľnosti prezentovanej odpovedi úplne odstraňujú možnosť jeho osobnej zaujatosti. Sú preto ešte objektívnejšie ako úlohy so stručnou odpoveďou. Práve preto sa venujeme týmto druhom úloh a to predovšetkým testom s dichotomickými otázkami a s výberom odpovede. Ďalšie uzavreté úlohy: priradovacie a usporiadacie sú menej často využívané a ich skórovanie je obdobné, preto sa im venujeme len okrajovo.

Bontis et al. (2009)<sup>[4]</sup> zhrňuje základné výhrady voči úlohám s výberom odpovede do troch základných oblastí:

1. sú ovplyvnené možnosťou uhádnutia správnej odpovede,
2. sú diskriminujúce voči určitým sociálno-ekonomickým a kultúrnym skupinám,
3. merajú deklaratívnu úroveň poznania a nie hĺbku učenia sa.

Sám Bontis et al. (2009)<sup>[4]</sup>, ktorý úlohy s výberom odpovede považuje (v manažérskom vzdelávaní) „za najsilnejšie prediktory celkovej výkonnosti študenta ak sú doplnené aj inými formami hodnotenia ako sú aktívna účasť na výučbe, prípadové štúdie, písomné zadania a simulačné hry“ odpovedá na jednotlivé výhrady. My sa venujeme predovšetkým prvej výhrade a uvádzame viacero možností ako v testoch odradiť probantov od hádania. K druhej výhrade Bontis uvádza zásadu dopriať dostatok času tým, ktorí môžu mať problémy s jazykom. V tejto súvislosti je často uvádzané<sup>[1]</sup>, ako príklad takto diskriminujúcej otázky, nájsť analógiu

<sup>1</sup> [http://en.wikipedia.org/wiki/SAT#Cultural\\_bias](http://en.wikipedia.org/wiki/SAT#Cultural_bias)

k výrazom „bežec a maratón“ v teste SAT. Správnu voľbou bolo „vesliar a regata“, ktorú vedelo označiť významne viac bielych ako čiernych študentov. Kritici upozorňovali na to, že študenti z chudobného prostredia nemajú možnosť zoznámiť sa s termínmi súvisiacimi so snobskými športami ako je jachting. Vo všeobecnosti táto druhá výhrada upozorňuje na možnú nejednoznačnosť pochopenia otázky. Ak nie je probant schopný interpretovať čo otázkou skúšajúci myslel, môže byť jeho odpoveď vyhodnotená ako nesprávna aj keď je potenciálne správna. V tejto súvislosti sa používa termín "multiple guess"<sup>[2]</sup> ("viacnásobný odhad"), pretože probant sa musia pokúsiť odhadnúť, nie určiť správnu odpoveď. Široká odpoveď by v tomto prípade umožnila probantovi uviesť argument na podporu svojej voľby a tým by prípadne získal lepšie hodnotenie.

Tretou a najväčšou nevýhodou úloh s výberom odpovede je obmedzenie typu znalostí, ktoré možno posudzovať. Úlohy s výberom odpovede sú najvhodnejšie pre testovanie dobre definovaných a menej náročných zručností. Riešenie zložitých problémov, logické myslenie, sú lepšie testovateľné úlohami so stručnou alebo širokou odpoveďou. Nezostáva iné ako mať túto nevýhodu na pamäti alebo v teste kombinovať viaceré druhy úloh.

### **1.2 Terminológia pre hodnotenie otázok s výberom odpovede**

Turek (2002) neodporúča, aby pri zatvorených úlohách s výberom odpovede bolo možné vyberať medzi viacerými správnymi možnosťami / voľbami. Môžeme sa len domnievať, že dôvodom odmietania týchto takzvaným „zaškrtávacích“ testov Turekom, môže byť aj ich problematické vyhodnocovanie. So vstupom počítačových systémov do testovania sa z písomných testov stali v súčasnej dobe najčastejšie formuláre na web stránkach, ktoré po odoslaní systém samočinne vyhodnotí. Napriek automatizácii výpočtu, samotný problém ako hodnotiť práve úlohy s viacerými správnymi odpoveďami nie je definitívne vyriešený. Výhodou takýchto úloh je omnoho vyššia **rozlišovacia / diskriminačná** schopnosť (ako ukážeme neskôr) a preto sa nimi zapodieваме podrobne. Mimochodom, s príchodom počítačov sa objavili aj nové typy testových úloh, ako sú: úlohy na dopĺňanie textu (anglicky: Multiple Gap-Filling alebo Cloze), identifikácia chýb v texte (Error Recognition - Štulrajterová 2002)<sup>[2]</sup> a výpočtové otázky s náhodne dynamicky generovanými číslami<sup>[3]</sup>. Aj napriek tomu delenie uvedené na začiatku ostáva v platnosti a tieto typy otázok by sa doň mohli zaradiť ako doplňovacie.



Vyhodnotenie otázok typu "Vyberte aspoň jednu (správnu) odpoveď" (anglicky: Multiple Choice Multiple Response – MC-MR) nie je také jednoduché, ako by sa na prvý pohľad mohlo zdať. Kohoe (1995a)<sup>[5]</sup> upozorňuje, že často to nie sú ani otázky ale nekompletné tvrdenia a preto je jednotlivé položky testu lepšie nazývať jednoducho **položkami a nie otázkami**. V čom spočíva problém spravodlivého hodnotenia? Napriek tomu, že tieto testovacie úlohy sú na rozdiel od otvorených úloh objektívne skórovateľné, problémom je vytvoriť didaktický test a nastaviť systém hodnotenia (**prideľovania bodov = skórovania**) tak, aby test dostatočne rozlišoval (diskriminoval) tých, ktorí vedia odpovede od tých, ktorí vedia uhádnuť / odhadnúť správnu odpoveď a pritom bol test primerane dlhý - teda krátky. V ďalšom texte rozoberáme možnosť prideľovania negatívneho skóre a spôsob jeho výpočtu. Výhodou otázok typu MC-MR oproti otázkam s jednou správnou voľbou (Vyberte jednu odpoveď - Multiple Choice Single Answer MC-SA) je práve ich **diskriminačná schopnosť** spôsobená variabilným počtom správnych volieb. Porovnáваме tu schopnosť oddeliť tých ktorí „len hádajú“ (v ďalšom texte sú pre zjednodušenie označení ako „hádači“) od tých, ktorí vedia (sú si istí) aspoň niekoľko málo odpovedí a to porovnaním MC-MR a MC-SA. Podmienkou, podľa Kehoe, 1995b Štulrajterová je

<sup>2</sup> Wikipedia, Multiple\_choice [http://en.wikipedia.org/wiki/Multiple\\_choice#Disadvantages](http://en.wikipedia.org/wiki/Multiple_choice#Disadvantages)

<sup>3</sup> Moodle2, Simple calculated question type  
[http://docs.moodle.org/20/en/Simple\\_calculated\\_question\\_type](http://docs.moodle.org/20/en/Simple_calculated_question_type)

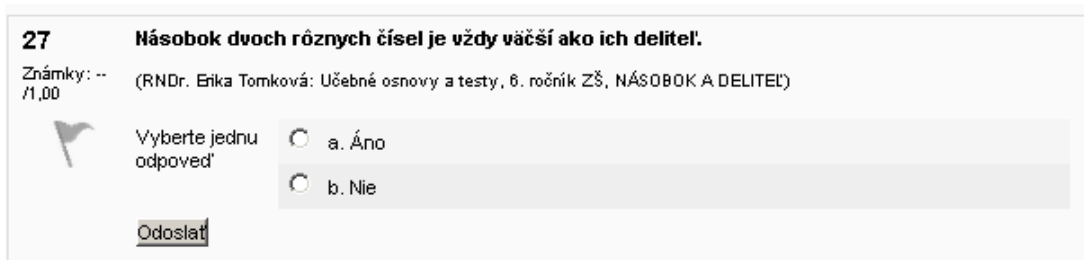
aby obsah testovania pokrýval len jednu špecifickú (konzistentnú) oblasť vedomostí, kde sa nepredpokladá, že študent, ktorý vie, vie napríklad len jednu polovicu a druhú musí hádať.

### **Otázky s výberom správnej možnosti (jednej správnej voľby – MC-SA)**

V týchto úlohách vyberá probant jednu možnosť z viacerých. Pretože v ďalšom texte budeme počítať pravdepodobnosť náhodného uhádnutia a pri týchto výpočtoch budeme často používať slovné spojenia „počet možností“ v úplne inom význame budeme sa snažiť rozlíšiť tieto dva významy tým, že **možnosti v otázkach** budeme nazývať aj **voľby**. Probant má teda práve jednu správnu voľbu. V počítačovej reprezentácii bývajú voľby v tomto prípade znázornené zvyčajne krúžkom , ktorý sa po výbere stane aktivovaným  (takzvané rádio tlačidlo = radio button).

### **Otázka s jednou správnou voľbou s výberom z dvoch možností (dichotomická úloha)**

Najjednoduchšou možnosťou MC-SA je výber odpovede len z dvoch volieb. Takáto otázka je dichotomická (s dvoma voľbami) alebo v anglickej literatúre nazývaná True-False, Yes- No alebo Binary Item.




**Obr. 1** Príklad dichotomickej otázky. Zdroj: RNDr. Erika Tomková: Učebné osnovy a testy, 6. ročník ZŠ. Reprezentácia pomocou modulu Quiz/Test v Moodle 2.0.

### **Otázka s jednou správnou odpoveďou s výberom z viac ako dvoch volieb (polytomické)**

**Základ otázky (stem = koreň; analogicky ku koreňu slova)** je úvodná otázka alebo neúplné tvrdenie na začiatku každej položky, po ktorom nasleduje viacero možností odpovedí (volieb). Medzi voľbami je práve jedna jednoznačne (Turek 2002) správna a viacero nesprávnych (tzv. distraktorov) ale rovnako alebo ešte viac hodnoverných / príťažlivých (Turek 2002), lákavých / pravdepodobných (autor).

**Príklad 1:** Základ tvorí úplná otázka. (Reprezentácia v Moodle 2. Slová: „Vyberte jednu odpoveď“ sú doplnené automaticky Moodle pri výbere otázok tohto typu).



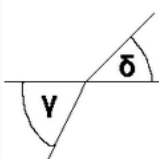
**Obr. 2** Príklad otázky s jednou správnou voľbou, ktorej základ tvorí úplná otázka

Základ môže obsahovať rozšírené alebo doplnkové informácie, ako sú obrázok, graf, tabuľka, prípadová štúdia alebo podrobný opis, ktorý má viac častí. Je tiež možné využiť jeden základ vo viacerých položkách (otázkach), napríklad sa na jeden obrázok odvoláva viacej otázok, v ktorých treba identifikovať zobrazené objekty.

### Príklad 3: Základ obsahuje obrázok.

21 Uhly vyznačené na obrázku nazývame

Známky: --  
/1,00



(RNDr. Erika Tomková: Učebné osnovy a testy, 5. ročník ZŠ, Uhly) [Nnt](#)

Vyberte jednu odpoveď

a. vrcholové

b. ani jedna z uvedených možností nie je správna

c. pravé

d. susedné

Obr. 3 Príklad otázky s jednou správnu voľbou, ktorej základ obsahuje obrázok

V tomto príklade je v základe otázky využité html formátovanie textu, čo umožňuje vloženie obrázku a ale aj odkazu na stránky autorky, ktorý je pod slovom „*hint*“.


### Otázka s viacerými správnymi voľbami (MC – MR alebo multiple-selection)

Na rozdiel od prechádzajúcich majú otázky typu MC – MR nie jednu ale viac možných správnych volieb. **Odpoveďou** budeme nazývať výber volieb. Odpoveď je správna ak sú vybrané všetky správne voľby. V prezentácii týchto otázok v Moodle je odpoveď správna ak súčet známok / bodov pri vybraných voľbách je 100%. Ak je viac správnych volieb je samozrejmé, že musí byť aj celkovo viac volieb ako dve. Zvyčajne býva volieb viac ako 4, najčastejšie 5 alebo 6. Často študentom nebýva známe koľko volieb je správnych – ba dokonca môže byť, že žiadna z ponúkaných volieb nie je správna. V počítačovej reprezentácii bývajú voľby v tomto prípade znázornené zvyčajne štvorčekom , ktorý sa po výbere zaškrtnie  (takzvané zaškrťavacie pole = chcek box).

### Príklad 4. Otázka typu MC – MR

8 Označ výroky, ktoré sú **nepravdivé** :

Známky: --/6,00 (RNDr. Erika Tomková: Učebné osnovy a testy, 6. ročník ZŠ, Celé čísla) [Nnt](#)



Vyberte aspoň jednu odpoveď.

a. Kladných čísel je viac ako záporných.

b. Súčet dvoch záporných čísel je číslo kladné.

c. Číslo, ktoré vyjadruje počet, nie je nikdy záporné.

d. Podiel kladného čísla a záporného je číslo záporné.

e. Podiel nuly a záporného čísla je číslo záporné.

f. Podiel čísla 1 000 000 a čísla -1 je číslo väčšie ako -1 111 111.

Obr. 4 Príklad otázky s viacerými správnymi voľbami

Z povahy otázok takéhoto typu vyplýva, že najmä pri neznámom počte správnych volieb je počet všetkých možností volieb (ich kombinácií) omnoho väčší ako pri otázkach s výberom len jednej správnej voľby, čo dokážeme matematicky v nasledujúcich častiach.

### 1.3 Skórovanie odpovedí (položiek) a účely testovania - rozlišujúce a overujúce testy

(Didaktické) testovanie je dobre teoreticky rozpracované (napríklad Štulrajterová, 2002 alebo Turek, 2002). Tu uvádzame len niektoré jeho aspekty, ktoré, nazdávame sa, priamo súvisia s využitím testovania (s použitím skórovania otázok s viacerými správnymi voľbami) v e-learningu a našim príspevkom do tejto oblasti.

Priradenie počtu bodov alebo výslednej známky k odpovediam závisí predovšetkým od účelu testu. V anglosaskej literatúre sa rozlišujú predovšetkým dva účely testov: formatívne („výchovné“) a sumatívne („celkové“) hodnotenia. Rozlíšenie medzi nimi vysvetľuje Burjan (online)<sup>[6]</sup> nasledovne: (**formatívne**) „Ide o priebežné hodnotenie, ktoré má plniť najmä diagnostickú funkciu a byť vodidlom pre prípadné korekcie výučby, alebo ide o záverečné, **sumatívne** hodnotenie, ktoré má byť bilanciou už dokončenej výučby“. Z pohľadu skórovania je pri formatívnom hodnotení lepšie ak aj najhorší žiak získa aspoň jeden bod a preto v ňom, zdá sa, nevzniká potreba **negatívneho hodnotenia** nesprávnych odpovedí. Pod pojmom negatívne hodnotenie rozumieme pridelenie zápornej hodnoty buď jednej voľbe alebo aj celej odpovedi na otázku.

Testy možno rozdeliť podľa porovnávania a interpretácie výkonov na (podľa Turek 2002):

- Rozlišujúce, tzv. NR testy (Norm - Referenced) alebo tiež testy relatívneho výkonu, v ktorých je výkon študenta porovnávaný s výkonom iných študentov, spravidla s priemerným výkonom preverovaných študentov.
- Overujúce, tzv. CR testy (Criterion - Referenced) alebo tiež testy absolútneho výkonu, v ktorých je výkon študenta porovnávaný s určitou dopredu stanovenou normou.

Aj tu je pravdepodobné, že normatívne testy sa zaoberajú bez negatívneho hodnotenia. Preto sa toľko zaujímate o negatívne hodnotenie? Pretože je jednou z možností ako odradiť od hádania odpovedí, ktorému venujeme nasledujúcu časť.

## 2 Pravdepodobnosť uhádnutia správnej odpovede (random guess)

Pravdepodobnosť uhádnutia správnej odpovede za predpokladu, že študent háda úplne náhodne (vôbec nepozera na voľby a úplne náhodne ich vyberá) je riadená v takomto prípade binomickým rozdelením. Diskrétna náhodná veličina  $K$  (uhádne práve  $k$  otázok / položiek z  $n$  = počet otázok) s binomickým rozdelením môže nadobúdať celočíselné hodnoty od nuly po  $n$ . Pravdepodobnosť, že jav nastane práve  $k$  - krát z  $n$  pokusov pri pravdepodobnosti  $p$  (v tomto prípade  $p = 1 / \text{počet volieb v otázke}$ ), je určená rozdelením hustoty pravdepodobnosti binomického rozdelenia:

$$f(k;n,p) = P[K = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

pre  $k = 0, 1, \dots, n$

kde  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  je binomický koeficient

### 2.1 Dichotomické otázky

Pravdepodobnosť uhádnutia správnej odpovede pri otázke s práve jednou správnou voľbou s výberom z dvoch volieb, bez jej znalosti je  $p=1/2$ . V takto zostavenom teste dosiahne približne polovicu bodov aj žiak, ktorý otázky vôbec nečíta. Presnejšie povedané: V priemere, ten, ktorý háda dosiahne polovicu správnych odpovedí, a to nezávisle od toho ako dlho učiteľ test konštruoval a koľko je v ňom otázok typu Áno a Nie. Ale len v priemere. Ak má test napríklad len 10 otázok, potom približne jeden z 1000 hádačov uhádne správne dokonca všetky odpovede ( $f(10, 10, 1/2) = 1 \cdot 1/2^{10} \cdot 1 = 1/1024 \sim 0,001$ ). Tieto úvahy sú za predpokladu, že

každá otázka má práve jednu správnu odpoveď a všetci respondenti / žiaci odpovedajú na každú otázku (správne či nesprávne).

Z uvedeného vyplýva:

- testy len s otázkami s dvoma voľbami sú slabo diskriminačné, teda na odlišenie hádačov treba z binomického rozdelenia zistiť aký je minimálny počet otázok, pri ktorých pri danom počte účastníkov testu ani jeden hádač sa nemôže ani len priblížiť k počtu bodov, ktoré získa študent, ktorý odpovedá na základe vedomostí. Napríklad pre dosiahnutie úspešnosti 80% až 90%, ktoré Burjan (2002) uvádza ako hranicu úspešného zvládnutia testu je pre úroveň náhodného dosiahnutia tejto hranice s pravdepodobnosťou 1% (jeden zo sto hádačov) alebo 0,1% (jeden z tisíc hádačov) potrebných viac ako **7** (pre 90%, 1%), **15** (90%, 0,1%), **17** (80%, 1%) a **20** (80%, 0,1%) otázok.
- Zostavenie takéhoto testu je jednoduché, pretože netreba vymýšľať ostatné (nesprávne) voľby. Tieto voľby sa nazývajú **distraktory**.

## **2.2 Otázky s jednou správnou voľbou z viac ako dvoch**

Pridaním tretej voľby k dichotomickej otázke sa teoretická pravdepodobnosť uhádnutia odpovede zníži na približne 33% a tým sa zníži aj počet otázok potrebných pre rozlíšenie dosiahnutia úspešnosti **80%**. Pre úroveň významnosti náhodného dosiahnutia tejto hranice na úrovni 1% a 0,1% poklesne počet otázok na **7** a **11** a pri otázkach so 4 voľbami už len na **4** a **8** otázok. To je základný dôvod pre používanie otázok s viacerými voľbami. Problémom je potreba distraktorov.

Ak by učiteľ nezobral do úvahy predchádzajúce úvahy a testoval by žiakov testom s 10 otázkami s výberom jednej správnej voľby a známky by im prideloval na základe rovnomerného rozdelenia výsledkov v intervaloch 0%, 20%, 40%, 60%, 80% hádači by testom prešli s priemernými známками od 3 do 4,3 v závislosti od počtu voľieb pri každej otázke. V teste s len 10 otázkami s dvoma alebo troma voľbami viac ako 1% probantov dosiahne úspešnosť 80% hádaním. Zdvojnásobenie počtu otázok na 20 má len malý vplyv na výsledky hádačov. Z podstaty procesu vyplýva, že priemerný počet úspešne uhádnutých odpovedí je stredná hodnota binomického rozdelenia  $E[K] = n \cdot p$ . Priemerný počet bodov v % je u hádačov stále  $1/\text{počet možností}$ , teda  $p$ , teda rovnaký. V prípade len dvoch možností odpovede je to 50%, troch možností 33% atď. Hádači tak majú rovnaký relatívny počet bodov nezávisle od počtu otázok, teda aj rovnakú priemernú známku. S rastom počtu otázok len klesá počet hádačov, ktorí získajú extrémne dobré ale aj zlé známky. **Z uvedených príkladov vyplýva, že v takýchto testov je nevhodné stanovovať rovnomerne hranice pre známky.**

Skúšajúci teda môže zvyšovať hranicu pre úspešné absolvovanie alebo zvyšovať počet otázok a ich zložitosť. Ak by chcel pôsobiť veľmi motivujúco a hranicu posunul dole na 50% (vyhodil by len tých, ktorí majú menej ako polovicu odpovedí správne) potreboval by pre otázky s 3 voľbami vyrobiť test aspoň so štyridsiatimi otázkami a s 28 otázkami pre  $p=1/4$  aby vylúčil 99,9% hádačov. Uvedené výpočty upozorňujú na veľmi reálnu možnosť vysokého skórovania tých, ktorí hádajú správne odpovede v testoch typu MC-SA, kde je možnosť výberu jednej odpovede z viacerých voľieb. Testujúci má viacero alternatív ako vyplynulo z teoretického rozboru. Môže zvyšovať hranicu na úspešné absolvovanie testu na 80% až 90% alebo zvyšovať počet otázok alebo počet možných voľieb. Ak predpokladáme, že na prečítanie a zodpovedanie jednej otázky je potrebná jedna minúta, potom so spomínaným počtom otázok trvá celý test pol až celú vyučovaciu hodinu (45 minút). Skúšajúci môže použiť inú formu testovania alebo to, čo je predmetom tejto časti: namiesto dichotomických otázok použiť otázky s viacerými voľbami, ktoré významne sťažujú možnosť uhádnutia správnej odpovede. Všetky výpočty

v predchádzajúcej časti boli pre otázky s práve jednou správnou voľbou a tiež predpokladali, že študent háda odpoveď úplne náhodne. V praxi to však tak nie je. Nielen že študenti „odhadujú“ správnú odpoveď až po prečítaní všetkých možností, čím zvyšujú pravdepodobnosť rozpoznania správnej voľby od distraktorov ale často sú testy zostavené tak že distraktory sú nedôveryhodné. Práve preto neodporúčame zvlášť pridávať počet volieb (na viac ako 5 alebo 6) ale skôr použiť viac správnych volieb, teda otázky typu MC-MR.

### 3 Možnosti odradenia od hádania odpovedí

Cieľom testovania je čo možno najobjektívnejšie ohodnotiť výkon testovaného. Z povahy testov vyplýva, že nás nezaujímajú myšlienkový postup, ale len výsledok a preto nemôžeme hádanie zakázať. (Samozrejme že predpokladáme, že sme zabránili podvádzaniu a pod.).

K odradeniu od hádania môžeme použiť tieto prístupy:

- **Direktívny** – zvyšujeme obtiažnosť testu (vrátane jeho úprav na základe predchádzajúcich skúseností) aby sme odstránili možnosť hádať.
- **Kooperatívny** – zavedieme pravidlá, ktoré budú umožňovať získať viac bodov v prípade ak si je probant istý odpoveďou (neháda). Toto je založené predovšetkým na korekcii skórovania.

**Korekcia skórovania** môže byť pozitívna alebo negatívna v tom zmysle, že pozitívna korekcia zvyšuje počet bodov, tých ktorí nehádajú a negatívna korekcia znižuje počet bodov tých, ktorí hádajú. V oboch prípadoch však treba umožniť **negatívne skórovanie**, teda aby k určitej voľbe mohlo byť priradené aj záporné skóre. Toto vysvetlenie je pokusom o novú definíciu týchto pojmov. Snažíme sa odlíšiť celkovú korekciu hodnotenia a skórovanie jednotlivých položiek (odpovedí). Aj v prípade negatívneho skórovania totiž výsledná korekcia môže byť pozitívna ak je vyvážená zvýšením pozitívneho skórovania správne zodpovedaných otázok (pozri alternatívu 4 v nasledovnom zozname).

Korekcia skórovania je dobrý nápad, avšak nie tak ľahko realizovateľný. Ťažké je rozhodnutie o tom, či probant hádal alebo nie alebo aká je miera hádania. K tomuto rozhodnutiu môžeme dospieť na základe buď štatistického vyhodnotenia jeho odpovedí alebo tým, že mu dáme možnosť vyjadriť svoju istotu správnosti odpovede (najjednoduchšie napríklad možnosťou neodpovedať).

Prehľad možností realizácie odradenia od hádania zostavený autorom tejto práce na základe literatúry a vlastných úvah:

1. Zvyšovať hranicu úspešnosti alebo zvyšovať počet otázok a ich obtiažnosť (na základe štatistických ukazovateľov uvedených v predchádzajúcej časti).
2. Využívať otázky s viacerými správnymi voľbami (typu MC – MR) s negatívnym skórovaním.
3. Umožniť neodpovedať a zaviesť korekciu pre nesprávne odpovedané úlohy negatívnym skórovaním.
4. Umožniť stanoviť hodnotu presvedčenia o správnosti odpovedi (pozitívna korekcia).
5. Analýza predchádzajúceho testovania a optimalizácia testov.

Postupne jednotlivé návrhy okrem prvého rozoberieme v samostatných častiach.

#### **3.1 Otázky s viacerými správnymi voľbami (MC-MR) s negatívnym skórovaním a pravdepodobnosť ich uhádnutia**

Doteraz to nebolo explicitne povedané ale zo všeobecnej (aspoň slovenskej) skúsenosti vyplýva, že pod skórovaním si priemerní študenti predstavujú pridelenie jedného bodu za správnú odpoveď (výber správnej voľby) a žiadneho bodu za nesprávnu odpoveď (výber



správnej voľby alebo neodpovedal). Otázky typu MC-MR nemajú zmysel bez negatívneho skórovania / hodnotenia (**penalty score = trestné body** za nesprávnu odpoveď). Ak nie je nastavené negatívne skórovanie, potom študenti rýchlo prídu na to, že zaškrtním všetkých volieb v otázke MC-MR získajú v Moodle plný počet bodov. Otázkou je ako správne nastaviť negatívne bodovanie...

**34** Sú dané dve desatinné čísla 15,8 a 0,69. Označ výroky, ktoré sú **pravdivé**:

Známky: 2,00/2,00 (RNDr. Erika Tomková: Učebné osnovy a testy, 5. ročník ZŠ, Desatinné čísla) [Moz](#)

Vyberte aspoň jednu odpoveď:

- a. V čísle 15,8 je číslica 5 rádu jednotiek. ✓
- b. Číslo 0,69 po zaokrúhlení na desatiny je 0,7 ✓
- c. Rozdiel týchto čísel je 15,11. ✓
- d. Číslo 15,8 po zaokrúhlení na jednotky je 20 ✗
- e. V čísle 0,69 je číslica 9 rádu stotín. ✓
- f. Súčet týchto čísel je 22,7. ✗

[Odoslať](#)

**Správny**  
Známky za odoslaný test: 2,00/2,00.

**Obr. 5 Príklad MC-MR testovacej otázky (Zdroj Tomková, reprezentácia Moodle )**

Obtiažnosť uhádnutia správnej odpovede pre otázky MC-MR, je dva (pre dve voľby) až desaťkrát (pre 6 volieb) nižšia ako pri výbere len jednej správnej voľby. Ide o teoretické / vypočítané hodnoty, ktoré zahrňujú všetky možnosti. Napríklad pre dve možné voľby a. b. sú to 4 možnosti: 0, a, b, ab, pre 3 možné voľby už 8 možností: 0, a, b, c, ab, ac, bc, abc. Tento počet možností teda zahrňuje aj: žiadna správna voľba a všetky voľby sú správne. Počet možností narastá podľa Pascalovho trojuholníka, teda ako  $2^m$ , kde m je počet možných volieb v položke. Ak sú všetky kombinácie volieb rovnako pravdepodobné (aspoň študent ich tak vníma), potom pravdepodobnosť náhodného uhádnutia správnej kombinácie volieb v otázke typu MC-MR je:

$$p = \frac{1}{2^m}$$

Z výpočtov vyplynie, že už pri 5 položkách tohto typu je možnosť uhádnutia správnych odpovedí relatívne zanedbateľná. (V prípade, že nie je vhodné aby možnosti „žiadna odpoveď“ alebo „všetky správne“ boli použité, poklesne počet kombinácií len o 2, čo zvýši počet otázok na dosiahnutie 0,1% v prípade dvoch možností na 10 ale v prípade 3, 4, 5, 6 len o jednu).

### 3.2 Možnosti negatívneho skórovania otázok typu MC-MR

Dobрым príkladom negatívneho skórovania otázok typu MC-MR je prijímací pohovor na Lekársku fakultu Univerzity Komenského (LF UK), v ktorom je využitý systém MC-MR tak, že za každú správne zaškrtnutú voľbu je pridelený +1 bod. Za nesprávne zaškrtnutú voľbu je -1 bod. Za správne nezaškrtnutú voľbu je pridelený +1. Neexistuje voľba neodpovedať (= nezaškrtnuté). Ak je nezaškrtnutá voľba, ktorá mala byť zaškrtnutá pridelený je tiež -1 bod.

Uchádzači sa pripravujú z materiálov, ktoré obsahujú stovky otázok s 8 voľbami s označenými správnymi odpoveďami.

Testy sú administrované ako anonymizované - odpoveďový hárok je zviazaný s osobnými údajmi 6 miestnym identifikačným číslom. Získaný počet bodov v teste je jediným kritériom pre prijatie. Testy sú skenované a bodovanie prebieha automatizovane. Výsledky (prijímacej skúšky) sú zverejnené v deň konania alebo najneskôr nasledujúci deň na internetovej stránke fakulty, kde si ich študenti môžu pozrieť (každý po zadaní svojho ID). Aj tak musia väčšinou čakať až do konania posledných testov, pretože rozhodnutie o tom či budú prijatí alebo nie závisí od toho koľko bodov získa prvých s študentov(CR test), pričom s je plánovaný počet

prijatých, ktorý stanoví dopredu dekan. V šk. roku 2010/2011 to bolo okolo 500 študentov pre všeobecné lekárstvo, čo znamenalo dosiahnuť 630 bodov z 800. Zaujímavosťou je, že výsledky môžu nadobúdať len párne hodnoty. V menovanom roku jeden študent/študentka dosiahol plný počet bodov a traja sa pomýlili len raz. Systém je transparentný a nie je nám známe, že by sa študenti sťažovali priamo na systém testovania. Ba práve naopak, je to jeden z výdobytkov zmeny politického režimu po roku 1989. Zdá sa teda, že takýto systém skórovania MC-MR testov môže byť dobrý. Možnosti jeho implementácie v Moodle rozoberieme v nasledovnom podrobnejšie.

### 3.3 Negatívne skórovanie v systéme Moodle

Z pohľadu Moodle je test (v Moodle anglicky nazývaný quiz) aktivita (na rozdiel od zdroja). Vo verzii 2.0 sú štandardne možné nasledovné typy otázok / úloh (zobrazené ako názov, ktorý je prekladom anglického originálu<sup>[4]</sup> a v zátvorke komentár autora):

- Krátka odpoveď
- Numerická (odpoveď)
- **Pravda/Nepravda**
- **Viaceré odpovede (MC aj MC-MR)**
- Esej
- Zhoda
- Náhodná zodpovedajúca krátka odpoveď
- Vložené odpovede (Cloze - dopĺňovačky)
- Drag-and-Drop Matching
- Jednoduchá výpočtová
- Výpočtová
- Výpočtová s viacerými odpoveďami

V Moodle sú teda podporované všetky typy otázok / položiek, o ktorých sme doteraz uvažovali. Od verzii 2.0 je možné negatívne **skórovanie**, nie však negatívna **korekcia** (výsledok bodov za položku je vždy kladný). V najnovších verziách 2.1 , 2.2 bola pridaná aj táto vlastnosť (pozri oznam 10 júna 2011 Tima Hunta – správcu modulu Quiz v Moodle Tracker).<sup>[7]</sup>

Vlastnosti testov v Moodle (rozdiely oproti písomnému testovaniu). Zdroj: Fikar, Moodle 1.9,<sup>[8]</sup> online dokumentácia Moodle na moodle.org a vlastné skúsenosti:

- Otázky sú uchované v (data)banke otázok, ktorú je možno využiť vo všetkých kurzoch a rozdeliť na kategórie a podkategórie.
- Testy môžu byť aj náhodne generované. Stačí vybrať požadovaný počet otázok a kategóriu databanky. Poradie otázok, ale aj odpovedí môže byť stále alebo premenlivé (zamiešané).
- Je možné povoliť aby študent mohol ten istý test aj opakovať (retest), pričom je možné zvoliť metódu známkovania viacerých pokusov toho istého probanta ako napríklad najvyššia alebo priemerná známka a pod.
- Možno presne nastaviť „správanie“ sa testu tak, aby test vyhovoval svojmu účelu (formatívne/sumatívne testovanie). Možno dovoliť adaptívny režim, ktorý umožňuje viacnásobné a interaktívne odpovedanie na každú otázku. Možno povoliť alebo zakázať zobrazovanie rôznych informácií (správna odpoveď, skóre a pod. ) hneď pri pokuse, neskôr alebo až po uzatvorení celého testu.

---

<sup>4</sup> Preklad tvorí skupina dobrovoľníkov vedená prof Fikarom na stránkach moodle.sk

- Je možné limitovať možnosť odpovedať nastavením časového limitu alebo bezpečnostnými opatreniami a tak znížiť možnosti neférového správania sa študentov.
- Je možné ku každej voľbe pripraviť (všeobecnú) spätnú väzbu, ktorá sa zobrazí v stanovenom čase. Dokonca ku každej voľbe je možné nastaviť spätnú väzbu, ktorá sa zobrazí ak študent danú voľbu vyberie.
- Databanku a samotné testy možno importovať/exportovať do/z Moodle. K dispozícii je mnoho formátov aj napríklad XML, ktorý umožňuje prevod aj do MS Excel.
- Zaznamenávanie všetkých odpovedí (aj pokusov s možnosťou ich kedykoľvek prezrieť) a spracovanie výsledkov a štatistík.
- Body sa pretransformujú na známky podľa stanovených pravidiel a automaticky sa zapíšu do žiackej knižky/indexu/výkazu užívateľa (alebo celkového prehľadu bodov).

Celková spätná väzba umožňuje nasmerovať automaticky ďalšie štúdium podľa výsledkov testov. Je možné zasielať emailom správy o testoch, nastaviť možnosť aby si študenti hodnotili testy sami. Je možné nastaviť stránkovania a ďalšie možnosti.

### 3.4 Nastavenie skórovania / hodnotenia v Moodle

Na nasledovnom obrázku je zobrazená časť formulára pre vytváranie alebo opravovanie jednej otázky (položky testu). Pre výpočet skóre (pozri nasledovnú časť) sú dôležité nasledovné položky:

- *PHZ* = prednastavená hodnota známky (anglicky Default question grade alebo Default mark od verzie 2.1) znamená celkový počet bodov za zodpovedanie otázky na 100%. (Prednastavená je hodnota 1, teda 1 bod za každú úplne správnu odpoveď. Môže nadobúdať len celočíselné hodnoty).
- *TF* = trestný faktor sú body ktoré sa odrátajú v interaktívnom režime ak nie je otázka zodpovedaná celkom správne. (Interaktívny režim je možnosť opakovať odpoveď viac krát. Prednastavená hodnota 0.1 je primeraná prednastavenej hodnote *PHZ* 1).
- *g* = známka. Každé voľbe máme možnosť priradiť známku ako percentuálnu hodnotu dosiahnutia celkovej správnej odpovede. Predvolená je hodnota „Žiadne“. Známkou možno zvoliť výberom z menu v rozsahu +/- 100% (pozri obrázok vpravo). Počet volieb je limitovaný z praktického hľadiska tým, že známky možno nastaviť ako zlomky:  $\frac{2}{3}$ ,  $\frac{1}{3}$ ,  $\frac{1}{6}$ ,  $\frac{1}{7}$ ,  $\frac{1}{8}$ ,  $\frac{1}{9}$  a percentuálne hodnoty 5%, 10% atď. Z tohto výberu možno zostaviť otázku s dvoma až desiatimi voľbami s rovnakými známkami (súčet pozitívnych známok sa musí rovnať 100%) alebo aj s viacerými voľbami s rôznymi známkami. Pre otázku s jednou správnou voľbou stačí nastaviť jednu známku na 100% a ostatné na „Žiadne“ alebo na záporné hodnoty, čím sa dosiahne negatívne bodovanie. Pre otázku s viacerými voľbami treba nastaviť viacero pozitívnych hodnotení, tak aby ich súčet bol 100%. Pozitívne známky zároveň indikujú že ide o správne voľby.

Spätná väzba	Známka
83.33333 %	Žiadne
80 %	
75 %	
70 %	
66.66667 %	
60 %	
50 %	
40 %	
33.33333 %	
30 %	
25 %	
20 %	
16.66667 %	
14.28571 %	
12.5 %	
11.11111 %	
10 %	
5 %	
	Žiadne

Text otázky

- Typ písma - - Veľkosť písma - - Formát -

**B** *I* U ABC x<sub>2</sub> x<sup>1</sup>

Najmenší spoločný násobok výrazov  $(3a - 3b)$ ,  $(a^2 - 2ab + b^2)$  je:

(RNDr. Erika Tomková: Učebné osnovy a testy, 9. ročník ZŠ, Lomené výrazy)

Cesta:

HTML formát

Prednastavená hodnota známky\* **1**

Trestný faktor\* 0.1

Všeobecná spätná väzba

- Typ písma - - Veľkosť písma - - Formát -

**B** *I* U ABC x<sub>2</sub> x<sup>1</sup>

Cesta:

HTML formát

Jedna alebo viaceré odpovede? **Iba jedna odpoveď**

Zamiešať odpovede

Číslovať voľby? a, b, c, ...

**Voľba 1**

Odpoveď  $(3a - 3b)(a - b)$

Známka **100 %**

Spätná väzba

Moodle auto-formát

Obr. 6 Príklad vytvorenia / nastavenie otázky v systéme Moodle

Nastavenie jednej voľby s hodnotou známky 100%, označuje jedinú správnu voľbu v otázke s iba jednou správnou odpoveďou / voľbou. (Obrázok nie je úplný – volieb môže byť viac, nesprávne voľby už nie sú zobrazené).

### 3.5 Výpočet bodovej hodnoty odpovede v Moodle

Ako vyplýva z prechádzajúceho označenia je celkové skóre ( $x$ ) za odpoveď na jednu otázku súčtom čiastkových známok pri voľbách, ktoré probant vybral vynásobené prednastavenou hodnotou známky ( $PHZ$ ) a od tohto je odrátaný trestný faktor ( $TF$ ) násobok  $PHZ$ , ak je dovolené viacnásobné odpovedanie na otázku:

$$x = PHZ \cdot \left( \sum_{i=1}^k g_i - TF \right)$$

kde  $k$  je počet vybraných voľieb a  $g_i$  je známka vybranej voľby.

Poznámky:

- Predchádzajúci vzorec platí pre otázky s jednou aj viacerými správnymi voľbami ako aj pre pozitívne aj negatívne známky, za predpokladu že „žiadne“ znamená 0. Vo verzii Moodle nižšej ako 2.1 je však  $x \geq 0$ . Od verzie 2.1 môže byť  $x$  aj záporné.

- V súčasných verziách musí byť počet vybraných volieb aspoň 1. Nie je možné nevybrať žiadnu odpoveď. Otázka bez vybranej voľby je vyhodnotená ako nezodpovedaná a pridelené skóre je nula.
- Terminológia používaná v Moodle ešte nie je celkom ustálená. Podľa nášho názoru by bolo vhodnejšie nazývať čiastkovú známku pri voľbách ako **% dosiahnutia celkovej správnej odpovede** alebo **percentuálna (%) váha**, celkové skóre, ktoré probant dosiahol za jednu odpoveď ako **body** a **známkou** až výslednú známku, ktorá sa zapisuje od klasifikácie.

Ďalšie podrobnosti možno nájsť v terminológii článku zo sekcie Dev(elopment) Moodle 2 s názvom Quiz statistics calculations<sup>[9]</sup>.

### 3.6 Odporúčania ako vyhodnotiť položkovú analýzu

Kehoe<sup>[5]</sup> uvádza nasledovné odporúčania pre hodnotenie výsledkov: „Podiel študentov ktorí odpovedali správne alebo nesprávne na danú položku ovplyvňuje jej diskriminačnú silu. Otázky na ktoré odpovedala správne (alebo nesprávne) veľkú časť skúmanej populácie (viac ako 85%) majú výrazne zníženou diskriminačnú schopnosť. V dobrom teste, väčšina položiek je zodpovedaná správne 30% až 80% respondentami.“ a ďalej „Položky, ktoré prakticky každý zodpovie správne sú zbytočné“. „Distraktory, ktoré nikto nevyberie by mali byť nahradené alebo odstránené. Netreba sa však znepokojovať, ak sú distraktory vybrané rôznym počtom probantov. Tiež skutočnosť, že väčšina študentov zodpovie položku nesprávne, neznamená, že by sa mala zmeniť.. ak je správna .. a jej korelácia s celkovým skóre je záporná.“

Autori Elektronickej učebnice pedagogického výskumu (Gavora et al. online)<sup>[10]</sup> uvádzajú tri nástroje ako vyhodnotiť spoľahlivosť testu (ako výskumného nástroja):

- Opakovanie administrácie výskumného nástroja
- Rozpolovanie výskumného nástroja
- Odhad vnútornej konzistencie výskumného nástroja

Podobne ako Gavora (online) tak aj Kehoe ako najčastejšie nástroje pre hodnotenie spoľahlivosti uvádzajú metódu odhadu vnútornej konzistencie a to pomocou vzorcov Kudara a Richardsona č. 21 (KR-21) alebo Cronbachovo alfa. Tiež zhodne uvádzajú, že koeficient vyjadrujúci spoľahlivosť by mal byť väčší ako 0,8 (1,0 je maximum). Gavora (online) upozorňuje, že „výskumný súbor, ktorý je homogénny, spravidla spôsobuje nižšiu reliabilitu výskumného nástroja ako výskumný súbor, ktorý je menej homogénny“. Kehoe zase na to, že aj „hodnoty ako 0,5 sú uspokojivé pre krátke testy (10 - 15 položiek)“ a tiež, že „nízke hodnoty sú zvyčajne pri nadmiere veľmi ľahkých alebo veľmi ťažkých položiek alebo keď sú položky nie sú dostatočne diskriminačné alebo keď položky sú z viacerých obsahov (oblastí)“. Kehoe naopak zdôrazňuje dôležitosť homogenity testu (testovacieho nástroja). Ako príklad uvádza rozdiely medzi vysokými spoľahlivosťami testov slovnej zásoby alebo matematiky, pretože bývajú viac homogénne ako testy v oblasti spoločenských vied. Uvádza paradoxný príklad, keď odstránením 7 nekorelujúcich položiek z 30 položkového testu sa zvýšila spoľahlivosť z 0,79 na 0,88 – teda kratší test poskytoval spoľahlivejšie výsledky ako test dlhší.

### 3.7 Implementácia testu s viacerými správnymi voľbami a negatívnym skórovaním v Moodle

Táto časť je venovaná najčastejšie využívanému spôsobu ako odradiť probantov od hľadania odpovedí a to využitím otázok s viacerými správnymi voľbami a negatívnym skórovaním. Príkladom takéhoto testu je spomínaná prijímacia skúška na LF UK. Ako bolo uvedené, Moodle umožňuje priradiť každej voľbe inú známku (percentuálnu váhu). V ďalšom vysvetlíme ako implementovať test podobný prijímacej skúške na LF UK v Moodle. Pretože v Moodle sa

skórujú len položky, ktoré boli označené probantom je to pomerne zložité. Problém spočíva v tom, že neoznačeným položkám sa v Moodle nepriradujú žiadne body napriek tomu, že majú priradenú určitú percentuálnu váhu. Dá sa to však obísť individuálnym nastavením váh jednotlivých volieb (modle.sk prekladaných ako známky).

V Moodle je možné zrealizovať aj test typu LF UK, treba však v závislosti od počtu správnych volieb meniť hodnotu „Prednastavená hodnota známky“ (anglicky Default Question Grade vo verzii 2.1 a Default mark vo verzii 2.2) ako aj Nastavenie bodovania danej voľby v Moodle (Známka). Je teda možné zrealizovať v Moodle test s viacerými správnymi voľbami a negatívnym skórovaním, ktorý je tak diskriminujúci ako prijímací test na LF UK. Stačí k tomu zvoliť správne váhy koeficientov známok.

Existujú však aj iné, menej drastické, spôsoby ako sprísiť bodovanie.

### 3.8 Umožniť neodpovedať a zaviesť regresiu pre zle zodpovedané úlohy negatívnym známkovaním

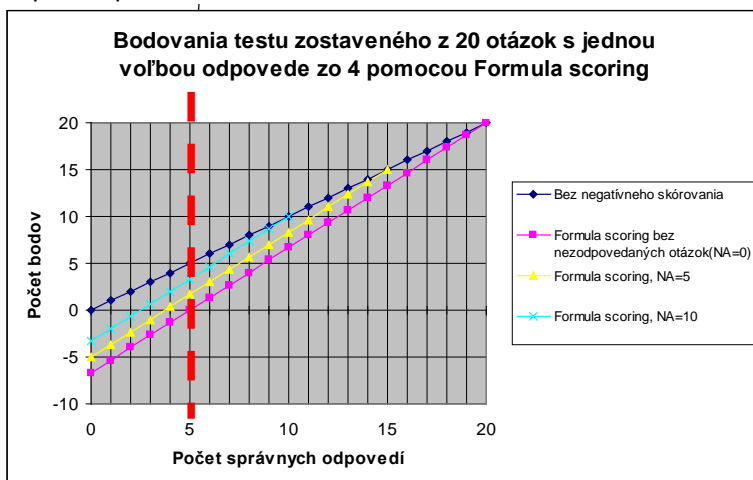
Na zvýhodnenie tých čo nehádajú je možné zaviesť možnosť neodpovedať a následne trestať nesprávne odpovede napríklad podľa metodológie R. Fraryho<sup>[11]</sup>[online], ktorú nazýva Formula Scoring (skórovanie podľa vzorca). Frary navrhuje pri otázkach s jednou správnou voľbou redukovať počet bodov za správne odpovede o počet nesprávnych odpovedí \* (počet volieb v otázke -1). Trestný faktor úmerný počtu možností v otázkach -1 zabezpečuje, že práve probant, ktorý zodpovie správne presne celkový počet otázok / počet volieb v otázke a pritom odpovie na všetky otázky dostane práve 0 bodov. Tu je dôkaz:

Ak je počet volieb  $m$  a probant háda náhodne, potom uhádne  $n/m$  otázok zo všetkých  $n$  otázok. Neuhádne  $n.(m-1)/m$  otázok. Ak je záporná známka nastavená na  $-1/(m-1)$ , potom suma jeho bodov je:

$$\frac{n}{m} + n \cdot \frac{(m-1)}{m} \cdot \frac{-1}{(m-1)} = \frac{n}{m} - \frac{n}{m} = 0$$

čo je základ tejto metódy a zároveň predpis na nastavenie veľkosti trestného faktoru.

Tento systém využíva napríklad aj najrozšírenejší test vzdelávacích predpokladov v USA - SAT Reasoning Test<sup>[12]</sup> (Collegeboard Test Tips online), kde je  $\frac{1}{4}$  trestného bodu za nesprávnu odpoveď pri otázkach s 5 voľbami.



Obr. 7 Graf: Bodovanie testu s 20 otázkami

Z grafu je vidieť, že záporná známka zmení strmú bodovania tak, že pre NA=0 prechádza nulou pri počte správnych odpovedí zodpovedajúcich hádaniu ( $20/4 = 5$ ). Každá otázka, ktorú probant vynechá namiesto toho aby ju zodpovedal nesprávne mu prinesie zisk tretiny bodu. Ak

je teda schopný rozpoznať tri takéto otázky získa o bod viac. Napríklad študent, ktorý 10 otázok vynechal a podarilo sa mu zodpovedať ostatných 10 otázok správne získa viac bodov (10) ako študent, ktorý odpovedal na 12 otázok správne a na 8 nesprávne (ten má len 9,33 bodu).

V prípade takto nastaveného bodovania však treba dostatočne presne informovať probantov. Frary uvádza inštrukcie Davisa (Davis 1967, str43): „Vaše skóre bude zodpovedať počtu správne zvolených odpovedí mínus malý násobok počtu nesprávnych odpovedí. Odpovedajte na otázku aj keď nie ste si úplne istí že odpovede sú správne a to najmä vtedy keď viete eliminovať jednu alebo viacero možností ako nesprávnych alebo máte tušenie alebo pocit, ktorá možnosť je správna. Avšak je lepšie vynechať otázku ako keby ste len hádali naslepo medzi všetkými možnosťami.“

Pokiaľ inštrukcie majú odradiť od úplne náhodného hádania, naopak ľudia na internete si vymieňajú skúsenosti ako hádať (získať najlepší výsledok). Napríklad pre SAT, v ktorom je 5 volieb v otázke (záporná známka  $-\frac{1}{4}$ ), sa odporúča (Collegeboard Test Tips. [14]): vylúčenie jednej voľby zvyšuje pravdepodobnosť uhádnutia na 25% (a očakávaný zisk o 1 / 16 bodu), vylúčiť dve voľby zvyšuje pravdepodobnosť na 33,3% (a zisk o 1 / 6 bodu), vylúčiť tri voľby znamená 50% pravdepodobnosť (a 3 / 8 bodu).

### 3.9 Umožniť stanoviť hodnotu presvedčenia o správnosti odpovede

Príklad vytvorenia pozitívnej diskriminácie tých, ktorí sú presvedčení o správnosti odpovede voči tým, ktorí nie sú je Známkovanie založené na istote alebo Známkovanie založené na sebadôvere (CBM), ktoré bližšie popíšeme.

#### Confidence Based Marking

Confidence-Based Marking alebo Certainty-Based Marking (CBM) je založené na možnosti testovaného neodpovedať alebo vyjadriť svoje presvedčenia, že odpoveď je správna na 3-bodovej stupnici vlastnej istoty s hodnotami C(confidence): 1 (nízka), 2 (stredná) alebo 3 (vysoká). Zámerne nie sú používané slová ako "určite" alebo "veľmi určite", pretože tieto majú rôzny význam pre rôznych ľudí.

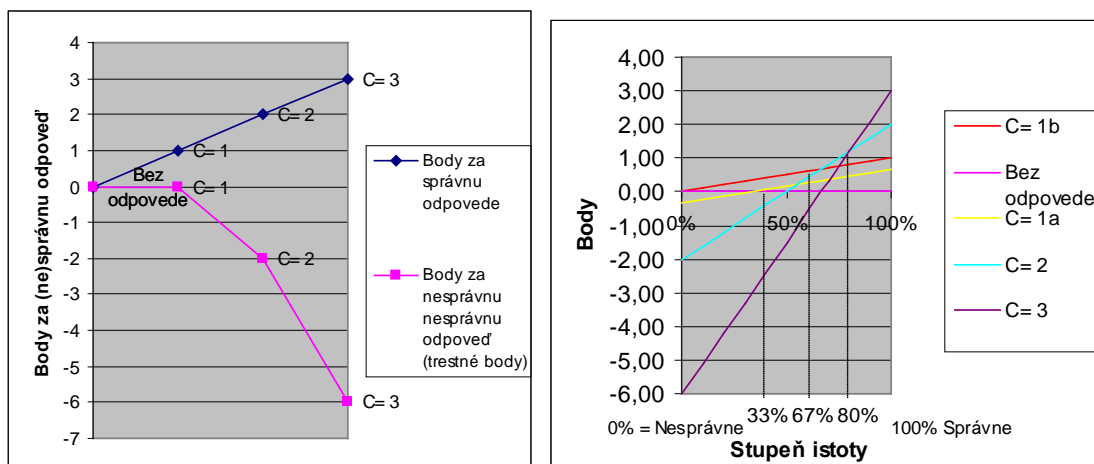
CBM už existuje niekoľko desaťročí (napríklad na Medical School at University College Londýn). Jeho propagátorom je A.R. Gardner-Medwin<sup>[13]</sup>(GARDNER-MEDWIN [online]), ktorý na internete zverejnil viacero svojich relevantných publikácií. CBM využíva viac britských univerzít. Na stránke LAPT [14] je implementovaná verzia takýchto testov.

**Obr. 8 Príklad známkovanie založené na istote / sebadôvere. Zdroj:** <http://www.ucl.ac.uk/lapt/laptlite/>  
Vyhodnotenie / bodovanie CBM je nasledovné:

	Stupeň istoty			
	Bez odpovede	C= 1	C= 2	C= 3
Body za správnu odpoveď	0	1	2	3
Body za nesprávnu odpoveď (trestné body)	0	0	-2	-6
Rozdiel: zisk-strata	0	1	4	9

**Obr. 9 Bodovanie CBM**

Bodovanie je založené na tom, že s narastajúcim stupňom istoty nielen narastá počet možných pozitívnych bodov ale narastá aj rozdiel medzi ziskom a stratou bodov. Nie je to však priama úmernosť. Rozdiely zisk-strata narastajú s mocniteľom 2 (risk =  $C^2$ ), čo znamená, že stupeň riskovania sa umocňuje, čo je základom tejto metódy. Pokiaľ pre  $C = 1$  je risk minimálny (dokonca výsledok je vždy lepší ako pri voľbe „Bez odpovede“) pre  $C = 2$  je risk vyjadrený rozpäťnásobným zisk – strata už štvornásobný a pre  $C = 3$  deväťnásobný.



**Obr. 10 Grafy: Body pre správne a nesprávne odpovede a percentá vyjadrujúce stupeň vlastnej istoty, kedy sa oplatí meniť rozhodnutia.**

Zdroj údajov <http://www.ucl.ac.uk/lapt/>, graf: autor (C=1a je nami navrhovaná zmena).

Z prvého grafu vyplýva, že vzhľadom na možnú stratu, označiť odpoveď stupňom 2 je dvakrát tak riskantné ako ju označiť stupňom 1. Označiť stupňom 3 je dvakrát riskantnejšie ako použiť stupeň 2. Miera ochoty podstúpiť risk závisí na probantovi. Môže získať ale aj stratiť. Z druhého grafu vyplýva, že označiť odpoveď stupňom 1 sa oplatí do istoty 50% až 67%, Odpoveď 2 sa oplatí voliť ak je istota od 67 do 80%. S väčšou istotou je vhodný stupeň 3.

Poznámka: Daňou za jednoduchosť (celočíselnosť) koeficientov je nie celkom presne možná interpretácia možností. Uvedené hranice istoty nie sú rovnomerne distribuované. Neoplatí sa neodpovedať, pretože odpoveď C1 j vždy lepšia možnosť. Navrhujeme úpravu aby pri C=1 a nesprávnej odpovedi bola pridelená penalizácia -0,33 bodu (na grafe označené ako C1a). Potom by mohli byť tlačidlá premenované aj na C1= 33% , C2= 55% a C3= 80%, čo je podľa nášho názoru zrozumiteľnejšie vyjadrenie miery potrebnej istoty a to napriek tomu, že aj tieto intervaly medzi nimi nie sú konštantné (33%, 22%, 25%).

#### 4 Odporúčania ako vytvárať otázky s jednou a viacerými správnymi voľbami

Turek 2002 vôbec neodporúča používať otázky s viacerými správnymi voľbami (str. 63: „Volíť úlohy len s jednou správnou odpoveďou“). Jeho ďalšie stručné odporúčania pre otázky s jednou správnou voľbou sú:

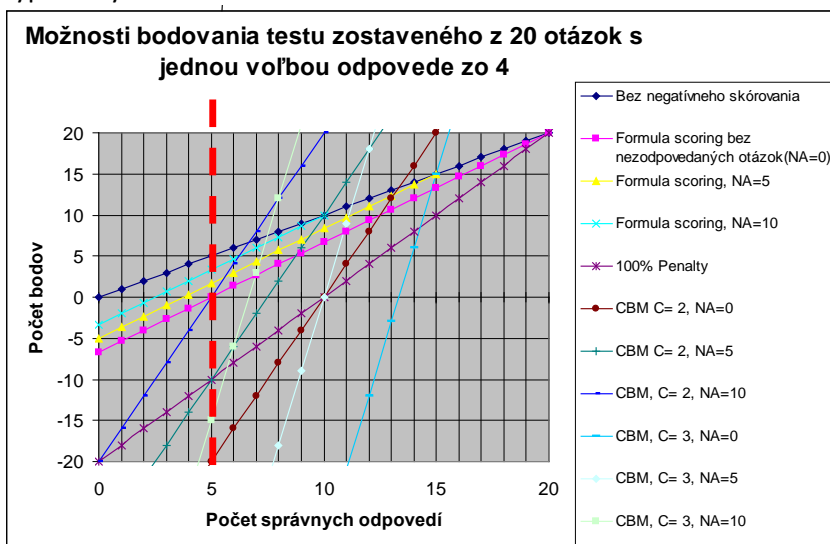
1. Najmenší počet ponúkaných odpovedí je 4, optimálny počet je 5.
2. Správne odpovede majú byť umiestnené rovnomerne na všetkých pozíciách.
3. Ponúkané odpovede majú mať rovnaký gramatický tvar a po stránke gramatickej sa majú hodiť do tvrdenia - textu úlohy.
4. V ponúkaných odpovediach sa vyhýbať formuláciám: „všetky vyššie uvedené odpovede“, „ani jedna z vyššie uvedených odpovedí“ a pod.
5. Dĺžka ponúkaných odpovedí by mala byť približne rovnaká.
6. Vyhýbať sa negatívnym formuláciám v texte úlohy.



Tieto rady sú vo veľkej miere podobné na inštrukcie z Virginia Polytechnic Institute z roku 1995 od Jerarda Kehoe (1995a) zverejnené v časopise Practical Assessment, Research & Evaluation. O dva roky neskôr – v roku 1997 – boli zverejnené návody od Roberta Frayho (1997),<sup>[15]</sup> ktoré sú rozdelené do troch skupín: obsah, štruktúra a voľby (distraktory). Na rozdiel od predchádzajúceho je Fray zástancom odpovedí "nič z vyššie uvedeného" a tiež uvádza ako vhodný počet odpovedí už dva a viac s odôvodnením, že zbytočné distraktory študenti ľahko odhalia. Fray podobne ako Turek odporúča vyhýbať sa negatívnym formuláciám, ak nie je účelom testovania práve schopnosť rozpoznať negáciu výrokov ako napríklad v teste z matematiky ako je to na obrázku Obr. 4 Príklad otázky s viacerými správnymi voľbami. Aj v tomto prípade je však vhodné grafické zvýraznenie negatívnej formulácie napríklad podčiarknutím.

Vynikajúce sú odporúčania Roberta Frayho [14], ktoré sa týkajú obsahu, štruktúry testu a možnosti odpovedí (distraktory). Okrem odporúčaní upozorňuje na štyri druhy možných chýb. Pre limitáciu rozsahu ich tu neuvádzame.

Pokiaľ Turek predpokladá, že nie je problém pri ľubovoľnej otázke pridať 4 alebo 5 dôveryhodných distraktorov odporúča Fray radšej ponechať menší počet možností ako umelo vymýšľať možnosti, ktoré sú veľmi málo pravdepodobné. Podobne je to aj s možnosťou „ani jedna z vyššie uvedených odpovedí“, ktorú Turek neodporúča, ale Fray vidí jej význam pri výpočtových úlohách.



**Obr. 11 Graf - Porovnanie výsledkov bodovania testu zostaveného z 20 otázok s jednou správnou voľbou zo 4 pomocou rôznych metód. Výsledky sú v rozsahu -20 až + 20 bodov.**

Vysvetlivky: NA – počet vynechaných otázok, C vyjadruje mieru istoty pri metóde CBM, CBM C=1 zodpovedá metóde bez negatívneho skórovania a preto nie je uvedené na grafe. Zvislá čiarkovaná čiara prechádzajúca 5 značí počet odpovedí, ktoré sú dosiahnuteľné náhodným hádaním.

### 5 Záver: výber optimálnej metódy bodovania

Všetky uvedené metódy sú lineárne: s počtom správnych odpovedí lineárne rastie počet bodov. Vzhľadom na tradíciu a jednoduchosť inštrukcií pre probantov je to optimálne aj keď nie jediné riešenie. Na grafe je vynesena čiarkovaná čiara hodnota 5, ktorá zodpovedá počtu správnych odpovedí získaných náhodným hádaním. Z grafu je vidieť, že do počtu 7 správnych odpovedí získa najviac bodov probant ak je využitá metóda bez negatívneho skórovania. Okolo hodnôt 5 až 7 možno sledovať rozdiely v počte bodov pri bodovaní pomocou formula scoring

v závislosti na počte vynechaných odpovedí (viacej vynechaných znamená viacej bodov). Ako bolo uvedené, jednotlivé hodnotenia zoradené od najbenevolentnejšieho po najprísnejšie sú (toto je vidieť aj na grafe ako nárast smernice priamok, ktorá je uvedená v zátvorke dole):

- Bez negatívneho hodnotenia (smernica: 1)
- Formula scoring (1,3)
- Individuálne nastavenie váhy distraktorov (nie je v grafe, smernica od 1 po 2)
- 100% penalty (2)

Metóda CBM sa nedá priamo porovnať, pretože sa nepredpokladá, že by probanti vo všetkých otázkach volili rovnakú mieru istoty C. Z výskumov známkovanie založené na istote[14] však vyplýva, že väčšina (10 000) respondentov volí takú mieru istoty, že získava len pozitívne hodnotenia, ktoré smerovaním zodpovedajú skórovaniu bez negatívneho hodnotenia. Predpokladáme však, že absolútne hodnoty bodov, ktoré nie sú uvedené, sú ešte vyššie a preto táto metóda môže byť hodnotená ako najbenevolentnejšia.

Hodnotenie obsahuje protichodné požiadavky:

- Skúšajúci potrebuje rozdeliť (diskriminovať) a probanti sa snažia získať čo najviac bodov.
- Najslabším študentom vyhovuje najmiernejšie hodnotenie (umožňujúce hádanie).
- Zvyšovanie stupňa prísnosti hodnotenia má pravdepodobne opačný účinok na motiváciu.
- Hodnotenie musí byť spravodlivé a transparentné. Presnejší systém bodovania je zložitejší a vyžaduje si podrobnejšie inštrukcie

Z uvedených dôvodov **neexistuje optimálny univerzálny systém hodnotenia** vhodný pre všetky okolnosti testovania. Ako najdôležitejšie odporúčame vybrať metódu podľa účelu hodnotenia (pozri časť Skórovanie odpovedí (položiek) a účely testovania), tak aby boli využité direktívne alebo kooperatívne možnosti odradenia probantov od hádania správnych odpovedí (časť Možnosti odradenia od hádania odpovedí). Určite sa prihovárime za kombinovanie rôznych typov otázok a využitie otázok typu MC-MR s negatívnym skórovaním.

### ***Záverčné odporúčania***

Stručné a zhrňujúce odporúčania na záver:

1. Podľa účelu hodnotenia vyberte metódu hodnotenia
2. Kombinujte typy otázok/položiek. V prípade dichotomických otázok a otázok s jednou správnou voľbou je potrebné zvoliť dostatočný počet otázok, preto sú vhodné aj otázky s viacerými voľbami.
3. Niekedy je vhodné stanoviť pevné poradie otázok. V niektorých testoch je poradie od jednoduchých otázok po zložité. Toto môže zvýšiť motiváciu avšak aj navodiť nereálne očakávania.
4. Implementácia v Moodle umožňuje nastaviť váhy ako jednotlivých volieb tak aj ľubovoľnej otázky. Možno teda nastaviť váhu otázok tak aby zodpovedali náročnosti (napríklad z predchádzajúceho testovania). Pomocou diskriminačného indexu možno objaviť otázky, ktoré sú nesprávne odpovedané. Je otázkou voľby či takéto „chytáky“ z testu vylúčiť alebo ponechať.
5. Venujte pozornosť vytváraniu distraktorov.

Tento text je skrátenou verziou článku zverejneného na [www.webuniverzita.com](http://www.webuniverzita.com), ktorý obsahuje výpočty ako aj nastavenia Moodle testov pre negatívne skórovanie. Autor sa teší na prípadné komentáre, doplnenie a návrhy.

## Literatura

---

1. TUREK, I. 2002. Základy pedagogického výskumu. [disketa] Publikované 2002. [Dátum aktualizácie 6.7.2006], str. 63
2. ŠTULRAJTEROVÁ, M. RAFAJLOVIČOVÁ, R.,: Skúšanie, testovanie a hodnotenie v edukačnom procese (2002), Dostupné na [FWWW.FHV.UMB.SK%2FAPP%2FSAVEDATAFILEPUBLIC.PHP](http://FWWW.FHV.UMB.SK%2FAPP%2FSAVEDATAFILEPUBLIC.PHP)
3. DEPALMA, A. 1990. Revisions Adopted in College Entrance Exams, The New York Times, November 1, 1990 [online] [cit 9.7.2011] Dostupné na internete: <http://www.nytimes.com/1990/11/01/us/revisions-adopted-in-college-entrance-tests.html>
4. BONTIS, N., HARDIE, T., SERENKO, A. (2009). Techniques for assessing skills and knowledge in a business strategy classroom International Journal of Teaching and Case Studies, 2, 2, 162-180. [online] [cit 9.7.2011] Dostupné na internete: <http://www.business.mcmaster.ca/mktg/nbontis/ic/publications/IJTC SBontis2-2.pdf>
5. KEHOE, J. 1995a. Writing multiple-choice test items. In Practical Assessment, Research & Evaluation, 4(9). [online] [cit 9.7.2011] Dostupné na internete: <http://PAREonline.net/getvn.asp?v=4&n=9> . Upravené zo zdroja: Testing Memo 4: Constructing Multiple-Choice Tests -- Part I, Office of Measurement and Research Services, Virginia Polytechnic Institute and State University, Blacksburg,
6. BURJAN V., Tvorba a využívanie školských testov (časti 1-7.) Exam Info. [online] [cit 9.8.2011] Dostupné na internete: [http://www.burjanoskole.sk/wp-content/uploads/documents/Tvorba\\_testov\\_komplet.pdf](http://www.burjanoskole.sk/wp-content/uploads/documents/Tvorba_testov_komplet.pdf)
7. HUNT T. 10 júna 2011 1:43 AM v Moodle Tracker, [online] [cit 9.8.2008] Dostupné na internete: <http://tracker.moodle.org/browse/MDL-1647>
8. FIKAR M. 2010, Moodle 1.9: Príručka učiteľa. FCHPT STU v Bratislave, 2010. Dostupné v elektronickej verzii priamo od autora na: [www.kirp.chtf.stuba.sk/~fikar](http://www.kirp.chtf.stuba.sk/~fikar)
9. MOODLE 2 DEV(elopment), Quiz statistics calculations, [online] [cit 9.8.2008] Dostupné na internete: [http://docs.moodle.org/dev/Quiz\\_statistics\\_calculations](http://docs.moodle.org/dev/Quiz_statistics_calculations)
10. GAVORA P., et al. Elektronická učebnica pedagogického výskumu [online] [cit 9.8.2008] Dostupné na internete: <http://www.e-metodologia.fedu.uniba.sk/index.php/kapitoly/ziskanie-hodnotnych-dat/zistovanie-reliability.php?id=i9p3>
11. FRARY R., Formula Scoring In Instructional Topics in Educational Measurement (ITEMS) [online] [cit 9.8.2008] Dostupné na internete: [http://www.ncme.org/pubs/items/ITEMS\\_Mod\\_4.pdf](http://www.ncme.org/pubs/items/ITEMS_Mod_4.pdf)
12. Collegeboard Test Tips. [online] [cit 9.8.2008] Dostupné na internete: [http://www.collegeboard.com/student/testing/sat/prep\\_one/test\\_tips.html](http://www.collegeboard.com/student/testing/sat/prep_one/test_tips.html) , <http://en.wikipedia.org/wiki/SAT> SAT predtým the Scholastic Aptitude Test or Scholastic Assessment Test
13. GARDNER-MEDWIN A.R. Zoznam publikácií venovaný CBM [online] [cit 9.8.2008] Dostupné na internete: <http://www.ucl.ac.uk/~ucgbarg/pubteach.htm>
14. LAPT: London Agreed Protocol for Teaching Dostupné na: <http://www.ucl.ac.uk/lapt/laptlite/>
15. FRARY, R. More Multiple-Choice Item Writing Do's and Don'ts. In ERIC Digest. ERIC Digest TM 95-3 (ED 398 236) [online] [cit 9.8.2008] Dostupné na internete: <http://www.ericdigests.org/1997-1/more.html> . Adapted with permission from "Testing Memo 10: Some Multiple-choice Item Writing Do's And Don'ts," Office of Measurement and Research Services, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060

## Informace o autorovi

**Miroslav Babinský**

Telefon: +421 917 669 548

E-mail: [miroslav.babinsky@gmail.com](mailto:miroslav.babinsky@gmail.com)

Zastávaná funkce: vedoucí Centra ďalšieho vzdelávania

Název inštituce: Slovenská technická univerzita v Bratislave, Vazovova 5, 812 43 Bratislava, Slovensko