Vím, že vím, aneb: Testy s mírou jistoty pro hodnocení znalostí studentů zdravotnických oborů

I know that I know: A certainty based marking tests designed for evaluating knowledge of healthcare students

Vít Blanař¹, Jan Pospíchal²

¹Katedra ošetřovatelství, Fakulta zdravotnických studií, Univerzita Pardubice, Pardubice, Česká republika

²Katedra klinických oborů, Fakulta zdravotnických studií, Univerzita Pardubice, Pardubice, Česká republika

ABSTRAKT

Východiska: Předkládaný způsob hodnocení počítačových testů vychází z potřeby snížit riziko tipování správných odpovědí při testování velkého množství studentů zdravotnických oborů. Systém vede studenty k uvědomování si svých chyb a odečítá body za neuvědomělé chyby.

Cíl: Porovnat vhodnost metody testu s mírou jistoty oproti testu bez korekce tipování při hodnocení znalostí studentů zdravotnických oborů.

Metody: Znalostní test průběžného hodnocení látky z anatomie podstoupilo 150 studentů. Jednalo se o počítačem hodnocených 40 dichotomických úloh s korekcí na hádání pomocí míry jistoty. Výsledky byly přepočítány na korekci pomocí odečtu bodu a na výsledky bez korekce.

Výsledky: Průměrná hodnota obtížnosti testových položek Q dosahovala hodnoty 22,3. Přesto s korekcí výsledku pomocí míry jistoty byl medián 49,2 % z požadovaných 70 %. Bez korekce byl medián výsledku testu 77,5 %.

Závěry: Hodnocení pomocí míry jistoty se zdá být ke studentům příliš přísné, avšak jestliže by studenti testovanou látku skutečně dobře ovládali, byli by úspěšní. Velkým přínosem oproti jiným korekčním mechanismům na hádání je větší stratifikace výsledků podle míry neuvědomělých chyb a otipovaných odpovědí.

ABSTRACT

Background: The presented method of evaluating computer tests is based on the requisite to reduce the risk of guessing the correct answers while testing a large number of Healthcare students. The system guides students to realize their own mistakes and deducts points for their mistakes made unconsciously.

Aim: To compare the suitability of a method dealing with the certainty based marking tests in contrast to the tests without correction of guessing the correct answers while marking the knowledge of students of health studies.

Methods: 150 students undertook a test evaluating their level of knowledge from anatomy which is taken on regular basis. This test had a form of 40 computer evaluated dichotomous tasks with the correction of guessing the correct answers with a degree of certainty. The results were recalculated for correction using a negative marking and for results without correction.

Results: The average value of the difficulty of test items Q reached the level of 22.3. Yet corrected using the degree of certainty the outcome median was 49.2% of the required 70%. Without corrections the median was 77.5%.

Conclusion: This evaluation using the degree of certainty seems to be too strict for the students, however if the students actually had the required knowledge, they would have been more successful. A great benefit compared to other correction mechanisms designed for guessing is greater stratification of the results according to the number of mistakes made unconsciously and guessed answers.

KLÍČOVÁ SLOVA

Hodnocení s mírou jistoty, didaktický test, hodnotící systém, edukace v ošetřovatelství, student, všeobecná sestra

KEY WORDS

Certainty-Based Marking, didactic test, evaluation system, nursing education, student, nurse

ÚVOD

Neoddělitelnou součástí výuky studentů zdravotnických oborů je hodnocení znalostí. Díky němu ověřujeme zvládnutí předneseného a procvičeného učiva. V České republice je zlatým standardem hodnocení pomocí ústního zkoušení, které bývá u zkušených akademických pracovníků kvalitní a komplexní. Nevýhodou je ovšem vysoká časová náročnost, sklon studentů k memorování velkého množství vědomostí a nižší objektivita při střídání zkoušejících stejného předmětu (1). Proto je v případech zkoušení většího množství studentů vhodné využít kvalitní didaktický test.

Podle Chrásy (2) se rozumí pojmem didaktický test zkouška, která je orientována na objektivní zjišťování úrovně zvládnutí učiva u určité skupiny osob. Jednotlivé testy a testové úlohy se však liší svou obtížností a validitou, tedy shodou obsahu testu s cíli a obsahem výuky. Nalezením vhodného nástroje pro rychlé a efektivní hodnocení znalostí studentů můžeme dosáhnout úspory času a energie vynaložené na ověřování těchto znalostí. Další výhodou je lepší kvantifikace, díky které můžeme studenty hodnotit objektivně.

Jednotlivé testy, které jsou v současné době používané, se liší nejčastěji počtem a druhem testových položek (jednotlivých otázek) nebo metodou hodnocení. Ideální podoba testu musí odpovídat cílům a obsahu výuky, měla by mít dostatečný počet testových položek a test by měl být vyplňován za přijatelných podmínek, stejných pro všechny studenty. Vhodnými podmínkami pro vyplňování testu rozumíme klidné prostředí, dostatek času, v okolí by neměly být materiály, které by mohly zlepšit výsledky studentů, zasedací pořádek by měl znemožňovat opisování a nahlížení do testu ostatním studentům. Při tvorbě testových položek bychom se měli zaměřit na vhodné použití distraktorů, tedy nabízených nesprávných odpovědí, které by neměly být příliš těžké, ani příliš „průhledné.“ Dalším prvkem je počet správných odpovědí, způsob výběru či přiřazování odpovědí, čas na jednotlivé položky a délka textu zadání (3).

Často používané druhy testových položek, kterými jsou ověřovány vědomosti studentů, jsou otázky uzavřené, s mnohočetným výběrem odpovědi (multiple choice questions, MCQ). V tomto případě testovaní vybírají z nabízených možností jednu či více odpovědí. Další možností jsou otázky s jednou nejlepší odpovědí (single-best answer, SBA) a položky s mnohočetným výběrem odpovědi typu ano/ne (multiple true/false, MTF), tedy dichotomické otázky. Mezi novější testové formáty patří přiřazovací otázky, či rozšířené přiřazovací otázky (extended-matching questions, EMQ) (3, 4). Problém u výše popsanych typů testových položek nastává, pokud

potřebujeme zabránit hádání či tipování správných možností. U testů s výběrem z nabídnutých možností existuje riziko tipování správné odpovědi vždy.

Vhodnou možností je využití korekce bodového hodnocení, za účelem snížení vlivu hádání správné odpovědi na samotný výsledek testu. Samotné „hádní,“ tedy označení odpovědi, kterou si zkoušený ve skutečnosti není jistý nebo pouze tipuje je velmi rozšířené, stejně jako čím dál častější využívání testů s uzavřenou nabídkou možností odpovědí (2).

Karp (5) ve své publikaci, jako jeden z prvních autorů, zmiňoval riziko tipování při přijímacích testech na lékařskou fakultu s patnácti položkami. V závislosti na dalších proměnných by se podle jeho výzkumu až 68 % uchazečů tipem trefilo do 1 až 8 položek z patnácti. 16 % tipujících studentů by se netrefilo ani jednou, ale 13,5 % uchazečů by správně tiplo 8 až 12 položek a 2,5 % by se trefilo do 12 až 15 správných odpovědí. Tradičně se v testech používají položky se čtyřmi možnostmi odpovědi. Při ponechání testu v této podobě může student tipovat u každé položky s 25% pravděpodobností úspěchu. Pokud se mu podaří odhalit nejjednodušší distraktor, tak má až 33,3% šanci, že se trefí do správné odpovědi. Pokud jsou v testu použity položky typu MTF s výběrem vždy ze dvou možností, má zkoušený šanci uspět až na více než 50 %, když označí všechny odpovědi „Ano“.

Existuje celá řada korekčních mechanismů, které by měly snížit vliv hádání či tipování správné odpovědi. Většina těchto principů je založena na zvýšení obtížnosti uhodnutí nebo určité znevýhodnění studentů, kteří odpoví na testové položky špatně, tedy penalizace za chyby. Princip zvýšení obtížnosti může fungovat ve formě přiřazení většího množství distraktorů k testové položce, například pět až šest nebo zvýšení počtu správných odpovědí u každé položky. Nevýhodou této metody je vyšší obtížnost vyplňování testu a větší časová náročnost. Zároveň klade vyšší nároky na tvorbu a kvalitu distraktorů (6).

Na principu penalizace funguje odečítání bodu za nesprávnou odpověď. Obvykle je místo jednoho získaného bodu jeden bod odečten. Tento způsob je označován, jako Negative Marking (NM). Student, pokud nezná správnou odpověď, má při použití tohoto způsobu hodnocení dvě možnosti. Buď riskovat odečet bodu a zkusit odpověď tipovat, nebo odpověď vynechat a nezískat žádný bod.

Druhou možností, které se věnujeme v našem výzkumu, je použití hodnocení s mírou jistoty (Certainty-Based Marking, CBM), které pracuje také s odečítáním bodu nebo jeho části na základě označení míry jistoty správné odpovědi samotným studentem (7).

Podle Chrásky (1) může být přínosem použití tzv. „korekce na hádání“, což je také jednou z metod fungujících na principu penalizace. Ta je využitelná za předpokladu, že se student, který hádá, dopouští většího množství chyb než ten, kdo úlohy skutečně řeší. V těchto případech díky korekci odečítáme určité procento správných odpovědí z výsledného skóre podle celkového počtu odpovědí špatných. Autor také uvádí, že při využití korekce na hádání je nutno testované osoby na tuto skutečnost upozornit, aby raději nechali položku nezodpovězenou, než aby tipovali. Tento systém může být nespravedlivý ke studentům, kteří odpověď netipují, ale dochází u nich k prostému chybování a odečítají se jim další body navíc.

Testování s mírou jistoty

Námi vybraná a prezentovaná varianta pro zkvalitnění testování vědomostí s omezením tipování je využití počítačově vyhodnocovaných testů s mírou jistoty (Certainty-Based Marking, CBM). Zapojení míry jistoty je možné do testů s dichotomickými úlohami (pravda, nepravda), přiřazovacími úlohami, uspořádávacími úlohami a úlohami s výběrem odpovědí. Tato metoda není vhodná pro všechny druhy otevřených testových úloh z důvodu limitace hodnocení úloh počítačem. Hodnocení s touto korekcí je možné využít při testování vstupních, průběžných i výstupních znalostí studentů ve většině oborů.

Předností testů s mírou jistoty je podpořit znalosti studentů a eliminovat tipování odpovědí při testech s výběrem možností hodnocených počítačem. Test je nejvhodnější pro znalosti na úrovni prvních dvou okruhů Bloomovy taxonomie (pamatovat, pochopit). Je ale samozřejmě možné sestavit test i pro vyšší stupně Bloomovy taxonomie a použít u nich korekci za pomoci míry jistoty. Výhodou je, že test zahrnuje sebehodnocení a tím rozvíjí schopnosti studentů v oblasti sebereflexe. Samotné uvědomění si vlastních nedostatků ve znalostech vytváří příležitost pro jejich doplnění. Někteří studenti mohou tipovat i nevědomě. Označí správnou odpověď v testu a sami si často myslí, že znají správnou odpověď, přitom většinou tipují nejpravděpodobnější možnost

a věří ve štěstí. Úspěšné tipování nemá přímou souvislost se znalostí, není tedy vhodné za to studenty hodnotit (7).

V testech s CBM vypadá každá položka vizuálně stejně, jako při běžném nastavení testu, ale student navíc formou sebehodnocení označí, nakolik si je se svou odpovědí jistý a dochází k bodové úpravě za zodpovězenou položku. Ukázka testové dichotomické úlohy je na obrázku 1.

Systém hodnocení je shrnut v tabulce 1. Míra jistoty, kterou student označí, je převedena na procentuální hodnotu a bod za správnou odpověď na položku je touto hodnotou vynásoben. Procentuální hodnoty přičítaných či odečítaných bodů vychází z výzkumů autora Gardner-Medwin používaných na University College London pro hodnocení znalostí studentů medicíny a biomedicíny (7, 8). Výsledek každé zodpovězené položky je modifikován dle zaškrtnuté jistoty studenta, zda označil správnou odpověď. Možnost korekce výsledku testů pomocí CBM je dostupná v programu Moodle od verze 2.1 a LAPT (London Agreed Protocol for Teaching) (9). Jednotlivé stupně jistoty jsou v české verzi označovány jako „velmi jistý“, „docela jistý“ a „ne moc jistý.“ V případě chybné odpovědi jsou odečítány body pouze, pokud si je student *velmi jistý* nebo *docela jistý* chybnou odpovědí. Student není veden k přeskačování odpovědí, ale uvědomění si, že odpověď není jistě. Proto při chybné odpovědi s označením *ne moc jistý* nedochází k odečítání bodů. Z tabulky 1 ale jasně vyplývá značná ztráta bodů při označení chybné odpovědi, kterou si je student *velmi jistý*. Tato srážka

Tabulka 1 Shrnutí hodnocení odpovědí testů s mírou jistoty

Stupeň jistoty	Ne moc jistý	Docela jistý	Velmi jistý	Bez odpovědi
Hodnocení při správné odpovědi	33 %	67 %	100 %	0 %
Penalizace za chybu	0 %	– 67 %	– 200 %	0 %

Obrázek 1

Ukázka položky z testu s korekcí výsledku mírou jistoty

Os scaphoideum a os lunatum patří mezi distální řadu ossa carpi.

Vyberte jednu z nabízených možností:

☐ Pravda

☒ Nepravda ✓

Jak jste si jistý? ☐ Ne moc (méně než 67%) ☐ Docela (více než 67%) ☒ Velmi (více než 80%)

Správná odpověď je 'Nepravda'

vychází z filozofie, že není možné si dovolit dělat nevědomé chyby. Nevýhodou CBM je ovšem potřeba vyšší vyspělosti zkoušených, kteří by měli být schopní kritického zhodnocení vlastních znalostí (7).

Příklad z praxe studentů může být ukázkou přínosu tohoto systému testování. Pokud si student není jistý se svou chybou, má možnost se v praxi zeptat a ověřit si svůj postup. V případě, že si je student *velmi jistý*, že v 1ml Heparinu je 500 m.j. a chtěl by tak podat pacientovi desetinásobek dávky, je srážka dvou bodů v rámci teoretické přípravy namístě. Při ústní zkoušce by byl student také hodnocen „nevyhověl“. Pokud by si však student v průběhu přípravy léčiva uvědomil, že si není jistý správností svého postupu, mohl by si svůj předpoklad včas ověřit. Stejně tak je důležité s jistotou umět látku základních předmětů. Například aby si nebyl student *velmi jistý*, že „*sinister*“ znamená „*pravá*“ a nepletl si tedy stranové operace u pacientů.

CÍL PRÁCE

Porovnat vhodnost metody testu s mírou jistoty oproti testu bez korekce při hodnocení znalostí studentů zdravotnických oborů.

Porovnat vhodnost metody testu s mírou jistoty oproti korekci pomocí negative marking při hodnocení znalostí studentů zdravotnických oborů.

METODIKA

Výzkumný soubor tvořilo 152 studentů nelékařských zdravotnických oborů. Všichni studenti byli z prvních ročníků bakalářských studijních oborů Všeobecná sestra, Zdravotnický záchranář nebo Porodní asistentka. Dva studenti nevyplnili v testu všechny odpovědi a nejsou zahrnuti ve výsledcích z důvodu nemožnosti výpočtu citlivosti testu.

Pilotáž předcházející předkládaným výsledkům probíhala na souboru 28 studentů stejných oborů metodou testu a re-testu s 20 položkami. Každý ze studentů tedy vyplnil test dvakrát v časovém odstupu 15 minut. Zvolena byla forma testu se čtyřmi možnostmi a jednou správnou odpovědí. Pilotáží byla ověřena nezbytnost edukace před psaním testu s CBM. Mezi testem a re-testem proběhla edukace o systému CBM, studenti před druhým testem neměli možnost doplnit si znalosti. Následně psaný stejný test (s jiným pořadím úloh) koreloval s prvním testem sice statisticky významně ($p = 0,05$), ale korelační koeficient Spearmanovy neparametrické korelace činil pouze $r = 0,677$. Dále byl po pilotáži zvýšen počet úloh z 20 na 40 a byl změněn charakter testu na dichotomické úlohy s cílem zjednodušit vyplňování testu a připravit odpovědi více jednoznačné. Pilotáž probíhala v prosinci 2014.

Samotný výzkum probíhal pomocí znalostního testu při průběžném hodnocení v předmětu anatomie. Studenti test vyplnili v období ledna 2015. Všem studentům byla látka obsažená v testu odpřednášena při běžné výuce v průběhu předchozího semestru. S korekcí pomocí míry jistoty byli studenti předem seznámeni a měli možnost si vyzkoušet ukázkový test se 20 úlohami, které nebyly obsaženy v hodnoceném testu. Tuto možnost využilo 132 studentů z výzkumného souboru.

Všichni zkoušení podstoupili počítačem hodnocený test s dichotomickými úlohami doplněný o míru jistoty odpovědi u každé položky. Test pro každého studenta byl stejný a obsahoval celkem 40 položek. Jednotlivé úlohy byly mezi sebou zamíchány (jiné pořadí úloh). Samotné úlohy byly zobrazovány v zabezpečeném okně. Studenti měli možnost volného pohybu mezi úlohami. Čas pro vyplnění testu byl stanoven na 30 minut. Časový limit vychází z pilotáže.

Testové úlohy byly vytvořeny z celého průřezu učiva přednášejícími předmětu. Následně proběhla oponentura otázek týmem čtyř odborníků za účelem odhalení a odstranění nedostatků testových úloh. Samotné testové otázky byly koncipovány tak, aby pro studenty nebyly příliš náročné. Test obsahoval šest jednoduchých motivačních úloh, které měl s jistotou znát každý student, za účelem testované motivovat k dalšímu vyplňování testu. Celkem bylo vytvořeno 200 testových úloh, ze kterých počítač před započítáním testu podle předem stanovených kritérií vybral 40 položek.

Před spuštěním byl test zabezpečen heslem a byl zpřístupněn na intranetu těsně před začátkem. V průběhu byli studenti kontrolováni dozorem, aby bylo vyloučeno napovídání či používání nedovolených pomůcek. Žádný ze studentů v průběhu testu neopustil své místo.

Klasifikace vycházela z hodnocení na základě procenta správných odpovědí podle návrhu Sedláčkové (10), tzv. přísné klasifikace. Místo hodnocení známkou jedna až pět jsme hodnotili pouze test, jako splnil a nesplnil. V rámci procentuálního hodnocení odpovídá 100 % až 70 % správných odpovědí úspěšnému splnění a 69,9 % a méně neúspěchu.

Výsledek testu bez korekce pomocí CBM byl exportován ze stejného testu. Počítač také převedl výsledky na hodnocení pomocí odečtu bodu, tzv. Negative-Marking (NM), kde se za správnou odpověď přičítá jeden bod a za chybnou se jeden bod odečítá. Výsledky byly ve všech případech převedeny na procentuální hodnocení výsledků testu.

Hodnota obtížnosti testové položky Q určuje procentuální část počtu testovaných, kteří úlohu řešili

chybně nebo ji vynechali (2). Byla počítána podle vzorce

$Q = 100 \left(\frac{n_N}{n} \right)$, kde n_N je počet žáků, kteří řešili úlohu chybně a n je celkový počet žáků. Tato hodnota byla tím vyšší, čím obtížnější testová položka byla (11). Koeficient citlivosti ULI slouží k výpočtu schopnosti úlohy rozlišovat mezi studenty s horšími a lepšími vědomostmi. Byl počítán dle vzorce

$d = \left(\frac{n_L - n_H}{0,5n} \right)$, kde d vyjadřuje velikost koeficientu ULI, n_L je počet osob ze skupiny žáků s lepšími výsledky, kteří úlohu řešili správně a n_H je počet osob ze skupiny žáků s horšími výsledky, kteří úlohu řešili správně (2, 12). Za účelem ověření reliability jsme použili výpočet Cronbach alfa.

VÝSLEDKY

Test s mírou jistoty při průběžném hodnocení znalostí anatomie splnilo na stanovených 70 % pouze 27 ze 150 studentů. Stejný test při přepočtu počítacem na hodnocení Negative-Marking by splnilo 48 studentů. Při ponechání dichotomického testu bez jakékoliv korekce by ho splnilo 120 studentů. Průměrný výsledek testu CBM byl 45,1 % s mediánem 49,2 %, minimum -35,0 % a maximum 97,5 %. Rozložení výsledku všech testů je zobrazeno na histogramu (Graf 1). Průměrná délka vyplňování testu byla 20,1 minut.

Graf 1 Výsledky testu CBM v intervalech po 20 % bodů

Průměrná hodnota obtížnosti testových položek Q dosahovala hodnoty 22,3 s minimem 0 až po maximum 48. Takto nízká úroveň je způsobena volbou jednoduché formy dichotomických úloh. Do testu bylo zapojeno šest jednoduchých motivačních úloh. Hodnota obtížnosti Q u těchto úloh byla nižší, než pět.

Podle koeficientu citlivosti ULI bylo označeno 23 úloh, jako dostatečně citlivých pro rozlišení mezi studenty s lepšími a horšími znalostmi. Šest úloh bylo motivačních a podle očekávání neměly dostatečnou citlivost na rozlišení studentů, kteří látku uměli a netipovali ji. Zbývajících jedenáct úloh bylo pod hranicí citlivosti pro rozlišení studentů a pro další testování nebudou použity. Hranice byla stanovena na $d \geq 0,25$ při obtížnosti $Q = 30 - 70$, pro úlohy s obtížností $Q = 20 - 30$ byla hranice $d \geq 0,15$ (12).

Porovnání výsledků jednotlivých alternativ korekce na hádání je graficky znázorněno v grafu 2. Při psaní nejjednodušší verze testu (dichotomické úlohy) je zřejmá dobrá stratifikace odpovědí studentů podle počtu neuvědomělých chyb nebo od tipovaných úloh u CBM. Systém hodnocení NM odlišoval studenty bez dostatečných znalostí také dobře, ale vede studenty k přeskakování úloh bez označení dopovědi.

Bylo ověřeno, že není možné používat dichotomický druh testových úloh bez jakékoliv korekce na hádání při absolutní standardizaci hodnocení (fixní hranice splnění testu na 70 %). I studenti s nejhorskými znalostmi měli při tomto způsobu hodnocení úspěšnost kolem 50 %, což odpovídá téměř náhodnému vyplnění bez znalostí správných odpovědí.

Graf 2 Porovnání jednotlivých alternativ dichotomického testu

Odhad vnitřní konzistence položek testu byl počítán pomocí Cronbachova alfa. Skupina studentů byla z pohledu přednesené látky homogenní, jednalo se pouze o studenty prvních ročníků. Prostředí pro psaní testu bylo studentům známé. Testovaná látka byla také homogenní. Výsledek Cronbach alfa byl 0,7, položky testu byly dostatečně vnitřně konzistentní.

DISKUZE

Průměrná hodnota obtížnosti testových položek Q dosahovala 22,3 a Chráska (1) doporučuje minimálně 20, jinak jsou úlohy příliš snadné. Ideálně by se hodnota obtížnosti měla pohybovat kolem 50. Přestože testové úlohy byly záměrně tvořené s nízkou obtížností, při hodnocení podle CBM se ukázalo, že nejvíce studentů (50 %) se pohybovalo v rozmezí úspěšnosti 27,5–63,3 %. Při použití CBM se průměrná úspěšnost v testu změnila ze 77,8 % na 45,1 %. I přes změnu principu bodového hodnocení měli stále skutečně úspěšní studenti vysoké procentuální ohodnocení. Méně úspěšná byla po přepočtu část studentů, která se při hodnocení bez korekce pohybovala mezi dolním a horním kvantilem procentuálního vyjádření výsledků.

V našem výzkumu se ukázalo hodnocení pomocí metody NM jako mírnější než CBM navzdory předpokladu, že studenti, kteří si odpovědi nebudou skutečně jistí, označí při CBM možnost docela jistý a tím při špatné odpovědi ztratí pouze 67 % bodu místo 100 %, které by ztratili za použití NM.

Potvrdilo se, že před použitím testu s korekcí CBM je nutné studenty dobře seznámit s principem hodnocení testu, dát jim možnost si způsob vyplňování úloh předem vyzkoušet např. zpřístupněním ukázkového testu (13).

Velice důležitý aspekt, který ovlivňuje výsledek testu, je u testovaných schopnost adekvátního sebehodnocení. Metoda CBM předpokládá určitou vyspělost studentů, kteří by neměli své znalosti nadhodnocovat, ani podhodnocovat. Právě nižší sebedůvěra, víra ve své znalosti a s tím spojené podhodnocování by mohlo způsobit neúspěch u některých studentů, kteří by například měli všechny otázky zodpovězeny správně, ale označili by všude možnost docela jistý a tím vždy získali pouze 67 % bodu.

Jako vyvážené hodnocení, které eliminuje vliv podhodnocování znalostí, se jeví snížení přísnosti klasifikace pro splnění testu o 4 % na 66 %. Po tomto snížení by i ti studenti, kteří by měli všechny odpovědi správně, ale všude by označili míru jistoty slovy docela jistý, uspěli. V našem výzkumu by po snížení hranice uspělo o 7 studentů více.

Další důležitou částí plánování testu s použitím míry jistoty je rozumné rozvržení počtu položek v testu. Při nižším počtu (např. deset) je pro studenta nemožné v testu uspět při jedné chybné odpovědi, se kterou si je jistý. Místo získání deseti procent se mu 20 % odečítá. Je tedy ihned na hranici 70 %. Je vhodné nastavit minimální počet položek v testu na 40. Neoptimálnější strategie vyplňování při tomto počtu je velmi jistě odpovědět na 28 položek a na zbý-

vajících 12 položek odpovědět ne moc jistě. Eliminuje se tím riziko odečítání bodu a student si uvědomí, které oblasti látky opravdu neumí suverénně. Při šesti chybných velmi jistých odpovědích (a 34 správných odpovědích) je student na hranici splnění testu (70 %).

V současné době je nejpreferovanější alternativou bodové korekce, která snižuje možnost tipování, zvýšení obtížnosti samotných úloh. Vyšší obtížnosti docílíme použitím většího množství distraktorů nebo použitím několika správných odpovědí. I tyto metody mají svá negativa.

O nevýhodách použití prostého testu s jednou správnou možností (ze šesti) v každé úloze píše i Kubitinger (6). V jeho studii proběhlo porovnání oproti testu se dvěma správnými možnostmi a třemi distraktory. Ač studie proběhla u žáků ZŠ, generalizuje své závěry a nedoporučuje používání jednoduchých testů s jednou správnou možností. Pro dostatečné zvýšení obtížnosti samotných úloh je podle jeho výsledků třeba použít koncept alespoň dvou správných odpovědí a tří až čtyř distraktorů. Nevýhodou v tomto případě je náročnost tvorby úloh a časová náročnost čtení textu úlohy samotným studentem.

Poslední možností, jak lze zvýšit náročnost testu a tím zabránit tipujícím ve splnění, je větší přísnost klasifikace, například zvýšení požadované úspěšnosti na 80 %. Tato metoda je však nejméně vhodná, protože znevýhodňuje především studenty, kteří netipují, ale přitom více chybují. Otázkou zůstává, zda se z pedagogického hlediska snažit eliminovat tipování v testech. Vaughn (14) ve své studii na čtyřech experimentech nepotvrzuje lepší výsledky studentů učících se metodou pokus-omyl, tedy opakovanými pokusy o splnění testu, namísto klasického studia látky.

U testů CBM je náročnější výpočet reliability. Nejjednodušší a nejpoužívanější Kuderův-Richardsonův vzorec není možné použít, protože studenti nejsou stejně hodnoceni za stejnou odpověď (15). Z důvodu přidělování pouze částí bodu je nutné provádět výpočty metodou půlení (split-half) nebo s použitím koeficientu Cronbachova alfa (1, 16, 17).

Díky míře jistoty je možné do testu dát jednoznačné otázky s jednou správnou odpovědí a zabránit možnosti tipováním uspět v testu. Vyhne se tím nejasným a složitě koncipovaným testovým položkám, které je někdy nutné použít u testů s více možnými odpověďmi. Gardner-Medwin (7) publikoval zkušenosti při testování dichotomických CBM na 345 studentech medicíny na Imperial College. Gardner-Medwin (7) shrnul ve své práci mýty o testech s CBM. Popisuje možnost použití mimo obory zdravotnictví, s různými typy testů, a o nemožnosti vyhrát nad systémem bez znalosti testované látky.

V práci jsme z nich vycházeli. Ve studii Bereby-Meyer (18) na amerických a izraelských studentech se čtyřmi experimenty v téměř identické podobě CBM měli 27 % přeskočených odpovědí, aby studenti neztratili body za chybnou odpověď. V našem šetření pouze dva studenti přeskočili odpovědi, ostatní využili hodnocení míry jistoty jako ne moc jistý a tím neztratili žádný bod za uvědomělou chybnou odpověď. Karandikar (19) ověřoval vhodnost testů s jednou správnou odpovědí oproti testu se zapojením srážky bodů při chybné odpovědi. V jeho studii byla použita srážka pouze 33 % bodu a nebylo dosaženo signifikantního výsledku. Proto mají CBM dle doporučení autora Gardner-Medwina (7, 8) stanovenou srážku bodu na 67 % z jednoho bodu a 200 % z hodnoty bodu tj. dva body.

LIMITY VÝZKUMU

Nízká úspěšnost v testu může být způsobena tím, že šlo o průběžné hodnocení, které nebylo samo o sobě předpokladem pro pokračování v navazujícím semestru. Studenti tedy nebyli dostatečně motivováni k přípravě na test. Jednalo se o první test, který studenti psali na vysoké škole. Zjištěné výsledky mohou být ovlivněny předchozí zkušeností studentů ze středních škol (možnost setkat se s elektronickými testy, obsahová náročnost učiva).

ZÁVĚR

Na první pohled se může zdát, že hodnocení pomocí míry jistoty je ke studentům příliš přísné, ale jestliže by studenti testovanou látku skutečně ovládali, byli by úspěšní. Velkým přínosem oproti jiným korekčním mechanismům na hádání je větší stratifikace výsledků podle míry neuvědomělých chyb a otipovaných odpovědí. Systém je vhodný především při velkém počtu studentů k hodnocení klíčových znalostí a zároveň slouží jako sebehodnotící a sebevzdělávací nástroj. Při použití testů s mírou jistoty došlo k významnému snížení úspěšnosti studentů, kteří nebyli na test vědomostně připraveni. Nadpolovičního procentuálního výsledku při odstranění korekce na hádání dosáhli studenti tipováním, které je pro dichotomické úlohy velkou limitací. Edukace o systému bodování v rámci CBM se ukázala jako velmi významná, což dokládá provedená pilotáž v našem šetření.

REFERENČNÍ SEZNAM

1. Chráška M. Didaktické testy. Brno: Paido; 1999.
2. Chráška M. Metody pedagogického výzkumu. Praha: Grada; 2007.
3. Štuka Č, Martinková P, Vejražka M, Trnka J, Komenda M. Testování při výuce medicíny: konstrukce a analýza testů na lékařských fakultách. Praha: Karolinum; 2013.
4. Slavík M. Vysokoškolská pedagogika. Vyd. 1. Praha: Grada publishing; 2012.
5. Karp J, Šafář Z. Kvizový způsob testování odborných znalostí středoškoláků při přijímacích zkouškách na lékařské fakulty. Sociologický Časopis. 1965;1(4):424–430.
6. Kubinger KD, Holocher-Ertl S, Reif M, Hohen-sinn Ch, Frebort M. On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. International Journal of Selection and Assessment. 2010;18(1):111–115.
7. Gardner-Medwin T, Curtin NA. Certainty-Based Marking (CBM) For Reflective Learning And Proper Knowledge Assessment. From the REAP International Online Conference on Assessment Design for Learner Responsibility [Internet]. 2007 May 29-30; Available from: http://www.ucl.ac.uk/lapt/REAP_cbm.pdf
8. Gardner-Medwin AR, Gahan M. Formative and Summative Confidence-Based Assessment, Proc. 7th International Computer-Aided Assessment Conference, Loughborough, UK, July 2003. p. 147–155.
9. Nozica B, Carapina M, Stanicic O. Certainty based marking: The use of modified grading schema in multiple choice quizzes. In: 6th International conference of education, research and innovation; 2013 Nov; Seville, Spain, p. 2160–2165.
10. Sedláčková J. Diagnostické metody ve vyučování matematice. Olomouc: Univerzita Palackého; 1993.
11. Koníček L, Malčík M, et al. Hodnocení výsledků vzdělávání – praktická část. Ostrava: Ostravská univerzita v Ostravě; 2007.
12. Jeřábek O, Bílek M. Teorie a praxe tvorby didaktických testů. Olomouc: Univerzita Palackého; 2010.
13. Cisar SM, Cisar P, Pinter R. True/false questions analysis using computerized Certainty-Based Marking tests. Intelligent Systems and Informatics, . SISY ,09. 7th International Symposium; 2009 Sep 25–26; p. 171–174, doi: 10.1109/SISY.2009.5291171.

14. Vaughn KE, Rawson KA. When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review*. 2012;(19): 899–905. doi 10.3758/s13423-012-0276-0.
15. Hrabal V, Lustigová Z, Valentová L. Testy a testování ve škole. Praha: SVI PedF UK. 1992;
16. Komenda S, Zapletalová J. Analýza didaktického testu a její počítačová podpora. Olomouc: Lékařská fakulta UP: 1996.
17. Chráška M, Kočvarová I. Kvantitativní design v pedagogických výzkumech začínajících akademických pracovníků. Zlín: Academia centrum; 2014.
18. Bereby-Meyer Y, Meyer J, Flaschner O M. Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making* [Internet]. 2002;15(4):313–327. doi: 10.1002/bdm.417. Available from: <http://doi.wiley.com/10.1002/bdm.417>
19. Karandikar RL. On multiple choice tests and negative marking. *Current Science*. 2010;99(8): 1042–1045.

KONTAKT NA HLAVNÍHO AUTORA

Vít Blanař, Mgr.
Fakulta zdravotnických studií
Univerzita Pardubice
Studentská 95
CZ-532 10 PARDUBICE
vit.blanar@upce.cz