

Fórum:Testy2/Klasifikace

Klasifikace, čili oznámkování výkonu studentů v testu, je pokračováním či rozšířením standardizace testu. Konstrukce klasifikační stupnice, respektive nastavení relace mezi výkonem v testu a klasifikačním stupněm, je **jediný subjektivní prvek, který do celého testování vstupuje**. Je mu tedy třeba věnovat náležitou pozornost **Cite error: Invalid <ref> tag; invalid names, e.g. too many**.

Pro nastavení relace mezi výkonem v testu a klasifikačními stupni je nutno si ujasnit, co je vlastně v daném případě smyslem vysokoškolského vzdělávání.

První z možných pohledů nahlíží na vysokoškolské medicínské vzdělávání jako na šestiletý test inteligence, který třídí jednotlivce podle intelektuálních schopností a pracovních návyků. Tento přístup odráží zájem potenciálních zaměstnavatelů vybrat nejvhodnější kandidáty na omezený počet míst a pomáhá zajistit, aby na klíčová místa byli vybráni intelektuálně nejschopnější. Při studiu staví tento přístup studenty proti sobě, nechává je mezi sebou soutěžit.

Druhý pohled je velmi odlišný. Předpokládá, že smyslem vzdělávání je osvětit, posilovat a socializovat občany. Pedagog by se podle tohoto pohledu neměl tolik soustředit na rozřazování studentů podle schopností, ale pomoci jim najít správnou představu o světě a sobě samotných s cílem vybavit je znalostmi, intelektuálními nástroji a návyky, které z nich učiní informované a kulturně gramotné členy společnosti.

Tyto dva pohledy na vzdělávání mají své paralely ve dvou hlavních přístupech ke klasifikaci.

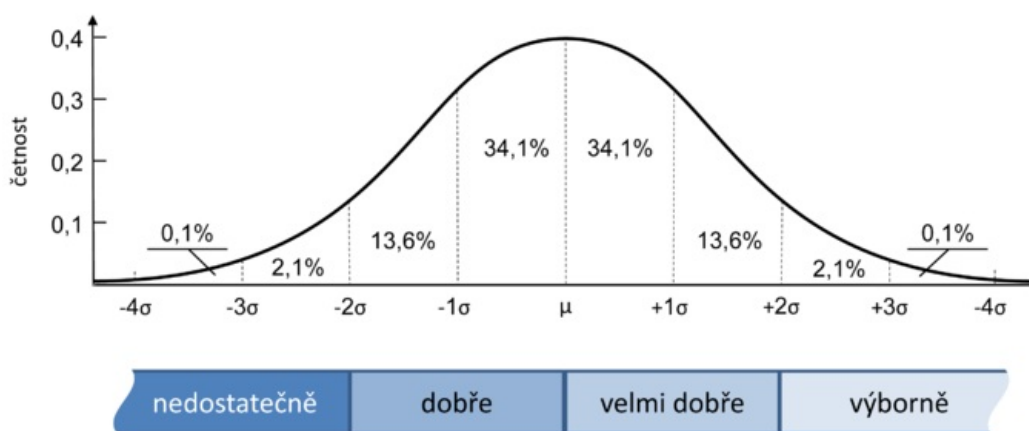
Klasifikace porovnávající výkonnost ve skupině (relativní klasifikace)

Prvním přístupem je klasifikace založená na **relativním výkonu ve skupině**. Stanovuje známku jako funkci pořadí studenta v rámci určité skupiny. Pomocí testu **vytvoříme pořadí studentů a tím rozdělíme klasifikační stupně** podle předem stanovených procentuálních mezí. Relativní klasifikace je založena na předpokladu, že výkonnost všech studijních skupin (napříč prostorem a časem) je v zásadě stejná. Z pohledu studenta obsahuje tento způsob klasifikace zjevnou nespravedlivost, protože hodnocení nezávisí jen na výkonu studenta, ale i na výkonech ostatních. Je tedy možné, že se stejnou mírou znalostí by student byl v jednom roce klasifikován lépe než v roce jiném.

Chceme-li použít hodnocení založené na relativní standardizaci, je třeba učinit dvě rozhodnutí. Nejprve je třeba stanovit, jaký klasifikační stupeň přiřadíme k průměrnému výkonu. Pro často používaný pětistupňový klasifikační systém A, B, C, D, F (1, 2, 3, 4, 5) můžeme intuitivně zvolit C jako klasifikaci odpovídající průměrnému výkonu; není to však jediná možnost.

Dále je nutné rozhodnout o hranicích oddělujících jednotlivé klasifikační stupně. Narazíme při tom na otázky typu „kde je hranice mezi mírně a silně nadprůměrným?“ Rozhodnutí je na zkušenosti a úsudku. Pomoci však mohou obecně platná doporučení.

Pro stanovení konkrétních známek se používá např. z-skór nebo percentilové pořadí podobným způsobem, jako je popsáno v kapitole Relativní standardizace testu. Při čtyřstupňové klasifikační stupnici pak hranicím mezi jednotlivými klasifikačními stupni odpovídají např. z-skóry -2, 0, 2, jak je naznačeno na obrázku 7.7:



Obr. 7.7 Příklad klasifikace výsledku studenta v testu při použití relativní standardizace.

Skupina se rozdělí podle dosaženého výsledku a oznamuje podle příslušné klasifikační stupnice – v tomto případě čtyřstupňové.

Klasifikace založená na kritériích (absolutní klasifikace)

Naproti tomu klasifikace založená na kritériích měří úspěšnost studenta **ve vztahu ke kritériím** vyžadovaným pro dosažení toho kterého klasifikačního stupně. Na rozdíl od normativní klasifikace není předem stanoveno rozdělení klasifikačních stupňů a studentovo oznamkování není ovlivněno výkonností ostatních. Pokud bychom teoreticky zkoušeli výrazně nadprůměrnou skupinu studentů, mohou všichni dostat dobré známky, a naopak, pokud by se náhodou sešla velmi slabá skupina studentů, nemusí dobré známky dostat nikdo.

Cílem klasifikace založené na kritériích je zhodnotit objektivní výsledky studenta nezávisle na výsledcích ostatních studentů ^[1].

- Hodnocení studentů podle **relativní standardizace** nám umožňuje zjistit, kteří studenti se naučili nejvíce, ale nezjistíme, jestli se studenti ve skutečnosti naučili dost.
- Naopak při hodnocení podle **absolutní standardizace** zjistíme, zda-li se studenti naučili dost, ale nemusíme být schopni určit, kteří studenti se naučili nejvíce ^[2].

Obě metody jsou ve skutečnosti do jisté míry provázané, protože při relativní klasifikaci jsou implicitně zahrnuta i absolutní kritéria, která měli na mysli pedagogové připravující test.

Nejjednodušším způsobem klasifikace je stanovit, kolik procent z celkového počtu bodů je třeba k dosažení určité klasifikace. Například pro známku A požadovat 90 % a více, pro B 80–90 % a tak dále. Problém tohoto přístupu je v arbitrárním nastavení hranic jednotlivých klasifikačních stupňů. Proto je vhodné u důležitějších zkoušek nastavit hranice pomocí expertního odhadu, například podle Angoffovy nebo Ebelovy metody.

Někdy je na závalu, že klasifikace neodráží různorodou obtížnost otázek. Studenti pak mohou dosáhnout vysokého počtu bodů i tím, že správně odpoví na mnoho snadných, méně důležitých otázek.

Propracovanější forma absolutní klasifikace proto rozlišuje mezi různými typy či úrovněmi znalostí a dovedností, které student prokazuje na různých úkolech. Větší důraz se klade na ty z nich, které odrážejí vyšší úroveň osvojení látky. Tento přístup bere v úvahu jak množství látky, tak úroveň její kognitivní komplexity. Můžeme například vzdělávací cíle svého kurzu rozdělit do dvou skupin: na základní a pokročilé. Základní cíle se týkají minimálních nezbytných znalostí a dovedností, které si studenti musí osvojit. Pokročilé cíle naproti tomu představují vyšší úroveň dovedností, jako použití kritického myšlení, řešení komplexních problémů a podobně.

Pro zjištění, nakolik se podařilo dosáhnout základních a pokročilých výukových cílů, může být přinejmenším pro začátek jednodušší použít dva zcela oddělené testy. Zjednoduší se tím vyhodnocení zkoušky a uchovávání záznamů o ní. Při oddělení testů je snazší se zaměřit na jednotlivé cíle výuky a zpracovat pro ně testové otázky. Pro posouzení základních cílů výuky to bývá poměrně snadné. Posuzování, nakolik se podařilo dosáhnout pokročilých cílů výuky, bývá obvykle obtížnější, neboť je těžší vymyslet testové otázky postihující i schopnost nabyté vědomosti aplikovat.

Pro absolvování obou druhů testu lze nastavit odlišné požadavky na výkon studentů, jak je naznačeno v tabulce 7.8.

Tab. 7.8 Příklad možného nastavení absolutní standardizace pro klasifikaci základního a pokročilého testu v pětibodové klasifikační stupnici.

Klasifikační stupeň	Základní test	Pokročilý test
A	90 % nebo více	85 % nebo více
B	90 % nebo více	75–84 %
C	80 % nebo více	60–74 %
D	80 % nebo více	50–59 %
F	méně než 80 %	méně než 50 %

V uvedeném příkladu požadujeme, aby studenti prokázali zvládnutí alespoň 80 % základních vzdělávacích cílů a 50 % pokročilých cílů. Pokud požadujeme, aby nastavení hranic úspěšnosti bylo objektivnější, můžeme použít některou z metod expertního odhadu popsaných výše.

Klasifikace v praxi

V kapitole, která se věnuje praktické klasifikaci (znamkování) studentských výsledků v testu, se omezíme na diskuzi o testech s uzavřenými položkami s výběrem odpovědi (*multiple choice*). Klasifikace dalších typů testových položek je diskutována v odborné literatuře (např. ^[3] nebo ^[4]).

Většina zahraničních univerzit má vypracovaná standardní klasifikační schémata, která pak umožňují srovnávání výsledků jednotlivých studentů uvnitř i mezi obory. Výsledek daného testu či celého souboru písemných i jiných hodnocení je tedy přepočítán na standardní škálu, podle níž jsou potom rozděleny známky.

Jako příklad můžeme vzít klasifikační schémata Univerzity v Edinburgu (<http://www.ed.ac.uk/schools-departments/student-administration/exams/regulations/common-marking-scheme>). Pro pregraduální lékařské obory je relevantní schéma CMS3 (CMS3: Bachelor of Medicine and Bachelor of Surgery) :

Tab. 7.9 CMS3 schéma

Počet bodů	Známka	Popis
90–100	A	Výborně
80–89	B	Velmi dobře
70–79	C	Dobře
60–69	D	Uspokojivě (Pass)

Přepočítávání pak funguje následovně: tvůrci testu pomocí jedné ze standardizačních metod stanoví minimální hranici pro úspěch v daném testu (tzv. *pass mark*) například na 80 % z maximálního možného počtu bodů. Jelikož známka D (pass) v CMS3 odpovídá 60–69 bodům bude pro účely známkování 80 % z maximálního možného počtu bodů odpovídat hodnocení 60 bodů. Pro známku A bude třeba v testu dosáhnout 95–100 % (tedy 90–100 bodů) z maximálního možného počtu bodů, pro B 90–94 %, pro C 85–89 % a pro D 80–84 %. Jinými slovy, nejprve určíme, kteří studenti v testu uspějí, a pak je mechanicky rozdělíme do jednotlivých klasifikačních stupňů. Přepočet lze

matematicky vyjádřit následovně: $Z = 60 + \frac{40}{100 - P} \cdot (p - P)$, kde Z je výsledný počet bodů (z kterého dle CMS3 schématu určíme známku), P je minimální procento nutné k úspěšnému absolvování daného testu (*pass mark*) a p jsou procenta dosažená daným studentem.

Odkazy

Reference

1. JAMES, Richard. *A comparison of norm-referencing and criterion-referencing methods for determining student grades in higher education* [online]. Centre for the study of higher education, ©2002. [cit. 2013-04-16]. <<http://www.cshe.unimelb.edu.au/assessinglearning/docs/AssessingLearning.pdf>>.
2. The University of North Carolina at Chappel Hill. *Grading systems* [online] . Center for Faculty Excellence, 2012. 6 s. Dostupné také z <<http://cfe.unc.edu/pdfs/FYC10.pdf>>.
3. MEZERA, Antonín. *Školní měření a evaluace výsledků vzdělávání ve škole : Studijní materiál pro interní potřebu učitelů základních a středních škol* [online]. [cit. 2012-12-18]. <<http://www.ppppraha7a8.cz/files/zaklady%20skolniho%20mereni.pdf>>.
4. MCLACHLAN, John C a Susan C WHITEN. Marks, scores and grades: scaling and aggregating student assessment outcomes. *Medical Education* [online]. 2000, roč. 34, vol. 34, no. 10, s. 788-797, dostupné také z <<http://doi.wiley.com/10.1046/j.1365-2923.2000.00664.x>>. ISSN 0308-0110. DOI: 10.1046/j.1365-2923.2000.00664.x (<http://dx.doi.org/10.1046%2Fj.1365-2923.2000.00664.x>).