

K METODICE STATISTICKÉHO VYHODNOCENÍ TESTŮ ZNALOSTÍ Z FYZIKY U STUDENTŮ LÉKAŘSKÝCH FAKULT V ČR

Běláček J, Komarc M, Kyplová J

Ústav biofyziky a informatiky 1.LF UK Praha

Formální vyhodnocení znalostních testů má řadu aspektů, z nichž některé lze zvládnout standardními statistickými nástroji, ale jiné zasluhují specifickou pozornost i speciální metodické a SW ošetření. Cílem toho příspěvku je shrnout metodiku použitou na ÚBI 1.LF UK Praha při zpracování výsledků testů z fyziky u studentů 1. ročníku čtyř lékařských fakult v ČR - viz abstrakt Kyplová J a kol. (2013) tohoto Sborníku - a vyhodnotit zkušenosti získané při analýze systému všech 30 položek použitého testu z pohledu možnosti jeho dalšího eventuálního zhodnocení.

Standardní analýzu výsledků můžeme založit na porovnání průměrů souhrnných skóre podle vhodných třídících skupin (fakulta, pohlaví, typ střední školy nebo kraj, kde student maturoval) prostřednictvím ANOVA. Vnitřní konzistenci testovacích baterií 30 otázek a 8 tématických oblastí (obecně položek) testu jsme ověřovali variantním použitím Cronbachova alfa (CA), prostředky explorativní a konfirmativní faktorové analýzy (FA) a několika IRT modely (Item Response Theory Models).

Vedle transparentních a standardně interpretovatelných komparací výběrových průměrů (testy a LSD post-hoc testy ANOVA) jsme analýzou korelačních matic a výsledků FA v datech identifikovali jeden majoritní faktor, který interpretujeme jako „obecné znalosti z fyziky“. Přijatelnou shodu tohoto jednofaktorového modelu s daty (měřeno nejpoužívanějšími indexy: RMSEA = 0,038; Goodness of Fit = 0,97) jsme ale obdrželi až po vyloučení dvou testovaných otázek z baterie s negativní faktorovou zátěží („F2: Kde se nepoužívají radioaktivní látky?“ a H4: „Co měří aneroid?“); hodnota CA se tím zvýšila z původních 0,65 na 0,68 [95%CI 0,65-0,71]. V rámci IRT analýzy byla otestována vhodnost dvou parametrických modelů (1PL resp. 2PL logistický model), které popisují pravděpodobnost správné odpovědi na danou položku jako funkci charakteristiky testovaných osob (úroveň znalostí) a charakteristiky položek (obtížnost, případně také diskriminativnost u 2PL). Výsledky poukázaly na statisticky signifikantní rozdíl mezi 1PL a 2PL modelem – diferenční $\chi^2_{28}(\text{LR}) = 386,6$; $p < 0,001$ - s tím, že 2PL model byl pro popis získaných dat vhodnější. Můžeme tedy tvrdit, že jednotlivé položky se liší v tom, jak dobře diskriminují osoby s vysokou resp. nízkou úrovní znalostí biofyziky.