

## JAK ZAČÍT S TEORIÍ ODPOVĚDI NA POLOŽKU? S POMOCÍ KNIHY „APPLYING THE RASCH MODEL: FUNDAMENTAL MEASUREMENT IN THE HUMAN SCIENCES“

HYNEK CÍGLER

*Katedra psychologie, Fakulta sociálních studií Masarykovy univerzity*

Teorie odpovědi na položku (Item Response Theory, IRT) je souhrnné označení pro moderní psychometrické postupy, které již nejsou výhradní doménou tzv. výkonových testů. Ačkoliv v zahraničí jde o značně rozšířený přístup, v České republice jsou zatím metody standardizované prostřednictvím IRT jen ojedinělými pokusy. Přitom využití IRT (zejména v případě Raschových modelů) neklade zvláštní nároky ani na psychometrika při tvorbě testu, ani na psychologa interpretujícího výsledky – spíše naopak. Za touto situací možná stojí nedostatek domácích textů, které by v krátkosti představily základní výhody a možnosti IRT v podobě srozumitelné pro běžného psychologa bez zvláštního statistického vzdělání. Proto nabízím recenzi na knihu autorů Bonda a Foxové „Applying the Rasch Model: Fundamental Measurement in Human Sciences“ (2001; 2007), která o těchto tématech pojednává velmi přístupnou formou.

V českém jazyce již existují některé základní texty, které teorii odpovědi na položku popisují po technické stránce<sup>1</sup>, pro podrobnější popis však i dnes musí čtenář sáhnout po literatuře zahraniční<sup>2</sup>. Jisté množství logaritmických rovnic a grafů pravděpodobnostních funkcí však běžného čtenáře od další četby spíše odradí, smysl a význam IRT pro standardizaci psychologických i didaktických testů se v nich ztrácí. Důkazem, že logiku IRT testů lze vysvětlit stejně jednoduše (ne-li ještě jednodušeji), jako v případě klasické testové teorie, je kniha Bonda a Foxové (2001; 2007), psaná čtivým, srozumitelným jazykem s názornými ukázkami a zejména konkrétními příklady použití.

Ačkoliv Raschův přístup patří mezi jednoparametrové IRT modely a kniha je vhodným úvodem do celé problematiky, je přesto vhodné mít při četbě na paměti, že mezi Raschovými a ostatními IRT modely existují další nevyřčené rozdíly, a že kniha tak představuje jen výsek problematiky teorie odpovědi na položku<sup>3</sup>.

---

<sup>1</sup> Jako příklad lze uvést knihu Martina Jelínka, Petra Květoně a Dalibora Vobořila „Testování v psychologii: Teorie odpovědi na položku a počítačové adaptivní testování“ (2011), případně pak krátký přehledový článek Tomáše Urbánka a Michala Šimečka (2001).

<sup>2</sup> Osobně doporučuji dvě základní učebnice, v první řadě *The Theory and Practice of Item Response Theory* od de Ayala (2009), případně pak *Item Response Theory for Psychologists* autorů Susan E. Embretson a Stevena P. Reise (2000).

<sup>3</sup> Rozdíl není daný jenom tím, že Raschovy modely zásadně nepracují s tzv. diskriminačním parametrem a s parametrem pseudouhádnutelnosti. Odlišnosti jsou zejména ve filozofickém pozadí, kdy jednoparametrový model je zaměřený na posouzení vhodnosti a přizpůsobení modelu pro konkrétní

Knihla začíná dvěma kapitolami o tom, co je to vlastně měření a proč klasická testová teorie selhává. Současné pak nabízí odlišná filozofická hlediska a východiska, od kterých se logika měření v rámci Raschových modelů odvíjí. Popisují, co to je fundamentální měření<sup>4</sup>, a proč právě Raschovy modely fundamentální jsou. K pochopení přispívají bohaté analogie k přírodním vědám, například historicky dlouhé cestě k měření času a teploty tak, jak je známe dnes. Tyto kapitoly neobsahují vůbec žádnou matematickou teorii, a proto jsou vhodné i pro úplné začátečníky.

Prvnímu vydání (2001) navíc předchází osmnáctistránkový úvod, který velmi přesvědčivě popisuje cestu, kterou se k Raschovým modelům oba autoři nezávisle na sobě dostali. Trevor Bond v něm mimo jiné vypráví o výsledcích, které získal s týmem dalších výzkumníků v roce 1984 po spuštění jednoduchého IRT programu DICOT:

*„... objevili jsme úžasné výstupy. O datech, která jsme důvěrně znali, jsme s použitím Raschova modelu zjistili více, než kolik bychom si mysleli, že může být možné. Za první jsme měli v ruce důkaz jednodimanzionality každého datasetu. Z dat ‚vystoupilo‘ pořadí položek podél dimenzí stejně jako způsob, jakým se mění jejich obtížnost. Dále jsme zjistili, že schopnosti respondentů jsou zobrazeny na stejné jediné škále jako obtížnost položky, a to se stejným smysluplně interpretovatelným měřítkem.“*

Zde je nutno podotknout, že o 30 let později moderní IRT programy, z nichž jeden je přílohou knihy na CD (viz níže), nabízejí ještě rozsáhlejší a podrobnější výstupy.

V následujících třech kapitolách pak autoři stále stejně snadno uchopitelným jazykem popisují základní principy Raschových modelů, způsob, jakým je vhodné konstruovat položky nových testů, nebo jaké požadavky a omezení na data tento přístup klade. Tyto kapitoly jsou velmi přínosné pro úplného začátečníka a pro uživatele psychologických testů, protože obsahují konkrétní návody, jak výstupy Raschových modelů klinicky interpretovat. IRT totiž neposkytuje pouze vážený či odvozený skóre, založený na skóre hrubém, ale o položkách i respondentech toho dovede říci mnohem více: dokáže například určit klinicky významné (rozuměj nepravděpodobné) odpovědi na konkrétní položky a také to, zda respondent celkově odpovídal podle předpokladů, nebo zda je jeho odpověďová matice něčím (a čím) pozoruhodná. Vstupujícími faktory pak mohou být například dílčí znalosti (v případě didaktického testu), nebo nějaké individuální zvláštnosti v osobnosti respondenta (v případě osobnostního testu).

Kapitoly 6, 7 a 8 pak popisují rozšíření a další aplikace Raschových modelů například na Likertovy škály (tzv. Rasch Rating Scale model, RSM), na data s různým počtem možných odpovědí na každou položku, využívaných například při hodnocení esejí (Partial Credit

---

data („exploratorní“ přístup), kdežto Raschův model naopak posuzuje vhodnost dat vůči teoretickému modelu („konfirmatorní“ přístup). Podrobněji viz například de Aylu (2009, s. 19).

<sup>4</sup> Maximálně zjednodušeně je to takové měření, které vytváří poměrové a intervalové škály, a umožňuje tak tvrzení typu „dvakrát větší“, což CTT nedokáže – nelze totiž říct, že percentil 80 (nebo třeba z-skóre 2) znamená dvakrát vyšší úroveň schopnosti oproti percentilu 40 (nebo z-skóre 1).

Rasch model, PCM), a také multifasetová data (ve kterých jsou zvažovány další proměnné nezávislé na respondentovi, jako například přísnost hodnotitele eseje či náročnost vylosovaného tématu).

Velká pozornost je věnována právě hodnocení esejí – dočteme se o příkladech amerických univerzit s přijímacími i průběžnými zkouškami nebo například Eight Grade Writing Testu, který je každoročně administrován všem žákům končících základní školu v Georgii. Univerzity i autoři zmíněné metody shodně využívají IRT vyvažování hodnotitelů a témat a posuzování reliability celé metody.

Další kapitola ukazuje, jak lze popisované metody aplikovat na měření vývoje schopností, učení a rehabilitace. Její poznatky jsou aplikovatelné na řadu oblastí, jako například efektivitu psychotherapeutického působení, nebo analýzu longitudinálních dat včetně způsobů identifikace nepravděpodobných vývojových trajektorií. Kromě toho představuje Saltus model (z latinského „skákat“), vhodný pro posouzení stadiálnosti vývoje, a kvalitativních změn na zdánlivě jednodimenzionální škále. Vše pak demonstruje na příkladu piagetovských úloh, testujících konzervace prostřednictvím přelévání tekutiny mezi sklenicemi (obdobu u nás dostupného testu TEKO).

V desáté kapitole se dočteme o příkladech využití Raschových modelů napříč obory, které stojí již poněkud dále od klasické psychodiagnostiky – například marketingovou analýzu spokojenosti zákazníků, hodnocení studentské spokojenosti na univerzitě (včetně zahrnutí demografických charakteristik do výpočtu), předpovědi výsledků sportovních zápasů, diagnostiku v medicíně či fyzioterapii atd.

Jedenáctá a dvanáctá kapitola je spolu se dvěma přílohami určena zejména psychometrikům, konstruujícím vlastní testy. Předkládá podrobný návod k testování jednodimenzionality a přiléhavosti modelů, konstrukci posuzovacích škál i klasických testů správně–špatně, ano–ne. Poslední kapitola pak nabízí ohlédnutí a zamyšlení nad smyslem testování v Raschově pojetí.

Součástí knihy jsou i dvě přílohy. První obsahuje technické aspekty Raschových modelů, a teprve zde (str. 277) se ve vyšší míře objevují nezbytné matematické vztahy. Druhá příloha je pak spíše návodem, jak začít s Raschovými modely – jmenuje vhodné programy, krok za krokem provádí čtenáře analýzou jejich výstupů a upozorňuje na informace, které je při hodnocení modelů vhodné vzít v potaz.

Nesporným kladem knihy je také přiložený software. IRT analýzy totiž bohužel není možné provádět v běžných statistických programech, jako je například SPSS či Statistica. Psychometrik se proto zpravidla musí spolehnout na jiné komerční programy, jako je třeba Winsteps či Xcalibre, nebo na některé z opensourcových aplikací. V tomto případě s knihou kupujeme však i CD s programy Bond & Fox Steps a Bond & Fox Facets, zjednodušenými verzemi programů Winsteps a Facets. Jsou sice ochuzené o některé funkce, zato však obsahují data k většině příkladů jmenovaných v knize včetně podrobného návodu k provedení analýz. Hned během četby si tak lze veškeré výpočty

Hynek Cígler: Jak začít s Teorií odpovědi na položku?

S pomocí knihy „Applying The Rasch Model: Fundamental Measurement in the Human Sciences“

vyzkoušet „naživo“, což značnou měrou přispívá ke snazšímu ovládnutí problematiky Raschových modelů.

Jak už bylo zmíněno, kniha „Applying the Rasch Model: Fundamental Measurement in Human Sciences“ není reprezentativním shrnutím teorie odpovědi na položku, předkládá pouze jeden z mnoha různých přístupů. Přesto je díky snadným formulacím, nepředpokládajících žádné matematické a statistické znalosti, vhodná pro první seznámení s IRT. Díky mnoha příkladům, vlastnímu programu a datům s podrobným tutoriálem je vysoce užitečným nástrojem pro psychometriky, kteří se chtějí naučit využít IRT pro zlepšení vlastních testů.

Na své si ale přijdou i ostatní psychologové z řad běžných uživatelů psychologických testů. Raschovy modely jsou aplikovatelné při měření širokého spektra psychologických konstruktů a v mnoha ohledech nabízejí určité výhody oproti klasické testové teorii. Lze proto očekávat, že jejich četnost jejich používání se spolu s dalšími IRT přístupy bude v budoucnu rozšiřovat, a je vhodné umět adekvátně interpretovat naměřená data. K šíření IRT mohou také přispět samotní uživatelé testů vyššími nároky na nové metody a poptávkou po moderně standardizovaných testech. Je proto v zájmu psychologické komunity se s nimi seznámit blíže – například právě při četbě knihy Bonda a Foxové.

## Zdroje

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model : Fundamental Measurement in the Human Sciences*. Mahwah (NJ): Lawrence Erlbaum Associates.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah (NJ): Lawrence Erlbaum Associates.

de Ayla, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah (NJ): Lawrence Erlbaum Associates.

Jelínek, M., Květoň, P., & Vobořil, D. (2011). *Testování v psychologii: Teorie odpovědi na položku a počítačové adaptivní testování*. Praha: Grada Publishing.

Urbánek, T., & Šimeček, M. (2001). Teorie odpovědi na položku. *Československá psychologie: Časopis pro psychologickou teorii a praxi*, 45(5), stránky 428-440. Dostupné online <http://www.psychodiagnostika.cz/urbanek.pdf>.

Podpořeno z projektu OPVK:

SOVA-21 – Internacionalizace, inovace, praxe: sociálně-vědní vzdělávání pro 21. století,  
CZ.1.07/2.2.00/28.0225



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ