

Příloha A: Vytváření deskriptorů úrovně ovládnání jazyka

Tato příloha se zabývá technickými aspekty popisování úrovně ovládnání jazyka. Pojednává o kritériích pro vytváření deskriptorů a o metodických postupech pro sestavování stupnic. Následuje anotovaná bibliografie.

Formulace deskriptorů

Zkušenosti se stupnicemi pro jazykové testy, širší teorie škálování v oboru aplikované psychologie i názory učitelů, kteří se účastnili konzultací (např. britského systému odstupňovaných cílů nebo švýcarského projektu), naznačují, že při vytváření deskriptorů by měla platit následující pravidla:

- **Pozitivní přístup:** Stupnice úrovně ovládnání zaměřené na hodnotitele a známkovací stupnice se často vyznačují tím, že deskriptory nižších úrovní ovládnání bývají formulovány negativně. U nižší úrovně znalostí je vždy obtížnější vyjádřit to, co student umí, než to, co neumí. Mají-li však jednotlivé úrovně ovládnání jazyka sloužit jako učební cíle, nikoliv jako pouhé nástroje ke třídění kandidátů, je pozitivní formulace hodnocení žádoucí. Někdy je možné tutéž položku formulovat pozitivně i negativně, například pokud jde o rozsah ovládnání jazykových prostředků (viz tab. A1).

Snahu vyhnout se negativnímu vyjadřování dále komplikuje skutečnost, že některé aspekty komunikativního ovládnání jazyka nepřispívají k jeho dalšímu rozvíjení. Čím méně se projevují, tím lépe. Příkladem je kategorie označovaná jako „nezávislost“, čili nakolik je student závislý a) na přizpůsobení řeči ze strany partnera v komunikaci; b) na možnosti požádat o vysvětlení a c) na pomoci druhé osoby při formulování obsahu sdělení. Tyto záležitosti lze často upravit formou výhrad k pozitivně formulovaným deskriptorům, například:

- Ve většině případů rozumí jasně artikulovanému spisovnému jazyku, který je mu/jí adresován a který se týká běžných záležitostí, pokud může občas požádat o zopakování nebo přeformulování.
- Rozumí tomu, co mluvčí přímo jemu/jí říká zřetelně a pomalu v rámci jednoduchého rozhovoru na každodenní téma. Porozumění mu/jí může být usnadněno, pokud se mu/jí mluvčí opravdu snaží pomoci.

Nebo:

- Dokáže se poměrně snadno zapojit do interakce v situacích, které mají jasnou strukturu, a do krátkých rozhovorů za předpokladu, že mu/jí druhá strana pomůže v případě nutnosti.

Tabulka A1 *Hodnocení: pozitivní a negativní formulace*

Pozitivní formulace	Negativní formulace
<ul style="list-style-type: none"> • ovládá základní jazykový repertoár a strategie, což mu umožňuje porozumět se ve všedních předvídatelných situacích. (Certifikát Eurocentres, úroveň 3) • ovládá základní jazykový repertoár a strategie nutné pro uspokojení většiny každodenních potřeb, většinou však hledá vhodná slova a přizpůsobuje obsah sdělení. (Tabulka pro hodnotitele Eurocentres, úroveň 3) 	<ul style="list-style-type: none"> • má velmi omezený jazykový repertoár, neustále hledá vhodná slova a často vyžaduje přeformulování obsahu sdělení. (ESU, úroveň 3) • z důvodu omezených jazykových schopností se velmi často neorientuje a nedorozumí v nerutinních situacích. (Finská úroveň 2) • dochází k poruchám komunikace, omezené schopnosti narušují obsah sdělení. (ESU, úroveň 3)
<ul style="list-style-type: none"> • slovní zásoba pokrývá oblast základních předmětů, míst a příbuzenských vztahů. (ACTFL, začátečník) 	<ul style="list-style-type: none"> • ovládá pouze omezenou slovní zásobu. (Holandsko, úroveň 1) • omezená zásoba slov a výrazů neumožňuje sdělovat myšlenky a názory. (Göteborg U)
<ul style="list-style-type: none"> • student reprodukuje a rozpoznává soubor naučených slov a krátkých frází. (Trim 1978, úroveň 1) 	<ul style="list-style-type: none"> • student produkuje pouze mechanicky naučené skupiny vět a výřky. (ACTFL, začátečník)
<ul style="list-style-type: none"> • dokáže použít krátké běžné výrazy a vyhovět tak jednoduchým konkrétním potřebám - pozdravy, výměna informací atd. (Elviri; úroveň 1, Miláno 1986) 	<ul style="list-style-type: none"> • ovládá pouze nejzákladnější jazykový repertoár, funkční porozumění jazyku je minimální nebo nulové. (ESU, úroveň 2)

- **Určitost:** Deskriptory by měly popisovat konkrétní úlohy nebo konkrétní míru dovednosti při plnění úloh. Je třeba pamatovat na dvě věci. Zaprvé, formulace deskriptorů by neměla být vágní, například „Student ovládá řadu přiměřených strategií“. Co je míněno strategií? Přiměřených čemu? Co znamená „řada“? Problémem vágních deskriptorů je, že se hezky čtou, ale přestože je každý lehce přijme za své, bývají rozdílně interpretovány. Za druhé, již od 40. let platí zásada, že dva po sobě následující stupně by se neměly odlišovat pouze v tom, že výrazy „někdy“ nebo „několik“ budou o stupeň výše nahrazeny výrazy „často“ nebo „většina“, nebo že „dosti široký“ se změní na „velmi široký“ či z „prostřední“ na „dobrý“. Rozdíly musejí být skutečné, nikoliv pouze verbální, a proto tam, kde nelze rozdíly smysluplně a konkrétně vyjádřit, mohou vzniknout mezery.
- **Jasnost:** Deskriptory mají být formulovány srozumitelně, bez použití profesního žargonu. Žargon nejenže ztěžuje chápání, ale někdy se navíc stává, že jestliže jej odstraníme, napohled dokonalý deskriptor přestane dávat smysl. Deskriptory mají také mít jednoduchou větnou stavbu a explicitní logickou strukturu.
- **Stručnost:** Existují dvě názorové školy. První, rozšířená především v Americe a Austrálii, pracuje s holistickými stupnicemi a snaží se každý stupeň charakterizovat v jediném dlouhém odstavci, který má vyčerpávajícím způsobem pokrývat všechny důležité aspekty. Takové stupnice vyhovují požadavku „určitosti“ tím, že obsahují úplný výčet charakteristických rysů, který hodnotitelům umožňuje vytvořit si detailní portrét typického studenta na dané úrovni ovládnutí, a jsou tedy bohatým zdrojem pro charakterizaci. Tento přístup má však dvě nevýhody. Zaprvé, neexistuje „typický“ jedinec. Jednotlivé rysy se

vyskytují v různých konfiguracích. Za druhé, není realisticky možné, aby deskriptor delší než souvětí o dvou větách sloužil v procesu hodnocení jako praktické vodítko. Učitelé zásadně upřednostňují krátké deskriptory. V projektu, jehož cílem bylo formulovat vzorové deskriptory, učitelé většinou odmítli nebo rozdělili na více částí deskriptory delší než přibližně 25 slov (asi dva řádky standardního písma).

- **Nezávislost:** Krátké deskriptory mají navíc další výhody. Je u nich pravděpodobnější, že budou popisovat něco, co se dá jednoznačně hodnotit slovy „ano, tento student to dokáže“. V důsledku toho mohou být kratší a konkrétní deskriptory využívány pro výčet nezávislých kritérií v kontrolních seznamech nebo dotaznících, které slouží učitelům k průběžnému hodnocení a studentům k sebehodnocení. Tato nezávislá platnost je signálem, že deskriptor slouží jako vyjádření cíle, že jeho smysl není pouze relativní a odvozený od ostatních deskriptorů na stupnici. To otevírá řadu možností, jak využívat nejrůznější formy hodnocení (viz kapitola 9).

Uživatelé Rámce mohou zvážit a, kde je to vhodné, uvést:

- *která z uvedených kritérií jsou v jejich kontextu nejdůležitější a podle jakých dalších explicitních či implicitních kritérií se sami řídí;*
- *nakolik je žádoucí a praktické, aby vytváření deskriptorů v jejich systému probíhalo podle uvedených kritérií.*

Metodické postupy při sestavování stupnic

U stupnice, čili posloupnosti úrovní, se předpokládá, že určité prvky jsou vlastní pouze jednomu určitému stupni a že pro každý stupeň existuje charakteristická míra dovedností. Z toho plyne, že tvorba stupnic se musí řídit pevnými zásadami. Deskriptory ovládnutí jazyka lze přiřazovat k jednotlivým stupňům mnoha způsoby. Metody, které se za tímto účelem používají, se dělí do tří skupin na metody intuitivní, kvalitativní a kvantitativní. Většina užívaných stupnic ovládnutí jazyka a jiných souborů úrovně ovládnutí byla vypracována jednou ze tří intuitivních metod v první skupině. Nejlepší metody kombinují komplementárním a kumulativním postupem všechna tři hlediska. Kvalitativní metody vyžadují intuitivní přípravu, intuitivní výběr materiálu a intuitivní interpretaci výsledků. Kvantitativní metody by měly kvantitativně vyhodnocovat kvalitativně pretestovaný materiál a také ony vyžadují intuitivní interpretaci výsledků. Společné referenční úrovně byly tedy vypracovány kombinovanou metodou, která zahrnovala intuitivní, kvalitativní i kvantitativní postupy.

Zvolíme-li si pro sestavení stupnice kvalitativní či kvantitativní metodu, existují dva možné výchozí body: deskriptory a vzorky výkonů.

Východiskem jsou deskriptory: Lze začít tím, že uvážíme, co chceme popsat, a poté jako základ pro kvalitativní fázi sepíšeme, vybereme nebo upravíme deskriptory pro žádané kategorie. Příkladem tohoto postupu jsou metody 4 a 9, první a poslední v níže uvedené kvalitativní skupině metod. Hodí se zvláště pro vytváření deskriptorů pro kategorie vycházející z kurikula, například pro komunikativní jazykové činnosti, ale může být užitečný i při vytváření deskriptorů různých aspektů kompetence. Začneme-li u kategorií a deskriptorů, má to tu výhodu, že lze definovat teoreticky vyvážené pokrytí aspektů užívání jazyka.

Východiskem jsou vzorky výkonů: Alternativou, kterou lze využít pouze při vytváření deskriptorů pro hodnocení výkonu, je vyjít od reprezentativních vzorků výkonu. Zde se můžeme dotázat reprezentativní skupiny posuzovatelů, čeho si všímají, když se vzorky pracují (kvalitativní metoda). Metody 5–8 pracují s variantami tohoto přístupu. Lze také postupovat tak, že posuzovatelé vzorky pouze ohodnotí a pak jsou za použití příslušných statistických postupů určeny klíčové faktory, které jejich rozhodování motivovaly (kvantitativní metoda). Příkladem tohoto přístupu jsou metody 10 a 11. Výhodou analýzy vzorků výkonu je, že jejím prostřednictvím lze získat velmi konkrétní popis založený na konkrétních údajích.

Poslední metoda, metoda 12 jako jediná používá při sestavování deskriptorů do stupnic přísně matematické postupy. Touto metodou byly, spolu v metodou 2 (intuitivní) a metodami 8 a 9 (kvalitativní), vypracovány Společné referenční úrovně a vzorové deskriptory v Rámci. Stejného statistického postupu lze však využít i po vypracování stupnice pro ověření její validity v praxi s cílem odhalit případné nedostatky.

Intuitivní metody:

Tyto metody nevyžadují strukturovaný sběr dat, pouze zásadovou interpretaci zkušeností.

1. **Metoda experta:** Stupnici vytváří určitá osoba, která může čerpat z již existujících stupnic, z kurikulárních dokumentů či jiných vhodných materiálů, pokud možno po předchozí analýze potřeb dané cílové skupiny. Stupnice je pak ověřována a upravována s pomocí informátorů (laických posuzovatelů).
2. **Metoda komise:** Postup je stejný jako u předcházející metody, ale na vytváření stupnice spolupracuje menší tým se širší skupinou konzultantů, kteří posuzují jednotlivé návrhy. Konzultanti mohou pracovat intuitivně na základě vlastních zkušeností nebo na základě srovnávání se studenty či vzorky výkonů. Nedostatky stupnic v kurikulech pro výuku moderních jazyků na středních školách, které vypracovaly komise ve Velké Británii a Austrálii, komentují Gipps (1994) a Scarino (1996; 1997).
3. **Metoda zkušenostní:** Stejná jako předchozí, ale proces tvorby trvá delší dobu, dokud se v dané instituci nebo u specifického kontextu hodnocení nedospěje k absolutnímu konsensu. Skupina lidí si mezi sebou ujasní význam jednotlivých úrovní ovládní a kritérií a následuje jejich systematické ověřování v praxi a analýza zpětné vazby s cílem zpřesnit jejich formulaci. Skupinky posuzovatelů mohou srovnávat výkony studentů s jednotlivými definicemi a definice se vzorky výkonů. Touto tradiční metodou byly vypracovány mnohé používané stupnice ovládní jazyka (Wilds 1975; Ingram 1985; Liskin-Gasparro 1984; Lowe 1985, 1986).

Kvalitativní metody:

Všechny tyto metody vyžadují spolupráci menšího pracovního týmu se skupinkami informátorů a vyznačují se tím, že získané informace jsou interpretovány spíše kvalitativně než statisticky.

4. **Metoda formulace klíčových pojmů:** Jakmile je stupnice navržena, rozstříhá se na jednotlivé části a informátoři (laičtí posuzovatelé), kteří by měli být typickými představiteli uživatelů této stupnice, jsou požádáni, a) aby uspořádali jednotlivé definice podle toho, jak mají jít podle jejich názoru za sebou; b) aby vysvětlili, jak k uvedenému pořadí dospěli, a c) jakmile se seznámí s rozdílem mezi svým a původně navrženým pořadím, aby označili klíčové body, které jim pomohly nebo naopak působily nejasnosti. Další možností je odstranit celý stupeň a dodatkovým úkolem pro informátory je pak zjistit, mezi kterými dvěma stupni je mezera. Takto byly vypracovány hodnotící stupnice pro Eurocentres.
5. **Metoda klíčových pojmů v rámci jazykového výkonu:** Deskriptory jsou srovnávány s typickými výkony v pásmech jednotlivých úrovní ovládnutí, aby bylo zaručeno, že popis koresponduje se skutečností. Tímto procesem provádějí učitelé některé průvodce ke Cambridgeským zkouškám, v nichž jsou charakteristiky jednotlivých úrovní ovládnutí srovnány se známkami, které byly přiděleny určitým pracím. Deskriptory pro IELTS (International English Language Testing System) vznikly tak, že skupinky zkušených posuzovatelů byly požádány, aby pro každou úroveň určily „klíčové vzorky prací“ a poté společně rozhodli o „klíčových rysech“ každého vzorku. Ve společné diskusi jsou pak stanoveny rysy, které se zdají být charakteristické pro jednotlivé úrovně ovládnutí, a na základě nich se formulují deskriptory (Alderson 1991; Shohamy a kol. 1992).
6. **Metoda primárního znaku:** Informátoři nejprve samostatně seřadí záznamy výkonů studentů (většinou písemné práce) podle kvality a poté ve společné diskusi rozhodnou o konečném pořadí. Pak je stanoven a pro každý stupeň popsán princip, na jehož základě byly práce roztrženy, a to s důrazem na prominentní rysy jednotlivých stupňů. Výsledný popis se stává znakem, který určuje pořadí (Mullis, 1980). Běžnou alternativou sestavování pořadí prací bývá roztrždit je na několik hromádek. Tento klasický přístup má také zajímavou vícerozměrnou variantu, která vypadá tak, že nejprve se prostřednictvím identifikace klíčových pojmů (metoda 5) určí nejdůležitější znaky a poté se sestaví pořadí prací samostatně pro každý znak. Tímto postupem lze tedy získat analytickou vícerozměrnou stupnici znaků, nikoliv pouze holistickou stupnici primárního znaku.
7. **Metoda binárního rozhodování:** Další obměnou předchozí metody je nejprve roztrždit reprezentativní vzorky na hromádky podle úrovní ovládnutí. V diskusi, jejímž hlavním účelem je stanovit hranice mezi jednotlivými stupni, se poté určí klíčové rysy (jako u metody 5). Každý z nich je formulován do podoby krátké zjišťovací otázky, na niž lze odpovědět ano/ne. Vznikne stromové schéma binárních rozhodnutí, podle něhož může hodnotitel při hodnocení postupovat jako podle algoritmu (Upshur a Turner 1995).
8. **Metoda posuzování:** Skupina posuzovatelů porovnává vždy dvojici výkonů, určí, který je lepší a proč. Takto jsou prostřednictvím metajazyka posuzovatelů identifikovány kategorie a prominentní rysy jednotlivých úrovní ovládnutí, na jejichž základě lze pak formulovat deskriptory (Pollitt a Murray 1996).
9. **Metoda třídění:** Informátoři třídí návrhy deskriptorů podle kategorií, které mají popisovat, a/nebo podle úrovní ovládnutí. Mají také možnost některé deskriptory upravit nebo zamítnout a označit ty, které považují za zvláště jasné, užitečné, důležité atd. Tímto způsobem byla editována databáze deskriptorů, z níž vzešel soubor vzorových stupnic Rámce (Smith a Kendall 1963; North 1996, 2000).

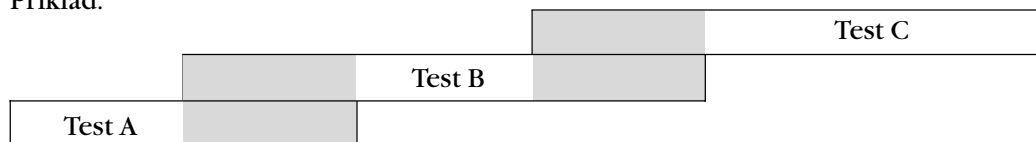
Kvantitativní metody:

Tyto metody se vyznačují širokým využitím statistické analýzy a pečlivou interpretací výsledků.

10. **Metoda diskriminační analýzy:** Vzorky výkonů, které již byly ohodnoceny (pokud možno týmově), jsou podrobeny detailní analýze promluvy. Tato kvantitativní analýza odhalí jejich kvalitativní rysy a četnost jejich výskytu. Mnohonásobnou regresí se pak zjistí, které z těchto rysů byly pro rozhodnutí hodnotitelů určující, a na jejich základě jsou vypracovány deskriptory pro jednotlivé úrovně ovládnání (Fulcher 1996).
11. **Metoda multidimenzního škálování:** Navzdory názvu jde o popisnou metodu umožňující stanovit klíčové rysy a jejich vzájemné vztahy. Výkony jsou ohodnoceny podle analytické stupnice sestávající z několika kategorií. Výsledek této analýzy ukáže, které z kategorií byly pro stanovení úrovně ovládnání určující, a umožní sestavení diagramu, který znázorní míru vzájemné blízkosti či vzdálenosti mezi různými kategoriemi. Tento postup tedy umožní stanovit podstatná kritéria a ověřit jejich validitu (Chaloub-Deville 1995).
12. **Item Response Theory (IRT) neboli analýza „skrytého znaku“:** IRT nabízí sérii modelů pro měření či škálování. Nejméně komplikovaný je Raschův model, pojmenovaný podle dánského matematika George Rasche. IRT byla vyvinuta na základě teorie pravděpodobnosti a využívá se především ke stanovování obtížnosti jednotlivých testovacích položek v rámci jejich databáze. U pokročilého studenta je pravděpodobnost správné odpovědi na elementární otázku značně vysoká, u začátečníka je pravděpodobnost správné reakce na otázku předpokládající pokročilejší znalosti jazyka velmi nízká. Tato jednoduchá skutečnost se stala základem metodologického postupu, který Raschův model využívá například ke kalibraci položek v rámci jedné stupnice. Další rozpracování této metody umožňuje její aplikaci při sestavování stupnic deskriptorů komunikativního ovládnání jazyka a stupnic testových položek.

Raschova analýza probíhá tak, že se různé testy či dotazníky spojí do řetězce, jehož jednotlivé články se překrývají za pomoci tzv. spojovacích položek, které jsou společné dvěma na stupnici po sobě následujícím testům. V níže uvedeném diagramu jsou spojovací položky vyznačeny šedým stínováním. Tímto způsobem lze testy cíleně zaměřovat na určité skupiny studentů, a přesto vycházet z téže stupnice. Je však nutné uvědomit si, že model zkresluje u každého testu výsledky s nejvyšším a nejnižším bodovým hodnocením.

Příklad:



Výhodou Raschovy analýzy je, že umožňuje měření, které se obejde bez měření vzorových výkonů i stupnic, jinými slovy umožňuje hodnocení, které je nezávislé na vzorových testech či dotaznících použitých k analýze. Stanovená hodnota obtížnosti jednotlivých stupňů zůstává konstantní a lze ji aplikovat i na další skupiny studentů za předpokladu, že budou náležet k téže statistické skupině populace. Systematický posun hodnot v čase (např. následkem změn kurikula nebo školení hodnotitelů) je možné kvantifikovat a začlenit do měření stejně jako systémové rozdíly v typu studentů či hodnotitelů (Wright a Masters 1982, Lincare 1989).

Raschovu analýzu lze při sestavování stupnic deskriptorů využít mnoha způsoby:

- a) Údaje získané kvalitativními metodami 6, 7 a 8 lze uspořádat pomocí Raschovy analýzy do aritmetické stupnice.
- b) Test lze pečlivě připravit tak, aby jeho jednotlivé položky přesně odrážely deskriptory ovládnutí jazyka. Testové položky pak mohou být pomocí Raschovy analýzy sestaveny do stupnice a jejich stupnicová hodnota určí relativní obtížnost příslušných deskriptorů. (Brown a kol. 1992; Carroll 1993; Masters 1994; Kirsch 1995; Kirsch a Mosenthal 1995).
- e) Deskriptory lze využít jako položky dotazníku, podle něhož hodnotí učitel své studenty (např. Dokáže se student zeptat na cestu?). Takto mohou být deskriptory kalibrovány přímo na aritmetické stupnici a seřazeny podle obtížnosti stejně jako položky v databázi testových položek.
- g) Tímto postupem byly sestaveny stupnice deskriptorů v kapitolách 3, 4 a 5. Všechny tři projekty popisované v přílohách B, C a D použily Raschovu metodu k seřazení deskriptorů do stupnic a k vzájemnému vyvážení výsledných stupnic.

Kromě toho, že lze Raschův model úspěšně využít k sestavení stupnice, může posloužit také k analýze skutečného využití jednotlivých pásem na hodnotící stupnici. Může poukázat na nejednoznačné slovní vyjádření nebo na příliš velké či příliš malé využití některého pásma, čímž dá podnět k nápravě (Davidson 1992; Milanovic a kol. 1996; Stansfield a Kenyon 1996; Tyndall a Kenyon 1996).

Uživatelé Rámce mohou zvážit a, kde je vhodné, uvést:

- *nakolik mají známky přidělované v jejich systému jednotný význam daný společnými definicemi;*
- *které z výše uvedených metod nebo které další metody jsou používány při vytváření těchto definic.*

Vytváření stupnic ovládnání jazyka: vybraná bibliografie s anotacemi

Alderson, J. C.: Bands and scores. In: Alderson, J. C. & North, B. (ed.): *Language testing in the 1990s*. London: British Council/Macmillan, Developments in ELT, 71–86, 1991.

Brindley, B.: Defining language ability: the criteria for criteria. In: Anivan, S. (ed.): *Current developments in language testing*. Singapore, Regional Language Centre 1991.

Brindley, B.: Outcome-based assessment and reporting in language learning programmes, a review of the issues. *Language Testing* 15 (1), 45–85, 1998.

Brown, A., Elder, C., Lumley, T., McNamara, T. & McQueen, J.: *Mapping abilities and skill levels using Rasch techniques*. Referát přednesený na 14th Language Testing Research Colloquium, Vancouver. Publikováno v *Melbourne Papers in Applied Linguistics* 1/1, 37–69, 1992.

Carroll, J. B.: Test theory and behavioural scaling of test performance. In: Frederiksen, N., Mislevy, R. J. & Bejar, I. I. (ed.): *Test theory for a new generation of tests*. Hillsdale, N. J. Lawrence Erlbaum Associates, 297–323, 1993.

Chaloub-Deville, M.: Deriving oral assessment scales across different tests and rater groups. *Language Testing* 12 (1), 16–33, 1995.

Davidson, F.: Statistical support for training in ESL composition rating. In Hamp-Lyons (ed.): *Assessing second language writing in academic contexts*. Norwood, N. J. Ablex: 155–166, 1992.

Fulcher: Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing* 13 (2), 208–238, 1996.

Gipps, C.: *Beyond testing*. London, Falmer Press 1994.

Kirsch, I. S.: Literacy performance on three scales: definitions and results. In: *Literacy, economy and society: Results of the first international literacy survey*. Paris, Organisation for Economic Cooperation and development (OECD): 27–53, 1995.

Zabývá se problémy, které vznikají následkem nepřesně definovaného účelu a zaměření testů, a tvorbou stupnic IELTS pro mluvení.

Autor polemizuje s tvrzením, že stupnice ovládnání jazyka představují kritériální hodnocení.

Kritizuje skutečnost, že hodnocení studentů je zaměřeno na to, co podle výsledků testu „dokážou“, nikoliv jednotlivé aspekty jejich rozvíjejících se dovedností.

Klasické využití Raschovy metody škálování testových položek k vytvoření stupnice ovládnání jazyka; položky testu zde mají podobu různých úloh ve čtení.

Zásadní článek doporučující Raschovu metodu sestavování testových položek do stupnic ovládnání jazyka.

Studie zkoumající kritéria, na něž rodilí mluvčí arabštiny berou zřetel při posuzování studentů. Zcela ojedinelá aplikace vícerozměrného škálování v testování jazyka.

Velmi srozumitelný návod, jak prostřednictvím Raschovy analýzy ověřit validitu hodnotící stupnice v cyklickém procesu. Vyslovuje se pro „sémantický“ přístup k vytváření stupnic, nikoliv pro přístup „konkrétní“, který představují např. deskriptory.

Systematický přístup k vytváření deskriptorů a stupnic, na jehož počátku stojí důkladná analýza procesu jazykového výkonu. Časově velmi náročná metoda.

Podporuje hodnocení učitele zaměřené na „standardní požadavky“ vycházející ze společných referenčních bodů vytvořených na základě spolupráce mezi školami. Uvádí problémy, jež způsobují nejednoznačné deskriptory v Anglickém národním kurikulu. Mezipředmětový přístup.

Jednoduchá zpráva bez technických údajů o složitém využití Raschovy analýzy při vytváření stupnic na základě údajů z testů. Metoda vyvinutá pro předvídání obtížnosti nových testových položek z určitých úloh a dovedností – tedy ve vztahu k referenčnímu rámci.

Příloha A: Vytváření deskriptorů úrovně ovládnání jazyka

Kirsch, I. S. & Mosenthal, P. B.: Interpreting the IEA reading literacy scales. In: Binkley, M., Rust, K. & Winglee, M. (ed.): *Methodological issues in comparative educational studies: the case of the IEA reading literacy study*. Washington D. C.: US Department of Education, National Center for Education Statistics: 135–192, 1995.

Linacra, J. M.: *Multi-faceted Measurement*. Chicago, MESA Press 1989.

Liskin-Gasparro, J. W.: The ACTFL proficiency guidelines: Gateway to testing and curriculum. In: *Foreign Language Annals* 17/5, 475–489, 1984.

Lowe, P.: The ILR proficiency scale as a synthesising research principle: the view from the mountain. In: James, C. J. (ed.): *Foreign Language Proficiency in the Classroom and Beyond*. Lincolnwood (Ill.), National Textbook Company 1985.

Lowe, P.: Proficiency: panacea, framework, process? A Reply to Kramsch, Schulz, and particularly, to Bachman and Savignon. *Modern Language Journal* 70/4, 391–397, 1986.

Masters, B.: Profiles and assessment. *Curriculum Perspectives* 14, 1: 48–52, 1994.

Milanovic, M., Saville, N., Pollitt, A. & Cook, A.: Developing rating scales for CASE: Theoretical concerns and analyses. In: Cumming, A. & Berwick, R.: *Validation in language testing*. Clevedon, Avon, Multimedia Matters: 15–38, 1996.

Mullis, I. V. S.: *Using the primary trait system for evaluating writing*. Manuscript No. 10-W-51. Princeton N. J., Educational Testing Service 1981.

North, B.: *The development of descriptors on scales of proficiency: perspectives, problems, and a possible methodology*. NFLC Occasional Paper, National Foreign Language Center, Washington D. C., April 1993.

North, B.: *Scales of language proficiency: a survey of some existing systems*. Strasbourg, Council of Europe CC-LANG (94) 24, 1994.

Podrobnější a techničtější verze předchozí zprávy popisující vývoj metody v průběhu tří příbuzných projektů.

Převratný postup ve statistice umožňující brát při prezentaci výsledků hodnocení v úvahu přísnost examinatorů. Aplikováno v projektu zaměřeném na vypracování modelových deskriptorů sloužících pro ověření vztahu mezi úrovněmi ovládnání a školními ročníky.

Nástin účelu a vývoje americké stupnice ACTFL z původní stupnice Foreign Service Institute (FSI).

Podrobný popis vytváření americké stupnice Interagency Language Roundtable (ILR) z původní stupnice FSI. Funkce stupnice.

Obhajoba systému, který se osvědčil ve specifickém kontextu, proti kritice akademických kruhů vyvolané rozšířením užívání stupnice a její metodikou vedení pohovoru do oblasti vzdělávání (s ACTFL).

Krátká zpráva pojednávající o využití Raschovy metody pro vytváření stupnic výsledků testů a hodnocení učitelů za účelem vypracovat systém pro hodnocení kurikulů v Austrálii.

Klasické pojednání o využití Raschovy metody ke zpřesnění hodnotící stupnice pro test mluvení (redukce stupnice na počet úrovní), s nímž mohou hodnotitelé efektivně pracovat.

Klasické pojednání o metodě primárního znaku při vytváření hodnotící stupnice pro psaní v mateřském jazyce.

Kritika obsahu tradičních stupnic ovládnání jazyka a způsobu jejich zpracování. Návrh projektu na vypracování deskriptorů za pomoci učitelů a na jejich sestavení do stupnic pomocí Raschovy metody na základě hodnocení učitelů.

Obsažný přehled kurikulárních a hodnotících stupnic, které byly později analyzovány a použity jako východní bod v projektu zaměřeném na vypracování modelových deskriptorů.

Příloha A: Vytváření deskriptorů úrovně ovládnání jazyka

North, B.: *The development of a common framework scale of language proficiency* New York, Peter Lang 2000.

Diskuse o stupnicích ovládnání jazyka a o vztahu mezi nimi, užíváním jazyka a modely kompetencí. Podrobný popis postupu, jímž byly v projektu vypracovány modelové deskriptory. Vzniklé problémy a jejich řešení.

North, B.: *Scales for rating language performance in language tests: descriptive models, formulation styles and presentation formats*. TOEFL Research Paper. Princeton NJ, Educational Testing Service (v tisku).

Podrobná analýza a historický přehled typů hodnotících stupnic používaných pro testy mluvení a psaní. Jejich výhody, nevýhody, problémy atd.

North, B. & Schneider, G.: Scaling descriptors for language proficiency scales. *Language Testing* 15/2: 217–262, 1998.

Souhrnná zpráva o projektu, jehož cílem bylo vypracovat modelové deskriptory. Rozbor výsledků a stability stupnice. V příloze příklady testovacích nástrojů a produktů.

Pollitt, A. & Murray, N. L.: What raters really pay attention to. In: Milanovic, M. & Saville, N. (ed.): *Performance testing, cognition and assessment*. Studies in Language Testing 3. Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem, 2–4 August 1993. Cambridge: University of Cambridge Local Examinations Syndicate: 74–91, 1996.

Zajímavý metodologický článek uvádějící do souvislostí tabulkovou analýzu repertoáru studenta a jednoduchou techniku škálování s cílem zjistit, na co se hodnotitelé zaměřují u různých stupňů ovládnání jazyka.

Scarino, A.: Issues in planning, describing and monitoring long-term progress in language learning. In: *Proceedings of the AFMLTA 10th National Languages Conference*. 67–75, 1996.

Kritizuje nejasné formulace a nedostatek informací o kvalitě výkonu studentů u typických hodnotících výroků pro hodnocení učitelem v rámci britských a australských kurikulárních profilů.

Scarino, A.: Analysing the language of frameworks of outcomes for foreign language learning. In: *Proceedings of the AFMLTA 11th National Languages Conference*. 241–258, 1997.

Jako u předchozího.

Schneider, G. & North, B.: „In anderen Sprachen kann ich“...*Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Bern/Aarau: NFP 33/SKBF (Umsetzungsbericht), 1999.

Krátká zpráva o projektu, jehož výsledkem byly modelové stupnice. Uvedena je také švýcarská verze Portfolia (40 stran A5).

Schneider, G. & North, B.: „Sans d'autres langues, je suis capable de ...“ Echelles pour la description, l'évaluation et l'auto-évaluation des compétences en langues étrangères. Berne/Aarau PNR33/CSRE (hodnotící zpráva), 2000.

Jako u předchozího.

Schneider, G. & North, B.: *Fremdsprachen können – was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Chur/Zürich, Verlag Rüegger AG, 2000.

Souhrnná zpráva o projektu, jehož výsledkem byly modelové stupnice. Obsahuje kapitolu srozumitelně pojednávající o vytváření stupnic v anglickém jazyce. Seznamuje čtenáře se švýcarskou verzí Portfolia.

Skehan, P.: Issues in the testing of English for specific purposes. *Language Testing* 1/2, 202–220, 1984.

Kritizuje normativní hodnocení a nejednoznačnost formulací u stupnic ELTS.

Shohamy, E., Gordon, C. M. & Kraemer, R.: The effect of raters' background and training on the reliability of direct writing tests. *Modern Language Journal* 76: 27–33, 1992.

Smith, P. C. & Kendall, J. M.: Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47/2, 1963.

Stansfield, C. W. & Kenyon, D. M.: Comparing the scaling of speaking tasks by language teachers and the ACTFL guidelines. In: Cumming, A. & Berwick, R.: *Validation in language testing*. Clevedon, Avon, Multimedia Matters: 124–153, 1996.

Takala, S. & Kaftandjieva, F.: Council of Europe scales of language proficiency: A validation study. In: Alderson, J. C. (ed.): *Case studies of the use of the Common European Framework*. Strasbourg, Council of Europe (v tisku).

Tyndall, B. & Kenyon, D.: Validation of a new holistic rating scale using Rasch multifaceted analysis. In: Cumming, A. & Berwick, R. *Validation in language testing*. Clevedon, Avon, Multimedia Matters: 9–57, 1996.

Upshur, J. & Turner, C.: Constructing rating scales for second language tests. *English Language Teaching Journal* 49 (1), 3–12, 1995.

Wilds, C. P.: The oral interview test. In: Spolsky, B. & Jones, R. (ed.): *Testing language proficiency*. Washington D. C., Center for Applied Linguistics, 29–44, 1975.

Jednoduché pojednání o základní kvalitativní metodě pro vypracování analytické stupnice pro psaní. Vedla k překvapivě vysoké spolehlivosti hodnocení ve skupině neškolených laiků.

První pokus o škálování deskriptorů místo pouhého sestavování stupnic. Zásadní článek, čte se velmi obtížně.

Využití Raschovy metody škálování k ověřování pořadí obtížnosti úloh, které se objevují v pokynech ACTFL. Zajímavá metodologická studie, jejímž přístupem se inspiroval projekt vypracování modelových deskriptorů.

Zpráva o využití a dalším rozpracování Raschova modelu pro sestavování stupnic sebehodnocení z adaptovaných modelových deskriptorů v kontextu projektu DIALANG. Ověřováno ve vztahu k finštině.

Jednoduchá zpráva o ověřování validity stupnice ESL pro ústní rozřazovací zkoušky z angličtiny jako vyučovacího jazyka při vstupu na univerzitu. Klasické využití vícerozměrné Raschovy analýzy pro stanovení výukových potřeb.

Složité rozpracování metody primárního znaku s cílem vytvořit tabulky pro binární rozhodování. Velmi užitečné pro oblast školství.

Prvotní pojednání o původní stupnici hodnotící ovládnutí jazyka. Obsahuje nuance, které většina nových přístupů k ústnímu zkoušení opomíjí, stojí tedy za pečlivé přečtení.