

Ekvivalence položek v mezinárodních datech: základní vymezení a možnosti analýzy*

Petra Anýžová**

Filozofická fakulta Univerzity Karlovy a
Vysoká škola finanční a správní, o.p.s., Praha

Item Equivalence in Cross-National Survey Research: Basic Definition and Possibilities for Analysis

Abstract: *This article argues that the concept of equivalence is one of the most important methodological aspects of valid and reliable measurement in cross-national survey research. The important topic of survey measure equivalence has not been systematically in Czech social science publications to date and this article hopes to address this gap in the literature. Consequently, the two main goals of this article are (1) to acquaint the reader with techniques that are used to find questions that are interpreted in the same way across countries before data collection and (2) to describe the testing and evaluation of measurement indicators' equivalence or comparability after data collection. This study presents cognitive approaches to "good" question wording practices, best translation practices and the application of both 'emic' (culture specific) and 'etic' (culture universal) approaches to survey question design. After data collection a range of statistic techniques are usually employed ranging from basic statistics such as the mean to advanced approaches such as multi-group structural equation modelling, multilevel modelling, latent class modelling and Item Response Theory). This article describes some of these techniques in the context of measurement equivalence and its associated research literature.*

Keywords: *cross-national survey research, equivalence, measurement indicators, item comparability, latent variable and multilevel statistical techniques.*

Data a výzkum - SDA Info 2013, Vol. 7, No. 1: 29-56.

DOI: <http://dx.doi.org/10.13060/1802-8152.2013.7.1.2>

(c) Sociologický ústav AV ČR, v.v.i., Praha 2013.

* Autorka děkuje oběma anonymním recenzentům a kolegům z FF UK a VŠFS za podnětné připomínky.

** Veškerou korespondenci pošlejte na adresu: Mgr. Petra Anýžová, Vysoká škola finanční a správní, o.p.s., Estonská 500, 101 00 Praha 10 nebo na e-mail: petra.anyzova@gmail.com.

Úvod

Prvotním záměrem mezinárodních sociálněvědních výzkumů je srovnávat různé sociální jevy mezi zeměmi, kulturami nebo etnickými či jinými sociálními skupinami [Kohn 1989]. Významy jednotlivých sociálních fenoménů se ale liší pro různé lidi, v různých dobách a především v různých kulturách. Proto proces složité transformace zkoumaného sociálního jevu do měřitelných empirických položek v dotaznících (tzn. operacionalizace) představuje pro mezinárodní výzkumy dramatickou překážku. Výzkumníkům se jeví nezbytné pokládat otázky na první pohled „zázračným“ způsobem tak, aby vystihly pozorovaný jev typicky, s jeho vlastnostmi společnými pro všechny kultury a zároveň velmi specificky tak, jak se objevuje v každém sociálním systému odlišně.

Logicky se tak jako jedna z prvních zásadních otázek objevilo to, zda jsou sledované sociální fenomény v různých společnostech vůbec srovnatelné, či nikoli; zda lze sociální jevy měřit ve výzkumu adekvátním způsobem stejně, rovnocenně, ekvivalentně ve všech participujících zemích. Relevantnost této metodologické otázky se potvrzuje obzvláště tváří v tvář častým pochybnostem, zda mezinárodní data opravdu reflektují rozdíly v odpovědích respondentů a nikoliv pouze rozdíly v interpretaci otázek, a zda tak nejsou zjištěné odlišnosti mezi kulturami jen něčím uměle vykonstruovaným při analýze dat [podrobněji Jowell 1998].

Na druhou stranu globální provázanost světa a multikulturní snahy o vzájemné porozumění si mezi zeměmi činí v současnosti z mezinárodních komparativních výzkumů zaměřených na dlouhodobé mapování postojů a změn v lidském chování nezbytnost v oblasti sociálního zkoumání. Z tohoto důvodu nelze od mezinárodní komparace ustoupit pod tíhou metodologické kritiky, ale spíše přistoupit k různým variantám řešení dané problematiky. Ty zmíněný problém ekvivalence cizojazyčných otázek sice zcela neeliminují, ale přispívají k vyšší transparentnosti při vykazování zjištěných kulturních rozdílů v mezinárodních šetřeních.

Hlavními důvody pro pojednání o této metodologické problematice jsou především tyto dvě skutečnosti. Zejména v českých sociálních vědách není problematika ekvivalence¹ nijak kontinuálně rozebírána a prezentována v odborných publikacích, spíše se jen obsahově skrývá za témata kvality sociálněvědních výběrových šetření a chyb měření v komparativních výzkumech [např. Kostecký, Čermák 2003; Krejčí 2008]. Je to o to překvapivější, že se Česká republika po roce 1989 do řady kontinuálních mezinárodních projektů aktivně zapojila, jak také názorně ilustruje Krejčí [2006]. Za druhé, velký potenciál mezinárodních dat je spatřován v analytických možnostech sekundárních analýz, které však s sebou přinášejí i řadu komplikací, přednostně spojených právě se srovnatelností položek mezi různými zeměmi. Každý výzkumník, který má v úmyslu se sekundárním analýzám mezinárodních dat

1 Také často označována jako invariance měření (measurement invariance) – stálost, neměnnost měření.

věnovat, by se měl s danou problematikou seznámit a získat základní přehled o možnostech jejího řešení.

Cílem následující přehledové stati je předvést způsoby dosahování a testování ekvivalence otázek kladených v mezinárodních dotaznících. Logika členění technik, které disponují těmito možnostmi, do dvou souborů bude vycházet z jednoduché představy o tom, jaká pravidla platí pro nastolování ekvivalence otázek/odpovědí *před sběrem dat* a jak lze zkoumat dosaženou ekvivalenci položek *po sběru dat*. Součástí tohoto schématu bude i popis jejich základního účelu, přínosu a nedostatků. Této problematice se česká akademická obec zatím ještě nijak systematicky nevěnovala, proto bude nejprve třeba představit koncept ekvivalence, základní pojmosloví a navrhnout stručnou kategorizaci těch druhů ekvivalence, jejichž testování se následující část textu věnuje především. Pokud se podaří docílit vytyčených záměrů, po přečtení závěru článku by měl čtenář získat základní přehled o problematice ekvivalence na pozadí metodologie mezinárodních šetření a o metodických způsobech jejího řešení.

Koncept ekvivalence a jeho kategorizace

Ve většině odborných publikací věnujících se metodologii mezinárodních výzkumů se dočteme, že základem a nutnou podmínkou kvalitního mezinárodního šetření je kromě validity a reliability také ekvivalence měření.² Na rozdíl od validity (přesnosti) nebo reliability (spolehlivosti) měření se ekvivalencí rozumí stálost či neměnnost měření, respektive jeho srovnatelnost. Horn a McArdle [1992: 117] definují invarianci měření jako situaci, kdy operacionalizace konstruktů vyústí v měření zcela totožných znaků, a to i za různých okolností studia zkoumaného fenoménu.³ Již v definici ekvivalence se odráží tři různé úrovně srovnatelnosti, a to 1) srovnatelnost teoretického konceptu, 2) srovnatelnost položek, které daný koncept měří, a 3) srovnatelnost jejich škál měření (tj. kategorií odpovědí).

Srovnatelnosti měření je většinou dosaženo poté, kdy respondenti vnímají, chápou a interpretují daný koncept, měřící položku a škálu měření úplně stejným nebo alespoň velmi podobným způsobem i přesto, že pocházejí z různých kulturních prostředí a obsahy položek jsou přeloženy do různých jazyků. Ukazuje se, že právě toto je velmi podstatný problém, který se projevuje obzvláště při formulaci postojových otázek, které by měly zachycovat hodnoty, názory a stanoviska respondentů k různým společenským tématům. Protože na rozdíl od tzv. objektivních proměnných, jako je dotaz na vzdělání či příjem respondenta⁴, je percepce těchto položek mnohem více závislá na kontextu

2 Současně je to i detekce chyb měření, mapování rozsahu chybějících odpovědí, návratnosti atp.

3 Různými okolnostmi se má na mysli různý čas, různé populace anebo různé metody sběru dat.

4 Výzkumníci věnující se rozvoji konceptu ekvivalence položek nejen v psychologii, ale i v sociálních vědách čím dál častěji upozorňují, že je nezbytné věnovat pozornost vedle postojových otázek také těmto objektivním položkám (resp. sociodemografickým charakte-

dotazování – tzn. na kulturních, společenských, politických, právních a ekonomických poměrech dané země v daném čase, na rodném jazyku, tradicích, sociální struktuře atp. – a tudíž jsou tyto položky více náchylné k systematické chybě měření.⁵

Otázkou pak zůstává, zda v situaci, kdy se dotazujeme například na problematiku násilí na ženách, antisemitismu, individualismu a kolektivismu, žebříčku hodnot, sociálního kapitálu, důvěry v politické instituce, autorit či nepotismu, hovoříme o tom samém jako respondent. Ukazuje se, že tomu tak vždy není. Například pojem „být tradiční“ má v různých zemích jiný význam, například v Bavorsku religiozní („být katolík“), jinde může mít spíše charakter životního stylu („mít rodinu, dům, děti“). Jiným příkladem může být různé vnímání a hodnocení frekvence návštěv na základě územní rozlohy země. V územně rozsáhlých zemích (např. Německu nebo Spojených státech) dochází v porovnání s územně malými státy k jinému vnímání otázky „Navštěvujete často své děti?“, neboť zde většinou není možné tak často navštěvovat své příbuzné kvůli velké vzdálenosti. Odpověď „ano, často“ tak může pro Čecha znamenat jednou týdně a pro Američana čtyřikrát za rok [Davidov 2012].

Výjimečně choulostivé jsou v dnešní době často diskutované pojmy spojené s přistěhovalectvím nebo etnickými skupinami. Například interpretace slova „imigrant“ je obzvláště specifická podle kulturní situace v dané zemi. Nejčastěji se uvádějí rozdíly mezi evropskými zeměmi (kde imigrant je přistěhovalec a znamená pro většinu obyvatel sociální problémy) a Izraelem (kde imigrant je Žid se všemi právy a povinnostmi vůči své zemi) [Davidov, Schmidt 2007].⁶ Pojem „etnická skupina“ má pro většinu zemí ve střední a východní Evropě podtext romské sociální skupiny, která je zde také odlišně posuzována [více příkladů viz www.csdiworkshop.org].⁷ A i když respondenti v různých zemích

ristikám) z hlediska jejich kulturní srovnatelnosti [více Mohler, Johnson 2010: 26; Scheuch 1993: 115–116].

5 Většinou se jedná o systematické chyby měření (bias), v nichž se koncentruje skupina rušivých faktorů, které vyplývají ze strukturálních a kulturních rozdílů mezi zeměmi (tzn. vliv tazatelů, rozdílná administrace dotazníků, sociální desirabilita, znalost škály hodnocení předložených výroků, způsoby odpovědí respondentů (response style), odlišné překlady otázek, špatná formulace položek atp.). Tyto rušivé faktory pak zpochybňují srovnatelnost měření v různých kulturních prostředích a výsledky sociálního bádání, neboť měřicí nástroj nemá stejný psychologický význam ve všech zahrnutých kulturních skupinách [Van de Vijver 1998: 43; Harkness, Mohler, Van de Vijver 2003: 13]. Van de Vijver a Leung [1997] identifikovali tři zdroje systematických chyb měření, které ohrožují ekvivalenci položek v mezinárodním výzkumu; konstruktovou (construct bias), metodologickou (method) a položkovou (item) chybu měření.

6 Zjištěno na základě výzkumu ESS, na kterém Izrael již od prvních vln šetření opakovaně participuje.

7 Pomímám v tuto chvíli až nepřekonatelné rozdíly mezi kulturně zcela odlišnými zeměmi (Afrika versus Evropa, Amerika), kde výzkumy některých sociálních jevů, jako je například studium médií (četnost sledování TV, poslouchání rozhlasu, čtení novin a získávání informací z internetu), zcela pozbývá smysl.

konceptu a položené otázce podobně rozumějí, ještě může dojít k velmi rozdílnému použití měřicí škály, která slouží k ohodnocení dané položky (např. výroku) a poskytuje analytikovi číselný skór (odpověď) respondenta, protože respondenti z některých zemí mají silnější tendenci k volbě krajních hodnot škály a zdá se, že tato tendence může být kulturně daná [Smith, Fischer 2008].

Konceptu ekvivalence se tedy výzkumníci z oblasti psychologie, sociologie, vzdělávání, statistiky a výzkumu organizací věnují jak z teoretického hlediska, tak i z empirického již přes 40 let. Ze všech sumarizací lze ale nakonec rozdělit a definovat ekvivalenci na dvou úrovních, které se pohybují na kontinuu od teorie k praxi. Na *teoretické úrovni* se výzkumníci zaměřují především na teoretickou srovnatelnost sociálních jevů a na přenos sociálního, emocionálního a kognitivního významu jednotlivých konceptů do všech jazyků dotazníku. Nejčastěji dochází k důslednějšímu vymezování tzv. konceptuální a překladové ekvivalence. Konceptuální (conceptual) ekvivalence označuje situaci, kdy dochází k sociálnímu, emocionálnímu i kognitivnímu přenosu významu jednotlivých konceptů. Respondenti připisují konceptům stejný význam, což patrně značí, že jejich typické chování se v dané věci příliš kulturně neliší (např. vnímání a projevoování úcty k rodičům, chápání procesu seberealizace, moderní výchova dětí atp.) [Straus 1969; Lonner 1985; van de Vijver, Leung 1997]. To znamená, že zkoumaný sociální jev (resp. jeho reprezentace v podobě konceptu) je kulturně srovnatelný, může být smysluplně diskutován mezi všemi respondenty a je mu podobně rozuměno.

Překladová (translation) ekvivalence nebo jinak lingvistická (linguistic) ekvivalence se zabývá konkrétní formulací otázek, přenosem významu jednotlivých pojmů v procesu překladu, totožnou interpretací otázek napříč světovými jazyky a také čtivostí a přirozeností položek v jednotlivých jazykových verzích dotazníku [Lonner 1985; Scheuch 1993; Blair, Piccinino 2005; Ægisdóttir et al. 2008]. Obecně překladová ekvivalence většinou zaručuje, že položka i její hodnotící škála budou mít i po překladu stejný význam. Tento typ ekvivalence bývá zajišťován zejména kvalitním zpětným překladem a psychometrickými technikami (viz níže), v některých případech bývá při kontrole překladu v pilotáži dotazníku testována podobnost faktorové struktury dat anebo aplikováno statistické třídění dat prvního a druhého stupně.⁸

Na *úrovni měření* jde zejména o to, ověřit si míru podobnosti vybraných indikátorů a jejich hodnoticích škál v mezinárodních datech, která se zákonitě odrazí v možnostech srovnání dat mezi kulturami. Nejmenší míru podobnosti vykazují měřicí indikátory, které jsou tzv. konstruktově (construct) ekvivalentní, jinak též konfigurálně (configural) ekvivalentní. Pokud jsou data konfigurálně ekvivalentní, předpokládá se, že ačkoli v jednotlivých dotaznicích

⁸ Druhů ekvivalence, které označují, že v překladu byly přeneseny adekvátní významy pojmů a že otázky budou respondenti shodně interpretovat, je mnoho, avšak jejich definice nejsou vždy jednoznačné a mohou se i značně překrývat [viz Johnson 1998]. To může být důvod, proč není o těchto druhích ekvivalencí natolik systematicky v odborné literatuře pojednáváno; většinou se shrnují pod jeden obecný pojem: konceptuální ekvivalenci.

není použit stejný indikátor (např. nestejně znějící otázka) a ani měřicí škála o stejném rozsahu a stejné jednotce (záleží na kulturně vhodném použití), všechny jeho varianty jsou totožně interpretovány ve všech kulturních skupinách a odkazují tak ke stejnému sociálnímu jevu. V případě konfigurální ekvivalence dat není zatím možné porovnávat vztahy latentních či manifestních proměnných s jinými proměnnými v datových souborech a průměrné skóry jedinců z takovéto škály mohou být opět srovnávány pouze uvnitř jedné skupiny a ne mezi nimi.

Typy měření, které vykazují metrickou (metric) ekvivalenci, jinak také ekvivalenci měřicí jednotky (measurement unit equivalence) [Van de Vijver, Leung 1997], se vyznačují stejnou jednotkou na škále měření a stejným rozsahem škály (např. Lickertova škála na měření postojů), ale subjektivně jsou rozsah a jednotky škály vnímány respondenty z různých kulturních skupin odlišně. Tato situace nastává například při obecném hodnocení souhlasu s výrokem na stupnici od jedné do pěti – Čechům může tato škála připomínat klasické známkování známé ze školních lavic, kde jedna znamená nejlepší hodnocení a pět nejhorší, pro Rusy by ale tato škála byla převrácená, protože jejich školní klasifikace má opačnou logiku.⁹ Jiným příkladem je hodnocení riskantnosti užívání alkoholu před jízdou automobilem na škále od jedné do pěti – skóre dva může znamenat pro Němce mnohem nebezpečnější chování řidiče než pro Čecha. Měřicí indikátory mají ale větší míru podobnosti než konfigurálně ekvivalentní indikátory. Průměrné skóry jedinců z takovéto škály mohou být přímo porovnávány opět pouze uvnitř jedné kulturní skupiny (např. t-test), ale ne mezi kulturami [Ægisdóttir et al. 2008]. Mezi kulturami ale mohou být porovnávány vztahy testovaných manifestních či latentních proměnných s jinými proměnnými z datového souboru (např. korelační a regresní analýza, rozdíly mezi skóry jedinců, rozdíly různých skupin či kategorií atp.).

Nejvyšší míru podobnosti mají tzv. skalárně (scalar) ekvivalentní indikátory. V tomto případě má škála měření indikátoru v dotazníku ve všech zemích stejný rozsah a stejnou jednotku měření a navíc mají jednotlivé skóry na škále pro respondenty stejný význam a jsou totožně interpretovány; škála je zbavena všech (subjektivních) vlivů a není systematicky zkreslena. Pokud jsou splněny všechny podmínky skalární ekvivalence, lze přímo srovnávat průměrné skóry jedinců v daném indikátoru uvnitř i mezi kulturními skupinami (např. t-test, analýza rozptylu) [van de Vijver, Leung 1997; Steenkamp, Baumgartner 1998; Vandenberg, Lance 2000; Ægisdóttir et al. 2008].

Techniky nastolování a testování ekvivalence ve výzkumu

Problematika ekvivalence na teoretické úrovni, která se dotýká zmíněného srovnávání sociálních jevů v multikulturním prostředí, se především odráží

⁹ Tradiční ruské známkování má tuto podobu (5 = nejlepší známka/výborně, 1 = nejhorší známka/nedostatečně). Některé modernější systémy známkování v Rusku dokonce používají stupnici od 10 do 1 (1 = nejhorší známka).

v praktické úrovni výzkumu při výběru indikátorů (resp. otázek v dotazníku), které mají za úkol tyto teoretické koncepty měřit. Nalézt konstruktově ekvivalentní indikátory do dotazníku není triviální záležitost. Vyžaduje to podle Mohlera a Johnsona [2010: 23] navržení kvalitního výzkumného nástroje, jeho pilotáž a případnou modifikaci ve všech participujících zemích, což je přirozeně časově i finančně značně náročné. Stejně tak i testování dosažené ekvivalence indikátorů při analýze dat není většinou pro výzkumníka jednoduchým úkolem. Jaké techniky lze tedy v těchto případech uplatnit?

K nastíněné problematice lze přistoupit laicky řečeno ze dvou stran. Jeden soubor technik se pokouší již při navrhování dotazníku nalézt teoreticky srovnatelné indikátory, přenést v procesu překladu co nejpodobnější významy jednotlivých pojmů a zajistit jejich shodnou interpretaci respondenty. Tyto snahy jsou vyvíjeny ve fázi formulování a překladu otázek před sběrem dat v terénu. Nejčastěji se v tomto kontextu hovoří o dosahování ekvivalence na konceptuální úrovni, ale bez kvalitní přípravy dotazníku nemůže být dosaženo ani ekvivalence položek na úrovni měření. Druhý soubor technik pracuje až s vytvořeným datovým souborem a nejčastěji ve fázi analýzy dat přímo testuje reálně dosaženou úroveň ekvivalence daných indikátorů ve výzkumu; jak na konceptuální úrovni, tak na úrovni měření.

Jak lze nalézt srovnatelné indikátory před sběrem dat?

Koncept ekvivalence je často spojován (ne-li zaměňován) s problematikou dosahování celkové kvality v mezinárodních výzkumech. K budování kvalitního šetření ale existují tři poměrně odlišné přístupy – statistický, technicko-administrativní a psychometrický –, které dokážou až ve vzájemné kombinaci vygenerovat kvalitně provedený výzkum. Řešení problematiky ekvivalence je v kompetenci až tohoto třetího, psychometrického přístupu ke kvalitě dat.¹⁰ Abychom tedy mohli v datovém souboru pracovat se srovnatelnými indikátory a komparovat tak mezinárodní data, musí být podstatný díl úsilí věnován právě fázi formulování a překladu otázek před sběrem dat v terénu. Jedná se konkrétně o navržení kvalitního výzkumného nástroje, jeho pilotáž a případnou modifikaci ve všech participujících zemích, což je přirozeně časově i finančně značně náročné [Mohler, Johnson 2010: 23]. Důraz je kladen především na přenesení co nejpodobnějších významů jednotlivých pojmů položek v procesu překladu a na zajištění jejich shodné interpretace respondenty. Při řešení problematiky ekvivalence před sběrem dat se nejčastěji hovoří o třech technikách, které by měly vyprodukovat při ideální souhrě kvalitní a srovnatel-

¹⁰ Zde platí, že není možné dosáhnout ekvivalentního měření bez jistoty, že výběrový design a sběr dat neproběhl ve všech zemích srovnatelným způsobem – to vše by totiž mohlo být dalším potenciálním zdrojem systematického zkresení měření a příčinou nesrovnatelnosti škál položek. Na druhou stranu naplnění těchto kvalitativních parametrů organizace sběru dat ještě samo o sobě nezaručuje, že koncepty, položky a jejich škály měření budou kulturně srovnatelné.

né otázky. Tím je relevantní překlad dotazníku, rovnováha při výběru etických a emických indikátorů a použití kognitivních metod ověřování kvality otázek.

Technika překladu dotazníku

Jako jeden z prvních metodologických problémů se při sestavování co nejkvalitnějšího designu mezinárodního šetření objevil relevantní, precizní a ekvivalentní překlad otázek do cizích jazyků. Dodnes je to v odborné literatuře snad nejčastěji zmiňovaná překážka komparativních výzkumů [Harkness 2003: 38; Scheuch 1993: 107]. Proto byla také první pozornost věnována především lingvistické a funkční ekvivalenci v naději, že pečlivý překlad dotazníku povede k odstranění všech systematických zdrojů zkreslení výsledků, a tím pádem k plné ekvivalenci dat [van de Vijver 1998: 51]. Implicitně se tedy téma vývoje srovnatelných otázek a dotazníků promítá již po dlouhou dobu především do problematiky adekvátního překladu mezinárodních dotazníků [Johnson 1998; Harkness 2003; Smith 2003] a překrývá se s oblastí využití psychometrických přístupů při zvyšování kvality mezinárodních šetření.

Jowell [1998] uvádí, že každá druhá země musí dotazník přeložit do vlastních funkčně ekvivalentních slov a frází, a tento proces s sebou přináší nemalé problémy. V počátcích budování kvalitního designu mezinárodního výzkumu se stal výzkumníkům trnem v oku doslovný překlad otázek bez přenesení jejich sociálního a kognitivního významu, který se liší pro různé sociální skupiny, což podstatně komplikuje vzájemné srovnávání otázek. Scheuch [1993: 107–114] uvádí řadu konkrétních příkladů z vlastních výzkumů, kdy je doslovný překlad pojmů zcela nepřipustný. Například Bogardova škála sociální distance, která umožňuje měřit distanci jedince vůči sociální (většinou etnické či národnostní) skupině, musela být v německé verzi dotazníku modifikována. Jeden ze stupňů Bogardovy škály je položka „přijetí příslušníka určité skupiny za souseda“¹¹, přičemž aby mohl být daný význam sociální distance vyjádřen ekvivalentně v němčině, bylo zapotřebí změnit formulaci dané položky na „přijetí příslušníka určité skupiny za známého (se kterým se sice zdravím, ale moc dobře ho neznám)“¹²; pojem souseď měl totiž pro Němce jiný sociální a emocionální význam.¹³

Nejnámější překladatelský přístup používaný v mezinárodních výzkumech je tzv. technika zpětného překladu (back-translation model), která velmi pečlivě kontroluje překladovou ekvivalenci¹⁴ [Harkness, Schoua-Glusberg 1998: 97; Harkness 2003: 41]. Základní logika zpětného překladu spočívá v přeložení

11 V originálním znění: „have as a neighbor in the same street“.

12 V originálním znění: „have as a greeting acquaintance“.

13 Jak dále Scheuch uvádí, daná položka sice neměla v Německu zcela totožnou pozici na stupnici sociální distance, ale měla v analýze dat stejnou ordinální pozici na škále (tzn. stejné umístění v pořadí sedmi položek Bogardovy stupnice), a proto byla celá škála pro potřeby analýzy dat tzv. konstruktově ekvivalentní, a tudíž částečně srovnatelná.

14 Termínu „překladová ekvivalence“ se překladatelé velmi brání a spíše preferují termín kvalita překladu, vhodnost překladu či adekvátnost překladu [Harkness, Schoua-Glusberg 1998: 94].

originálního (zdrojového) dotazníku do rodného jazyka dané země. Poté požaduje zpětný překlad přeložené verze dotazníku do původního jazyka, porovnat zpětný překlad s originálem a případně opět pokračovat v procesu překládání z této druhotně vytvořené verze dotazníku. V procesu překládání se má takto postupovat až k dosažení nejlepšího, nejbližšího možného výsledku překladu. Navíc každý tento krok překladu má ideálně za povinnost provést jiný překladatel, resp. tým spolupracovníků, což umožňuje snížit možná zkreslení. O osobách překladatelů se předpokládá, že jsou opravdu znalé kultury země, do jejíhož jazyka se dotazník překládá, podchytí všechny významy zkoumaného jevu a vyhnou se tak doslovnému překladu; někdy se dokonce zvažuje účast bilingvních překladatelů [Brislin 1976: 221].

Dnes již existují různé variace a vylepšení tohoto modelu zpětného překladu, např. metoda TRAPD – akronym pro translation (překlad), review (revize), adjudication (posouzení), pretesting (předběžné testování), documentation (dokumentace), která je využívána kupříkladu v projektu European Social Survey (ESS). Tato metoda spočívá v provedení dvou nezávislých překladů, jejich revizi a finálním posouzení. Proces překladu postupuje podle specifického protokolu, je pečlivě zaznamenáván každý krok a jednotlivé rozhodovací úseky pro budoucí citace, odkazy a diskuze překladatelů a výzkumníků. Standardně bývá finančně zajištěn tým tří zkušených překladatelů pro překlad, kontrolu a úpravu překladu [Harkness 2003; European Social Survey 2010].

Zda se podařilo překladatelům přenést význam jednotlivých otázek do všech jazykových verzí dotazníků, se většinou zjišťuje pilotáží. Způsobů, jak ohodnotit pravděpodobnou srovnatelnost přeložených otázek v pilotáži, je několik. Lze předložit originální a přeloženou verzi dotazníku bilingvním osobám nebo malé skupině respondentů a porovnávat průměrné skóry položek v obou verzích dotazníku statistickými metodami (např. t-test). Pokud jsou výsledky významně rozdílné, dochází pak k přeformulování otázek nebo jejich vyloučení. Stejně tak může být posuzována souběžná, obsahová a konstruktová validita měřicího nástroje a mezipoložková reliabilita nebo mohou být jednotlivé otázky v obou verzích překladu hodnoceny na základě podobného významu nebo své jednoznačnosti či (kulturně) vhodného použití za pomoci kognitivních rozhovorů [Mullen 2005; Ægisdóttir et al. 2008; více o moderních přístupech k překladu Harkness, Villar, Edwards 2010].

Výběr etických a emických indikátorů

S překladem dotazníku úzce souvisí také problematika výběru emických a etických položek.¹⁵ Emické koncepty či položky se týkají způsobu myšlení,

15 Pojmy vycházejí z lingvistické teorie o specifických znacích jazyka používaných v určitém kulturním prostředí. Fonemika (phonemics) se zabývá hláskami, které existují pouze v jediném konkrétním jazyce (tj. emický) a fonetika (phonetics) hláskami existujícími univerzálně ve všech jazycích (tj. etický) [Harris 1976]. Ve stejném duchu rozlišovali typy otázek i Przeworski a Teune [1970]: „core items/culture general“ (pro etické otázky) a „cultural specific items“ (pro emické otázky).

vzorů chování a témat, které jsou specifické pro dané konkrétní kulturní prostředí (např. postoje k imigrantům, vymezení blízké sociální skupiny). Etické položky pak popisují názory, vzorce myšlení a chování společné všem lidem v jakémkoli kulturním prostředí (např. mezilidské vztahy, sociální distance) [více Davidson et al. 1976; Harris 1976; Ægisdóttir et al. 2008]. Van de Vijver a Leung [1997: 265] zdůrazňují rozdílné přístupy k překladu emických a etických otázek. Tři varianty překladu, které navrhují (tj. application – adaptation – assembly) se liší především v míře modifikace původní otázky ze zdrojového dotazníku. Doslovný překlad otázky je vhodný v případech šetření velmi podobných zemí se stejným kulturním zázemím a jazykem (tedy pro etické položky). Jakékoli další varianty jazykové a obsahové modifikace položek či dokonce tvoření nových, kulturně vhodných (emických) otázek jsou adekvátní ve výzkumu značně rozdílných zemích a tam, kde se očekává větší systematické zkreslení výstupů analýz.

Jak etický, tak i emický přístup k tvorbě dotazníku má několik výhod i nevýhod. Hlavní výhodou etického přístupu je snížení nákladů na náročnost překladu a v případě, že je daná otázka totožně interpretována ve všech kulturních skupinách a odkazuje ke stejnému sociálnímu jevu, také dosažení značně vysoké míry ekvivalence měřicích položek a otevření možností neomezené statistické komparace dat. Emický přístup naproti tomu dokáže zachytit kulturně odlišné aspekty zkoumaného jevu a je významným přínosem z hlediska kulturního porozumění, především pokud zkoumáme velmi rozdílné země; tehdy etický přístup selhává. Bohužel emické otázky většinou nedosahují příliš vysokých úrovní ekvivalence na úrovni měření, což značně omezuje statistické srovnávání mezinárodních dat.

Kvůli množství nevýhod obou přístupů se v současnosti navrhuje kombinovaný „emicko-etický“ přístup k mezikulturnímu transferu dotazníkových metod, kde je patřičná pozornost věnována oběma typům konceptů. Tento přístup preferuje k souboru etických otázek společných všem přeloženým dotazníkům vybrat v každé zemi kulturně specifické otázky, protože ve správném poměru dokreslí kulturní obrázek o daném jevu [viz výzkumná praxe ISSP; <http://www.issp.org/>]. Ve stejné míře se v současnosti doporučuje využívat potenciálu baterií otázek (spíše než jednoduchých otázek) sestávajících i z několika indikátorů, protože jen tímto způsobem je možné statisticky zachytit chyby měření. Brislin [1976: 220] navíc zdůrazňuje aktivní spolupráci expertů v multikulturních týmech, kteří díky hluboké znalosti vlastní země reprezentují svou kulturu při přípravě šetření a snadněji tak nalézají etické a emické indikátory a možné zdroje jejich systematického zkreslení.

Metody využívající kognitivních přístupů

V průběhu pilotáže nemusí docházet pouze k testování kvality překladu otázek do cizích jazyků, lze zkoumat daleko širší problematiku – kvalitu celého výzkumného nástroje. Dostáváme se tak až za hranice sestavování tzv. dobrých otázek („good“ questions), pro které se vymezují obecně stylistická, syntaktická

a gramatická doporučení, k oblasti kognitivních přístupů. Kognitivní přístupy se zabývají psychologickými stránkami získávání informací od respondentů ve standardizovaných dotazníkových šetřeních, v centru jejich pozornosti stojí problematika interpretace otázky na individuální a interpersonální úrovni [Vinopal 2008: 11]. Nejen že díky kognitivním přístupům je možné mapovat, jakým způsobem respondent reaguje na výzkumné otázky a jak chápe a zpracovává položky v dotazníku, ale také lze identifikovat reálné či potenciální zdroje zkreslení odpovědí na úrovni interakce a komunikace respondenta s tazatelem; v mezinárodních šetřeních dokonce i na úrovni kulturních a jazykových kontextů [Braun, Harkness 2005; Schwarz et al. 2010].¹⁶

V mezinárodních výzkumech je tak tematiku kognitivních přístupů obzvláště třeba chápat v širším kontextu výzkumné chyby, neboť jak již bylo řečeno výše, chyby měření jsou významným zdrojem narušování různých úrovní ekvivalence na úrovni měření a podstatně tak ztěžují vzájemné srovnávání indikátorů. Vinopal [2008, 2009] uvádí, že kognitivní přístupy se věnují chybám pozorování, které vznikají na úrovni výzkumného nástroje, tazatele, respondenta a módu sběru dat.¹⁷ A jejich praktickým účelem je mimo jiné tuto chybu redukovat a pokusit se zvýšit kvalitu otázek i celých dotazníků [Vinopal 2009: 399]. Jaké konkrétní metody testování dotazníku tedy kognitivní přístupy využívají?

Vinopal [2008: 49] podotýká, že každá technika může sloužit různým cílům, přináší jiný typ poznatků a má různý potenciál ve smyslu možného vylepšení cílových otázek. Techniky, které lze v současnosti použít k ověření kvality dotazníku, jsou kognitivní rozhovor, debriefing, kódování chování účastníků rozhovoru, posouzení kognitivními experty, měření reakčního času atd. [více Vinopal 2008: 51–57]. Pro účely mezinárodního srovnávání je nejčastěji využíváno kognitivních rozhovorů a metody debriefingu.

Kognitivní rozhovor odkrývá výzkumníkům to, co je ukryté pod běžně sledovaným povrchem odpovědí na věcný dotaz [Vinopal 2008: 51], umožňuje poznat, proč respondent odpověděl na otázku daným způsobem. Otázky typu „Jakým způsobem jste dospěl ke své odpovědi?“, „Jak byste otázku vyjádřil vlastními slovy?“, „Co podle Vás tato otázka/výrok znamená, jaký má podle Vás význam?“ [Pan et al. 2010: 92] jsou základním stavebním kamenem kognitivních rozhovorů, které mají většinou podobu hloubkových rozhovorů a využívají techniky „přemýšlení nahlas (thinkaloud)“ či „zpětného dotazování (retrospective probing)“ [více Vinopal 2008: 51–53]. V mezinárodních výzkumech se tyto rozhovory většinou realizují s bilingvními tazateli za účasti kombinace bilingvních a monolingvních respondentů [více Goerman, Caspar

16 Zvláštní zřetel je brán na překladovou ekvivalenci škál odpovědí, efekt kontextu nabídky odpovědí a preferenci jedné kategorie odpovědí před jinými.

17 Chyby pozorování mohou mít jak systematický, tak i nesystematický charakter, přičemž v problematice ekvivalence je zdůrazňován především vliv systematických chyb měření na srovnatelnost jednotlivých otázek [více van de Vijver, Leung 1997].

2010], neboť se nejčastěji testuje překladová ekvivalence. Navíc se předpokládá, že sladění tazatele s respondentem obzvláště v sociodemografických charakteristikách, jazyku a etniku hraje při kognitivních rozhovorech ještě větší roli z hlediska systematického zkreslování výsledků než při standardizovaném dotazníkovém šetření v terénu. Proto se na tuto souhru klade o to větší důraz [Goerman, Caspar 2010: 79].

Metoda debriefingu pracuje již většinou přímo s tazateli, kteří mají za úkol zhodnotit průběh rozhovoru a odhalit slabá místa v dotazníku [více Vinopal 2008: 54]. Tazatelé sdělují výzkumníkům své zkušenosti z rozhovorů, komentují interpretaci jednotlivých slov v otázce, hodnotí, zda je otázka kulturně vhodně zvolená, uvádějí své dojmy z respondentů a z jejich reakcí na otázky. V mezinárodních výzkumech jsou to většinou jazykoví experti anebo rodilí mluvčí, kteří zaznamenávají kulturní normy mluvení a preferovaný styl komunikace respondentů v různých zemích [Pan et al. 2010: 99].

Jak lze testovat ekvivalenci měřících indikátorů po sběru dat?

Nutno podotknout, že i fázi analýzy dat z hlediska testování ekvivalentních otázek byla v zahraniční odborné literatuře věnována již velká pozornost, avšak až o mnoho let později než překladu dotazníků. Hlavní příčina by se dala přisoudit především nedostatečnému rozvoji výpočetní techniky v 60. letech a bleskovému zdokonalování softwarových možností v současnosti, neboť většina technik testování dosažené ekvivalence indikátorů na úrovni měření vyžaduje mnohorozměrnou, složitou a velmi komplexní statistickou analýzu. Časová a finanční náročnost takového testování a nejistota většiny sociálních výzkumníků ohledně možností jeho realizace posunula analytické a metodologické práce na tomto tématu až do posledních přibližně 30–40 let [Johnson 1998: 2]. Zájem o to, jak nejlépe mezinárodní šetření kontrolovat, se ale stále zvyšuje, z tohoto důvodu se také testování ekvivalence dat postupně stává v současnosti nedílnou součástí mnoha mezikulturních studií.

Obecně lze konstatovat, že v případě testování srovnatelnosti indikátorů až po sběru dat hrají primární roli statistické techniky. Zde uvedený přehled je řadí od základních až po ty nejsložitější analytické přístupy; byl sestaven na základě výčtu technik Brauna a Johnsona [2010: 377] a Johnsona [1998: 25–29] s podporou informací z řady dalších empirických studií, kde byly dané techniky aplikovány v praxi při analýze dat nebo vzájemně porovnávány. Kritériem pro výběr vhodné techniky může být kromě obvyklé volby, zda do analýzy vybrat kategoriální, či kardinální proměnnou, také její analytická a potažmo časová náročnost, počet zemí a počet proměnných, které lze zahrnout do analýzy, anebo míra podrobností, které potřebujeme zjistit o nesrovnatelných položkách v dané zemi a příčinách této situace [Braun, Johnson 2010: 376].

Statistická analýza prvního a druhého stupně

Jednoduchá analýza položek není při testování jejich srovnatelnosti mezi zeměmi příliš běžnou volbou, ale velmi lehce se na data aplikuje a umožňuje

bezprostřední a rychlé pochopení toho, zda obecně pracujeme se srovnatelnými či nesrovnatelnými daty. Toto je obzvláště vhodné pro výzkumníky, které zajímá především zkoumaný problém a ne metodologická analýza [Braun, Johnson 2010: 375], ale přirozeně mají tyto techniky svá slabá místa. Konkrétně se jedná například o analýzu procentuálního rozložení (distribuce) jednotlivých kategorií odpovědí a jeho variability napříč zeměmi. Důležitou informací je také údaj o respondentech, kteří na danou otázku neodpověděli (tzn. chybějící případy), protože právě tento údaj může poukazovat na problém s interpretací či překladem dané otázky [Braun, Johnson 2010: 379]. Tyto varianty testů jsou přirozeně ošidné při větším množství zkoumaných zemí, protože výstupy jsou velmi nepřehledné [Braun, Johnson 2010: 378].

Dále je možné porovnávat průměry ordinálních a kardinálních proměnných mezi dvěma zeměmi (t-test) anebo mezi více zeměmi (analýza rozptylu). Braun a Johnson [2010: 380] ale důrazně upozorňují na to, že tento test lze realizovat, pouze pokud jsme si jisti tím, že dané položky jsou skalárně ekvivalentní. Proto navrhuji porovnávat průměry různých proměnných mezi zeměmi a na základě toho sledovat pouze vytvořené pořadí zemí. Pokud mají dané proměnné měřit stejný jev (tzn. latentní proměnnou), měly by také podle průměrů vytvořit stejné pořadí zemí. Jestliže tomu tak není, indikátory pravděpodobně neměří stejný jev, a mohou tak výrazně zkreslit výsledky kulturního srovnávání, protože ty se budou lišit v závislosti na vybraném indikátoru [Braun, Johnson 2010: 380].¹⁸ Třetí varianta připadající v úvahu je srovnávání korelace měřících indikátorů reprezentujících latentní proměnnou s referenčními položkami anebo socio-demografickými proměnnými (či jinými vysvětlujícími proměnnými), s nimiž by měly vykazovat podobné vztahy napříč zeměmi, pokud mají mít ve všech kulturách stejný význam [více Johnson 1998: 25; Braun, Johnson 2010: 380–381].

Mnohorozměrné statistické analýzy

Složitější statistické techniky vyžadují od výzkumníka daleko více trpělivosti a matematických dovedností. Většinou na zjištění toho, že máme v datovém souboru nesrovnatelné otázky, připadá mnohem větší díl analytické práce než při aplikaci jednoduchých technik, což ještě v součtu s nepříliš uživatelsky příjemnými statistickými programy činí z těchto metod velmi nepopulární analytický nástroj. Navíc nedodržování poměrně striktních pravidel při testování dat mnohorozměrnými technikami může vést až k takovému zkreslení výsledků, že na problém s neekvivalentními daty jednoduše nepřijdeme [Braun, Johnson 2010: 376]. Čím jsou pro nás tedy užitečné? Jejich hlavní výhoda tkví především v podrobnosti zjištění, která máme o (ne)srovnatelných položkách, a v počtu zemí a proměnných, které můžeme do analýzy zahrnout.

¹⁸ Podobným způsobem lze analyzovat další míry centrální tendence (modus, medián), střední chybu průměru i jiné míry variability (resp. míry rozptýlenosti), interval spolehlivosti atp.

Konkrétních analýz, které lze pro účely testování ekvivalence indikátorů použít, je mnoho: explorační faktorová analýza [Welkenhuysen-Gybels, van de Vijver 2001], testování reliability škál [Devins et al. 1997], mnohorozměrná korespondenční analýza [Blasius, Thiessen 2006], mnohorozměrné škálování [Braun, Scott 1998; Fontaine 2003], strukturální modelování, víceurovňové modelování, analýza latentních tříd a teorie položkové analýzy (IRT). Avšak z praktických důvodů bude nadále vhodnější podrobněji hovořit jen o těch analýzách, které jsou nejčastěji a s nejlepšími výsledky aplikovány na mezinárodních datech. Z tohoto důvodu udávám u prvních čtyř metod pouze odborné studie, ve kterých se lze o povaze jejich metodologického využití při testování ekvivalence otázek dozvědět více, aniž bychom se nadále zabývali jejich hlubšími technickými popisy (souhrnně viz Tabulka 1).

Strukturální modelování (multigroup structural equation modeling, SEM)

Strukturální model je formální (grafickou a matematickou) reprezentací určité teorie (komplexní hypotézy) o kauzálních vztazích mezi proměnnými, přičemž se často očekává, že sledované jevy a kauzální vztahy mezi nimi reflektují hlubší souvislosti z reality [Matějů 1989: 402]. Výhodou strukturálního modelování je především to, že umožňuje pracovat nejen s manifestními (přímo měřenými) proměnnými (v tom se příliš neliší od regresní analýzy), ale též s latentními proměnnými, které jsou definovány přímo v modelu (podobně jako ve faktorové analýze), proto lze na strukturální modelování pohlížet jako na kombinaci faktorové a regresní analýzy [Matějů 1989; Hox, Bechger 1998: 1] s tím, že je obvykle posuzováno spíše jako konfirmační než explorativní technika.¹⁹

Modely jsou typicky zobrazovány grafickým diagramem vycházejícím z pěšinkové analýzy (path analysis), kde jsou měřicí indikátory umístěny do čtverců, latentní proměnné do ovalů a chyby měření do kroužků. Základem je odlišený modelu měření (measurement model), který přesně stanovuje, který měřicí indikátor náleží k dané latentní proměnné, a strukturálního modelu (structural model), který popisuje vzájemné vztahy mezi těmito latentními proměnnými a je již oproštěn od variance chyb měření daných indikátorů, což je jeho velká výhoda [více Matějů 1989: 408]. Tento statistický model je pak obvykle matematicky reprezentován řadou strukturálních (lineárních regresních) rovnic, které představují dané kauzální vazby. V průběhu analýzy se odhadují nejen parametry modelu (zpravidla regresní koeficienty, chyby odhadů, vyčerpané variance atp.), ale posuzuje se i jeho celková vhodnost, obvykle na základě chí-kvadrát testů, které jsou ale příliš náchylné na velikost výběrového souboru.²⁰ Z tohoto důvodu se dnes prosazují i jiné statistiky vhodnosti modelu

19 Do souboru technik strukturálního modelování lze zahrnout také několik tradičních mnohorozměrných analýz jako speciální případy modelování, např. korelační analýzu, regresní analýzu, kanonickou korelační analýzu, diskriminační analýzu a faktorovou analýzu [Hox, Bechger 1998: 1].

20 To je značný problém při zařazení velkého množství zemí do analýzy, protože i když náš model měří data velmi dobře, testy vhodnosti založené na chí-kvadrát testech mají

Tabulka 1. Vybrané mnohorozměrné statistické techniky pro testování ekvivalence měřících indikátorů po sběru dat

Metoda	Vhodný typ proměnné	Empirické studie – aplikace metody na datech	Typ ekvivalence
Explorační faktorová analýza (Principal component analysis)	Kardinální data Ordinální data*	Welkenhuyesen-Gybelis, van de Vijver 2001	Konstruktová
Testování reliability škál	Nominální data (CATPCA) Kardinální data Ordinální data*	Devins et al. 1997	Překladová
Mnohorozměrná korespondenční analýza	Kategorální data (grafická metoda)	Blasius, Thiessen 2006	Konstruktová
Mnohorozměrné škálování	Ordinální data (grafická metoda)	Braun, Scott 1998; Fontaine 2003	Konstruktová
Víceúrovňové (hierarchické) modelování	Kardinální data Ordinální data*	van de Vijver, van Hemert, Poortinga (eds.) 2008	Koncepтуální (ítem bias)
Strukturální modelování	Kardinální data Ordinální data* Nominální data (za podmínky vypočítání matice jiných než Pearsonovských korelací, např. polychorických; PRELIS)	Mullen 1995; Billiet 2003; Byrne, Watkins 2003 van Hemert et al. 2002; Davidov et al. 2012	Konstruktová (konfigurální) Metrická Skalární
a) strukturální modelování		Mullen 1995; Billiet 2003; Byrne, Watkins 2003	
b) víceúrovňové strukturální modelování		van Hemert et al. 2002; Davidov et al. 2012	
c) konfirmační faktorová analýza		Hsueh et al. 2005; Lievens et al. 2007; Davidov 2010	
d) Multitrait-multimethod design (MTMM)		Saris 2003; Saris, Gallhofer 2007; Revilla, Saris 2011	Koncepтуální (method bias)
Analýza latentních tříd	Kategorická i kardinální data	Kankaras, Moors 2009	Konstruktová Metrická Skalární
Teorie odpovědi na položku (Item response theory)	Dichotomická data Ordinální proměnné	Wang, Russel 2005; May 2006; Woehr, Meriac 2010	Překladová Skalární (ítem bias)

Zdroj: vlastní zpracování podle Brauna a Johnsona [2010: 377]. Poznámka: * Ordinální proměnné jsou přípustné, pouze pokud je splněna základní podmínka, že vzdálenosti (intervaly) mezi jednotlivými kategoriemi (čísly) je možné považovat za stejné, rovnocenné. Takových proměnných není v sociologii mnoho, ale občas se vyskytují; v ideálním případě jde o proměnné intervalové.

(např. BIC, CFI, RMSEA), které zohledňují za prvé (ještě) akceptovatelnou jednoduchost modelu, chyby měření a také míru, do jaké testovaný model reprodukuje výchozí kovarianční matici [více Meade, Johnson, Braddy 2006].

Ve strukturních modelech je ale také navíc možné otestovat hypotézu o ekvivalenci položek na úrovni měření prostřednictvím několikavýběrové konfirmační faktorové analýzy (MGCFAs), což je v současnosti nejvyužívanější postup ověřování ekvivalence dat.²¹ V prvním kroku analýz je klasicky ověřena konstruktová, resp. konfigurální ekvivalence dat. Konkrétně se testuje podobnost modelů měření, které by měly mít stejnou faktorovou strukturu v jednotlivých skupinách/výběrech, tzn. statisticky významně velké faktorové zátěže měřicích položek a stejný počet faktorů ve všech zemích [Meredith 1993; Steenkamp, Baumgartner 1998; Vandenberg, Lance 2000]. Je to základní typ ekvivalence mezi daty, který ale ještě neumožňuje přímé statistické porovnávání zemí mezi sebou v daných proměnných; pouze z něj vyplývá, že daný teoretický jev je podobně chápán ve všech kulturách a je také podobně měřen v každé zemi.

Druhým krokem je analýza metrické ekvivalence, kdy se testuje shoda velikostí faktorových zátěží (směrnic) položek v modelech měření ve všech zemích [Vandenberg, Lance 2000: 37]. Pokud je metrická ekvivalence dat potvrzena, škály měření daných indikátorů jsou srovnatelné mezi zeměmi z hlediska jednotky a rozsahu, ale ještě stále není jisté, zda je škála měření používána respondenty stejným způsobem. Proto přichází v úvahu pouze porovnávání vztahů latentních proměnných a měřicích položek s ostatními proměnnými z datového souboru napříč zeměmi (např. korelační a regresní analýza). Nejvyšší úroveň, tzv. skalární ekvivalence dat, se ověřuje testem shody velikosti konstant měřicích indikátorů v modelu měření.²² Skalární ekvivalence dat je nutná podmínka pro komparaci takových statistických ukazatelů, jako je aritmetický průměr měřicí položky či průměr latentních proměnných (resp. faktorů). Posledním typem ekvivalence je tzv. striktní ekvivalence, která vypovídá o shodě variancí chyb měření položek v daných skupinách a na základě ní lze ověřit podobnost reliability položek.²³

Velmi zvláštním případem je ve strukturním modelování model MTMM (multitrait-multimethod model), který má v případě testování ekvivalence tendenci náš model zamítat [Hox, de Leeuw, Brinkhuis 2010: 397–398].

21 Strukturní analýzy, které testují, zda jsou parametry daného modelu pro danou populaci skutečně signifikantně odlišné od parametrů získaných na jinak definované populaci, se nazývají několikavýběrové analýzy (multisample analysis) [Matějů 1989: 406].

22 V tuto chvíli je zřejmá kombinace regresní a faktorové analýzy, neboť dochází k testování parametrů v regresních rovnicích (tj. směrnic a konstant), které reprezentují vztahy mezi faktory a měřicími položkami [více Hox, Bechger 1998: 1].

23 Je možné dále v analýzách pokračovat a testovat srovnatelnost strukturních částí modelů, jako je ekvivalence kovariancí a variancí faktorů, ale tyto testy vypovídají více o konstruktové validitě měřicího nástroje spíše než o ekvivalenci položek na úrovni měření.

položek speciální úlohu, a to odhalit možný zdroj systematického zkreslení položek v metodě sběru dat.²⁴ Za prvé tak lze vytvořit ve strukturním modelování model měření sestávající z jedné latentní proměnné, která je měřena nejméně třemi měřicími položkami pocházejícími ze stejného sběru dat (např. CAPI), a porovnat jej na základě konfirmační faktorové analýzy s modelem měření založeném na položkách z jiného sběru dat (např. CAWI). Za druhé lze za podmínky měření položek různými metodami sběru dat ověřit ekvivalenci a chyby měření tzv. jednoduchých položek, které nejsou součástí baterie položek; latentní proměnnou potom měří tři stejné položky sebrané v šetření různými metodami sběru dat. Tento přístup je poměrně novátorský a konečně umožňuje ověřovat srovnatelnost jednoduchých položek [více Revilla, Saris 2011].

Nevýhoda strukturních modelů tkví v tom, že do modelu nejsou zahrnuty různé úrovně měření tak, jak je tomu v případě víceúrovňového modelování, čímž výzkumník přichází o možnost vysvětlovat rozdíly mezi průměry latentních proměnných nějakou kontextuální proměnnou (např. zemí, kulturou) a nacházet tak možné příčiny systematického zkreslení položek [Hox, de Leeuw, Brinkhuis 2010: 400]. Z tohoto důvodu se začínají v současnosti prosazovat pro testování ekvivalence indikátorů víceúrovňové strukturní modely (multilevel structural equation modeling, MSEM), kde se testuje ekvivalence modelů měření a strukturních modelů na každé hierarchické úrovni dat zvlášť (tzn. položka, respondent, země).

Víceúrovňové modelování (multilevel modeling)

Víceúrovňové (hierarchické) modelování je metoda speciálně vyvinutá pro statistickou analýzu dat, která mají hierarchickou či shlukovou strukturu [Hox, de Leeuw, Brinkhuis 2010: 401].²⁵ Principiálně vycházejí z konceptu kontextuální analýzy, resp. ze skutečnosti, že mnohé proměnné nemá smysl měřit na individuální úrovni (micro level), ale až na úrovni větších celků (macro level), a že působení faktorů na těchto dvou úrovních může být různé [Soukup 2006:

24 Poprvé byl MTMM design představen Campbellem a Fiskem v roce 1959, kteří zjistili, že otázky v otaznících interagují s konkrétní metodou sběru dat, což má pak za následek rozdílnou interpretaci otázek či narušení vztahů mezi proměnnými [Saris 2003: 268]. Vznikají tedy náhodné i systematické chyby ve výzkumu, které od sebe nelze odlišit, pokud se nepřistoupí k více experimentálnímu designu výzkumu. Základem tohoto nového MTMM designu jsou opakovaná měření několika proměnných různými metodami současně (např. dotazováním pomocí papírových dotazníků, telefonu či počítače), většinou v kombinaci minimálně tří otázek v dotazníku a tří využitých metod sběru dat [Řehák 1998]. Postupně se provádí metaanalýza dosavadních poznatků z MTMM experimentů (které k současnému datu otestovaly více než 3 000 otázek), na jejímž základě se v současnosti buduje webová aplikace Survey Quality Prediction [SQP 2.0; více www.sqp.nl]. Ta umožňuje výzkumníkům vytvářet otázky a předpovídat následně jejich kvalitu z hlediska validity, reliability a efektu metody z různých parametrů daných otázek [více Saris, Gallhofer 2007].

25 Většinou se jedná o data, která vycházejí z vícestupňového náhodného výběru, a jejich analýza tak musí plně respektovat strukturu (design) výzkumu [Soukup 2006: 993].

988]. Soukup [2006: 989–994] navíc zdůrazňuje, že až v případě, že zohledníme hierarchickou strukturu v datech a budeme využívat v analýze hodnot měřených na mikro- i mikroúrovni zároveň, lze správně zobecňovat závěry analýzy na celý základní soubor (např. jedince, třídy a školy).

Víceúrovňové modely lze chápat ze statistického hlediska jako zobecnění klasické lineární regresní analýzy, kdy ale nezůstaneme jen při modelování jedné (první) úrovně (např. žák), ale modelujeme každou relevantní úroveň zvlášť (třída, škola, země atp.), protože nemůžeme jednotlivá pozorování v rámci jednotlivých skupin považovat za vzájemně nezávislá [Soukup 2006: 990]. Jsou tedy založeny na víceúrovňové regresní analýze a reprezentuje je hierarchický systém regresních rovnic [Hox, de Leeuw, Brinkhuis 2010: 401]. Jejich velkou výhodou je možnost zařadit do analýzy velké množství zemí.

Pokud chceme testovat srovnatelnost proměnných víceúrovňovým modelováním, postupujeme při modelování vztahů tak, že odpověď na testovanou položku považujeme za závislou proměnnou. V tomto případě pak naše položka představuje první úroveň, respondent druhou úroveň a zkoumané země třetí úroveň modelu. Tímto způsobem lze zjistit, co nám způsobuje varianci odpovědí, resp. položkovou systematickou chybu (item bias) [Braun, Johnson 2010: 387]. Nevýhoda této metody tkví ve špatné identifikaci odlehklých pozorování – například ve které zemi došlo k neekvivalentními překladu otázky –, proto je pro účely testování ekvivalence vhodnější zkombinovat strukturní a víceúrovňové modelování dohromady [podrobněji Braun, Johnson 2010: 386–390].

Pro přehled metodologické aplikace těchto metod doporučuji následující empirické studie: konfirmační faktorová analýza (Hsueh et al. [2005]; Lievens et al. [2007]; Davidov [2010]), strukturní modelování (Mullen [1995]; Billiet [2003]; Byrne, Watkins [2003]), víceúrovňové modelování (van de Vijver et al. [2008]), víceúrovňové strukturní modelování (van Hemert et al. [2002]; van de Vijver, Poortinga [2002]; Davidov et al. [2012]).

Teorie odpovědi na položku (item response theory, IRT)

Teorie odpovědi na položku (item response theory, IRT)²⁶ je moderní metodologický přístup, který je již dlouhodobě využíván v psychometrii [Stout 2002]. Přispívá ke zvyšování kvality sociologických výzkumů tím, že na základě modelů umožňuje uspokojivě zodpovídat otázky funkce každé položky v dotazníku, její rozlišovací schopnosti a informačního přínosu pro test (tj. dotazník) [Rudá 2012: 7]. V oblasti metodologie mezinárodních výzkumů ji lze efektivně využít při sledování toho, zda byl zachován v průběhu překladu dotazníku význam daných položek, při posuzování relevance výběru kategorií a jejich optimálního počtu do škály odpovědí a především

26 V české odborné literatuře také známá jako teorie položkové analýzy.

při analýze položkové systematické chyby [Van de Vijver 2003: 216; Meade, Lautenschlager 2004].

Matematické modely vniklé na základě IRT znázorňují vztah mezi latentní proměnnou a odpovědí na konkrétní položku, což IRT značně odlišuje od klasické teorie testů (CTT), které hodnotí validitu a reliabilitu na základě jednoduchých statistických ukazatelů součtů a průměrů. IRT modely jsou založeny na pravděpodobnosti (tj. šancích), s jakou určitý respondent odpoví na konkrétní položku v souladu s jeho úrovní měřené latentní proměnné; tím se podobají logistické regresi [Stout 2002; Rudá 2012]. V modelech jsou opět odhadovány určité parametry, stejně jako je tomu u konfirmační faktorové analýzy. Hlavní rozdíl mezi nimi ale spočívá v tom, že IRT pracuje většinou s každou položkou odděleně, ale CFA odhaduje parametry všech měřících položek (tj. celé baterie položek) v dané latentní proměnné najednou.

IRT sleduje ekvivalenci položek především z hlediska tzv. odlišného fungování položek (differential item functioning, DIF), což je v terminologii IRT vyjádření pro systematickou chybu měření (bias). Systematické zkreslení může být v kontextu teorie odpovědi na položku identifikováno dvěma různými způsoby. Jedním z nich je porovnání parametrů pomoci klasického chí-kvadrátu²⁷, druhým je porovnání plochy oblasti mezi charakteristickými křivkami položek (ICC) v různých kulturních skupinách. V praxi je také důležitá vizuální analýza, porovnání tvarů funkce pro každou skupinu odděleně [Jelínek, Květoň, Vobořil 2011: 89]. Tyto křivky reprezentují pravděpodobnost označení dané kategorie odpovědi v rámci sledované položky, a pokud si křivky vzájemně odpovídají ve vybraných populacích (tzn. pravděpodobnost výběru odpovědi na danou otázku je totožná ve všech skupinách pro danou latentní proměnnou), je potvrzeno, že jednotlivé položky měří pro každou ze skupin stejný konstrukt, a že jsou tudíž srovnatelné [Bhalla, Lin 1987: 282; Johnson 1998: 26; Woehr, Meriac 2010: 423].

Tato metoda pracuje se statistickým modelem s jedním základním předpokladem o jednodimenzionalitě latentní proměnné reprezentované několika měřícími indikátory [Jelínek, Květoň, Vobořil 2011]. Právě striktní požadavek na jednodimenzionalitu latentní proměnné je hned vedle dalšího požadavku na velký počet položek zařazených do analýzy, které by měly být měřeny ideálně dichotomickou škálou odpovědí, podnětem řady kritik na účet IRT.²⁸ Navíc je teorie položkové analýzy krajně nepraktická v případě velkého množství

27 Tzv. odlišné fungování položky se prokáže, pokud má položka statisticky významně odlišný odhadnutý parametr a , který vyjadřuje schopnost diskriminace položky a někdy je také označován jako směrnice charakteristické křivky položky, nebo odlišný parametr b , který značí obtížnost položky a někdy je také připodobňován ke konstantě položky v CFA modelu, i když jeho funkce a interpretace je zcela odlišná.

28 Zahraniční literatura na téma různých statistických variací tohoto modelu IRT je ale v současnosti už poměrně bohatá. Nejznámější IRT model pro ordinální proměnné je model stupňovaných odpovědí (GRM), viz Stout [2001, 2002], May [2006] atp.

zemí zařazených do analýzy [Braun, Johnson 2010: 390].²⁹ Na druhou stranu několik studií již srovnávalo výsledky testování ekvivalence prostřednictvím CFA i IRT a dochází k tomu, že a) tyto dvě metody mají jiné předpoklady (linearitu/nelinearitu modelů, jednodimenzionalitu/vícemimenzionalitu faktorů), b) testy mohou identifikovat i velmi rozdílné problematické položky, c) v některých případech může IRT generovat podrobnější a užitečnější informace o povaze systematického zkreslení položek, d) žádná z těchto metod není bez nevýhod a e) nejlepších výsledků lze dosáhnout při aplikaci jich obou [Meade, Launtenschlager 2004; Raju, Laffitte, Byrne 2002]. Pro přehled metodologické aplikace této metody doporučuji následující empirické studie: Wang, Russel [2005]; May [2006]; Woehr, Meriac [2010].

Analýza latentních tříd (latent class modeling, LCM)

Mnohorozměrná analýza latentních tříd, která vychází z analýzy latentních struktur, významné metodologické inovace P. Lazarsfelda [více Jeřábek, Soukup 2008], obecně hledá latentní třídy podobných respondentů ve sledované populaci a odhaduje jejich ideální počet. Respondenti jsou rozřazováni do latentních tříd na základě kategorických měřicích indikátorů, i když v současnosti se možnosti této analýzy rozrůstají i o volbu ze spojitých proměnných; v tom je její obrovská výhoda oproti jiným technikám [Hox, de Leeuw, Brinkhuis 2010: 403–404]. Analýza latentních tříd dokáže zručně kombinovat některé výhody strukturního modelování a víceúrovňového strukturního modelování tím, že modeluje rozdíly mezi zeměmi jako rozdíly mezi latentními třídami respondentů, jejichž počet je ale menší než počet zemí [Hox, de Leeuw, Brinkhuis 2010: 396].

V případě, že tuto vícerozměrnou metodu používáme na testování srovnatelnosti indikátorů, postup ověřování ekvivalence se značně shoduje s přístupem konfirmační faktorové analýzy v modelu měření. Předpokládáme, že latentní model má víceméně podobu konfirmačního faktorového modelu, ve kterém zjišťujeme rozdíly mezi latentními třídami [Hox, de Leeuw, Brinkhuis 2010: 404]. Dále už tato metoda funguje z metodologického (ne statistického) hlediska na obdobném principu jako strukturní modelování. Dochází k postupnému testování shody několika modelů (a jejich parametrů) – homogenního, částečně homogenního a heterogenního –, lišících se mírou shody latentních struktur, silou vztahu mezi latentní třídou a indikátory a vlivem proměnné země (příslušnosti k dané skupině) na latentní proměnnou a potažmo měřicí položky. Pro přehled metodologické aplikace této metody doporučuji následující empirickou studii: Kankaras, Moors [2009].

29 Kritika někdy zasnívá i ze strany psychometrie, například podle Borga [1998: 147] nemůže žádný statistický model zajistit správný překlad otázek, protože položky s podobnými statistickými charakteristikami ještě nemusejí zachycovat zkoumaný sociální jev se všemi specifickými kulturními kontexty.

Závěr

Pokud bychom měli odpovědět na otázku, zda je lepší řešit problematiku ekvivalence otázek před sběrem dat, anebo až po sběru dat, jako vyčerpávající odpověď se jeví to, že jsou obě varianty nezbytné pro validní mezinárodní srovnávání. Problematiku ekvivalence je zcela jistě nezbytné řešit *před sběrem dat*, neboť bez toho jen těžko dosáhneme kvalitních výsledků. Ukazuje se, že první fáze výzkumu – vývoj srovnatelných otázek, jejich překlad, tvorba dotazníků – většinou přináležejí centrálním či národním koordinačním týmům daných projektů, které se problematice ekvivalence již po dlouhou dobu vytrvale a systematicky věnují v přípravné a realizační fázi výzkumu.³⁰

Na druhou stranu i přes velké úsilí vytvořit a předložit respondentům (metodologicky) kvalitní dotazník s plně ekvivalentními otázkami a srovnatelnou nabídkou odpovědí se tyto snahy nemusejí setkat se stoprocentní úspěšností, obzvláště v případě velkého počtu participujících zemí. Položková systematická chyba může zkreslit škálu měření do různé míry, a rozsah statistické srovnatelnosti dat tak nelze předem odhadnout. Z mnoha příkladů z praxe také plyne, že i mezinárodní výzkumy se kontinuálně rozvíjejí a kvalita jejich výsledků roste společně s množstvím opakovaných šetření.³¹ Pokud tedy výzkumník přistoupí k sekundární analýze prvních sérií sběru dat či se pokusí získat časový vývoj sociálního jevu, je možné, že se začne potýkat s komparací ne zcela ekvivalentních indikátorů.³² Z tohoto důvodu je velmi důležité otestovat reálně dosaženou úroveň ekvivalence i *po sběru dat*. Ukazuje se, že i tam, kde to nečekáme, nelze občas připustit určitý druh komparace mezinárodních dat nebo je nezbytné vyřadit určitou zemi z komparativní analýzy, aby nedocházelo ke zkreslování celkového kulturního obrazu.

Tato stať předložila čtenáři téma ekvivalence měřicích indikátorů v mezinárodních výzkumech především z metodologického hlediska. Testování srovnatelnosti jednotlivých proměnných napříč různými kulturami je přirozeně součástí komplexních snah o dosažení kvalitních výsledků komparativního

30 Příkladem může být struktura organizace realizace mezinárodního sociálního výzkumu ESS (European Social Survey) s centrálním koordinačním týmem ve Velké Británii [více <http://www.europeansocialsurvey.org>]. Stejně tak je tomu u projektů ISSP (International Social Survey Programme), PISA (Programme for International Student Assessment), PIAAC (Programme for International Assessment of Adult Competencies), ESV (European Social Value) atd. Mnoho mezinárodních projektů v současnosti již také rutinně spolupracuje se společností cApStAn [<http://www.capstan.be/content/home.html>], zaměřující se na kontrolu kvality překladu v případě multilingvních projektů a zajištění lingvistické a sémantické ekvivalence obsahu textu.

31 Názorným příkladem jsou čím dál kvalitnější výstupy mezinárodního projektu ESS, který se pravidelně realizuje od roku 2002.

32 Na tuto problematiku upozorňoval van Deth [2003: 296], podle něhož je nutné zdůraznit problémy práce se sekundárními daty, které lze nalézt ve třech metodologických oblastech, a to odlišné operacionalizaci položek, nevhodném měření klíčových jevů a neadekvátní dokumentaci.

výzkumu, a proto je zcela nezbytné věnovat tomuto tématu hlubší pozornost. Jako úvod do tématu měl čtenář nejprve šanci seznámit se s konceptem ekvivalence a s její základní kategorizací na teoretické úrovni a úrovni měření, přičemž tento počin je třeba chápat jako jednu z řady možností, jak jednotlivé druhy ekvivalence vykládat a jak chápat jejich vzájemné postavení. Předložená kategorizace měla posloužit jednak ke sblížení se specifickou terminologií, jednak k udržení logické linie při výkladu technik určených ke hledání a testování ekvivalentních otázek.

Tím se dostáváme k hlavnímu cíli této přehledové stati, kterým bylo představit co nejucelenější přehled dostupných a nejvyužívanějších technik, jejichž záměrem je buď (1) při vytváření mezinárodních dotazníků nalézt ekvivalentní otázky, které by byly adekvátně přeložené a stejně interpretovatelné napříč šetřenými zeměmi, anebo (2) v datovém souboru ověřit úspěšnost výzkumníků při hledání takovýchto indikátorů, tj. statisticky otestovat reálnou srovnatelnost indikátorů a jejich kvalitu. V tomto duchu byly techniky rozděleny na dva soubory; techniky, které jsou pravidelně využívány před sběrem dat (překlad dotazníku, výběr etických a emických indikátorů a použití kognitivních metod), a techniky, které je možné uplatnit až po sběru dat při analýze datového souboru (jednoduché a mnohorozměrné statistické techniky). Při jejich výkladu byl kladen důraz na úvodní představení samotné techniky s více či méně specifickými informacemi o jejím fungování podle její složitosti s tím, že následoval popis její vazby s řešením problematiky srovnatelných otázek ve výzkumu. Užitečné případným zájemcům snad budou také odkazy na empirické studie, které vysvětlují aplikaci dané techniky při testování ekvivalence položek na mezinárodních datech. I bohatě citovaná zahraniční a česká odborná literatura by měla především posloužit čtenáři, pokud by se chtěl danou problematikou detailněji zabývat.

Problematika ekvivalence v mezinárodních výzkumech je značně široké téma, bohužel nebylo možné se dotknout takových témat, jako například významu a důsledků tzv. částečné ekvivalence položek [více Byrne et al. 1989; Vandenberg, Lance 2002]. Také by bylo ideální čtenáři přímo předvést aplikaci jednotlivých technik při testování ekvivalence položek z mezinárodního výzkumu, ale na to řádky této stati již nestačí. Zbývá tedy do budoucna doufat, že metodologické analýzy tohoto druhu se na české akademické půdě již brzy objeví a budou velkým přínosem pro všechny, kteří usilují o kvalitní vědeckou práci s komparativními výzkumy.

PETRA ANÝŽOVÁ *vystudovala sociologii na Filozofické fakultě Karlovy Univerzity v Praze, kde je také od roku 2010 studentem doktorského programu stejného oboru. Je rovněž odborným asistentem na Katedře sociologie na Vysoké škole finanční a správní, o.p.s. Jejím profesním zaměřením je především metodologie sociologického výzkumu se specializací na mezinárodní komparativní šetření a sociologie vzdělání.*

Literatura

- Almond, Gabriel A., Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, NJ: Princeton University Press.
- Ægisdóttir, Stefánía, Lawrence H. Gerstein, Deniz Canel Çınarbas. 2008. „Methodological Issues in Cross-Cultural Counseling Research: Equivalence, Bias, and Translations.” *The Counseling Psychologist* 36 (2): 188–219. DOI: 10.1177/0011000007305384.
- Berry, John W., Ype H. Poortinga, Janak Pandey (eds.). 1997. *Handbook of Cross-Culture Psychology: Theory and Methods*. Boston: Allyn and Bacon.
- Billiet, Jaak. 2003. „Cross-cultural equivalence with structural equation modeling.“ Pp. 247–263 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross cultural survey methods*. New Jersey: John Wiley & Sons.
- Bhalla, Gaurav, Lynn Y. S. Lin. 1987. „Cross-Cultural Marketing Research: A Discussion of Equivalence Issues and Measurement Strategies.“ *Psychology & Marketing* 4 (4): 275–285.
- Blasius, Jörg, Victor Thiessen. 2006. „Assessing Data Quality and Construct Comparability in Cross-national Surveys.“ *European Sociological Review* 22 (3): 229–242. DOI: 10.1093/esr/jcio54.
- Blair, Johnny, Linda Piccinino. 2005. „The Development and Testing of Instruments for Cross-Culture and Multi-Cultural Surveys.“ Pp. 13–30 in J. H. P. Hoffmeyer-Zlotnic, J. A. Harkness (Eds.). *Zuma Nachrichten 11 Methodological Aspects in Cross-National Research*. Mannheim.
- Borg, Ingwer. 1998. „A Facet-Theoretical Approach to Item Equivalency.“ *ZUMA – Nachrichten Spezial* 3, January 1998: 145–158.
- Braun, Michael, Janet A. Harkness. 2005. „Text and Context: Challenges to Comparability in Survey Questions.“ Pp. 95–109 in J. H. P. Hoffmeyer-Zlotnic, J. A. Harkness (Eds.). *Zuma Nachrichten 11 Methodological Aspects in Cross-National Research*. Mannheim.
- Braun, Michael, Timothy P. Johnson. 2010. „An Illustrative Review of Techniques for Detecting Inequivalences.“ Pp. 375–393 in J. A. Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Braun, Michael, Jacqueline Scott. 1998. „Multidimensional Scaling and Equivalence: Is having a Job the Same as Working?“ Pp. 129–144 in J. A. Harkness (eds.). *Cross-Cultural Survey Equivalence*. Mannheim: *ZUMA-Nachrichten Spezial* 3.
- Brislin, Richard, W. 1976. „Comparative Research Methodology: Cross-cultural Studies.“ *International Journal of Psychology* 11 (3): 215–229. DOI: 10.1080/00207597608247359.
- Byrne, Barbara M., Richard J. Shavelson, Bengt Muthén. 1989. „Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance.“ *Psychological Bulletin* 105 (3): 456–466.
- Byrne, Barbara M., David Watkins. 2003. „The Issue Of Measurement Invariance Revisited.“ *Journal of Cross-Cultural Psychology* 34 (2): 155–175. DOI: 10.1177/0022022102250225.
- Davidov, Eldad. 2010. „Testing for comparability of human values across countries and time with the third round of European Social Survey.“ *International Journal of Comparative Sociology* 51 (3): 171–191. DOI: 10.1177/0020715210363534.

- Davidov, Eldad. 2012. „Comparability of measurement instruments across countries or time points.“ presentation, University of Tartu, 4. 10. 2012.
- Davidov, Eldad, H. Dülmer, E. Schlüter, P. Schmidt, B. Meuleman. 2012. „Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance.“ *Journal of Cross-Cultural Psychology* 43 (4): 558–575. DOI: 10.1177/0022022112438397.
- Davidov, Eldad, Peter Schmidt. 2007. „Values and attitudes towards immigrants: A comparison between Germany and Israel.“ prezentace přednesená na workshopu: *Immigration, public opinion and governmental policies*, Pompeu Fabra University, Barcelona, March 2007.
- Davidson, Andrew R., James Jaccard, Harry C. Triandis, Maria Luisa Morales, Rogelio Diaz- Guerrero. 1976. „Cross-cultural model testing: toward a solution of the etic-emic dilemma.“ *International Journal of Psychology* 11 (1): 1–13. DOI: 10.1080/00207597608247343.
- Devins, Gerald M., Morton Beiser, Rene Dion, Luc G. Pelletier, R. Gary Edwards. 1997. „Cross-Cultural Measurements of Psychological Well-Being: The Psychometric Equivalence of Cantonese, Vietnamese, and Laotian Translations of the Affect Balance Scale.“ *American Journal of Public Health* 87 (5): 794–799.
- European Social Survey. 2010. *ESS Round 5 Translation Guidelines*. Mannheim, European Social Survey GESIS.
- Fontaine, Johnny R. J. 2003. „Multidimensional scaling.“ Pp. 235–246 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.
- Goerman, Patricia L., Rachel A. Caspar. 2010. „Managing the Cognitive Pretesting of Multilingual Survey Instruments: A Case Study of Pretesting of the U.S. Census Bureau Bilingual Spanish/English Questionnaire.“ Pp. 75–90 in J. A. Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Harris, Marvin. 1976. „History and significance of the emic/etic distinction.“ *Annual Review of Anthropology* 5 (1): 329–350. DOI: 10.1146/annurev.an.05.100176.001553.
- Harkness, Janet. 2003. „Questionnaire translation.“ Pp. 35–56 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.
- Harkness, Janet (eds.). 1998. *Cross-Cultural Survey Equivalence*. Mannheim: ZUMA-Nachrichten Spezial 3.
- Harkness, Janet A., Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph Mohler, Beth-Ellen Pennell, Tom W. Smith (eds.). 2010. *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Harkness, Janet, Peter Ph. Mohler, Fons J. R. Van de Vijver. 2003. „Comparative Research.“ Pp. 3–16 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.
- Harkness Janet A., Alicia Schoua-Glusberg. 1998. „Questionnaires in Translation.“ Pp. 87–126 in J. A. Harkness (eds.). *Cross-Cultural Survey Equivalence*. Mannheim: ZUMA-Nachrichten Spezial 3.
- Harkness, Janet A., Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). 2003. *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.

- Harkness, Janet A., Ana Viilar, Brad Edwards. 2010. „Translation, Adaptation, and Design.“ Pp. 117–140 in Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Hoffmeyer-Zlotnic, Jürgen H. P., Janet A. Harkness (eds.). 2005. *Methodological Aspects in Cross-National Research*. Mannheim: Zuma Nachrichten 11.
- Horn, J. L., J. J. McArdle. 1992. „A practical and theoretical guide to measurement invariance in aging research.“ *Experimental Aging Research* 18 (3): 117–144. DOI: 10.1080/03610739208253916.
- Hox, Joop J., Timo M. Bechger. 1998. „An Introduction to Structural Equation Modeling.“ *Family Science Review* 11: 354–373.
- Hox, Joop J., Edith D. de Leeuw, Matthieu J. S. Brinkhuis. 2010. „Analysis Models for Comparative Surveys.“ Pp. 395–418 in J. A. Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Hsueh, Kuei-Hsiang, Linda R. Phillips, WenYun Cheng, Sandra J. Fulton Picot. 2005. „Assessing Cross-Cultural Equivalence Through Confirmatory Factor Analysis.“ *Western Journal of Nursing Research* 27 (6): 755–771. DOI: 10.1177/0193945905276585.
- Jelínek, Martin, Petr Květoň, Dalibor Vobořil. 2011. *Testování v psychologii: Teorie odpovědi na položku a počítačové adaptivní chování*. Praha: Grada Publishing.
- Jeřábek, Hynek, Petr Soukup (eds.) 2008. *Advanced Lazarsfeldian methodology*. Praha: Karolinum.
- Johnson, Timothy P. 1998. „Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research.“ Pp. 1–40 in J. A. Harkness (ed.). *Cross-Cultural Survey Equivalence*. Mannheim: ZUMA-Nachrichten Spezial 3.
- Jowell, Roger. 1998. „How comparative is comparative research?“ Centre For Research Into Elections And Social Trends *CREST. Paper No 66* [online]. *Maison Francaise*, Oxford: June 26/27 1997 [27/10/2009]. Dostupné z: <<http://www.crest.ox.ac.uk/papers.htm>>.
- Jowell, Roger, Rory C. Fitzgerald, Gillian Eva (eds.). *Measuring Attitudes Cross-Nationally*. London: Sage Publications.
- Kankaras, Milos, Guy Moors. 2009. „Measurement Equivalence in Solidarity Attitudes in Europe: Insights from a Multi-Group Latent-Class Factor Approach.“ *International Sociology* 24 (4): 557–579. DOI: 10.1177/0268580909334502.
- Kohn, Melvin L (ed.). 1989. *Cross-National Research in Sociology*. Newbury Park, CA: Sage.
- Kostecký, Tomáš, Daniel Čermák. 2003. „Výběrová šetření a analýza agregátních dat – diskuze na téma použitelnosti různých přístupů v komparativních analýzách politického chování.“ *Sociologický časopis* 39 (4): 529–550.
- Krejčí, Jindřich. 2006. „Mezinárodní sociálněvědní komparativní výzkum a Česká republika: Přehled výzkumů a dostupných dat.“ *Sociologický časopis* 42 (1): 149–173.
- Krejčí, Jindřich. 2008. *Kvalita sociálněvědních výběrových šetření v České republice*. Praha: Sociologické nakladatelství.
- Lievens, Filip, Frederik Anseel, Michael H. Harris, Jacob Eisenberg. 2007. „Measurement Invariance of the Pay Satisfaction Questionnaire Across Three Countries.“ *Educational and Psychological Measurement* 67 (6): 1042–1051. DOI: 10.1177/0013164406299127.

- Lonner, Walter J. 1985. „Issues in testing assessment in cross-cultural counseling.“ *The counseling Psychologist* 13: 599–614. DOI: 10.1177/0011000085134004.
- Matějů, Petr. 1989. „Metoda strukturního modelování: Přehled základních problémů.“ *Sociologický časopis* 25 (4): 399–418.
- May, Henry. 2006. „A Multilevel Bayesian Item Response Theory Method for Scaling Socioeconomic Status in International Studies of Education.“ *Journal of Educational and Behavioral Statistics* 31 (1): 63–79. DOI: 10.3102/10769986031001063.
- Meade, Adam W., Emily C. Johnson, Phillip W. Braddy. 2006. „The Utility of Alternative Fit indices in Tests of Measurement Invariance.“ [online] *Paper presented at the Antal Academy of Management Conference*. Atlanta, GA. [20/2/2013]. Dostupné z: <<http://www4.ncsu.edu/~awmeade/Links/Papers/AFIsMI%28AoMo6%29.pdf>>
- Meade, Adam W., Gary J. Lautenschlager. 2004. „Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance.“ *Organizational Research Methods* 7 (4): 361–388. DOI: 10.1177/1094428104268027.
- Meredith, William. 1993. „Measurement Invariance, Factor Analysis and Factorial Invariance.“ *Psychometrika* 58 (4): 525–543. DOI: 10.1007/BF02294825.
- Mohler, Peter Ph., Timothy P. Johnson. 2010. „Equivalence, Comparability, and Methodological Progress.“ Pp. 17–29 in J. A. Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Mullen, Michael R. 1995. „Diagnostic Measurement Equivalence in Cross-National Research.“ *Journal of International Business Studies* 26 (3): 573–596.
- Pan, Yuling, Ashley Landreth, Hyunjoo Park, Marjorie Hinsdale-Shouse, Alisú Schoua-Glusberg. 2010. „Cognitive Interviewing in Non-English languages: A Cross-Cultural Perspective.“ Pp. 91–113 in J. A. Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Przeworski, Adam, Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: John Wiley & Sons.
- Raju, Nambury S., Larry J. Laffitte, Barbara M. Byrne. 2002. „Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory.“ *Journal of Applied Psychology* 87: 517–529. DOI: 10.1037/0021-9010.87.3.517.
- Revilla, Melanie, Willem Saris. 2011. „How to evaluate the cross-cultural equivalence of single items.“ (RECSM, UPF) Mini conference: *Measurement invariance: Methods, problems, and further directions*. Zurich15.–16. 7. 2011
- Rudá, Eliška. 2012. „Teorie odpovědi na položku a její aplikace v sociologii.“ Nepublikovaná diplomová práce. Praha: FF UK.
- Řehák, Jan. 1998. „Kvalita dat II. (Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur)“. *Sociologický časopis* 34 (2): 195–204.
- Saris, Willem E. 2003. „Multitrait-Multimethod studies.“ Pp. 265–274 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.

- Saris, Willem, Irmtraud Gallhofer. 2007. „Can questions travel successfully?“ Pp. 53–79 in Robert Jewell et al. (eds.). *Measuring Attitudes Cross-Nationally*. London: Sage Publications.
- Scheuch, Erwin K. 1993. „The Cross-culture Use of Sample Surveys: Problems of Comparability.“ *Historical Social Research* 18 (2): 104–138.
- Schwarz, Norbert, Daphna Oyserman, Emilia Peytcheva. 2010. „Cognition, Communication, and Culture: Implications for the Survey Response Process.“ Pp. 177–190 in J. A. Harkness et al. (eds.). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.
- Smith, Tom W. 2003. „Developing Comparable Questions in Cross-national Survey.“ Pp. 69–91 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross cultural survey methods*. New Jersey: John Wiley & Sons.
- Smith, Peter B., Ronald Fischer. 2008. „Acquiescence, Extreme Response Bias and Culture: A Multilevel Analysis.“ Pp. 285–314 in J. R. Fons vad de Vijver, D. van Hemert, Y. H. Poortinga (eds.). *Multilevel analysis of individuals and cultures*. New York: Taylor & Francis Group.
- Soukup, Petr. 2006. „Proč užívat hierarchické lineární modely?“ *Sociologický časopis/ Czech Sociological Review* 42 (5): 987–1012.
- Steenkamp, Jan-Benedict E. M., Hans Baumgartner. 1998. „Assessing Measurement Invariance in Cross-National Consumer Research.“ *Journal of Consumer Research* 25 (1): 78–90. DOI: 10.1086/209528.
- Stout, Wiliam. 2001. „Nonparametric Item response Theory: A Maturing and Applicable Measurement Modeling Approach.“ *Applied Psychological Measurement* 25 (3): 300–306. DOI: 10.1177/01466210122032109.
- Stout, William. 2002. „Psychometrics: From Practice to Theory and Back (15 years of Nonparametric Multidimensional IRT, DIF/Test Equity, and Skills Diagnostic Assessment).“ *Psychometrika* 67 (4): 485–518. DOI: 10.1007/BF02295128.
- Straus, Murray A. 1969. „Phenomenal Identity and Conceptual Equivalence of Measurement in Cross-National Comparative Research.“ *Journal of Marriage and the Family* 31 (2): 233–241.
- Vandenberg, Robert J., Charles E. Lance. 2000. „A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research.“ *Organizational Research Methods* 3 (1): 4–70. DOI: 10.1177/109442810031002.
- Van Deth, Jan W. 2003. „Using published survey data.“ Pp. 291–309 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.
- Van de Vijver, Fons J. R. 1998. „Towards a Theory of Bias and Equivalence.“ Pp. 41–65 in J. A. Harkness (eds.). *Cross-Cultural Survey Equivalence*. Mannheim: ZUMA-Nachrichten Spezial 3.
- Van de Vijver, Fons J. R. 2003. „Bias and substantive analysis.“ Pp. 207–233 in Janet A. Harkness, Fons J. R. Van de Vijver, Peter Ph. Mohler (eds.). *Cross-cultural survey methods*. New Jersey: John Wiley & Sons.
- Van de Vijver, Fons J. R., Dianne van Hemert, Ype H. Poortinga (eds.). 2008. *Multilevel analysis of individuals and cultures*. New York: Taylor & Francis Group.

- Van de Vijver, Fons J. R., Kwok Leung. 1997. „Methods and Data Analysis of Comparative Research.“ Pp. 257–301 in John W. Berry, Ype H. Poortinga, Janak Pandey (eds.). *Handbook of Cross-Culture Psychology: Theory and Methods*. Boston: Allyn and Bacon.
- Van de Vijver, Fons J. R., Ype H. Poortinga. 2002. „Structural Equivalence in Multilevel Research.“ *Journal of Cross-Cultural Psychology* 33 (2): 141–156. DOI: 10.1177/0022022102033002002.
- Van Hemert, Dianne A., Fons J. R. van de Vijver, Ype H. Poortinga, James Georgas. 2002. „Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries.“ *Personality and Individual Differences* 33: 1229–1249. DOI: 10.1016/S0191-8869(02)00007-7.
- Vinopal, Jiří. 2008. *Kognitivní přístupy v metodologii výzkumných šetření: metoda okamžité validizace*. Praha: Sociologický ústav AV ČR.
- Vinopal, Jiří. 2009. „Situace standardizovaného dotazování z hlediska kognitivních přístupů.“ *Sociologický časopis/Czech Sociological Review* 45 (2): 397–420.
- Wang, Mo, Steven S. Russell. 2005. „Measurement Equivalence of the Job Descriptive Index Across Chinese and American Workers: Results from Confirmatory Factor Analysis and Item Response Theory.“ *Educational and Psychological Measurement* 65 (4): 709–732. DOI: 10.1177/0013164404272494.
- Welkenhuysen-Gybels, Jerry G. J., Fons J. R. van de Vijver. 2001. „A comparison of methods for the evaluation of construct equivalence in a multigroup setting.“ *Proceedings. American Statistical Association*.
- Woehr, David J., John P. Meriac. 2010. „Using Polytomous Item Response Theory to Examine Differential Item and Test Functioning: The Case of Work Ethic.“ Pp. 419–433 in J. A. Harkness et al. (eds). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New Jersey: John Wiley & Sons.