Nový přístup v testování ekvivalence na příkladu měření subjektivního blahobytu[1]

New Approach to Testing Measurement Invariance: An Example Using Subjective Well-Being

Petra Raudenská, Radka Hanzlová

ABSTRACT Survey-based measures of subjective well-being are more and more often analyzed cross-culturally. However, international comparison of these measures requires measurement invariance. Therefore, the major goal of this study was to investigate the cross-country comparability of the five-item subjective well-being scales used in the International Social Survey Programme (2011, 2017). This study applied both the traditional exact and the more recent Bayesian approximate approach to assess whether the subjective well-being scales were measurement invariant. The Bayesian approach detected several non-invariant items that were problematic for cross-national comparison and could be dropped from the scales. Consequently, measurement invariance was established in all countries for the reduced scales, allowing researchers to meaningfully compare their latent mean scores and the relationships with other theoretical constructs of interest. Thus, the study highlighted the advantages of using multiple indicators and the necessity of measurement invariance testing in subjective well-being research.

KEYWORDS subjective well-being, measurement invariance, International Social Survey Programme, MG-BSEM

Úvod

Výzkumy zabývající se měřením štěstí a životní spokojenosti, obecněji definované jako osobní pohoda nebo subjektivní blahobyt, jsou běžné především v empirické psychologii a sociálních vědách, ale v poslední době roste jejich popularita také v ekonomii. Hlavní nevýhodou mnoha publikovaných studií měřících subjektivní blahobyt je to, že ho většinou měří

Sociální studia / Social Studies 1/2021. S. 53–72. ISSN 1214-813X.

1 Petra Raudenská napsala tento článek s podporou určenou na dlouhodobý koncepční rozvoj výzkumné organizace RVO: 68378025. Radka Hanzlová mohla tento článek napsat s podporou projektu Specifického vysokoškolského výzkumu 2020–2020 (260682) a programu UK Progres č. Q15 s názvem „Životní dráhy, životní styly a kvalita života z pohledu individuální adaptace a vztahu aktérů a institucí“, řešených na Filozofické fakultě Univerzity Karlovy.

pomocí jedné položky/otázky, a nikoliv baterie obsahující více položek (Huppert et al. 2009), tedy jako vícedimenzionální koncept. Je však známo, že jednopoložková měření nejsou příliš přesná, mají omezenou obsahovou validitu, nemají vysokou reliabilitu (spolehlivost) a neumožňují kontrolu chyb měření (Davidov, Cieciuch a Schmidt 2018).

Navíc se s daty o subjektivním blahobytu často zachází, jako by byla automaticky srovnatelná jak mezi jednotlivci a regiony, tak i v čase (Diener 2009; Bjørnskov 2010), přičemž tento předpoklad je většinou založen na nepřímých důkazech. Fors a Kulin (2016: 325) dále poukazují na to, že v současném srovnávacím výzkumu zaměřeném na subjektivní blahobyt chybí analýzy systematicky kontrolující měření ekvivalence mezi různými kulturami. Metodologické studie škál blahobytu se potom obvykle zaměřují na srovnávání mezi národními skupinami a málokdy zkoumají porovnání mezi větším počtem zemí v čase (viz Emerson, Guhn a Gadermann 2017; Jovanović et al. 2018). Z toho plyne, že metodologické pokroky zaznamenané v posledních letech, které poskytují příležitost měřit ekvivalenci mezi velkým počtem skupin a časovými body, byly ve výzkumu blahobytu použity pouze zřídkakdy (např. Żemojtel-Piotrowska et al. 2017).

Metodologové tvrdí, že by ekvivalence měření neměla být předpokládána, ale spíše empiricky testována. Nicméně se zdá, že testování ekvivalence měření obvykle zůstává v rukou konkrétního uživatele dat (Seddig a Leitgöb 2018). Cílem této studie je proto prozkoumat mezinárodní srovnatelnost škál subjektivního blahobytu použitých v široce využívaném Mezinárodním programu sociálního výzkumu (dále jen ISSP) v roce 2011 v modulu Zdraví a v roce 2017 v modulu Sociální sítě.

V této studii jsou prezentovány dva přístupy k testování ekvivalence měření – tradiční přístup konfirmační faktorové analýzy k testování tzv. přesné ekvivalence měření (v angličtině *exact measurement invariance*, dále jen EMI) a novější přístup založený na moderní bayesovské statistice k testování tzv. přibližné ekvivalence měření (v angličtině *approximate measurement invariance*, dále jen AMI). Závěry bayesovského přístupu, který byl doposud používán ve výzkumu subjektivního blahobytu pouze ojediněle (např. Raudenská 2020), mohou lépe a spolehlivěji osvětlit rozsah nesrovnatelnosti jednotlivých položek v testovaných škálách. Zkoumáním toho, do jaké míry jsou škály subjektivního blahobytu srovnatelné mezi zeměmi ve výzkumu ISSP 2011 a 2017, se tato studie snaží významně přispět k diskusi o vlastnostech vícepoložkového měření subjektivního blahobytu a doufá, že bude motivovat k dalším podobně orientovaným metodologickým studiím na poli české sociální vědy.

Studie sestává z následujících částí. První část věnujeme konceptu subjektivního blahobytu, konkrétně jeho definici, možnostem měření a testovaným škálám. Poté jsou stručně shrnuty výsledky testu EMI a AMI daných škál s důrazem na srovnání výsledků obou přístupů. Na závěr jsou prezentovány výsledky srovnání latentních průměrů mezi zeměmi a tato zjištění jsou diskutována.

Teoretická východiska

Koncept subjektivního blahobytu

V literatuře se subjektivní blahobyt definuje především jako „subjektivní úsudek o vlastním životě“ (Diener, Lucas a Oishi 2018). Existují dva relativně odlišné, přesto se

překrývající přístupy k výzkumu blahobytu (více viz Ryan a Deci 2001). První z nich se široce označuje jako hédonismus (Kahneman et al. 1999). Hédonický přístup se zaměřuje především na prožitek štěstí a definuje blahobyt z hlediska maximalizace dosažení radosti a minimalizace nepříjemných pocitů (Diener a Lucas 1999). Druhým přístupem je eudaimonismus (Waterman 1993), který zdůrazňuje, že blahobyt spočívá v naplnění vlastního potenciálu a užitku a označuje tak „správně žitý život". O jednotné definici subjektivního blahobytu se však vedou mezi výzkumníky rozsáhlé diskuse, obzvláště zdali se jedná o jednodimenzionální, či vícedimenzionální koncept.

Od definice konceptu se následně odvíjí jeho měření, přičemž v mezinárodních výzkumech se bohužel nejčastěji setkáme s měřením subjektivního blahobytu pomocí jedné otázky na štěstí či životní spokojenost (např. European Social Survey ESS, European Values Study EVS, World Values Survey WVS, Eurobarometr EB, Gallup World Poll). Nicméně jak konstatuje např. OECD (2013), tak ve skutečnosti v sobě subjektivní blahobyt zahrnuje nejen koncept štěstí či životní spokojenosti, ale rovněž i dobré duševní stavy, hodnocení vlastního života a afektivní reakce jedinců na jejich zkušenosti. Stejně tak McGillivray a Clarke (2006: 4) tvrdí, že „subjektivní blahobyt zahrnuje vícedimenzionální hodnocení života včetně kognitivních úsudků o životní spokojenosti, citového hodnocení emocí a nálad". Měření pomocí jedné otázky je problematické zejména kvůli omezené obsahové validitě a nízké reliabilitě, neboť odpovědi na jednopoložkovou otázku jsou výrazně ovlivněny kontextuálními faktory (např. předcházející otázkou, aktuální náladou, okolnostmi dotazování, pamětí). V současnosti proto převažuje vícedimenzionální pojetí subjektivního blahobytu, v rámci kterého se tento koncept měří jako baterie obsahující více položek.

V současnosti existuje několik běžně používaných standardizovaných škál či indexů subjektivního blahobytu. Patří mezi ně např. pětipoložková škála spokojenosti se životem (SWLS) (Diener et al. 1985), dvanáctipoložková škála pozitivní a negativní zkušenosti (SPANE), osmipoložková „vzkvétající škála" (FS) (Diener et al. 2009), sedmipoložkový index osobního blahobytu (PWI) (Cummins et al. 2003) nebo pětipoložkový index emoční pohody (WHO-5) (WHO 1998). Doposud však existuje pouze omezené množství studií, které by se zaměřovaly na testování ekvivalence těchto škál či indexů, přestože se běžně používají i v mezinárodním kontextu (např. Emerson, Guhn a Gadermann 2017; Żemojtel-Piotrowska et al. 2016; Jovanović et al. 2019).

Výjimkou jsou studie, které čerpají z dat třetí vlny ESS, do které byl zahrnut speciální modul „Osobní a sociální blahobyt". Například Huppert a So (2013) v těchto datech identifikovali deset znaků pozitivního blahobytu a vytvořili podle toho profily 23 evropských zemí. Autoři mimo to také otestovali mezikulturní srovnatelnost položek navrženého latentního faktoru pozitivního blahobytu ve třech evropských regionech (severní, jižní/západní a východní Evropa).

Významnou prací v tomto směru je nová studie od Raudenské (2020), která na datech z třetí a šesté vlny ESS jako první testovala srovnatelnost čtyřpoložkové afektivní škály měřící pozitivní emoce a sedmipoložkové škály měřící negativní emoce mezi velkým počtem zemí a v čase, přičemž použila nový bayesovský přístup k měření přibližné ekvivalence. Autorka dospěla k závěru, že nová metoda odhalila několik nesrovnatelných položek, které jsou problematické pro mezinárodní srovnání, a navrhla jejich vyřazení z původní škály.

Následně pak také ukazuje profily 23 evropských zemích dle míry manifestovaných negativních a pozitivních emocí v obecné populaci a proměnu těchto afektivních reakcí v čase s odkazy na společenské a ekonomické změny odehrávající se v době sběru dat.

S ohledem na nedostatek metodologických studií týkajících se měření subjektivního blahobytu jako vícedimenzionálního konceptu mezi zeměmi vznikla i tato studie, která si klade za cíl prozkoumat ekvivalenci měření dvou škál subjektivního blahobytu použitých v ISSP 2011 a 2017 mezi velkým počtem zemí za pomoci přístupu k testování přibližné ekvivalence měření a přispět tak do české odborné literatury novými poznatky jak z hlediska mezinárodní srovnatelnosti subjektivního blahobytu, tak i co se týká aplikace nové metody testu přibližné ekvivalence založené na bayesovské statistice.

Testování přesné (EMI) a přibližné ekvivalence měření (AMI)

Pro smysluplné srovnání konceptů je vyžadováno, aby byl stejný koncept měřen stejným způsobem napříč různými skupinami (např. zeměmi, regiony, sociálními skupinami nebo časovými body). Stanovení ekvivalence neboli invariance měření měřicího nástroje znamená, že by měl měřit stejný koncept na stejné měřicí stupnici. To může zaručit, že pozorované rozdíly mezi skupinami skutečně odrážejí mezikulturní rozdíly, a ne jen chyby měření (Davidov, Cieciuch a Schmidt 2018). Pro testování ekvivalence měření byly v literatuře navrženy různé techniky (Davidov, Schmidt a Billiet 2011; v češtině viz Anýžová 2014, 2015).

Tradiční přístup k testování *přesné ekvivalence měření* (EMI) rozlišuje tři hierarchicky uspořádané úrovně, z nichž každá je definována vzrůstající restrikcí, která je kladena na testovaný model (např. Meredith 1993; Steenkamp a Baumgartner 1998; Vandenberg a Lance 2000; Davidov et al. 2014). Na nejnižší úrovni (*konfigurální ekvivalence*) se předpokládá, že obecná faktorová struktura měření je stejná napříč různými skupinami. Daný koncept je tedy měřen stejnými položkami stejně v různých skupinách.

Druhá úroveň (*metrická ekvivalence*) vyžaduje, aby faktorové zátěže byly ve všech porovnávaných skupinách totožné, což naznačuje, že význam latentní proměnné zůstává ve skupinách stejný. Metrická ekvivalence je považována za již dostatečnou pro věcné analýzy porovnávající faktorové kovariance a nestandardizované regresní koeficienty mezi zkoumanými populacemi (Steenkamp a Baumgartner 1998; Vandenberg a Lance 2000).

Třetí a zpravidla nejvyšší úroveň (*skalární ekvivalence*) vyžaduje, aby faktorové zátěže a položkové konstanty byly ve všech skupinách stejné, což naznačuje, že respondenti se stejnou hodnotou latentního faktoru mají i stejnou očekávanou odpověď bez ohledu na skupinu, do které patří (Davidov et al. 2014). Tato úroveň je považována za nezbytnou pro porovnání latentních průměrů nebo průměrů hrubých skórů jednotlivých položek mezi skupinami (Meredith 1993). Pro EMI byla doposud nejvíce používanou metodou víceskupinová konfirmační faktorová analýza (MG-CFA) (více viz např. Kim et al. 2017).

Rozhodnutí, zda bylo EMI dosaženo, je založeno na nejvíce používaných indexech vhodnosti modelu: chí-kvadrát test (LRT test, poměr věrohodností), CFI (Comparative Fit Index), RMSEA (Root Mean Squared Error of Approximation) a SRMR (Standardized Root Mean Squared Residual). U větších vzorků (více než 300 případů) se liberálnější obecné indexy vhodnosti modelu (CFI a RMSEA) upřednostňují před chí-kvadrátem, protože jsou

méně citlivé na velikost výzkumného vzorku (Chen 2007). Hodnoty CFI nad 0,9 jsou považovány za důkaz přijatelného modelu a hodnoty nad 0,95 signalizují dobrou shodu modelu s daty. Pro RMSEA je hranice pro přijatelný model 0,08 a pro dobrý model 0,05. Pokud porovnáváme nejméně deset skupin, Rutkowski a Svetina (2014) navrhují mezní hodnotu RMSEA až kolem 0,1.

Aby bylo možné posoudit, zda bylo v MG-CFA dosaženo příslušné úrovně ekvivalence měření, porovnávají se obecné indexy vhodnosti modelu mezi méně a více restriktivním modelem. Pokud je změna mezi modely přibližně 0,01 a 0,015 nebo menší u indexů CFI a RMSEA, více restriktivní model může být přijat (Chen 2007). Nicméně pokud pracujeme se spojitými daty pro velký počet skupin, jež mají rovněž i větší velikost, tak se zdá, že liberálnější mezní hodnoty 0,02 pro ΔCFI a 0,03 pro ΔRMSEA jsou vhodnější pro přechod z konfigurálního na metrický model a 0,01 pro ΔCFI i ΔRMSEA při přechodu z metrického na skalární model (Rutkowski a Svetina 2014). Pro kategoriální data Rutkowski a Svetina (2017) navrhují při přechodu z konfigurálního na metrický model hodnoty 0,004 pro ΔCFI a 0,05 pro ΔRMSEA a při přechodu z metrického na skalární model ΔCFI 0,004 a ΔRMSEA 0,01.

Je zvykem, že výzkumníci, kteří testují skalární ekvivalenci měření mezi mnoha zeměmi/skupinami, ji velmi často nedokážou podpořit daty (viz Davidov et al. 2014; Cieciuch et al. 2017). Pro takové případy Byrne, Shavelson a Muthén (1989) představili koncept částečné *ekvivalence*, kdy jsou nesrovnatelné parametry volně odhadovány (tj. vypočítány) a na těch zbývajících zůstávají restrikce shody. Navrhují, že pro dosažení částečné skalární ekvivalence je dostatečné, aby nejméně dvě položky z testovaného latentního faktoru měly stejné faktorové zátěže a konstanty. V praxi to však znamená nalezení lokální chybné specifikace parametrů, na něž nemůže být restrikce uvalena, neboť nejsou mezi skupinami totožné. Restrikce na totožnost parametru mezi skupinami (tj. faktorových zátěží či konstant) by pak měla být uvolněna z modelu jedna po druhé, dokud nebude rozdíl v obecných indexech vhodnosti modelu (CFI, RMSEA) mezi metrickým a částečně skalárním modelem dostatečně malý.

Největší slabina tohoto manuálního přístupu spočívá v tom, že každá změna v restrikci má za následek jinak specifikovaný model a konečný seznam nesrovnatelných parametrů je tak závislý na pořadí, ve kterém jsou změny zadávány do modelu (Żemojtel-Piotrowska et al. 2017). V případě velkého počtu skupin vedou výsledky testování částečné EMI k dlouhé sérii úprav a tento manuální přístup nemůže zaručit, že finální model bude optimální (Muthén a Asparouhov 2013; Rutkowski a Svetina 2014; Kim et al. 2017).

Tyto problémy vyvolávají otázku, zda jsou kritéria pro hodnocení EMI přiměřená (Van de Schoot et al. 2013). Aktuální metodologické studie proto navrhují testování ekvivalence pomocí jiných nástrojů než MG-CFA, jako například hierarchického modelování, analýzy latentních tříd, anebo pro testování tzv. přibližné ekvivalence (AMI) využít bayesovského modelování (MGBSEM) či alignment procedury (více viz např. Kim et al. 2017). My v této studii dále představíme koncept přibližné ekvivalence a budeme se zaobírat aplikováním bayesovského modelování pro potřeby testování ekvivalence měření, ale vzhledem k výčtu uvedenému výše není námi nastíněný přístup jediný možný.

Základní myšlenka metody testování přibližné ekvivalence (AMI) spočívá v nahrazení požadavku rovnosti faktorových zátěží a konstant mezi skupinami tím, aby všechny parametry byly „přibližně stejné“ (Muthén a Asparouhov 2013; Van de Schoot et al. 2013). Předešlý

přístup k testování přesné ekvivalence požadoval, aby byly rozdíly mezi faktorovými zátěžemi či konstantami mezi skupinami nulové. Oproti tomu přístup přibližné ekvivalence předpokládá, že rozdíly mezi těmito parametry budou téměř nulové (Muthén a Asparouhov 2013). Rozdíly jsou stále udržovány na minimální hodnotě, která má zajistit, že zkoumané koncepty zůstanou přibližně srovnatelné. Díky této možnosti není shoda modelů zatížena nepřiměřeně přísným předpokladem totožných parametrů mezi skupinami, který neodráží původní záměr výzkumníků srovnávat sociální jevy mezi kulturami (Pokropek, Schmidt a Davidov 2020). Testování přibližné ekvivalence již bylo úspěšně použito v mnoha současných srovnávacích studiích (Davidov, Cieciuch a Schmidt 2018; Seddig a Leitgöb 2018; Cieciuch et al. 2017; Davidov et al. 2015; Zercher et al. 2015; Cieciuch et al. 2014; Bujacz et al. 2014).

Bayesovský přístup k testování přibližné ekvivalence obvykle vyžaduje omezení průměrného rozdílu mezi parametry (tzv. apriorní průměr/*prior mean*) na nulu a odchylky parametrů (tzv. apriorní rozptyl/*prior variance*) na hodnotu větší než nula, ale stále dostatečně malou. Velikost apriorního rozptylu, kterou určuje sám výzkumník, pak odráží úroveň aproximace. Čím nižší hodnotu apriorního rozptylu výzkumník zvolí, tím více předpokládá, že parametry se v modelu mezi skupinami neliší a že tento model je více podobný modelu EMI (Cieciuch et al. 2017). Naopak s vyšší hodnotou apriorního rozptylu výzkumník předpokládá, že parametry mohou být ve skupinách více odlišné. Špatně zvolená hodnota apriorního rozptylu tak může mít vážné důsledky pro odhad latentních průměrů v datech a pro výpočet směrodatných odchylek samotných parametrů (Pokropek, Schmidt a Davidov 2020).

V simulační studii Van de Schoot a jeho kolegové (2013) navrhují nastavení apriorního rozptylu na hodnotu 0,05, aniž bychom se dle nich vystavovali riziku neplatných závěrů při srovnání latentních průměrů mezi skupinami. Nicméně Pokropek, Schmidt a Davidov (2020) na základě své simulační studie už zamítají apriorní rozptyl pohybující se okolo 0,05 nebo 0,1 jako hodnoty povolující příliš velkou meziskupinovou variabilitu mezi parametry. Autoři navrhují mechanismus testování série modelů s různě malými apriorními rozptyly (např. 0,001; 0,005; 0,01; 0,025 a 0,05), aby bylo možné identifikovat nejlépe vyhovující model s tím nejvhodnějším apriorním rozptylem reflektujícím data. Je podle nich vhodné začít testovat nejpřísnější model s apriorním rozptylem na hodnotě 0,000, který odpovídá kritériím EMI (tj. naprostá shoda velikosti parametrů mezi skupinami); model v této podobě většinou příliš dobře neodpovídá datům. Postupně se pak apriorní rozptyl zvyšuje do té doby, než indexy vhodnosti modelu prokážou dostatečně dobrý model či jeho významné zlepšení od předchozího (méně úspěšného) modelu.

V bayesovském testování AMI je nejpoužívanějším kritériem pro výběr modelu, který se nejlépe shoduje s daty, informační kritérium o odchylce (Deviance Information Criterion, dále jen DIC) a Bayesovo informační kritérium (BIC). Upřednostňuje se model s nejnižší hodnotou DIC a BIC, přičemž BIC na rozdíl od DIC penalizuje komplexní modely s více proměnnými.

K určení, zda model správně reflektuje data a zda tak lze v datech nalézt AMI, se obvykle používají dvě míry (Muthén a Asparouhov 2013; Kim et al. 2017). Jedná se o posteriorní prediktivní pravděpodobnost (Posterior Predictive Probability value, dále PPP) a 95% kredibilní interval (95% Credibility Interval, dále CI, jejich definice viz Muthén a Asparouhov 2013; Van de Schoot et al. 2013). Bayesovský model dobře odpovídá datům,

když PPP není statisticky významné a CI obsahuje nulu. Hodnota PPP okolo 0,5 nebo vyšší naznačuje, že model odpovídá datům velmi dobře.

Při hledání modelu s nejlépe zvoleným apriorním rozptylem se pak používají následující prahové hodnoty (porovnává se model s menším a vyšším apriorním rozptylem). Pokud se v modelu s vyšším apriorním rozptylem snížila hodnota BIC o 20 bodů a více, hodnota DIC se snížila o 14 bodů a více a hodnota PPP zvýšila o 0,01 a více, doporučuje se přijmout model s tímto apriorním rozptylem a již jej více nezvyšovat. Pokropek, Schmidt a Davidov (2020) zdůrazňují, že zmíněné indexy je lepší kombinovat, přičemž jako nejspolehlivější se ukazuje být změna indexu DIC a PPP. Tyto prahové hodnoty je dle autorů vhodné použít, pokud výzkumník pracuje s daty, kde se nachází v průměru 24 skupin o velikosti 1 500 jednotek (tedy klasický případ mezinárodního výzkumu v jedné vlně šetření). Pro práci s daty s malým počtem skupin a jednotek, anebo naopak větším počtem skupin a jednotek navrhují autoři mírně odlišné prahové hodnoty.

Software Mplus poskytuje výzkumníkům ještě další typ výstupu k vyhodnocení vhodnosti modelu, konkrétně tzv. výstup o odlišnosti (*difference output*). Tato část výstupu předkládá seznam všech parametrů odlišných od průměru v každé skupině. Pokud model přibližné ekvivalence nevyhovuje datům, výstup o odlišnosti pomůže k vyhledání možné problematické položky podle toho, která se příliš odlišuje od průměru v dané skupině v co největším počtu zemí. Pokud naopak model přibližné ekvivalence vyhovuje datům, výstup o odlišnosti ukáže ty položky, které jsou stále srovnatelné, ale mírně odlišné (Muthén a Asparouhov 2013; Cieciuch et al. 2017). Podrobné vysvětlení bayesovského přístupu pro testování přibližné ekvivalence měření je nad rámec této studie a lze jej nalézt ve studii od Muthéna a Asparouhova (2013), Van de Schoota et al. (2013) a ve studii Pokropka, Schmidta a Davidova (2020).

Nakonec je vhodné zmínit, že bayesovský přístup je aplikovatelný (stejně jako MG-CFA metoda s vhodnou metodou odhadu) i pro kategoriální nenormálně rozložená data (Muthén a Asparouhov 2013), je též vhodný pro testování velkého množství skupin a složité a velmi komplexní modely. S chybějícími daty zachází stejně jako metody odhadu parametrů určené pro testování dat s chybějícími informacemi (Asparouhov a Muthén 2010).

Data a metody

Data

Prezentovaná analýza je založena na datech z ISSP z roku 2011 (modul Zdraví) a 2017 (modul Sociální sítě). Celkem zahrnuje 90 634 osob ze 40 různých zemí (více viz tabulka A1 v online appendixu[2]). Z analýzy z roku 2011 byly z důvodu velkého počtu chybějících hodnot vyloučeny Čína a Francie.[3] V rámci této studie jsme testovaly dvě pětipoložkové škály subjektiv-

2 Appendix lze nalézt v online podobě na https://www.soc.cas.cz/publikace/novy-pristup-v-testovani-ekvivalence-na-prikladu-mereni-subjektivniho-blahobytu.

3 Data byla získána po registraci z webové stránky GESIS (www.gesis.org/issp/home). Další informace o sběru dat, výběru výzkumného vzorku, dotaznících a další metodologické dokumenty jsou rovněž k dispozici na tomto webu nebo přímo na stránkách ISSP (www.issp.org).

ního blahobytu. Obě škály byly námi vybrané tak, aby zachytily vícedimenzionalitu konceptu subjektivního blahobytu, avšak nejsou to typické standardizované škály měřící subjektivní blahobyt vyvíjené výzkumníky s ohledem na přínos každé jednotlivé položky do obsahové validity konstruktu (na rozdíl od standardizovaných psychologických škál SWLS nebo PWI). Ponechávají tedy určitý prostor k případné modifikaci. Obsahují položky zaměřené obecně na pozitivní a negativní hodnocení vlastního života (hédonismus), ale rovněž i položky týkající se vlastního potenciálu a schopnosti čelit problémům (eudaimonismus). Jednotlivé položky jsou hodnoceny na pěti- či sedmibodové škále odpovědí (více viz tabulka 1).

Tabulka 1: Koncept subjektivního blahobytu, jeho položky a formulace otázek

Kód položky	Formulace otázky	Původní škála	Redukovaná škála
ISSP 2011			
happy11	Kdybyste se měl(a) dnes obecně zamyslet nad svým životem, řekl(a) byste, že jste … (1) zcela šťastný(á); (2) velmi šťastný(á); (3) docela šťastný(á); (4) ani šťastný(á), ani nešťastný(á); (5) docela nešťastný(á); (6) velmi nešťastný(á); (7) zcela nešťastný(á)	x	
health11	Jak často jste v posledních 4 týdnech měl(a) kvůli zdravotním potížím problémy vykonávat zaměstnání či pracovat v domácnosti? (1) nikdy; (2) zřídka; (3) občas; (4) často; (5) velmi často	x	
sad11	Jak často jste v posledních 4 týdnech jste se cítil(a) nešťastný/á a sklíčený/á? (1) nikdy; (2) zřídka; (3) občas; (4) často; (5) velmi často	x	x
lost11	Jak často jste v posledních 4 týdnech pocítil(a) ztrátu sebedůvěry? (1) nikdy; (2) zřídka; (3) občas; (4) často; (5) velmi často	x	x
overcome11*	Jak často jste v posledních 4 týdnech cítil(a), že není ve Vašich silách překonat problémy, které máte? (1) nikdy; (2) zřídka; (3) občas; (4) často; (5) velmi často	x	x
ISSP 2017			
health17	Celkově vzato, řekl/a byste, že Vaše zdraví je .. (1) výborné; (2) velmi dobré; (3) dobré; (4) ucházející; (5) špatné	x	
sad17	Jak často v uplynulých 4 týden jste se cítil(a) nešťastný/á a sklíčený/á? (1) nikdy; (2) zřídka; (3) občas; (4) často; (5) velmi často	x	x
overcome17*	Jak často v uplynulých 4 týdnech jste se cítil(a), že není ve Vašich silách překonat problémy, které máte? (1) nikdy; (2) zřídka; (3) občas; (4) často; (5) velmi často	x	x
achieve17	Do jaké míry pro Vás následující tvrzení platí nebo neplatí? Snadno dosahuji svých cílů. (1) naprosto pravdivé; (2) většinou pravdivé; (3) spíše pravdivé; (4) ani pravda, ani nepravda; (5) spíše nepravdivé; (6) většinou nepravdivé; (7) naprosto nepravdivé	x	
satisfac17	Vezmete-li v úvahu všechny okolnosti, jak jste v současné době celkově spokojený/á se svým životem? (1) zcela spokojený; (2) velice spokojený; (3) docela spokojený; (4) ani spokojený, ani nespokojený; (5) docela spokojený; (6) velmi nespokojený; (7) zcela nespokojený	x	x

* Referenční položka v modelu měření, jejíž faktorová zátěž byla zafixována na jedničku.

Analytická strategie

Analýza měření ekvivalence obou škál probíhala v několika krocích (viz tabulka 2). V prvním kroku byly testovány dva modely pro původní pětipoložkové škály ve všech zemích za pomoci konfirmační faktorové analýzy (CFA) a následně víceskupinové konfirmační faktorové analýzy (MG-CFA) s metodou odhadu MLR (*maximum likelihood estimation with robust standard errors*) vhodnou pro analýzu spojitých nenormálně rozložených dat. Chybějící hodnoty byly v analýze ošetřeny FIML (*full information maximum likelihood*) metodou. V případě EMI byla původně využita odhadní metoda WLSMV (*weighted least squares means and variance adjusted*) vhodná pro kategoriální data, ale modely s touto odhadní metodou nekonvergovaly již v prvním kroku testování CFA v některých zemích. V ostatních zemích navíc index RMSEA ukazoval zkreslující hodnoty vyšší než 0,1–0,2 v porovnání s žádoucími výsledky indexu CFI>0,98, což potvrdilo někdy zmiňované nevýhody využití indexu RMSEA pro kategoriální proměnné nebo při posuzování ekvivalence modelů s odhadní metodou WLSMV (viz Muthén 1999; Xia 2016; Monroe a Cai 2015; Navruz 2016).

Z tohoto důvodu byla využita odhadní metoda MLR pro spojitá nenormální data, současně ale bylo sledováno, viz Byrne (2010), a) zda mají proměnné dostatečný počet kategorií roven nebo vyšší pěti; b) normalita dat; c) velikost vzorku (viz také Cieciuch a Davidov 2012; Meuleman, Davidov a Billiet 2009). Počet respondentů v analyzovaných zemí přesahoval 1 000 respondentů, což je obecně považováno za velký vzorek. V modelu byly použity proměnné s pěti a sedmi kategoriemi. Šikmost a špičatost jejich rozdělení nepřesahovaly hodnoty dva pro velikost šikmosti a sedm pro velikost špičatosti (viz tabulka A2a a A2b v online appendixu), což značí nepříliš vychýlená data od normálního rozdělení (viz Byrne 2010; Mîndrilă 2010).

V případě škály z roku 2017 ukázala EMI na nesrovnatelnost položek v modelu již v prvním kroku testování, jinými slovy nebylo dosaženo ani konfigurální ekvivalence, a proto došlo k odstranění dvou nejvíce problematických položek. Dále byla testována už jen redukovaná třípoložková škála. Každá analýza EMI obsahovala hodnocení konfigurální, metrické a skalární ekvivalence. Vhodnost modelu byla vyhodnocena na základě mezních kritérií navrhovaných v literatuře (více viz sekce o testování přesné a přibližné ekvivalence měření). Ve druhém kroku, pokud nebylo dosaženo přesné metrické či skalární ekvivalence, jsme testovaly částečnou ekvivalenci a uvolňovaly nesrovnatelné parametry za pomoci modifikačních indexů.

Ve třetím kroku jsme testovaly AMI s pomocí bayesovského modelování. Pro potřeby AMI bylo nezbytné testované položky nejdříve standardizovat, což by mělo dle autorů studie Pokropek, Schmidt a Davidov (2020) napomoci, pokud mají testované položky odlišné distribuce. Ideální hodnota apriorního rozptylu byla v bayesovské analýze nalezena pomocí porovnání modelů s různými hodnotami apriorního rozptylu. Hodnocení modelů bylo založeno na prahových hodnotách indexů BIC, DIC a PPP (více opět viz podkapitola o testování EMI a AMI).

Výsledky bayesovského přístupu byly shodné s výsledky EMI a rovněž ukázaly na nesrovnatelnost obou původních pětipoložkových škál z roku 2011 i 2017. Proto byly ve čtvrtém kroku, na základě výsledků z výstupu odlišnosti (tj. *difference output*) v Mplus,

uvolněny ty položky, které se mírně odlišovaly mezi nejvíce skupinami a u kterých jsme usoudily, že výrazně nenaruší obsahovou validitu konstruktu. U nově redukovaných třípoložkových škál subjektivního blahobytu byla opakovaně testována AMI. Na základě těchto výsledků bylo odhadnuto pořadí zemí odvozené ze srovnání latentních průměrů z dat ISSP 2011 a 2017. Všechny analýzy byly provedeny pomocí softwarového balíčku Mplus 7.2 (Muthén a Muthén 1998–2017; vzory pro Mplus inputy jsou k dispozici v tabulce A5 v online appendixu).

Tabulka 2: Shrnutí pořadí analytických kroků

Krok	Typ analýzy	Model (použitá škála)
Modely: *M1*: subjektivní blahobyt v ISSP 2011, *M2*: subjektivní blahobyt v ISSP 2017		
1a.	Konfirmační faktorová analýza (CFA)	M1, M2 (původní)
1b.	Testování přesné ekvivalence *EMI* – (MG-CFA) konfigurální ekvivalence metrická ekvivalence	M1 (původní), M2 (původní i redukovaná)
2.	Testování částečné ekvivalence *EMI* – MG-CFA (modifikační indexy) částečná metrická ekvivalence částečná skalární ekvivalence	M1 (původní), M2 (redukovaná)
3.	Testování přibližné ekvivalence *AMI* – MG-BSEM, bayesovský model s různými apriorními rozptyly, výstup odlišnosti skalární ekvivalence výstup odlišnosti – vyřazení pravděpodobně nesrovnatelných položek	M1 (původní), M2 (původní)
4.	Testování přibližné ekvivalence *AMI* – MG-BSEM, bayesovský model s různými apriorními rozptyly, výstup odlišnosti skalární ekvivalence	M1 (redukovaná), M2 (redukovaná)
5.	Porovnání latentních průměrů redukovaných škál	M1 (redukovaná), M2 (redukovaná)

Výsledky

V prvním kroku byla v každé zemi provedena separátní konfirmační faktorová analýza původních pětipoložkových škál z roku 2011 a 2017 (M1 a M2). Pro škálu subjektivního blahobytu z roku 2011 ukazovaly výsledky v jednotlivých zemích dobrou shodu modelu s daty; u většiny zemí se hodnoty RMSEA pohybovaly okolo 0,08 a hodnoty CFI nad 0,95. Bylo tak možné v dalších krocích testovat původní škálu se všemi pěti položkami. V případě škály z roku 2017 se však ukázalo, že model s danou faktorovou strukturou (pět položek) nereprezentuje data v některých zemích adekvátně. V každé zemi tak byly postupně vytipovány problematické položky identifikované především pomocí nízkých faktorových zátěží (*health17* a *achieve17*), vzájemných korelací a testem reliability dané škály, přičemž jsme dospěly k redukovanému modelu se třemi položkami, který ve všech jednotlivých zemích již dosáhl žádoucích (resp. perfektních) hodnot obecných indexů vhodnosti modelu. V dalších krocích byla testována již jen tato redukovaná třípoložková škála.

Ve druhém kroku jsme přistoupily k testování EMI pro všechny země dohromady za pomoci simultánní MG-CFA (viz tabulka 3). Ukázalo se, že v případě obou škál bylo

dosaženo konfigurální ekvivalence, což znamená, že modely s danou faktorovou strukturou adekvátně reprezentovaly data ve všech zemích, všechny položky dostatečně přispěly k vysvětlení daného faktoru a koncept subjektivního blahobytu byl v jednotlivých zemích podobně chápán. V případě metrické ekvivalence změna velikosti indexů RMSEA a CFI přesáhla mezní hodnoty, což naznačovalo příliš velký rozdíl mezi metrickým a konfigurálním modelem. Proto jsme přistoupily k testování částečné metrické ekvivalence, kdy jsme uvolňovaly nesrovnatelné faktorové zátěže za pomoci modifikačních indexů do té doby, než byl rozdíl mezi konfigurálním a částečně metrickým modelem dostatečně malý (tj. ΔCFI $\leq$ 0,01, ΔRMSEA $\leq$ 0,015).

Pro dosažení částečné metrické ekvivalence bylo nutné v případě škály z roku 2011 uvolnit dvě faktorové zátěže u položky *happy11* ve Finsku a položky *overcome11* v Koreji. V případě škály z roku 2017 pak čtyři faktorové zátěže u položky *overcome17* v USA a Francii a položky *satisfac17* na Islandu a v Thajsku. Jelikož bylo v případě obou škál dosaženo částečné metrické ekvivalence, v dalším kroku jsme testovaly již pouze částečnou skalární ekvivalenci.

Z výsledků testování škály z roku 2011 vyplývá, že po uvolnění dvou faktorových zátěží a celkem 44 konstant položek bylo dosaženo částečné skalární ekvivalence. Ze srovnání latentních průměrů škály tak musí být vyřazeno Japonsko, Itálie a Turecko, protože u těchto zemí bylo třeba uvolnit konstanty u tří položek, což je více než polovina položek testované škály. V případě ostatních 27 zemí bylo prokázáno, že lze validně srovnávat latentní průměr subjektivního blahobytu. Pro dosažení částečné skalární ekvivalence u redukované třípoložkové škály z roku 2017 musely být uvolněny čtyři faktorové zátěže a celkem 26 konstant položek. Pouze u Ruska bylo třeba volně odhadnout konstanty u dvou položek (tedy více než polovinu), a proto bylo nutné tuto zemi vyřadit ze srovnání latentních průměrů. U ostatních 29 zemí byly plně ekvivalentní alespoň dvě položky ze škály.

Kvůli problémům, které jsou spojeny s testováním částečné skalární ekvivalence v EMI (viz příslušná podkapitola výše), jsme ve třetím kroku přistoupily k jiné metodologii. Zaměřily jsme se na testování přibližné skalární ekvivalence daných škál, a to pomocí bayesovského modelování. Tabulka 4 zobrazuje indexy vhodnosti modelu pro testování AMI původních i redukovaných škál při použití různě velkých hodnot apriorního rozptylu. Ukázalo se, že oba modely původních pětipoložkových škál (M1, M2) při použití různých apriorních rozptylů (od 0,00 až po 0,05) vykazovaly špatnou shodu s daty (PPP < 0,05 a CI neobsahoval nulu) a nebylo tak dosaženo přibližné skalární ekvivalence. Tyto výsledky nejsou vzhledem k výsledkům testování EMI překvapivé.

V případě, že bayesovské modelování nepodpoří přibližnou skalární ekvivalenci měření ve všech zemích, lze použít dvě následující strategie (viz Muthén a Asparouhov 2013; Van de Schoot et al. 2013). Na základě výsledků z výstupů odlišnosti v Mplus (viz tabulka A3 v online appendixu) poskytujících informace o odchylkách konstant pro všechny položky ve všech zemích lze z analýzy odstranit buď 1) nesrovnatelné země, nebo 2) položky ze škály. V rámci analýzy jsme aplikovaly oba přístupy, nicméně první strategie výběru nesrovnatelných zemí nepřinesla žádná významná zlepšení v indexech vhodnosti modelu a přibližná skalární ekvivalence nebyla dosažena pro všechny položky škály ani v několika málo zemích (výsledky zde nejsou prezentovány).

Přistoupily jsme proto ke druhé strategii uvolňování položek s odchýlenými konstantami ve velkém počtu zemí ze škály a přepočítávání nových modelů. V případě obou škál jsme, podobně jako v EMI, došly k redukovaným škálám se třemi položkami, u kterých jsme rovněž zkoušely testovat modely s různými velikostmi apriorních rozptylů. V případě dat z roku 2011 bylo uspokojivých výsledků dosaženo při použití apriorních rozptylů 0,005, 0,01, 0,025 a 0,05, přičemž jako nejvhodnější byl vyhodnocen model s nejméně liberálním apriorním rozptylem 0,005 (nezbytné parametry modelu: $\Delta BIC \geq 20$, $\Delta DIC \geq 14$, $\Delta PPP \geq 0{,}01$, $PPP \geq 0{,}05$, CI obsahuje 0). U dat z roku 2017 byl jako nejvhodnější vyhodnocen model s apriorním rozptylem 0,01. Na rozdíl od testování EMI, výstup odlišnosti v Mplus odhalil odchylky faktorových zátěží od průměru, které však nemají vliv na srovnatelnost škály mezi zeměmi. V případě redukované škály z roku 2011 se jednalo pouze o položky *overcome11* v Jižní Koreji a *sad11* na Filipínách a z roku 2017 o položku *satisfac17* na Filipínách, v Rusku, Thajsku a Maďarsku a *overcome17* v USA a ve Francii. V případě konstant výsledky ukázaly, že se ve škále mírně odlišovaly téměř všechny konstanty položek v analyzovaných zemích, což naznačuje správnost použití AMI, ale nijak to nelimituje porovnání latentních průměrů mezi zeměmi.

Tabulka 3: Metoda přesné ekvivalence měření *EMI* původních a redukovaných škál subjektivního blahobytu v ISSP 2011 a 2017 – MG-CFA (metoda odhadu MLR); obecné indexy vhodnosti modelu

	Chi² (ΔChi²)	**df** (Δdf)	**CFI** (ΔCFI)	**RMSEA** (ΔRMSEA)	**C. I. RMSEA**
M1: ISSP 2011 (původní škála) – 30 zemí					
Konfigurální ekvivalence	1630,162	150	0,975	0,080	[0,077; 0,084]
Metrická ekvivalence	2464,772 (830,715)	266 (116)	0,963 (0,012)	0,074 (0,006)	[0,071; 0,076]
Částečná metrická ekvivalence	2326,264 (691,074)	264 (114)	0,965 (0,010)	0,071 (0,009)	[0,069; 0,074]
Částečná skalární ekvivalence	3012,621 (698,355)	335 (72)	0,955 (0,010)	0,072 (0,001)	[0,070; 0,075]
M2: ISSP 2017 (původní škála) – 30 zemí					
Konfigurální ekvivalence	6131,136	150	0,858	0,165	[0,161; 0,168]
M2a: ISSP 2017 (redukovaná škála) – 30 zemí					
Konfigurální ekvivalence	0,004	0	1,000	0,000	[0,000; 0,000]
Metrická ekvivalence	460,850 (460,847)	58 (58)	0,983 (0,017)	0,069 (0,069)	[0,063; 0,075]
Částečná metrická ekvivalence	305,399 (305,396)	54 (54)	0,990 (0,010)	0,056 (0,056)	[0,050; 0,062]
Částečná skalární ekvivalence	570,162 (279,583)	86 (32)	0,980 (0,010)	0,062 (0,006)	[0,057; 0,067]

Zdroj dat: ISSP 2011 a ISSP 2017.
Poznámky: Chi^2 = Satorra-Bentler Scaled χ^2, ΔChi^2 = Satorra-Bentler χ^2 Difference Test, df = stupně volnosti, Δdf = rozdíl df, CFI = Comparative Fit Index, ΔCFI = rozdíl CFI, RMSEA = Root Mean Square Error of Approximations, ΔRMSEA = rozdíl RMSEA, C. I. RMSEA = 95% interval spolehlivosti pro index RMSEA

Tabulka 4: Metoda testování přibližné (skalární) ekvivalence *AMI* původních a redukovaných škál subjektivního blahobytu v ISSP 2011 a 2017 – MG-BSEM; obecné indexy vhodnosti modelu a výběr apriorního rozptylu

Model	**Apriorní rozptyl**	**DIC**	**BIC**	**PPP**	**95% CI**
M1: ISSP 2011 (původní škála – 5 položek)					
30 zemí, N= 46,142	*0,000*	540499,408	542248,741	0,000	[8594,532; 8775,888]
	0,001	534911,630	539600,526	0,000	[2842,406; 3133.091]
	0,005	533981,505	538587,641	0,000	[1935.506; 2149,979]
	0,010	533904,847	538493,686	0,000	[1858,154; 2065,486]
	0,025	533876,454	538457,564	0,000	[1832,608; 2037,951]
	0,050	533859,285	538451551	0,000	[1828,644; 2031,960]
M2: ISSP 2017 (původní škála – 5 položek)					
30 zemí, N = 44,492	*0,000*	558839,410	560867,543	0,000	[15473,833; 15699,431]
	0,001	553555,921	558312,837	0,000	[9900,228; 10298,934]
	0,005	550444,730	555108,601	0,000	[6909,850; 7177,033]
	0,010	550233,063	554876,988	0,000	[6712,222; 6952,392]
	0,025	550160,654	554795,210	0,000	[6645,986; 6876,894]
	0,050	550137,735	554782,305	0,000	[6639,750; 6863,901]
M1a: ISSP 2011 (redukovaná škála – 3 položky)					
30 zemí, N= 46,142	*0,000*	312884,855	314096,669	0,000	[3809,697; 3928,872]
	0,001	309387,269	312456,670	0,000	[207,868; 371,913]
	0,005	309151,368	312178,410	0,223	[-36,407; 94,984]
	0,010	309125,510	312156,564	0,396	[-56,678; 74,293]
	0,025	309070,708	312150,139	0,466	[-60,816; 64,640]
	0,050	308969,932	312150,067	0,470	[-61,564; 65,572]
*Významné odchylky faktorových zátěží (LAM) od průměru (apriorní rozptyl = 0,005): Jižní Korea_ overcome11, Filipíny_sad11					
M2a: ISSP 2017 (redukovaná škála – 3 položky)					
30 zemí, N = 44,492	*0,000*	332555,954	333778,531	0,000	[4840,715; 4977,958]
	0,001	328574,606	331755,362	0,000	[872,557; 1099,203]
	0,005	327699,105	330830,316	0,002	[25,841; 170,783]
	0,010	327638,965	330757,399	0,217	[-40,476; 101,505]
	0,025	327575,995	330732,497	0,416	[-56,587; 73,835]
	0,050	327453,015	330730,422	0,450	[-58,492; 69,865]
*Významné odchylky faktorových zátěží (LAM) od průměru (apriorní rozptyl = 0,01): Francie_overcome17, USA_overcome17, Maďarsko_satisfac17, Filipíny_satisfac17, Rusko_satisfac17, Thajsko_satisfac17					

Poznámky: DIC = Deviance Information Criterion, PPP = Posterior Predictive *p* value, 95% CI = 95% kredibilní interval, ISSP 2011 (redukovaná škála: sad11, lost11, overcome11), ISSP 2017 (redukovaná škála: sad17, overcome17, satisfac17).

Na závěr bylo na základě výsledků EMI (částečná skalární ekvivalence) a AMI (přibližná skalární ekvivalence) příslušných škál subjektivního blahobytu odhadnuto pořadí zemí odvozené ze srovnání latentních průměrů z dat 2011 a 2017 a prostého aritmetického průměru škály, které prezentuje tabulka A4 v online appendixu. Ačkoliv v obou způsobech testování (EMI a AMI) byla podpořena (částečná) skalární ekvivalence, na základě které můžeme mezi zeměmi porovnávat latentní průměry daných škál, pořadí zemí v případě obou škál je poměrně rozdílné. U některých zemí je pořadí na základě latentních průměrů EMI a AMI až o sedm (Slovensko a Island v roce 2017), devět (Finsko, Indie, Švédsko a Litva v roce 2017), jedenáct (Velká Británie v roce 2011) či dokonce dvanáct míst odlišné (Austrálie v roce 2011). Například latentní průměr EMI označuje Slovensko v roce 2017 jako zemi s největším subjektivním blahobytem, avšak latentní průměr AMI ho posouvá až na osmé místo v pořadí. Rozdíly v pořadí názorně dokládají i hodnoty Pearsonova korelačního koeficientu (r = 0,839 pro rok 2011, resp. r = 0,842 pro rok 2017). Rovněž i pořadí zemí dle prostého aritmetického průměru je odlišné od pořadí zemí dle EMI (r = 0,877 pro rok 2011, resp. r = 0,786 pro rok 2017) a AMI (r = 0,947 pro rok 2011, resp. r = 0,946 pro rok 2017; viz tabulka A4b v online appendixu). Ale i v případě, kdy by si pořadí zemí dle latentních průměrů a prostého aritmetického průměru škály vzájemně odpovídaly, je nezbytné srovnatelnost škálových položek testovat. Stejně tak je vhodnější upřednostnit výpočet latentního průměru před prostým aritmetickým průměrem škály.

Diskuse a shrnutí

V mnoha výzkumech se subjektivní blahobyt neměří jako vícedimenzionální koncept pomocí baterie obsahující více položek, ale spíše jednopoložkovým měřením, které má velice omezenou obsahovou validitu, navíc neumožňuje kontrolu chyb měření a ověření reliability. V současné době sice existuje několik uznávaných škál pro měření subjektivního blahobytu (např. SWLS, PWI, WEMWBS, WHO), které jsou s úspěchem aplikovány především v národních výzkumech, nicméně tyto škály doposud nebyly nijak systematicky použity v mezinárodním výzkumu, a nemáme tedy k dispozici informace o jejich mezinárodní srovnatelnosti.

Cílem této studie bylo proto otestovat mezinárodní srovnatelnost baterie položek subjektivního blahobytu použitých ve výzkumu ISSP v roce 2011 v modulu Zdraví a v roce 2017 v modulu Sociální sítě. Sekundárním cílem bylo vzájemně porovnat dva přístupy k testování přesné (EMI) a přibližné (AMI) ekvivalence měření těchto položek. Tradiční přístup EMI uplatňující MG-CFA analýzu odhalil jako problematické dvě položky (*health17* a *achieve17*), které musely být již v prvním kroku vyřazeny ze škály ISSP 2017, a test částečné skalární ekvivalence po uvolnění několika desítek parametrů z testovaného modelu vyřadil navíc tři země z mezinárodního srovnání latentních průměrů škály ISSP 2011 a jednu zemi z mezinárodního srovnání latentních průměrů škály ISSP 2017. Tímto způsobem bylo ukázáno, že testování EMI je mezi mnoha skupinami velice obtížné, někdy nespolehlivé a často vede k zamítnutí plné skalární úrovně ekvivalence.

Novější přístup k testování AMI (viz Muthén a Asparouhov 2013), který dovoluje při testování modelu, aby se jeho parametry mohly od sebe mírně lišit mezi testovanými skupinami,

odhalil jako problematické dokonce čtyři položky; *happy11* a *health11* v případě škály ISSP 2011 a *health17* a *achieve17* v ISSP 2017. Bylo prokázáno, že vztahy mezi latentními proměnnými (resp. redukovanými třípoložkovými škálami) s jinými metricky ekvivalentními proměnnými z datového souboru je možné validně srovnávat ve všech participujících zemích výzkumu ISSP 2011 a 2017 a stejně tak je možné porovnat jejich latentní průměry. Ani jeden z přístupů neprokázal, že by bylo možné mezinárodně srovnávat prostý průměr hrubých skórů jednotlivých položek daných škál.

Zdá se tedy, že tradiční přístup k testování EMI neodhalil problematické položky *happy11* a *health11* v případě škály ISSP 2011, které označil za ještě částečně srovnatelné v mnoha zemích. Nicméně je nutné uvést, že na základě nezbytnosti uvolnění jejich parametrů problém srovnatelnosti s těmito položkami naznačil. Také navrhl vyloučit ze srovnání latentních průměrů tři země (tj. Japonsko, Itálii a Turecko) v případě srovnání škály ISSP 2011 a jednu zemi (tj. Rusko) ze srovnání škály ISSP 2017. Tyto země bayesovské modelování odhadlo jako ještě akceptovatelně odlišné pro srovnání latentních průměrů. Rovněž na základě srovnání latentních průměrů z EMI a AMI přístupu vyplývá, že EMI přístup detekoval odlišně pozici až osmi zemí mezi roky 2011 a 2017 na škále subjektivního blahobytu.

Redukce zkoumaných škál na menší počet manifestních položek je problematika, které by měli výzkumníci věnovat zvýšenou pozornost. Je to na jednu stranu vyjma vyřazení problematických zemí jedním z možných postupů při testování ekvivalence vybraných škál (viz Muthén a Asparouhov 2013; Van de Schoot et al. 2013), na druhou stranu tento zásah může významně narušit obsahovou validitu daného nástroje a vnitřní integritu dané škály. Námi vybrané baterie položek nejsou typické standardizované škály měřící subjektivní blahobyt vyvíjené výzkumníky s ohledem na přínos každé jednotlivé položky do obsahové validity konstruktu (na rozdíl od standardizovaných škál typu SWLS nebo PWI), ponechaly nám tedy určitý prostor k případné modifikaci.

V případě škály ISSP 2011 byla odstraněna jako kulturně nesrovnatelná položka reflektující zdravotní problémy respondenta a položka dotazující se na obecnou spokojenost, naopak ve škále zůstaly položky mapující frekvenci sklíčenosti respondenta, jeho pocitu ztráty sebedůvěry a pocitu, zda je v jeho silách překonat problémy. Z hlediska zachování obsahové validity dané škály nespatřujeme v redukované škále větší problém, neboť tři zbývající položky zachovaly věcnou podstatu konceptu zaměřujícího se na emoční stav respondenta v posledním měsíci, naopak otázky na zdravotní stav nebo obecné štěstí mohly daný koncept narušovat. Stejná situace platí i pro škálu ISSP 2017, kde byla opět vyjmuta otázka na zdravotní stav a dosahování cílů a byly ponechány položky měřící sklíčenost, překonávání problémů a obecnou spokojenost se životem. Opět redukovaná škála tak spíše ve výsledku měřila emoční stav respondenta místo jeho subjektivního pocitu zdraví či volních vlastností. Na druhou stranu v případě jiných škál může vyřazování položek ze škály způsobit značné problémy s věcným významem zkoumaného konstruktu, a je tak nutné pečlivě zvážit jejich případnou modifikaci.

Nevýhodou statistického přístupu k testování ekvivalence již sebraných dat je rovněž absence jakýchkoli podkladových materiálů k posouzení, proč jsou dané položky nesrovnatelné a musí být ze škály vyloučeny. Ačkoli diskuse o možném vysvětlení kulturně

specifického vnímání daných položek nemůžou být z hlediska rozsahu běžnou součástí takovýchto metodologických studií, jejich nedostatek je v této oblasti patrný a ztěžuje další práci na vývoji srovnatelných škál v mezinárodním výzkumu. Jak metodologické, tak i věcné důvody (sociální desirabilita, zkreslení souhlasem, styly odpovídání, nevhodný překlad, různá interpretace konceptu, výběr škály) mohou významně ovlivnit fungování položky v multikulturním prostředí (Davidov, Cieciuch a Schmidt 2018; Bjørnskov 2010),[4] a proto především kvalitativní přístupy jako kognitivní dotazování a online probing umožňují návrat do terénu a identifikování skutečných důvodů chybějící ekvivalence položky v daném kulturním prostředí (více viz Meitinger 2017).

Z využitelných kvantitativních přístupů pak lze jako příklad uvést zejména metodu odlišného fungování položky (*differential item functioning*, dále jen DIF), vycházející z teorie odpovědi na položku (*item response theory*, dále jen IRT) (viz Teresi 2006). Tato teorie se běžně využívá pro konstrukci a vývoj testů, škál, dotazníků a v psychometrii (viz Embretson a Reise 2000; Reeve a Fayers 2005) a primárně se zaměřuje na fungování jednotlivých položek, méně často pak na úroveň celého testu či modelu (DeMars 2000). Pokud navíc testujeme ekvivalenci mezi více skupinami najednou, je nutné aplikovat pokročilejší verzi IRT s kovariáty (IRT-C), která umožňuje zkoumat DIF napříč více proměnnými/skupinami současně (viz Tay, Huang a Vermunt 2016; Tay, Vermunt a Wang 2013).

Výsledky sice potvrdily, že lze mezinárodně validně srovnávat latentní průměry redukovaných třípoložkových škál subjektivního blahobytu z dat ISSP 2011 a 2017, na druhou stranu je nutné upozornit na skutečnost, že tyto výsledky nemusí být potvrzeny na jiných datech, v jiných časových bodech sběru dat či s jinými škálami. Potvrzení ekvivalence měření těchto škál v datech ISSP nemusí nutně znamenat, že bude ekvivalence měření daných škál dosaženo vždy za jakýchkoli okolností. Jak zdůrazňuje kolektiv autorů Cieciuch et al. (2014), ekvivalence měření položek musí být testována opakovaně mezi různými skupinami, neboť jejich nesrovnatelnost je způsobena řadou vnějších i vnitřních aspektů daného výzkumu ve specifický čas. Zercher et al. (2015) také navrhuje, že v prvním kroku by měla být vždy otestována přesná ekvivalence EMI, a pokud by bylo nutné přistoupit k testování částečné (skalární) EMI, pak je vhodné použít testování AMI, obzvláště pokud porovnáváme mnoho skupin ve velkých mezinárodních výzkumech (Van de Schoot et al. 2013). K tomuto testování, jak již bylo řečeno, lze využít buď výše popsané bayesovské modelování, ale rovněž i jiné slibné techniky jako alignment proceduru, pokud například očekáváme, že se ve škále poměrně hodně liší jen několik málo parametrů a ostatní jsou téměř shodné (více viz Van de Schoot et al. 2013; Pokropek, Davidov a Schmidt 2019). Doufáme, že naše výsledky budou stimulovat další sekundární analýzy, které využijí škály subjektivního blahobytu v mezinárodních komparacích, a případně podpoří i navazující metodologický výzkum příčin nesrovnatelnosti některých položek.

---

4 Z hlediska problematického překladu Bjørnskov (2010) upozorňuje na skutečnost, že ruský a francouzský překlad slova „happy“ znamená šťastný či mít štěstí, zatímco dánský překlad „lykkelig“ klade větší důraz na dimenzi úspěchu či dosažení. Navíc Benítez et al. (2018) ukazují, že není žádný rozdíl v interpretaci extrémních hodnot odpovědí na škále měřící štěstí či životní spokojenost, ale rozdíly se objevují v interpretaci středních/středových kategorií odpovědí.

Literatura

ANÝŽOVÁ, Petra. 2014. „Srovnatelnost Schwartzovy hodnotové škály v mezinárodních datech.“ *Sociologický časopis/Czech Sociological Review* 50(4): 547–580.

ANÝŽOVÁ, Petra. 2015. *Srovnatelnost postojových škál v komparativním výzkumu*. Olomouc: Univerzita Palackého.

ASPAROUHOV, Tihomir a Bengt MUTHÉN. 2014. „Multiple-Group Factor Analysis Alignment.“ *Structural Equation Modeling: A Multidisciplinary Journal* 21(4): 495–508.

ASPAROUHOV, Tihomir a Bengt MUTHÉN. 2010. *Bayesian Analysis of Latent Variable Models Using Mplus.* Muthén & Muthén. Technical report (www.statmodel.com). Los Angeles, CA.

BENÍTEZ, Isabel, José Luis PADILLA, Fons van de VIJVER a Amaya CUEVAS. 2018. „What Cognitive Interviews Tell Us about Bias in Cross-Cultural Research: An Illustration Using Quality-of-Life Items.“ *Field Methods* 30(4): 277–294.

BJØRNSKOV, Christian. 2010. „How Comparable Are the Gallup World Poll Life Satisfaction Data?“ *Journal of Happiness Studies* 11(1): 41–60.

BUJACZ, Aleksandra, Joar VITTERSØ, Veronika HUTA a Lukasz D. KACZMAREK. 2014. „Measuring Hedonia and Eudaimonia as Motives for Activities: Cross-National Investigation Through Traditional and Bayesian Structural Equation Modeling.“ *Frontiers in Psychology* 5: Article 984.

BYRNE, Barbara M. 2010. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Second Edition. New York, NY: Taylor & Francis Group.

BYRNE, Barbara M., Richard J. SHAVELSON a 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.

CIECIUCH, Jan a Eldad DAVIDOV. 2012. „A Comparison of the Invariance Properties of the PVQ-40 and the PVQ-21 to Measure Human Values across German and Polish Samples.“ *Survey Research Methods* 6(1): 37–48.

CIECIUCH, Jan, Eldad DAVIDOV, René ALGESHEIMER a Peter SCHMIDT. 2018. „Testing for Approximate Measurement Invariance of Human Values in the European Social Survey.“ *Sociological Methods & Research* 47(4): 665–686.

CIECIUCH, Jan, Eldad DAVIDOV, Peter SCHMIDT, René ALGESHEIMER a Shalom H. SCHWARTZ. 2014. „Comparing Results of an Exact vs. an Approximate (Bayesian) Measurement Invariance Test: A Cross-Country Illustration with a Scale to Measure 19 Human Values.“ *Frontiers in Psychology* 5: Article 982.

CUMMINS, Robert A., Richard ECKERSLEY, Julie PALLANT, Jackie van VUGT a RoseAnne MISAJON. 2003. „Developing a National Index of Subjective Wellbeing: The Australian Unity Wellbeing Index.“ *Social Indicators Research* 64(2): 159–190.

DAVIDOV, Eldad, Jan CIECIUCH, Bart MEULEMAN, Peter SCHMIDT, René ALGESHEIMER a Mirjam HAUSHERR. 2015. „The Comparability of Measurements of Attitudes toward Immigration in the European Social Survey.“ *Public Opinion Quarterly* 79(S1): 244–266.

DAVIDOV, Eldad, Bart MEULEMAN, Jan CIECIUCH, Peter SCHMIDT a Jaak BILLIET. 2014. „Measurement Equivalence in Cross-National Research.“ *Annual Review of Sociology* 40(1): 55–75.

DAVIDOV, Eldad, Jan CIECIUCH a Peter SCHMIDT. 2018. „The Cross-Country Measurement Comparability in the Immigration Module of the European Social Survey 2014–15.“ *Survey Research Methods* 12(1): 15–27.

DAVIDOV, Eldad, Peter SCHMIDT a Jaak BILLIET (eds.). 2011. *Cross-cultural Analysis: Methods and Applications*. New York, NY: Routledge.

DEMARS, Christine. 2010. *Item Response Theory*. Oxford: University Press.

DIENER, Ed. 2009. *The Science of Well-Being. The Collected Works of Ed Diener*. Dordrecht: Springer.

DIENER, Ed, Robert A. EMMONS, Randy J. LARSEN a Sharon GRIFFIN. 1985. „The Satisfaction With Life Scale." *Journal of Personality Assessment* 49(1): 71–75.

DIENER, Ed, Richard E. LUCAS a Shigehiro OISHI. 2018. „Advances and Open Questions in the Science of Subjective Well-Being." *Collabra: Psychology* 4(1): 15.

DIENER, Ed, Derrick WIRTZ, Robert BISWAS-DIENER, William TOV, Chu KIM-PRIETO, Dong-won CHOI a Shigehiro OISHI. 2009. „New Measures of Well-Being." Pp. 247–266 in Ed DIENER (ed.). *Assessing Well-Being. The Collected Works of Ed Diener*. Dordrecht: Springer Netherlands.

DIENER, Ed a Richard E. LUCAS. 1999. „Personality and Subjective Well-Being." Pp. 213–229 in Daniel KAHNEMAN, Ed DIENER a Norbert SCHWARZ (eds.). *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Found.

EMBRETSON, Susan E. a Steven P. REISE. 2000. *Item Response Theory for Psychologists*. Mahwah, NJ: L. Erlbaum Associates.

EMERSON, Scott D., Martin GUHN a Anne M. GADERMANN. 2017. „Measurement Invariance of the Satisfaction with Life Scale: Reviewing Three Decades of Research." *Quality of Life Research* 26(9): 2251–2264.

FORS, Filip a Joakim KULIN. 2016. „Bringing Affect Back in: Measuring and Comparing Subjective Well-Being Across Countries." *Social Indicators Research* 127(1): 323–339.

HUPPERT, Felicia A., Nic MARKS, Andrew CLARK, Johannes SIEGRIST, Alois STUTZER, Joar VITTERSØ a Morten WAHRENDORF. 2009. „Measuring Well-Being Across Europe: Description of the ESS Well-Being Module and Preliminary Findings." *Social Indicators Research* 91(3): 301–315.

HUPPERT, Felicia A. a Timothy T. C. SO. 2013. „Flourishing Across Europe: Application of a New Conceptual Framework for Defining Well-Being." *Social Indicators Research* 110(3): 837–861.

CHEN, Fang Fang. 2007. „Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 14(3): 464–504.

JOVANOVIĆ, Veljko, Robert A. CUMMINS, Melissa WEINBERG, Ljiljana KALITERNA a Zvjezdana PRIZMIC-LARSEN. 2019. „Personal Wellbeing Index: A Cross-Cultural Measurement Invariance Study Across Four Countries." *Journal of Happiness Studies* 20(3): 759–775.

KAHNEMAN, Daniel, Ed DIENER a Norbert SCHWARZ (eds.). 1999. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Found.

KIM, Eun Sook, Chunhua CAO, Yan WANG a Diep T. NGUYEN. 2017. „Measurement Invariance Testing with Many Groups: A Comparison of Five Approaches." *Structural Equation Modeling: A Multidisciplinary Journal* 24(4): 524–544.

McGILLIVRAY, Mark a Matthew CLARKE. 2006. *Understanding Human Well-being*. Tokyo: UNU Press.

MEITINGER, Katharina. 2017. „Necessary but Insufficient." *Public Opinion Quarterly* 81(2): 447–472.

MEREDITH, William. 1993. „Measurement Invariance, Factor Analysis and Factorial Invariance." *Psychometrika* 58(4): 525–543.

MEULEMAN, Bart, Eldad DAVIDOV a Jaak BILLIET. 2009. „Changing Attitudes toward Immigration in Europe, 2002–2007: A Dynamic Group Conflict Theory Approach." *Social Science Research* 38 (2): 352–365.

MÎNDRILĂ, Diana. 2010. „Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data." *International Journal of Digital Society* 1(1): 60–66.

MONROE, Scott a Li CAI. 2015. „Evaluationg Structural Equation Models for Categorical Outcomes: A New Test Statistic and a Practical Challenge of Interpretation." *Multivariate Behavioral Research* 50(6): 569–583.

MUTHÉN, Bengt a Tihomir ASPAROUHOV. 2012. „Bayesian Structural Equation Modeling: A More Flexible Representation of Substantive Theory.“ *Psychological Methods* 17(3): 313–335.

MUTHÉN, Bengt a Tihomir ASPAROUHOV. 2013. „BSEM Measurement Invariance Analysis.“ *Mplus Web Notes* 17: 1–48.

MUTHÉN, K. Linda a Bengt O. MUTHÉN. 1998–2017. *Mplus User's Guide (8.)*. Los Angeles, CA: Muthén & Muthén.

MUTHÉN, Bengt O. 1999. „Fit Indices for Categorical Outcumes.“ *Mplus Discussion*. 2. listopadu. Cit. 16. srpna 2020 (http://www.statmodel.com/discussion/messages/23/26.html).

NAVRUZ, Bilgin. 2016. „*The Behaviors of Robust Weighted Least Squares Estimation Techniques for Categorical/Ordinal Data in Multilevel CFA Models*.“ Doctoral dissertation. Texas A & M University.

OECD. 2013. *OECD Guidelines on Measuring Subjective Well-Being*. OECD.

POKROPEK, Artur, Eldad DAVIDOV a Peter SCHMIDT. 2019. „A Monte Carlo Simulation Study to Assess the Appropriateness of Traditional and Newer Approaches to Test for Measurement Invariance.“ *Structural Equation Modeling: A Multidisciplinary Journal* 26(5): 724–744.

POKROPEK, Artur, Peter SCHMIDT a Eldad DAVIDOV. 2020. „Choosing Priors in Bayesian Measurement Invariance Modeling: A Monte Carlo Simulation Study.“ *Structural Equation Modeling: A Multidisciplinary Journal* online: 1–15.

RAUDENSKÁ, Petra. 2020. „The Cross-Country and Cross-Time Measurement Invariance of Positive and Negative Affect Scales: Evidence from European Social Survey.“ *Social Science Research* 86: 102369.

REEVE, Bryce B. a Peter FAYERS. (2005). „Applying Item Response Theory Modelling for Evaluating Questionnaire Item and Scale Properties.“ Pp. 55–73 in Peter FAYERS a Ron HAYS (eds.). *Assessing Quality of Life in Clinical Trials: Methods and Practice,* 2nd Edition. Oxford: Oxford University Press.

RUTKOWSKI, Leslie a Dubravka SVETINA. 2014. „Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys.“ *Educational and Psychological Measurement* 74(1): 31–57.

RUTKOWSKI, Leslie a Dubravka SVETINA. 2017. „Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance.“ *Applied Measurement in Education* 30(1): 39–51.

RYAN, Richard M. a Edward L. DECI. 2001. „On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being.“ *Annual Review of Psychology* 52(1): 141–166.

SEDDIG, Daniel a Heinz LEITGÖB. 2018. „Approximate Measurement Invariance and Longitudinal Confirmatory Factor Analysis: Concept and Application with Panel Data.“ *Survey Research Methods* 12(1): 29–41.

STEENKAMP, Jan-Benedict E. M. a Hans BAUMGARTNER. 1998. „Assessing Measurement Invariance in Cross-National Consumer Research.“ *Journal of Consumer Research* 25(1): 78–107.

TAY, Louis, Quiming HUANG a Jeroen K. VERMUNT. 2016. „Item Response Theory With Covariates (IRT-C): Assessing Item Recovery and Differential Item Functioning for the Three-Parameter Logistic Model.“ *Educational and Psychological Measurement* 76(1): 22–42.

TAY, Louis, Jeroen K. VERMUNT a Chun WANG. 2013. „Assessing the Item Response Theory With Covariate (IRT-C) Procedure for Ascertaining Differential Item Functioning.“ *International Journal of Testing* 13(3): 201–222.

TERESI, Jeanne A. 2006. „Different Approaches to Differential Item Functioning in Health Applications: Advantages, Disadvantages and Some Neglected Topics.“ *Medical Care* 44(11): 152–170.

VAN DE SCHOOT, Rens, Anouck KLUYTMANS, Lars TUMMERS, Peter LUGTIG, Joop HOX a Bengt MUTHÉN. 2013. „Facing Off with Scylla and Charybdis: A Comparison of Scalar, Partial,

and the Novel Possibility of Approximate Measurement Invariance.“ *Frontiers in Psychology* 4: Article 770.

VANDENBERG, Robert J. a 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.

WATERMAN, Alan S. 1993. „Two Conceptions of Happiness: Contrasts of Personal Expressiveness (Eudaimonia) and Hedonic Enjoyment.“ *Journal of Personality and Social Psychology* 64(4): 678–691.

WHO. 1998. *Wellbeing Measures in Primary Health Care/The Depcare Project*. WHO Regional Office for Europe: Copenhagen.

XIA, Yan. 2016. *Investigating the Chi-Square-Based Model-Fit Indexes for WLSMV and ULSMV Estimators*. Doctoral disseration. Florida State University.

ZERCHER, Florian, Peter SCHMIDT, Jan CIECIUCH a Eldad DAVIDOV. 2015. „The Comparability of the Universalism Value Over Time and Across Countries in the European Social Survey: Exact Vs. Approximate Measurement Invariance.“ *Frontiers in Psychology* 6: Article 733.

ŻEMOJTEL-PIOTROWSKA, Magdalena et al. 2017. „Measurement Invariance of Personal Well-Being Index (PWI-8) Across 26 Countries.“ *Journal of Happiness Studies* 18(6): 1697–1711.

Data

ISSP 2011: ISSP Research Group (2015): International Social Survey Programme: Health and Health Care – ISSP 2011. GESIS Data Archive, Cologne. ZA5800 Data file Version 3.0.0, https://doi.org/10.4232/1.12252

ISSP 2017: ISSP Research Group (2019): International Social Survey Programme: Social Networks and Social Resources – ISSP 2017. GESIS Data Archive, Cologne. ZA6980 Data file Version 2.0.0, https://doi.org/10.4232/1.13322

Autorky

Petra Anýžová Raudenská pracuje jako vědecký pracovník v Sociologickém ústavu AV ČR. Jejími výzkumnými zájmy jsou metodologie mezinárodních výzkumů, vzdělanostní a sociální nerovnosti, lidský kapitál, životní spokojenost a problematika hodnotových orientací. Je autorkou monografie *Srovnatelnost postojových škál v komparativním výzkumu* (2015) a editorkou několika dalších knih. Publikovala v českém i anglickém čísle *Sociologického časopisu*, slovenské *Sociológii* a v mezinárodních časopisech *Social Science Research*, *International Sociology*, *International Journal of Sociology and Social Policy* a *Innovation: The European Journal of Social Science Research* a rovněž v řadě domácích monografií. Kontakt: petra.raudenska@soc.cas.cz

Radka Hanzlová je studentkou třetího ročníku doktorského studia sociologie na Filozofické fakultě UK a zároveň pracuje jako odborná pracovnice v Sociologickém ústavu AV ČR. Mezi její výzkumné zájmy patří metodologie kvantitativních výzkumů, dotazníková šetření, testování a vývoj škál, subjektivní blahobyt, štěstí, životní spokojenost a osobní hodnoty. Kontakt: radka.hanzlova@soc.cas.cz