

Testování psychometrických vlastností a ekvivalence české verze Škály spokojenosti se životem (SWLS) pomocí metod konfirmační faktorové analýzy, teorie odpovědi na položku a bayesovského modelování*

RADKA HANZLOVÁ[Ⓛ], PETRA RAUDENSKÁ[Ⓛ]**

Sociologický ústav AV ČR, v. v. i., Praha

Testing the Psychometric Properties and Equivalence of the Czech Version of the Satisfaction with Life Scale (SWLS) using Confirmatory Factor Analysis, Item Response Theory, and Bayesian Modelling

The Satisfaction with Life Scale (SWLS) is one of the most commonly used instruments for measuring life satisfaction. The aim of this study is to test the psychometric properties of the Czech version of the SWLS using Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT) and to test its invariance between social groups in terms of gender, age, and education using Bayesian modelling on a representative sample of the Czech online population, as the scale has not yet been tested on representative data in the Czech Republic. The research sample consists of 960 respondents aged 18 to 69 years. The results confirmed that the psychometric properties of the Czech version of the SWLS are very good, but, at the same time, it is evident that the fifth item shows worse results than the other four items. In terms of dimensionality, CFA and IRT confirmed its modified single-factor structure with correlated residuals between the fourth and fifth items as the most appropriate. Testing for approximate measurement invariance using Bayesian modelling showed that the SWLS measures comparably between groups based on gender, age, and education. In conclusion, the Czech version of the SWLS is a suitable, verified, and reliable instrument for measuring the life satisfaction of Czech citizens.

Keywords: SWLS, Confirmatory Factor Analysis (CFA), Item Response Theory (IRT), Bayesian modelling BSEM, psychometrics, measurement invariance

Sociologický časopis / Czech Sociological Review, 2023 (online first)

<https://doi.org/10.13060/csr.2023.047>

* Radka Hanzlová mohla tento článek napsat s podporou projektu „Dynamika postojů adolescentů k cizincům“, financovaného Grantovou agenturou České republiky (reg. č. 23-07283M). Petra Raudenská napsala tento článek s podporou projektu „Zdraví a štěstí v post-covidovém světě“, financovaného Grantovou agenturou České republiky (reg. č. 22-09220S).

** Veškerou korespondenci posílejte na adresu: Mgr. Radka Hanzlová, Ph.D., Sociologický ústav AV ČR, v. v. i., Jilská 1, 110 00 Praha 1, e-mail: radka.hanzlova@soc.cas.cz.

© Autorky, Sociologický ústav AV ČR, v. v. i., 2023

Tento článek je publikován v režimu tzv. otevřeného přístupu k vědeckým informacím (Open Access), který je distribuován pod licencí Creative Commons (CC BY-NC 4.0) (<http://creativecommons.org/licenses/by-nc/4.0/>).

Úvod

Subjektivní blahobyt či osobní pohoda je zejména v psychologii a sociologii široce používaný termín, který se objevil koncem padesátých let 20. století, kdy se začal používat jako jeden z indikátorů kvality života (Keyes et al., 2002). Subjektivní blahobyt v sobě obsahuje tři měřitelné komponenty, mezi které patří pozitivní afekty, negativní afekty a životní spokojenost (Andrews a Withey, 1976). Tato studie se zaměřuje pouze na jednu dílčí komponentu subjektivní osobní pohody, konkrétně životní spokojenost, kterou lze definovat jako „kognitivní a globální hodnocení kvality vlastního života jako celku“ (Pavot a Diener, 1993). Poznatky o životní spokojenosti populace jsou důležité, jelikož životní spokojenost má sociální, ekonomické, zdravotní a psychologické dopady na život jedinců a rovněž představuje důležitý základ pro intervence v oblasti různých politik (Dolan a Metcalfe 2012; Dolan a White 2007).

V mezinárodních komparativních výzkumech se životní spokojenost obvykle měří pomocí jedné jednoduché otázky, např. v Evropském sociálním výzkumu (ESS) či Mezinárodním programu sociálního výzkumu (ISSP): „*Když vezmeme v úvahu všechny okolnosti, jak jste v současnosti celkově spokojen(a) se svým životem?*“ s možnostmi odpovědi na 11bodové škále od 0 „naprosto nespokojený(á)“ do 10 „naprosto spokojený(á)“ nebo v případě Eurobarometru: „*Řekl(a) byste, že jsme celkově se svým životem (1) velmi spokojen(a), (2) spíše spokojen(a), (3) spíše nespokojen(a) nebo (4) velmi nespokojen(a)?*“

Způsob měření pomocí jednoduché otázky je sice nenáročný a snadno proveditelný, nicméně zároveň problematický jak z hlediska teorie, tak i metodologie. Z teoretického hlediska představuje hlavní problém samotná konceptualizace termínu životní spokojenosti ovlivňující obsahovou validitu, neboť každý jedinec si pod tímto pojmem může představit něco jiného a ve výsledku pak můžeme jednou otázkou měřit různé koncepty. S tím nepochybně souvisí i skutečnost, že životní spokojenost lze dále vztáhnout k různým dílčím oblastem (subdimenzím), např. práce, rodina, vztahy atp., přičemž ačkoliv se ve znění otázky ptáme na celkovou životní spokojenost, lidé svoji odpověď nejčastěji vztahují k současné situaci a problémům, které řeší. To neoddělitelně souvisí s hlavním metodologickým problémem měření pomocí jedné otázky, kterým je jeho nízká reliabilita (spolehlivost). Odpovědi na jednu otázku v rámci celého dotazníku jsou totiž výrazně ovlivněny kontextuálními faktory, např. předcházející otázkou, současnou náladou, okolnostmi, kdy, kde a s kým dotazování probíhá, pamětí, ale i počasím (Huppert et al., 2009; Schwarz, 1987; Schwarz a Clore, 1983; Schwarz a Strack, 1991).

Z vícepoložkového měření celkové životní spokojenosti je nejvíce používaným a ověřeným nástrojem Škála spokojenosti se životem (*Satisfaction with Life Scale*, dále jen „SWLS“) od Dienera et al. (1985), kterou testujeme i v této studii. SWLS obsahuje pět položek hodnocených na sedmibodové Likertově škále od 1 „rozhodně nesouhlasím“ do 7 „rozhodně souhlasím“ (Tabulka 2). Finální, pětipoložková podoba škály byla vyvinuta z původních 48 položek vztažených k hodnocení všech složek subjektivní osobní pohody, tedy pozitivních a negativních afektů a životní spokojenosti. Faktorová analýza všech položek vyextrahovala tři faktory jako tři komponenty subjektivní osobní pohody. Jelikož cílem bylo měřit pouze životní spokojenost, dva faktory vztahující se k afektům a položky s nízkými faktorovými zátěžemi z faktoru životní spokojenosti byly vyřazeny, což vedlo ke snížení počtu položek na 10. Některé z těchto položek měly stejný význam a byly tak duplicitní, polovina z nich byla vyřazena a

výsledkem je finální škála s pěti položkami, která se obzvláště v západoevropských zemích a Spojených státech používá od osmdesátých let (Diener et al., 1985). V následujících desetiletích bylo vyvinuto několik dalších nástrojů pro měření životní spokojenosti, nicméně všechny vycházejí z původní SWLS s pěti položkami. Jako příklad lze uvést 15položkovou *Temporal Satisfaction with Life Scale* (TSWLS) (Pavot et al., 1998) nebo šestipoložkovou *Riverside Life Satisfaction Scale* (RLSS) (Margolis et al., 2019).

SWLS byla a stále je intenzivně zkoumána a testována (Emerson et al., 2017; Pavot a Diener, 2008; Vassar, 2007). Naprostá většina studií využívajících metod klasické testové teorie (*Classical Test Theory*, dále jen „CTT“) potvrdila její dobré psychometrické vlastnosti včetně validity, vnitřní konzistentnosti a test-retest reliability (Diener et al., 1985; Lucas et al., 1996; Pavot et al., 1991; Pavot a Diener, 1993; Vassar, 2007). Z hlediska testování ekvivalence a dimenzionality SWLS se však výsledky liší (Clench-Aas et al., 2011; Emerson et al., 2017). Ekvivalence SWLS je testována nejčastěji z hlediska sociodemografických charakteristik (pohlaví, věk, vzdělání) a dalších znaků (např. etnicita, rodinný stav či časové body), přičemž na základě výsledků studií nelze stanovit jednoznačný závěr o dosažené úrovni ekvivalence (Emerson et al., 2017; Jovanović, 2017). Co se týká dimenzionality SWLS, většina studií potvrdila její jednodimenzionální strukturu s jedním faktorem, nicméně některé studie navrhuji dvoudimenzionální strukturu s faktorem „přítomnost“ (první až třetí položka) a „minulost“ (čtvrtá a pátá položka), případně další modifikace, jako např. jednofaktorová struktura s korelací chyb mezi čtvrtou a pátou položkou (Clench-Aas et al., 2011; Jovanović, 2017).

Ačkoliv SWLS byla testována v mnoha zemích po celém světě, v České republice jen omezeně. Pro Českou republiku existuje pouze jedna starší studie z roku 1999, testující její psychometrické vlastnosti na malém vzorku vysokoškolských studentů (Lewis et al., 1999), a na ni navazující studie, testující její stabilitu v čase, ale opět na velmi malém a homogenním vzorku (Navrátil a Lewis, 2006). Právě tyto charakteristiky (malý výzkumný vzorek a jeho homogenita) představují problém pro adekvátní validizaci škály a zobecnitelnost výsledků, nicméně zároveň jsou typickým znakem většiny dosud realizovaných studií (Clench-Aas et al., 2011; Hutlell a Gustavsson, 2008).

Cílem této studie je pomocí konfirmační faktorové analýzy (*Confirmatory Factor Analysis*, dále jen „CFA“) a Teorie odpovědi na položku (*Item Response Theory*, dále jen „IRT“) otestovat psychometrické vlastnosti českého překladu SWLS. Pro testování ekvivalence z hlediska pohlaví, věku a vzdělání byl použit novější přístup testování tzv. přibližné ekvivalence měření (*approximate measurement invariance*, dále jen „AMI“), vycházející z bayesovské statistiky (dále jen „BSEM“). Ve výzkumu subjektivní osobní pohody nebyl dosud bayesovský přístup příliš aplikován (např. Raudenská, 2020; Raudenská a Hanzlová, 2021), nicméně jeho závěry poskytují spolehlivější informace o nesrovnatelnosti jednotlivých položek ve škále.

Předpokládáme, že hlavním přínosem této studie bude otestování a validizace českého překladu SWLS a zhodnocení, zda je možné škálu použít pro srovnání mezi různými sociálními skupinami v české populaci. Zároveň se aplikací v českém prostředí nepříliš používaných metod (IRT, BSEM) budeme snažit ukázat jejich možné využití při modelování dat a podnítit zájem českých výzkumníků o použití těchto metod v dalších podobně orientovaných metodologických studiích.

Konfirmační faktorová analýza (CFA) a Teorie odpovědi na položku (IRT)

Konfirmační faktorová analýza (*Confirmatory Factor Analysis*, „CFA“), jejímž autorem je švédský psychometrik Jöreskog (1969), je statistická metoda spadající pod větší celek metodologie strukturního modelování. Jak je z jejího názvu patrné, jejím cílem je ověření (konfirmace) určitého apriorního (předem daného) modelu vztahů mezi položkami, který by měl být založen na teoretickém základě. Již na začátku analýzy má tak výzkumník představu o podobě a struktuře modelu a aplikací CFA na testovaná data se ho snaží potvrdit či vyvrátit. Vzhledem k tomu, že dimenzionalita SWLS škály není jasně prokázána a existuje několik teoretických modelů (Clench-Aas et al., 2011; Emerson et al., 2017), představuje CFA velmi vhodnou metodu pro jejich otestování.

Pro stanovení rozhodnutí o výsledcích CFA se používají chí-kvadrát test a indexy vhodnosti modelu, např. CFI (Comparative Fit Index) nebo RMSEA (Root Mean Squared Error of Approximation). Vzhledem k tomu, že chí-kvadrát je citlivý na velikost vzorku a u větších souborů vychází téměř vždy statisticky významný (Chen, 2007; Kline, 2005), bude v této studii použit pouze jako deskriptivní ukazatel a hodnocení bude založeno na hodnotách indexů CFI a RMSEA. V případě CFI hodnoty nad 0,90 ukazují na přijatelný model a hodnoty nad 0,95 signalizují dobrou shodu modelu s daty (West et al., 2012). Pro RMSEA je hranice pro přijatelný model 0,08 a pro dobrý model 0,06 (West et al., 2012).

Velmi vhodnou metodou pro hodnocení kvality a vývoj škál je pak Teorie odpovědi na položku (*Item Response Theory*, „IRT“), která však v případě SWLS nebyla dosud příliš aplikována (Avcu, 2021; Nima et al., 2020; Oishi, 2006; Vittersø et al., 2005). Jde o poměrně moderní metodologický přístup založený na matematickém modelování vztahu latentní proměnné a měřitelné odpovědi na konkrétní položku v rámci celého testu či škály, čímž se právě primárně odlišuje od CTT, která pracuje s celým testem/škálou. Na základě informací o fungování jednotlivých položek lze vylepšit psychometrické vlastnosti škály jako celku (De Ayala, 2009; DeMars, 2010). Právě zaměření na fungování jednotlivých položek patří mezi hlavní přínosy IRT. Tím dalším je vizuální zobrazení pomocí různých grafů, které přispívá k lepší interpretaci a pochopitelnosti výsledků. Aplikací IRT lze rovněž zkoumat vhodnost délky odpovědní škály a styl odpovídání respondentů. Jinými slovy můžeme zjistit, zda je počet bodů na odpovědní škále odpovídající nebo zda jsou některé kategorie nadbytečné, zda respondenti nevyužívají pouze krajní či prostřední body odpovědní škály a mnoho dalšího (De Jong et al., 2008; O'Connor et al., 2015).

SWLS obsahuje pět ordinálních položek hodnocených na sedmibodové Likertově škále, proto je na ně vhodné aplikovat model odstupňovaných odpovědí, tzv. *Graded Response Model* („GRM“). Na základě tohoto modelu má každá položka odhadnutý vlastní diskriminační parametr (a), který označuje, jak dobře či špatně daná položka diskriminuje a jak silně se vztahuje k latentnímu rysu¹, a určitý počet threshold parametrů (b_i), které lze definovat jako prahové hodnoty mezi jednotlivými kategoriemi na odpovědní škále (počet threshold parametrů je vždy o jeden nižší než počet bodů na odpovědní škále). Co se grafického zobrazení týká, pro zhodnocení fungování kategoriálních ordinálních položek je nejpřínosnější graf charakteristických křivek položky (tzv. *category characteristic curve*, „CCC“). Z těchto křivek můžeme vyčíst, jaká je pravděpodobnost, že respondent s určitou úrovní latentního rysu odpoví

¹ Jde o ekvivalent faktorových zátěží v CTT.

na položku konkrétní kategorií na odpovědní škále. Přechody mezi jednotlivými kategoriemi na odpovědní škále jsou hodnoty threshold parametrů (b). Tento graf lze zároveň využít pro analýzu toho, jak respondenti danou odpovědní škálu využívají, identifikovat jejich response style. Další užitečné zobrazení je pomocí grafu informačních funkcí položky (*item information function*, „IIF“), ze kterých lze vyčíst, které položky jsou pro danou škálu nejpřínosnější, tzn. poskytují nejvíce informace, a naopak identifikovat položky, které by mohly být ze škály vyřazeny. Pro zhodnocení škály jako celku se používá informační funkce testu (*test information function*, „TIF“) doplněná chybou měření. Při testování navržených teoretických modelů se používají různé metody odhadu daných parametrů, jejichž volba často závisí na druhu proměnných, se kterými výzkumník pracuje (např. spojitě či kategoriálně). V naší studii bude výpočet IRT modelů založen na metodě MML (*maximum marginal likelihood estimation*) s algoritmem EM (*expectation-maximization*).

Testování přibližné ekvivalence měření (AMI)²

Pro smysluplné porovnání konceptů (v případě této studie životní spokojenosti) je vyžadováno, aby byl studovaný koncept měřen stejným způsobem napříč různými skupinami (např. v našem případě sociálními skupinami). Dosažení určité úrovně ekvivalence (srovnatelnosti) měření nástroje měřícího daný koncept znamená, že nástroj měří stejný koncept na stejné měřicí stupnici shodně ve všech testovaných skupinách. Tím lze zaručit, že pozorované rozdíly mezi skupinami jsou skutečným odrazem sociálních rozdílů, a nikoliv pouze chybou měření (Davidov, Cieciuch et al., 2018). Ekvivalenci lze testovat pomocí různých metod (Anýžová, 2013, 2014, 2015; Davidov, Schmidt et al., 2018), které lze rozdělit do dvou skupin – na tradiční přístupy testující *přesnou ekvivalenci měření* (EMI) a modernější přístupy zaměřující se na testování *přibližné ekvivalence měření* (AMI). V rámci této studie bude aplikován novější přístup k testování přibližné ekvivalence založený na bayesovském modelování.

K představení této techniky zopakujeme několik již zmíněných faktů. Tradiční přístup k testování přesné ekvivalence je založen na testování několika hierarchicky uspořádaných úrovní, z nichž každá je definována vzrůstající restrikcí kladenou na testovaný model (Davidov et al., 2014; Meredith, 1993; Steenkamp a Baumgartner, 1998; Vanderberg a Lance 2000). Nejnižší úroveň je *konfigurální ekvivalence*, která předpokládá, že faktorová struktura měřeného konceptu je stejná napříč různými skupinami. Jinými slovy, že daný koncept je měřen stejnými položkami totožně v různých skupinách. Druhá úroveň (*metrická ekvivalence*) vyžaduje shodu faktorových zátěží ve všech porovnávaných skupinách, aby zůstal zachován stejný význam latentní proměnné. Dosažení metrické ekvivalence je dostatečným předpokladem pro věcné analýzy srovnávající faktorové kovariance a nestandardizované regresní koeficienty mezi zkoumanými populacemi (Steenkamp a Baumgartner, 1998; Vanderberg a Lance, 2000). Třetí úroveň je *skalární ekvivalence*, která ke shodě faktorových zátěží přidává i požadavek na shodu položkových konstant ve všech skupinách. Tím se předpokládá, že respondenti se stejnou hodnotou latentní proměnné mají i stejnou očekávanou odpověď bez ohledu na příslušnost ke konkrétní skupině (Davidov et al., 2014). Tato úroveň je považována za nezbytnou pro porovnání latentních průměrů či průměrů hrubých skóre

² Obecný popis této metody vychází z již publikovaného článku Raudenská a Hanzlová (2021), její použití na česká data je zde originální.

jednotlivých položek mezi skupinami (Meredith, 1993). Nejčastější metodou pro testování EMI je víceskupinová konfirmační faktorová analýza (MG-CFA) (Kim et al., 2017). Pro stanovení rozhodnutí, jaké úrovně EMI bylo dosaženo, se používají chí-kvadrát test a indexy vhodnosti modelu, např. CFI (Comparative Fit Index) nebo RMSEA (Root Mean Squared Error of Approximation). Dosažení příslušné úrovně ekvivalence je posuzováno porovnáním indexů vhodnosti mezi méně a více restriktivním modelem (ke stanovení jejich ideálních hodnot doporučujeme články West et al., 2012; Rutkowski a Svetina, 2014, 2017).

Naproti tomu testování AMI je založeno na uvolnění požadavku na shodu faktorových zátěží a konstant mezi skupinami a jeho nahrazení tím, aby všechny parametry byly „přibližně stejné“ (Muthén a Asparouhov, 2012; Van de Schoot et al., 2013). Předpokládá se, že rozdíly mezi parametry nebudou nulové jako v případě EMI, ale téměř nulové (Muthén a Asparouhov, 2012). Rozdíly mezi parametry jsou stále udržovány na minimální hodnotě, aby bylo zajištěno, že zkoumané koncepty zůstanou přibližně srovnatelné. Snížení požadavku naprosté rovnosti, který může být v mnoha případech vnímán jako nepřiměřený, neovlivňuje výzkumný záměr testovat a porovnávat sociální jevy mezi různými skupinami či kulturami (Pokropek et al., 2020). Testování AMI bylo aplikováno v mnoha srovnávacích studiích, přičemž zejména v posledních letech zájem o tuto metodu neustále roste (Bujacz et al., 2014; Cieciuch et al., 2018; Davidov, Cieciuch et al., 2018; Seddig a Leitgöb, 2018; Winter a Depaoli, 2020; Zercher et al., 2015). Pro testování AMI lze využít bayesovského modelování (MG-BSEM) nebo metodu alignment (více viz např. Kim et al., 2017), přičemž v této studii budeme aplikovat bayesovský přístup, který umožňuje využít apriorní informace distribucí proměnných vyúsťující jednak v lepší odhad velikosti parametrů a latentních proměnných, tak i v lepší odhad statistik celkového modelu. Stejně tak je vhodnou metodou generující reliabilnější výsledky v případě menších vzorků (více viz Lee, 2007).

Bayesovský přístup k testování přibližné ekvivalence obvykle požaduje omezení průměrného rozdílu mezi parametry (tzv. apriorní průměr / *prior mean*) na nulu a odchylky parametrů (tzv. apriorní rozptyl / *prior variance*) na hodnotu nenulovou, byť stále dostatečně malou. Rozhodnutí o velikosti apriorního rozptylu závisí na samotném výzkumníkovi, přičemž od této velikosti se odráží úroveň aproximace. S rostoucí hodnotou apriorního rozptylu se předpokládá, že se parametry v modelu budou více lišit a že tento model bude více odlišný od modelu EMI (Cieciuch et al., 2018). Jelikož nevhodně zvolená hodnota apriorního rozptylu může význačně ovlivnit odhad latentních průměrů a výpočet směrodatných odchylek jednotlivých parametrů, je doporučováno testovat sérii modelů s různou velikostí apriorních rozptylů (např. 0,001; 0,005; 0,01; 0,025 a 0,05), aby bylo možné identifikovat nejvhodnější model nejvíce odpovídající datům (Pokropek et al., 2020). Dle autorů je vhodné začít testování AMI s apriorním rozptylem 0,000, který odpovídá skalární ekvivalenci v EMI a dle kterého jsou faktorové zátěže a konstanty shodné mezi skupinami. Tento model však většinou nedává žádoucí výsledky, proto se dále pokračuje s postupným zvyšováním apriorního rozptylu až do chvíle, kdy model v porovnání s předchozím vykazuje výrazné zlepšení, a indexy vhodnosti modelu dosahují požadovaných hodnot pro dobrý model. Co se týká hranice apriorního rozptylu, u které lze poměrně bezpečně bez rizika nezakreslených hodnot porovnávat latentní průměry mezi skupinami, tak ta není jasně stanovena. Např. Van de Schoot et al. (2013) ve své simulační studii tuto hranici navrhuje na 0,05, nicméně podle Pokropka et al. (2020) již tato hodnota povoluje příliš velkou meziskupinovou variabilitu mezi parametry.

Výběr nejvíce vyhovujícího modelu je v případě bayesovského testování AMI založen na několika kritériích, přičemž mezi nejpoužívanější patří informační kritérium o odchylce (*Deviance Information Criterion*, dále jen „DIC“) a Bayesovo informační kritérium (BIC). Žádoucí jsou co nejnižší hodnoty DIC a BIC, přičemž v případě komplexnějších modelů s více proměnnými je DIC preferováno před BIC. Pro zhodnocení, zda model správně odpovídá datům, se používají posteriorní prediktivní pravděpodobnost (*Posterior Predictive Probability value*, dále jen „PPP“) a 95% kredibilní interval (95% *Credibility Interval*, dále jen „CI“) (Kim et al., 2017; Muthén a Asparouhov 2012). Pro konstatování, že bayesovský model dobře odpovídá datům, nesmí být PPP statisticky významné (hodnoty 0,05 či vyšší) a CI musí obsahovat nulu.

Při testování různých velikostí apriorního rozptylu a hledání toho nejlépe odpovídajícího, respektive porovnávání modelu s menším a vyšším apriorním rozptylem, jsou stanoveny následující prahové hodnoty pro rozdíly mezi DIC a BIC. Pokud došlo ke snížení hodnoty BIC o 20 bodů či více, u hodnoty DIC o 14 bodů či více a hodnota PPP se zvýšila o 0,01 v případě testování modelu s vyšším apriorním rozptylem, doporučuje se tento model přijmout a hodnotu apriorního rozptylu již nezvyšovat. Žádoucí je při hodnocení kombinovat výše uvedené indexy, přičemž za nejspolehlivější jsou považovány změny DIC a PPP (Pokropek et al., 2020).

Data a metody³

Data

Data pro analýzu byla sbírána v srpnu 2021 online metodou (CAWI) prostřednictvím panelu výzkumné agentury Data Collect, s. r. o. Výzkumný vzorek je reprezentativní pro online populaci ČR ve věku 18 až 69 let a byl získán kvótní metodou na základě pohlaví, věku, vzdělání, kraje a velikosti místa bydliště. Celkem bylo sebráno 1 048 platných dotazníků, přičemž 88 z nich bylo vyřazeno v průběhu fáze čištění a kontroly. Finální výzkumný vzorek čítá 960 respondentů, z nichž je 481 mužů (50,1 %, průměrný věk 43,3 let) a 479 žen (49,9 %, průměrný věk 44,0 let). Rozložení výzkumného vzorku z hlediska věkových skupin a vzdělanostních kategorií je uvedeno v Tabulce 1.

Tabulka 1. Rozložení výzkumného vzorku dle věku a vzdělání (N = 960)

	Četnosti	Procenta (%)
Věkové skupiny		
18–24	91	9,5
25–34	183	19,1
35–44	224	23,3
45–54	196	20,4
55–69	266	27,7
Vzdělanostní kategorie		
základní	58	6,0
středoškolské bez maturity/vyučení	331	34,5
středoškolské s maturitou	361	37,6
vyšší odborné, vysokoškolské	210	21,9

³ Dataset a skripty jsou k dispozici v online příloze (<https://doi.org/10.13060/csr.2023.047>).

Výzkumný nástroj (SWLS)

Pro analýzu byl použit český překlad SWLS⁴ s několika změnami a úpravami⁵ na základě pilotáže (viz Tabulka 2). Znění otázky v dotazníku bylo následující:

Níže je pět výroků, s nimiž můžete souhlasit či nesouhlasit. Na stupnici od 1 (rozhodně nesouhlasím) do 7 (rozhodně souhlasím) vyjádřete míru svého souhlasu či nesouhlasu s každým tvrzením. Odpovídejte, prosím, otevřeně a upřímně.

(1) rozhodně nesouhlasím – (2) nesouhlasím – (3) spíše nesouhlasím – (4) ani souhlasím, ani nesouhlasím – (5) spíše souhlasím – (6) souhlasím – (7) rozhodně souhlasím

1. Můj život je v mnoha ohledech blízko mému ideálu.
2. Podmínky mého života jsou vynikající.
3. Jsem spokojen(a) se svým životem.
4. Až dosud jsem dostal(a) od života téměř vše, co jsem chtěl(a).
5. Kdybych mohl(a) žít svůj život znovu, nezměnil(a) bych téměř nic.

⁴ Dostupný zde: <http://labs.psychology.illinois.edu/~ediener/SWLS.html>.

⁵ Hlavní změnou byl dle našeho názoru správný překlad prostředního bodu na odpovědní škále, který v originálním překladu není odpovídající, neboť odpověď „nemohu se rozhodnout“ lze interpretovat jako „nevím“ a tím dochází k porušení ordinality škály. Ve znění jednotlivých položek došlo pouze ke stylistickým opravám zejména ve slovosledu.

Tabulka 2. Znění položek SWLS a rozložení odpovědí (N = 960)

	1 Rozhodně nesouhlasím	2 Nesouhlasím	3 Spíše nesouhlasím	4 Ani souhlasím, ani nesouhlasím	5 Spíše souhlasím	6 Souhlasím	7 Rozhodně souhlasím
<i>(1) Můj život je v mnoha ohledech blízko mému ideálu.</i>							
Četnosti	65	93	153	283	255	84	27
Procenta	6,8	9,7	15,9	29,4	26,6	8,8	2,8
Kumulativní %	6,8	16,5	32,4	61,8	88,4	97,2	100
<i>(2) Podmínky mého života jsou vynikající.</i>							
Četnosti	59	81	134	261	305	97	23
Procenta	6,1	8,4	14	27,2	31,8	10,1	2,4
Kumulativní %	6,1	14,5	28,5	55,7	87,5	97,6	100
<i>(3) Jsem spokojen(a) se svým životem.</i>							
Četnosti	42	54	92	201	306	205	60
Procenta	4,4	5,6	9,6	20,9	31,8	21,4	6,3
Kumulativní %	4,4	10	19,6	40,5	72,3	93,7	100
<i>(4) Až dosud jsem dostal(a) od života téměř vše, co jsem chtěl(a).</i>							
Četnosti	75	83	146	233	259	137	27
Procenta	7,8	8,6	15,2	24,3	27	14,3	2,8
Kumulativní %	7,8	16,4	31,6	55,9	82,9	97,2	100
<i>(5) Kdybych mohl(a) žít svůj život znovu, nezměnil(a) bych téměř nic.</i>							
Četnosti	134	138	164	222	169	84	49
Procenta	14	14,4	17,1	23	17,6	8,8	5,1
Kumulativní %	14	28,4	45,5	68,5	86,1	94,9	100

Metody a analytická strategie

Příprava dat a základní analýzy byly provedeny ve statistickém softwaru SPSS 27, testování dimenzionality a testování ekvivalence v Mplus 7.2. IRT analýza byla realizována ve volně dostupném statistickém softwaru R za použití balíčku „mirt“⁶. Na data byl aplikován model odstupňovaných odpovědí (GRM) vhodný pro ordinální proměnné (Samejima, 1969).

Analýza testování dimenzionality a ekvivalence probíhala v několika krocích (viz Tabulka 3). V prvním kroku byly za pomoci konfirmační faktorové analýzy (CFA) s metodou odhadu MLR (*maximum likelihood estimation with robust standard errors*) a teorie odpovědí na položku (IRT) s metodou MML (*maximum marginal likelihood estimation*) s algoritmem EM (*expectation-maximization*)⁷ testovány tři varianty modelu SWLS (pro jejich grafické zobrazení viz Graf 1), které vychází z dosud realizovaných studií (např. Clench-Aas et al., 2011; Jovanović, 2017).

Za prvé jednofaktorový model (M1), který předpokládá, že všechny měřicí položky adekvátně a shodně přispívají k vysvětlení latentního faktoru „spokojenost se životem“. Za druhé, dvoufaktorový model (M2), ve kterém první tři položky (*swl01*, *swl02*, *swl03*) tvoří faktor „přítomnost/či aktuální spokojenost se životem“ a poslední dvě položky (*swl04*, *swl05*) „minulost/či celková spokojenost se životem, kde hodnotím i prožitou minulost“. A za třetí modifikovaný jednofaktorový model s korelovanými rezidui mezi *swl04* a *swl05* (M3), který opět předpokládá, že všechny měřicí položky shodně přispívají k vysvětlení latentního faktoru „spokojenost se životem“, ale navíc rovněž připouští, že poslední dvě položky jsou si velmi podobné, souvisí spolu a měří podobný konstrukt. Tato předpokládaná shoda mezi položkami se do modelu zakresluje jako korelace jejich chyb měření.

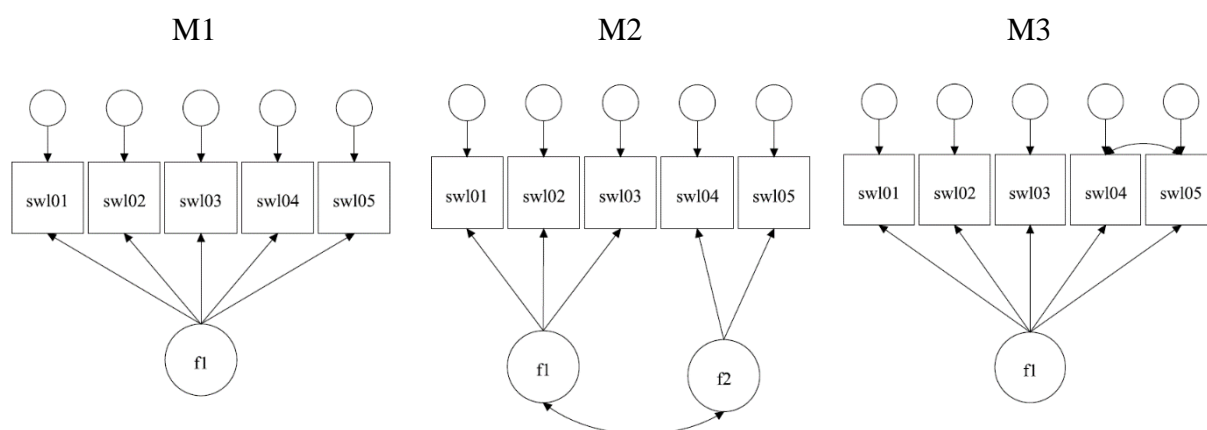
Na základě výsledků byl vybrán nejvhodnější model, který byl ve druhém kroku použit pro testování přibližné ekvivalence (AMI) za pomoci bayesovského modelování (BSEM) z hlediska pohlaví, věku a vzdělání⁸. Vzhledem k použití IRT pro testování psychometrických vlastností škály SWLS by se dále nabízelo přistoupit k testování DIF (*Differential Item Functioning*, tzv. odlišné fungování položky), a nikoliv k testování ekvivalence, nicméně k tomuto kroku jsme nepřistoupily z důvodu, že DIF je primárně určeno pro jednodimenzionální modely a porovnání mezi dvěma skupinami. Námi vybraný model M3 není plně jednodimenzionální, jelikož obsahuje korelaci chyb mezi dvěma položkami, a porovnáváme ho mezi čtyřmi až pěti skupinami dle proměnných věk a vzdělání. V případě porovnání více skupin najednou by bylo nutné aplikovat u IRT složitější techniky, což by bylo nad rámec tohoto článku.

⁶ Program Mplus rovněž nabízí možnost provést IRT analýzu, nicméně nenabízí tolik výpočetních a grafických možností jako balíčky v R, proto jsme se rozhodly pro tuto kombinaci softwarů.

⁷ Použití a komparace obou těchto metod v metodologických pracích zabývajících se psychometrií a testováním ekvivalence je dnes již běžnou praxí (viz např. Bean a Bowen, 2021; Buchholz a Hartig, 2020; Caycho-Rodríguez et al., 2020; Kim a Yoon, 2011; Yuan et al., 2021 atp.).

⁸ V úmyslu bylo otestovat ekvivalenci i z hlediska kraje, nicméně vzhledem k nízkému počtu případů v některých krajích, respektive malému N, nebyla tato analýza možná.

Graf 1. Tři testované modely SWLS



Zdroj: vlastní zpracování v Mplus.

Tabulka 3. Analytická strategie

Modely:		
<i>M1</i> : jednofaktorový model		
<i>M2</i> : dvoufaktorový model (přítomnost – <i>swl01</i> , <i>swl02</i> , <i>swl03</i> , minulost – <i>swl04</i> , <i>swl05</i>)		
<i>M3</i> : jednofaktorový model s korelovanými rezidui mezi <i>swl04</i> a <i>swl05</i>		
Krok	Typ analýzy	Model
1	Testování dimenzionality (CFA, IRT)	M1, M2, M3
2	Testování přibližné ekvivalence AMI (MG-BSEM), bayesovský model s různými apriorními rozptyly z hlediska pohlaví, věku, vzdělání	M3

Výsledky/Analýza

Deskriptivní statistiky

V Tabulce 4 jsou zobrazeny deskriptivní statistiky a výsledky testování normality pro celou škálu i pro jednotlivé položky. Průměr odpovědí na položky *swl01* až *swl05* na sedmibodové škále byl 3,97, 4,10, 4,59, 4,08 a 3,63, průměr celé škály pak 20,37. Detailnější rozložení odpovědí dle pohlaví, věku a vzdělání znázorňuje Tabulka 5.

Tabulka 4. Deskriptivní statistiky a testování normality (N = 960)

	Průměr	Směrodatná odchylka	Šikmost	Špičatost
<i>swl01</i>	3,97	1,42	-0,30	-0,31
<i>swl02</i>	4,10	1,40	-0,47	-0,19
<i>swl03</i>	4,59	1,45	-0,65	0,06
<i>swl04</i>	4,08	1,51	-0,39	-0,51
<i>swl05</i>	3,63	1,69	0,08	-0,82
SWLS	20,37*	6,18	-0,37	-0,11

Poznámka: *Průměr v rozmezí 5–35.

Tabulka 5. Deskriptivní statistiky podle pohlaví, věku a vzdělání – průměrné hodnoty

	Pohlaví		Věk					Vzdělání			
	Muži (481)	Ženy (479)	18–24 (91)	25–34 (183)	35–44 (224)	45–54 (196)	55–69 (266)	ZŠ (58)	SŠ bez mat. / vyučení (331)	SŠ s mat. (361)	VOŠ, VŠ (210)
<i>swl01</i>	4,04	3,90	3,81	4,10	3,99	4,13	3,79	4,16	3,75	3,98	4,25
<i>swl02</i>	4,14	4,05	4,22	4,32	4,21	4,11	3,80	3,91	3,82	4,18	4,45
<i>swl03</i>	4,65	4,54	4,48	4,67	4,70	4,66	4,44	4,47	4,40	4,61	4,90
<i>swl04</i>	4,15	4,01	3,96	4,34	4,16	4,14	3,83	3,93	3,82	4,12	4,46
<i>swl05</i>	3,66	3,60	3,47	3,96	3,67	3,59	3,45	3,33	3,39	3,68	4,00
SWLS*	20,64	20,09	19,95	21,39	20,72	20,64	19,31	19,79	19,19	20,56	20,07

Poznámky: *Průměr v rozmezí 5–35. V závorce uvedené N.

Testování dimenzionality pomocí CFA

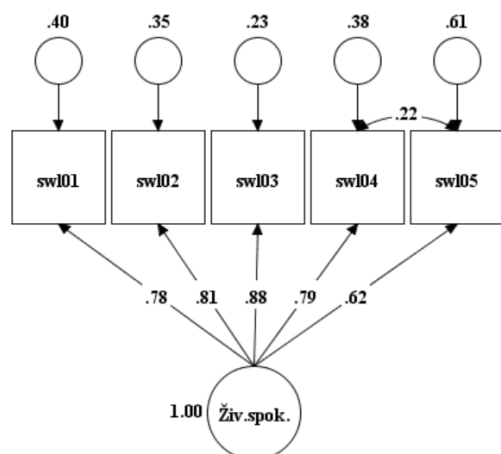
Prostřednictvím CFA s metodou odhadu MLR byly testovány tři modely (M1, M2, M3) faktorové struktury SWLS (viz Tabulka 3). Výsledky zobrazené v Tabulce 6 ukazují, že M2 a M3 vykazují lepší shodu s daty než základní jednofaktorový M1 (CFI = 0,995 vs. 0,980, RMSEA = 0,045 vs. 0,080). V případě M2 však dosahuje hodnota korelačního koeficientu mezi faktory téměř úplné shody ($r = 0,91$), což naznačuje, že faktory „přítomnost“ a „minulost“ nelze snadno rozlišit a tento model tak není úplně vhodný. Navíc oba tyto faktory jsou tvořeny příliš malým počtem položek (každý faktor by měl sestávat minimálně ze tří položek). Model M3 s korelovanými rezidui mezi *swl04* a *swl05* dosahuje shodných hodnot jako dvoufaktorový M2, jelikož jsou oba modely matematicky ekvivalentní (MacCallum et al., 1993), ale nezahrnuje kontroverzně vysokou korelaci mezi dvěma faktory a navíc věcně lépe odráží aspekt času související se zněním položek (tj. přítomnost a minulost) a vyznačuje se vyšší parsimonií (Clench-Aas et al., 2011). Rovněž je při měření psychologických konstruktů typickým znakem přítomnost sekundárních minoritních latentních faktorů, tzn. že měřený konstrukt pro důvěryhodnou interpretaci celkového skóre nemusí splňovat požadavek striktní jednodimenzionality, ale postačuje, aby byl primárně jednodimenzionální (Slocum-Gori et al., 2009). Tuto podmínku model M3 také splňuje. Pro následující metodologické testování byl z výše uvedených důvodů jako nejvhodnější vybrán M3, tedy modifikovaný jednofaktorový model s korelovanými rezidui mezi *swl04* a *swl05*. Ke stejnému závěru, tedy zvolení M3 jako nejvhodnějšího modelu, dospěli i další výzkumníci ve svých studiích (Bai et al., 2011; Clench-Aas et al., 2011; Glaesmer et al., 2011; Jovanović, 2017).

Tabulka 6. Testování modelů (CFA)

Model	Chi ² (df)	CFI	TLI	RMSEA	90% C.I. RMSEA	SRMR
M1	35,541 (5)	0,980	0,961	0,080	[0,056; 0,105]	0,022
M2	11,731 (4)	0,995	0,988	0,045	[0,016; 0,076]	0,011
M3	11,731 (4)	0,995	0,988	0,045	[0,016; 0,076]	0,011

Poznámky: Chi² = Chí-kvadrát, df = stupně volnosti, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, 90 % C.I. RMSEA = 90% interval spolehlivost pro RMSEA, SRMR = Standardized Root Mean Squared Residual.

Graf 2. Fit nejvhodnějšího modelu (M3), výsledky CFA, standardizované odhady



Zdroj: vlastní zpracování v Mplus.

IRT analýza

Psychometrické vlastnosti SWLS byly dále zkoumány pomocí IRT analýzy, která mimo standardní statistiky z CTT poskytuje další údaje o hodnotách parametrů a rovněž i velmi názorné vizuální zobrazení. Tato analýza je založena na testování SWLS jako celku, tedy jednofaktorové struktury⁹. V Tabulce 7 jsou zobrazeny vypočtené hodnoty diskriminačních (a) a threshold (b_i) parametrů z GRM pro jednotlivé položky a v Grafu 3 jejich informační funkce. Informační funkci celé škály včetně standardní chyby znázorňuje Graf 4.

Tabulka 7. Odhady položkových parametrů (GRM IRT)

	Diskriminační parametr	Threshold parametry					
	a	$b1$	$b2$	$b3$	$b4$	$b5$	$b6$
<i>swl01</i>	2,52	-1,85	-1,17	-0,53	0,36	1,39	2,31
<i>swl02</i>	2,82	-1,84	-1,20	-0,62	0,19	1,32	2,34
<i>swl03</i>	3,71	-1,96	-1,42	-0,92	-0,23	0,67	1,65
<i>swl04</i>	2,75	-1,70	-1,14	-0,54	0,19	1,09	2,25
<i>swl05</i>	1,74	-1,52	-0,78	-0,17	0,64	1,49	2,32

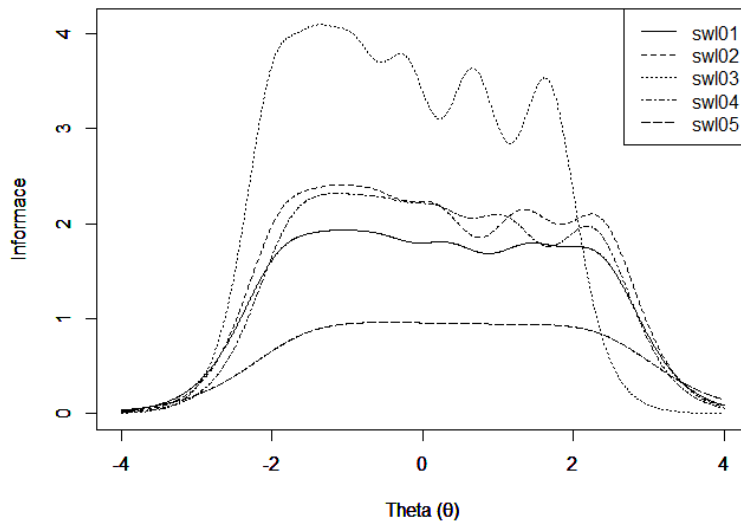
Dle výsledků v Tabulce 7 a Grafech 3 a 4 můžeme zhodnotit, že SWLS vykazuje velmi dobré fungování, a to zejména mezi úrovní -2 a 2,5 kontinua latentního rysu θ (životní spokojenosti). Hodnoty diskriminačního parametru dosahují pro všechny položky vysokých, a tedy žádoucích hodnot, dle Bakera (2001) se dají označit za „velmi vysoké“. Zároveň je evidentní, že hodnota tohoto parametru je nižší u položky *swl05*, což koresponduje s výsledky CFA a její nižší faktorovou zátěží a ukazuje na jistou problematičnost této položky. Co se týká hodnot threshold parametrů, tak ty se pohybují v rozmezí od -1,96 (*swl03*) po 2,34 (*swl02*). Například pro *swl02*, osoba s hodnotou latentního rysu -1,84 má 50% pravděpodobnost, že odpoví kategorií 2 (nesouhlasím) nebo vyšší; osoba s hodnotou -1,20 má 50% šanci, že odpoví kategoriemi 3 až 7 spíše než kategoriemi 1 nebo 2; až osoba s hodnotou 2,34 má 50% šanci, že zvolí kategorii 7 (rozhodně souhlasím) spíše než kategorie 1 až 6.

Z hodnot threshold parametrů lze dále vyčíst, že pravděpodobnost, že osoba s určitou úrovní latentního rysu vybere danou kategorii na odpovědní škále, se liší pro jednotlivé položky. To konkrétně znamená, že například na položku *swl03* kategorii 7 (rozhodně souhlasím) pravděpodobně zvolí osoba s úrovní latentního rysu 1,65, zatímco u položky *swl02* pro tuto odpověď musí respondent dosahovat vyšší úrovně latentního rysu (2,34). Rozdílnost ve vzdálenostech mezi jednotlivými kategoriemi se projevuje i uvnitř položek. Tím je myšleno, že např. pro položku *swl02* jsou rozdíly mezi thresholdy $b1/b2$ -0,64, mezi $b2/b3$ -0,59, mezi $b3/b4$ -0,80, mezi $b4/b5$ -1,13 a mezi $b5/b6$ pak -1,02, což signalizuje, že respondenti nevnímají kategorie mezi jednotlivými body na odpovědní škále stejně. Celková hodnota životní spokojenosti (latentního rysu) se tedy bude lišit od celkového skóre vypočteného dle CTT, ve které jsou rozdíly mezi kategoriemi považovány za stejné a jasně stanovené.

⁹ Pomocí IRT není možné graficky zobrazit modifikaci modelu (korelaci chyb), proto bylo testování založeno na jednofaktorovém modelu (M1), pro který jsou IRT modely primárně určeny.

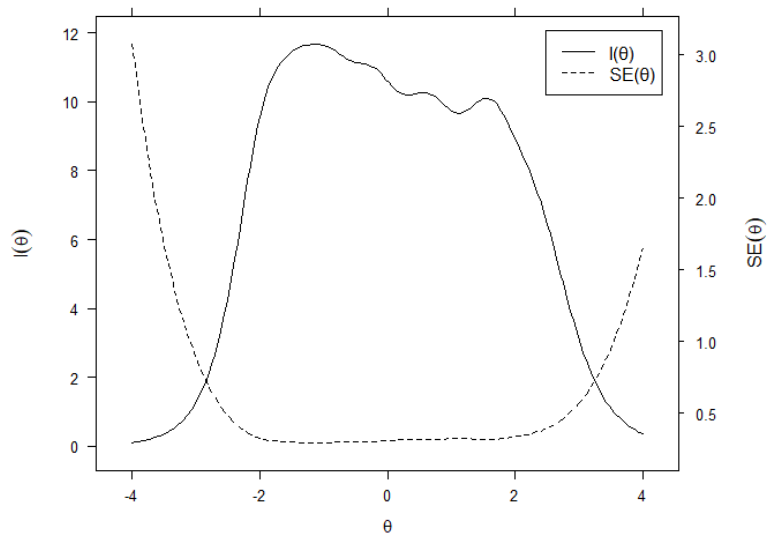
Z hlediska informačního přínosu jednotlivých položek do škály (viz Graf 3), vychází nejlépe položka *swl03*, která je v grafu položena nejvýše a zároveň vykazuje určitou rozlišitelnost patrnou z tvaru křivky (tj. není úplně plochá). Naopak *swl05*, jejíž křivka je velmi plochá a umístěna v grafu nejnižše, poskytuje nejméně informace a její přínos je tak do škály nejnižší. Tyto výsledky odpovídají hodnotám diskriminačního parametru (a), když pro *swl03* je jeho hodnota nejvyšší a pro *swl05* naopak nejnižší a rovněž korespondují s nejvyšší, respektive nejnižší hodnotou faktorové zátěže v CFA modelu.

Graf 3. Informační funkce položek SWLS (IIF)

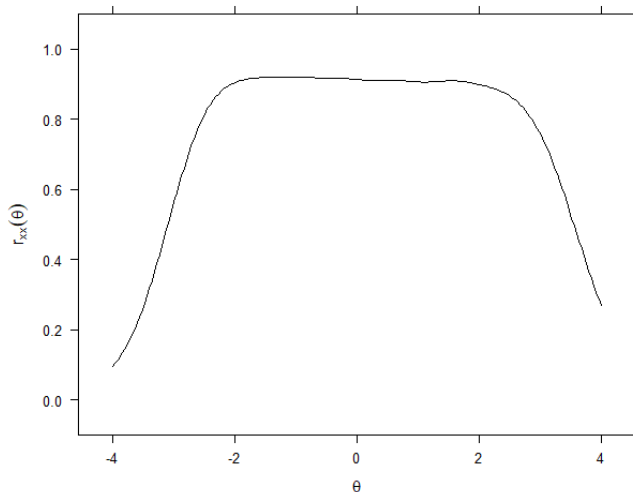


Z Grafů 4 a 5 můžeme vyčíst, že obecně SWLS vykazuje dobrou reliabilitu a malou chybu měření. Škála poskytuje velmi spolehlivé informace a dobře rozlišuje mezi respondenty ve velmi širokém rozmezí mezi -2 až $2,5$ úrovně latentního rysu θ , přičemž největší informační přínos má škála okolo -1 až $0,5$ úrovně latentního rysu θ , při které dosahuje zároveň nejnižší chyby měření. Naopak pod úrovní -3 a nad úrovní 4 latentního rysu θ standardní chyba výrazně roste a informace poskytované škálou tak již nejsou příliš spolehlivé, nicméně v celkovém měřítku jde o velmi malý podíl případů.

Graf 4. Celková informační funkce škály (TIF) včetně chyby měření



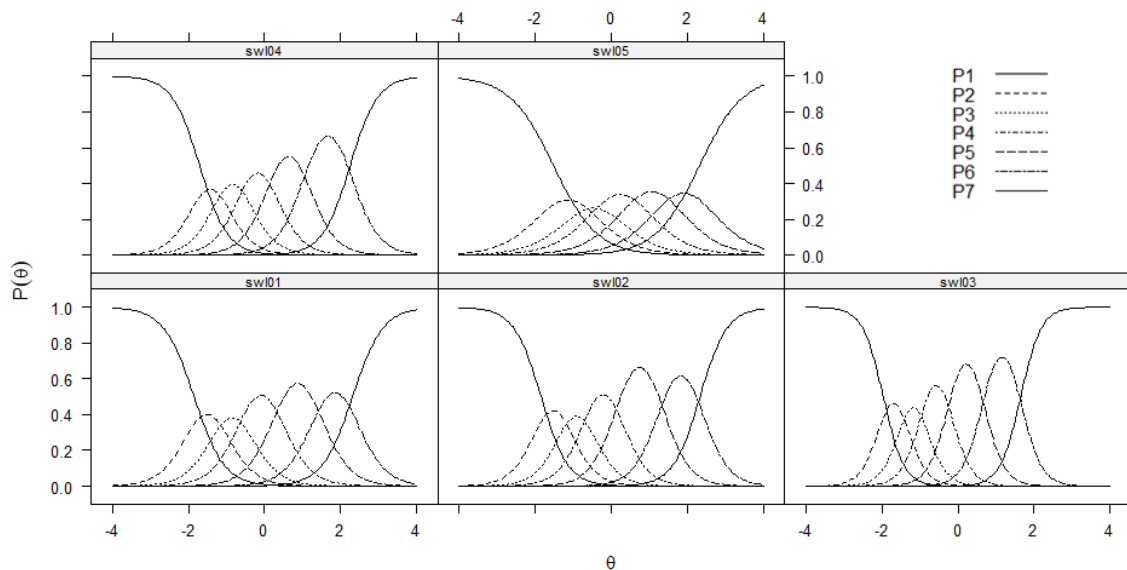
Graf 5. Reliabilita



Na základě IRT, konkrétně z grafů charakteristických křivek kategorií (CCC), můžeme rovněž zjistit, zda je dobře zvolena odpovědní škála, respektive zda je počet jejích kategorií odpovídající. Tyto křivky ukazují, jak dobře či špatně každá kategorie odpovědní škály funguje při přechodu z jedné kategorie na druhou. Na základě toho můžeme odhalit, že některé kategorie respondenti příliš nevyužívají nebo že jsou nadbytečné. Z výsledků zobrazených v Grafu 6 lze usuzovat, že odpovědní škála má nadbytečný počet kategorií. Ačkoliv krajní kategorie 7 (rozhodně souhlasím), v menší míře i 1 (rozhodně nesouhlasím), byly respondenty zřídka použity (viz také Tabulka 2), zároveň jsou tyto kategorie pro respondenty s vysokou, resp. nízkou úrovní latentního rysu θ nejpravděpodobnější (což je žádoucí). Naopak mezi kategoriemi 2 a 3, které jsou téměř překryty ostatními kategoriemi (nejviditelnější je tato skutečnost pro *swl05*), nejsou respondenti schopni dostatečně rozlišit. Podobné je to i v případě

položky *swl05* u kategorie odpovědi 6. Na základě těchto výsledků se domníváme, že by bylo možné snížit počet kategorií na pět či šest¹⁰.

Graf 6. Charakteristické křivky kategorií (CCC) jednotlivých položek SWLS



Pomocí IRT, respektive na základě výpočtu GRM, byly rovněž otestovány tři modely faktorové struktury SWLS s cílem porovnat výsledky s CFA a zhodnocením, zda byla volba M3 jako nejvhodnějšího modelu a jeho použití pro další analýzy oprávněná. Z výsledků, které zobrazuje Tabulka 8, můžeme vidět, že nejlepší hodnoty dosahuje M3 (CFI = 0,996, RMSEA = 0,061), tedy stejně jako v případě testování pomocí CFA. Volba tohoto modelu a jeho použití pro testování ekvivalence byla tedy dvojnásobně potvrzena pomocí různých metod.

Tabulka 8. Testování modelů (IRT)

Model	M2	CFI	TLI	RMSEA	90% C.I. RMSEA	SRMSR
M1	51,198	0,987	0,974	0,098	[0,074; 0,123]	0,041
M2	18,656	0,996	0,990	0,062	[0,035; 0,091]	0,033
M3	18,419	0,996	0,990	0,061	[0,034; 0,091]	0,033

Poznámky: M2 = hodnota statistiky, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, 90 % C.I. RMSEA = 90% interval spolehlivosti pro RMSEA, SRMSR = Standardized Root Mean Squared Residual.

Testování ekvivalence

Po otestování dimenzionality a výběru nejvhodnějšího modelu (M3) jsme v dalších krocích analýzy přistoupily k testování přibližné ekvivalence měření (AMI) prostřednictvím bayesovského modelování z hlediska různých skupin, konkrétně pohlaví (dvě kategorie), věku (pět kategorií) a vzdělání (čtyři kategorie).

¹⁰ Vzhledem k tomu, že neexistují jasně daná formální kritéria, na základě kterých lze s jednoznačností rozhodnout o vhodném snížení či zvýšení počtu kategorií odpovědní škály a jde tedy o explorativní techniku, rozhodnutí záleží na úsudku samotného výzkumníka.

Na základě Tabulky 9, která zobrazuje indexy vhodnosti modelu pro testování AMI při použití různě velkých hodnot apriorního rozptylu, lze obecně konstatovat, že výsledky vychází velmi dobře, respektive ukazují na velmi dobrou shodu modelu s daty ($PPP > 0,05$ a CI obsahuje nulu). Ve všech případech jsme začaly s testováním apriorního rozptylu 0,000, který znamená plnou skalární ekvivalenci, a následně pokračovaly se zkoušením více liberálních hodnot apriorního rozptylu (0,001; 0,002), až bylo dosaženo žádoucích hodnot indexů a rozdílů mezi nimi pro přijetí daného modelu a dosažení přibližné skalární ekvivalence měření. V případě pohlaví byla prokázána skalární ekvivalence, když model hned při testování apriorního rozptylu 0,000 splňoval nezbytné parametry. Z hlediska věku byl jako nejvhodnější vyhodnocen model při apriorním rozptylu 0,002 a z hlediska vzdělání byla prokázána přibližná ekvivalence měření při nejméně liberálním rozptylu 0,001. Lze tedy konstatovat velmi dobré dosažené výsledky ekvivalence, plné skalární v případě pohlaví a přibližné v případě věku a vzdělání s velmi malými apriorními rozptyly.

Tabulka 9. Testování přibližné ekvivalence měření (AMI) z hlediska pohlaví, věku a vzdělání – MG-BSEM; obecné indexy vhodnosti modelu a výběr apriorního rozptylu

	Apriorní rozptyl	DIC	BIC	PPP	95% CI
<i>Pohlaví (2 kategorie)</i>					
	0,000	11 215,752	15 459,355	0,118	[-8,433; 34,920]
<i>Věk (5 kategorií)</i>					
	0,000	12 908,792	16 895,285	0,001	[28,811; 94,145]
	0,001	9 569,397	15 594,412	0,054	[-5,514; 162,751]
	0,002	9 581,365	15 557,917	0,064	[-7,282; 156,488]
<i>Vzdělání (4 kategorie)</i>					
	0,000	12 987,261	15 103,788	0,000	[31,109; 92,494]
	0,001	13 365,879	15 243,495	0,161	[-17,880; 47,527]

Poznámky: DIC = Deviance Information Criterion, PPP = Posterior Predictive p value, 95% CI = 95% kredibilní interval.

Na závěr analýzy byly na základě výsledků AMI srovnány latentní průměry v jednotlivých skupinách včetně porovnání s prostým aritmetickým průměrem (viz Tabulka 10). Jak můžeme z tabulky vidět, výsledky pořadí se téměř neliší a můžeme tak konstatovat, že v případě této škály lze pro srovnání průměrné hodnoty životní spokojenosti použít prostý aritmetický průměr SWLS. Konkrétně výsledky ukázaly, že muži dosahují vyšší životní spokojenosti než ženy, stejně tak i lidé ve věku 25 až 34 let v porovnání s ostatními věkovými kategoriemi a z hlediska vzdělání pak lidé s vyšším odborným či vysokoškolským vzděláním.

Tabulka 10. Srovnání latentních průměrů SWLS z AMI a aritmetického průměru mezi jednotlivými skupinami

	Latentní průměr AMI (S.D.)	Aritmetický průměr (S.D.)
Pohlaví		
muži	1,572 (0,87)	20,64 (6,02)
ženy	0,000 (0,00)	20,09 (6,33)
Věkové skupiny		
18–24	0,000 (0,00)	19,95 (5,75)
25–34	0,038 (0,50)	21,40 (5,67)
35–44	0,006 (0,07)	20,72 (6,52)
45–54	0,003 (0,00)	20,64 (5,80)
55–69	0,000 (0,00)	19,31 (6,50)
Vzdělanostní kategorie		
základní	0,000 (0,00)	19,79 (6,77)
středoškolské bez maturity/vyučení	–0,060 (0,06)	19,19 (5,84)
středoškolské s maturitou	–0,012 (0,02)	20,56 (6,29)
vyšší odborné, vysokoškolské	0,000 (0,00)	22,07 (5,95)

Poznámky: Apriorní rozptyl v AMI pro pohlaví 0,001, pro věkové skupiny 0,002 a pro vzdělanostní kategorie 0,001.

Shrnutí a závěr

SWLS představuje nejčastěji používaný nástroj pro měření životní spokojenosti, který byl testován v mnoha zemích jak z hlediska jejich psychometrických vlastností, faktorové struktury, reliability, tak i srovnatelnosti mezi různými zeměmi či sociálními skupinami. Nicméně v ČR doposud chyběla studie, která by testovala její psychometrické vlastnosti a srovnatelnost mezi různými sociálními skupinami na reprezentativním vzorku české populace, což je nezbytný předpoklad pro zobecnitelnost daných výsledků.

Cílem této studie bylo proto na reprezentativním vzorku české online populace 18–69 let otestovat psychometrické vlastnosti české verze SWLS za pomoci CFA, IRT a její ekvivalenci z hlediska pohlaví, věku a vzdělání prostřednictvím bayesovského modelování. Záměrem bylo rovněž na konkrétním příkladu prakticky aplikovat v české sociologii doposud nepříliš používané metody IRT a BSEM, představit je dalším výzkumníkům a podnítit jejich zájem o použití těchto metod.

Psychometrické testování SWLS s využitím CFA a IRT ukázalo na její velmi dobré vlastnosti a fungování. V souladu s výsledky jiných studií (např. Garcia et al., 2021; Nima et al., 2020) se však ukázalo, že položka *swl05* v porovnání s ostatními čtyřmi vykazuje horší výsledky jak z hlediska diskriminační síly, resp. faktorových zátěží, tak z hlediska její obtížnosti. Naopak nejlepšího fungování z hlediska diskriminace i informačního přínosu do škály dosahuje *swl03*. Obecně však škála jako celek poskytuje velmi dobré informace se zvláště vysokou spolehlivostí a malou chybou měření mezi –2 až 2,5 úrovně latentního rysu, tedy životní spokojenosti.

Co se týká analýzy odpovědní škály, respektive vhodnosti počtu bodů na škále, tak dle výsledků lze usuzovat, že sedmibodová škála se zdá být nadbytečná, jelikož respondenti nejsou schopni dostatečně rozlišit mezi kategoriemi 2 a 3, v případě *swl05* se to týká i kategorie 6. Domníváme se, že by bylo možné a vhodné zkrátit odpovědní škálu na šesti- či pětibodovou.

Ostatně stejný závěr na základě výsledků svých studií formulovali i další autoři (např. Avcu, 2021). Pro adekvátní otestování a zhodnocení, zda je možné použít odpovědní škálu s menším počtem kategorií, je žádoucí provést nový metodologický výzkum. Nabízí se rovněž možnost využít současných dat a ex post sloučit kategorie a porovnat výsledky, nicméně tento postup nepovažujeme z metodologického hlediska za správný a vhodný.

Možné vysvětlení pro určitou problematičnost *swl05* lze nalézt ve znění této položky, jelikož významově je směřována do minulosti (Pavot a Diener, 1993). Nicméně to platí i pro *swl04*, proto je velmi často ve studiích testována i dvoufaktorová struktura této škály, ve které první tři položky tvoří faktor „přítomnost“ a poslední dvě „minulost“. V rámci naší studie jsme testovaly tři modely faktorové struktury, konkrétně základní jednofaktorový model (M1), dvoufaktorový model (M2) a modifikovaný jednofaktorový model s korelovanými rezidui mezi *swl04* a *swl05* (M3). Konfirmační faktorová analýza ukázala na velmi dobré výsledky všech třech modelů, nicméně statistiky vhodnosti byly lepší pro M2 a M3. Tyto výsledky potvrdila i IRT analýza založená na výpočtu GRM, nicméně vhodnost M1 již nepotvrdila – koeficient RMSEA byl vyšší než doporučená hranice. Vzhledem k počtu položek ve škále, problematické rozlišitelnosti mezi dvěma faktory v případě M2 a rovněž i na základě výsledků dalších studií (Bai et al., 2011; Clench-Aas et al., 2011; Glaesmer et al., 2011; Jovanović, 2017) byl jako nejvhodnější model, který byl dále použit pro testování ekvivalence mezi skupinami na základě pohlaví, věku a vzdělání, vybrán M3, tedy modifikovaný jednofaktorový model s korelovanými rezidui mezi *swl04* a *swl05*.

V případě testování přibližné ekvivalence za pomoci bayesovského modelování se potvrdila plná skalární ekvivalence v případě pohlaví, v případě věku a vzdělání pak jako nejvhodnější byly vyhodnoceny modely s velmi malými apriorními rozptyly, konkrétně 0,001 u vzdělání a 0,002 u věku. Celkově lze konstatovat velmi dobré dosažené výsledky testování ekvivalence, na základě kterých lze porovnávat latentní i aritmetické průměry životní spokojenosti mezi pohlavím, věkovými a vzdělanostními kategoriemi. Konkrétně dle výsledků lze říci, že vyšší životní spokojenosti dosahují muži v porovnání s ženami, stejně tak i lidé ve věku 25 až 34 let a lidé s vyšším odborným či vysokoškolským vzděláním.

Vzhledem k tomu, že ekvivalence měřicích nástrojů a škál není často testována a že pro porovnání ať už mezi různými skupinami, zeměmi či časovými body se využívá prostého aritmetického průměru, jsme na závěr naší analýzy provedly jeho srovnání s latentním průměrem z AMI. Ačkoliv srovnání neukázalo na výrazné rozdíly, i když výsledky prostého aritmetického průměru a latentních průměrů si odpovídají, ekvivalenci škálových položek je nezbytné testovat. Zároveň je považováno za vhodnější upřednostnit hodnoty latentních průměrů před prostým aritmetickým průměrem škály (např. Steinmetz, 2018).

Na základě výsledků našeho testování české verze SWLS, které potvrdilo vhodnost škály z hlediska její struktury a obecně psychometrických vlastností, tak rovněž i co se týká ekvivalence mezi skupinami na základě pohlaví, věku a vzdělání, se domníváme, že by bylo vhodné škálu využít i pro mezinárodní testování její kulturní srovnatelnosti. Těchto studií je obecně nedostatek, přičemž většina se zaměřuje na porovnání dvou či tří skupin a specifické cílové skupiny a jejich výsledky jsou různé (přehled viz Emerson et al., 2017). I přesto že v jednotlivých zemích funguje SWLS většinou velmi dobře a měří srovnatelně mezi různými sociálními skupinami, tak to automaticky neznamená, že bude srovnatelná mezi více zeměmi, jak dokládá Jang et al. (2017). Výsledky zmiňované studie jsou však založeny na malém

výzkumném vzorku a specifické cílové skupině (manažeri), což je typickým znakem většiny studií, proto se domníváme, že by bylo vhodné a přínosné provést mezinárodní srovnání SWLS mezi více zeměmi založené na reprezentativních datech. Například tedy zařadit SWLS do zavedených mezinárodních komparativních výzkumů, jako je Evropský sociální výzkum (ESS) či Mezinárodní program sociálního výzkumu (ISSP).

Doufáme, že se nám tímto článkem podařilo představit SWLS české odborné veřejnosti, zhodnotit její psychometrické vlastnosti a fungování jak z hlediska vnitřní struktury, tak i z hlediska srovnatelnosti mezi různými sociálními skupinami a rovněž nastínit možnosti pro další směřování a práci s touto škálou.

RADKA HANZLOVÁ vystudovala doktorské studium na Katedře sociologie Filozofické fakulty Univerzity Karlovy a v současné době pracuje jako postdoktorandka v Sociologickém ústavu AV ČR, v. v. i. Zabývá se metodologií kvantitativních výzkumů, komparativní analýzou, testováním a vývojem škál, především v oblasti životní pohody, štěstí a mentálního zdraví.
ORCID: 0000-0003-4245-5129

PETRA RAUDENSKÁ pracuje jako vědecká pracovnice v Sociologickém ústavu AV ČR, v. v. i. Jejími výzkumnými zájmy jsou metodologie mezinárodních výzkumů, vzdělanostní nerovnosti, životní spokojenost a problematika hodnotových orientací. Je autorkou monografie Srovnatelnost postojových škál v komparativním výzkumu (Univerzita Palackého v Olomouci, 2015) a editorkou několika dalších knih. Publikovala v Social Science Research, Poetics, International Sociology a Innovation: The European Journal of Social Science Research.
ORCID: 0000-0002-0330-139X

Literatura

- Andrews, F. M. a Withey, S. B. (1976). *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*. Springer US. <https://doi.org/10.1007/978-1-4684-2253-5>
- Anýžová, P. (2013). Ekvivalence položek v mezinárodních datech: základní vymezení a možnosti analýzy. *Data and Research – SDA Info*, 7(1), 29–56.
<https://doi.org/10.13060/1802-8152.2013.7.1.2>
- Anýžová, P. (2014). Srovnatelnost Schwartzovy hodnotové škály v mezinárodních datech. *Sociologický časopis / Czech Sociological Review*, 50(4), 547–580.
<http://dx.doi.org/10.13060/00380288.2014.50.4.108>
- Anýžová, P. (2015). *Srovnatelnost postojových škál v komparativním výzkumu*. Univerzita Palackého v Olomouci.
- Avcu, A. (2021). Item Response Theory-Based Psychometric Investigation of SWLS for University Students. *International Journal of Psychology and Education Studies*, 8(2), 27–37. <https://doi.org/10.52380/ijpes.2021.8.2.265>
- Bai, X., Wu, Ch., Zheng, R. a Ren, X. (2011). The Psychometric Evaluation of the Satisfaction with Life Scale Using a Nationally Representative Sample of China. *Journal of Happiness Studies*, 12(2), 183–197. <https://doi.org/10.1007/s10902-010-9186-x>

- Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
- Bean, G. J. a Bowen, N. K. (2021). Item Response Theory and Confirmatory Factor Analysis: Complementary Approaches for Scale Development. *Journal of Evidence-Based Social Work*, 18(6), 597–618. <https://doi.org/10.1080/26408066.2021.1906813>
- Buchholz, J. a Hartig, J. (2020). Measurement Invariance Testing in Questionnaires: A Comparison of Three Multigroup-CFA and IRT-based Approaches. *Psychological Test and Assessment Modeling*, 62(1), 29–53.
- Bujacz, A., Vittersø, J., Huta, V. a Kaczmarek, L. D. (2014). Measuring Hedonia and Eudaimonia as Motives for Activities: Cross-national Investigation through Traditional and Bayesian Structural Equation Modeling. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00984>
- Caycho-Rodríguez, T., Neto, J., Tomás, J. M., Valencia, P. D., Ventura-León, J., Neto, F. ... Vilca, L. W. (2020). Psychometric Properties of the Satisfaction with Job Life Scale in Portuguese Workers: A Systematic Study Based on the IRT and CFA Modeling. *Heliyon*, 6(6), e03881. <https://doi.org/10.1016/j.heliyon.2020.e03881>
- Cieciuch, J., Davidov, E., Algesheimer, R. a Schmidt, P. (2018). Testing for Approximate Measurement Invariance of Human Values in the European Social Survey. *Sociological Methods & Research*, 47(4), 665–686. <https://doi.org/10.1177/0049124117701478>
- Clench-Aas, J., Nes, R. B., Dalgard, O. S. a Aarø, L. E. (2011). Dimensionality and Measurement Invariance in the Satisfaction with Life Scale in Norway. *Quality of Life Research*, 20(8), 1307–1317. <https://doi.org/10.1007/s11136-011-9859-x>
- Davidov, E., Cieciuch, J. a Schmidt, P. (2018). The Cross-Country Measurement Comparability in the Immigration Module of the European Social Survey 2014-15. *Survey Research Methods*, 12(1), 15–27. <https://doi.org/10.18148/SRM/2018.V12I1.7212>
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P. a Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, 40(1), 55–75. <https://doi.org/10.1146/annurev-soc-071913-043137>
- Davidov, E., Schmidt, P., Billiet, J. a Meuleman, B. (eds.). (2018). *Cross-cultural Analysis: Methods and Applications* (2. vyd.). Routledge. <https://doi.org/10.4324/9781315537078>
- De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. Guilford Press.
- De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P. a Baumgartner, H. (2008). Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research*, 45(1), 104–115. <https://doi.org/10.1509/jmkr.45.1.104>
- DeMars, Ch. (2010). *Item Response Theory*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195377033.001.0001>
- Diener, E., Emmons, R. A., Larsen, R. J. a Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
- Dolan, P. a Metcalfe, R. (2012). Measuring Subjective Wellbeing: Recommendations on Measures for Use by National Governments. *Journal of Social Policy*, 41(2), 409–427. <https://doi.org/10.1017/S0047279411000833>

- Dolan, P. a White, M. P. (2007). How Can Measures of Subjective Well-Being Be Used to Inform Public Policy? *Perspectives on Psychological Science*, 2(1), 71–85.
<https://doi.org/10.1111/j.1745-6916.2007.00030.x>
- Emerson, S. D., Guhn, M. a Gadermann, A. M. (2017). Measurement Invariance of the Satisfaction with Life Scale: Reviewing Three Decades of Research. *Quality of Life Research*, 26(9), 2251–2264. <https://doi.org/10.1007/s11136-017-1552-2>
- Garcia, D., Nima, A. A., Kazemitabar, M., Amato, C., Lucchese, F., Mihailovic, M. a Kijima, N. (2021). Psychometric Properties of the Swedish Version of the Satisfaction with Life Scale in a Sample of Individuals with Mental Illness. *PeerJ*, 9.
<https://doi.org/10.7717/peerj.11432>
- Glaesmer, H., Grande, G., Braehler, E. a Roth, M. (2011). The German Version of the Satisfaction With Life Scale (SWLS): Psychometric Properties, Validity, and Population-Based Norms. *European Journal of Psychological Assessment*, 27(2), 127–132.
<https://doi.org/10.1027/1015-5759/a000058>
- Hultell, D. a Gustavsson, J. P. (2008). A Psychometric Evaluation of the Satisfaction with Life Scale in a Swedish Nationwide Sample of University Students. *Personality and Individual Differences*, 44(5), 1070–1079. <https://doi.org/10.1016/j.paid.2007.10.030>
- Huppert, F. A., Marks, N., Clark, A., Siegrist, J., Stutzer, A., Vittersø, J. a Wahrendorf, M. (2009). Measuring Well-Being Across Europe: Description of the ESS Well-Being Module and Preliminary Findings. *Social Indicators Research*, 91(3), 301–315.
<https://doi.org/10.1007/s11205-008-9346-0>
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504.
<https://doi.org/10.1080/10705510701301834>
- Jang, S., Kim, E. S., Cao, Ch., Allen, T. D., Cooper, C. L., Lapierre, L. M. ... Woo, J.-M. (2017). Measurement Invariance of the Satisfaction With Life Scale Across 26 Countries. *Journal of Cross-Cultural Psychology*, 48(4), 560–576.
<https://doi.org/10.1177/0022022117697844>
- Jöreskog, K. G. (1969). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, 34(2), 183–202. <https://doi.org/10.1007/BF02289343>
- Jovanović, V. (2017). Measurement Invariance of the Serbian Version of the Satisfaction With Life Scale Across Age, Gender, and Time. *European Journal of Psychological Assessment*, 35(4), 555–563. <https://doi.org/10.1027/1015-5759/a000410>
- Keyes, C. L. M., Shmotkin, D. a Ryff, C. D. (2002). Optimizing Well-Being: The Empirical Encounter of Two Traditions. *Journal of Personality and Social Psychology*, 82(6), 1007–1022. <https://doi.org/10.1037/0022-3514.82.6.1007>
- Kim, E. S., Cao, Ch., Wang, Y. a Nguyen, D. T. (2017). Measurement Invariance Testing with Many Groups: A Comparison of Five Approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524–544.
<https://doi.org/10.1080/10705511.2017.1304822>
- Kim, E. S. a Yoon, M. (2011). Testing Measurement Invariance: A Comparison of Multiple-group Categorical CFA and IRT. *Structural Equation Modeling*, 18(2), 212–228.
<https://doi.org/10.1080/10705511.2011.557337>

- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2. vyd.). The Guilford Press.
- Lee, S. Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. Wiley.
<https://doi.org/10.1002/9780470024737>
- Lewis, C. A., Shevlin, M. E., Smékal, V. a Dorahy, M. J. (1999). Factor Structure and Reliability of a Czech Translation of the Satisfaction with Life Scale among Czech University Students. *Studia Psychologica*, 41(3), 239–244.
- Lucas, R. E., Diener, E. a Suh, E. (1996). Discriminant Validity of Well-Being Measures. *Journal of Personality and Social Psychology*, 71(3), 616–628.
<https://doi.org/10.1037/0022-3514.71.3.616>
- MacCallum, R. C., Wegener, D. T., Uchino, B. N. a Fabrigar, L. R. (1993). The Problem of Equivalent Models in Applications of Covariance Structure Analysis. *Psychological Bulletin*, 114(1), 185–199. <https://doi.org/10.1037/0033-2909.114.1.185>
- Margolis, S., Schwitzgebel, E., Ozer, D. J. a Lyubomirsky, S. (2019). A New Measure of Life Satisfaction: The Riverside Life Satisfaction Scale. *Journal of Personality Assessment*, 101(6), 621–630. <https://doi.org/10.1080/00223891.2018.1464457>
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Muthén, B. a Asparouhov, T. (2012). Bayesian Structural Equation Modeling: A More Flexible Representation of Substantive Theory. *Psychological Methods*, 17(3), 313–235.
<https://doi.org/10.1037/a0026802>
- Navrátil, M. a Lewis, Ch. A. (2006). Temporal Stability of the Czech Translation of the Satisfaction with Life Scale: Test-Retest Data over One Week. *Psychological Reports*, 98(3), 918–920. <https://doi.org/10.2466/pr0.98.3.918-920>
- Nima, A. A., Cloninger, K. M., Persson, B. N., Sikström, S. a Garcia, D. (2020). Validation of Subjective Well-Being Measures Using Item Response Theory. *Frontiers in Psychology*, 10, 3036. <https://doi.org/10.3389/fpsyg.2019.03036>
- O'Connor, B. P., Crawford, M. R. a Holder, M. D. (2015). An Item Response Theory Analysis of the Subjective Happiness Scale. *Social Indicators Research*, 124(1), 249–258. <https://doi.org/10.1007/s11205-014-0773-9>
- Oishi, S. (2006). The Concept of Life Satisfaction across Cultures: An IRT Analysis. *Journal of Research in Personality*, 40(4), 411–423. <https://doi.org/10.1016/j.jrp.2005.02.002>
- Pavot, W. a Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5(2), 164–172. <https://doi.org/10.1037/1040-3590.5.2.164>
- Pavot, W. a Diener, E. (2008). The Satisfaction With Life Scale and the Emerging Construct of Life Satisfaction. *The Journal of Positive Psychology*, 3(2), 137–152.
<https://doi.org/10.1080/17439760701756946>
- Pavot, W., Diener, E., Colvin, C. R. a Sandvik, E. (1991). Further Validation of the Satisfaction With Life Scale: Evidence for the Cross-Method Convergence of Well-Being Measures. *Journal of Personality Assessment*, 57(1), 149–161.
https://doi.org/10.1207/s15327752jpa5701_17
- Pavot, W., Diener, E. a Suh, E. (1998). The Temporal Satisfaction With Life Scale. *Journal of Personality Assessment*, 70(2), 340–354. https://doi.org/10.1207/s15327752jpa7002_11

- Pokropek, A., Schmidt, P. a Davidov, E. (2020). Choosing Priors in Bayesian Measurement Invariance Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(5), 1–15. <https://doi.org/10.1080/10705511.2019.1703708>
- Raudenská, P. (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. <https://doi.org/10.1016/j.ssresearch.2019.102369>
- Raudenská, P. a Hanzlová, R. (2021). Nový přístup v testování ekvivalence na příkladu měření subjektivního blahobytu. *Sociální studia / Social Studies*, 18(1), 53–72. <https://doi.org/10.5817/SOC2021-1-53>
- Rutkowski, L. a Svetina, D. (2014). Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>
- Rutkowski, L. a Svetina, D. (2017). Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance. *Applied Measurement in Education*, 30(1), 39–51. <https://doi.org/10.1080/08957347.2016.1243540>
- Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika*, 34(4), 1–97. <https://doi.org/10.1007/BF03372160>
- Seddig, D. a Leitgöb, H. (2018). Approximate Measurement Invariance and Longitudinal Confirmatory Factor Analysis: Concept and Application with Panel Data. *Survey Research Methods*, 12(1), 29–41. <https://doi.org/10.18148/srm/2018.v12i1.7210>
- Schwarz, N. (1987). *Stimmung als Information Untersuchungen zum Einfluß von Stimmungen auf die Bewertung des eigenen Lebens*. Springer Verlag.
- Schwarz, N. a Clore, G. L. (1983). Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States. *Journal of Personality and Social Psychology*, 45(3), 513–523. <https://doi.org/10.1037/0022-3514.45.3.513>
- Schwarz, N. a Strack, F. (1991). Evaluating One's Life: A Judgment Model of Subjective Wellbeing. In F. Strack, M. Argyle a N. Schwarz (eds.), *Subjective Well-Being: An Interdisciplinary Perspective* (s. 27–48). Pergamon Press.
- Slocum-Gori, S. L., Zumbo, B. D., Michalos, A. C. a Diener, E. (2009). A Note on the Dimensionality of Quality of Life Scales: An Illustration with the Satisfaction with Life Scale (SWLS). *Social Indicators Research*, 92(3), 489–496. <https://doi.org/10.1007/s11205-008-9303-y>
- Steenkamp, J.-B. E. M. a Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, 25(1), 78–107. <https://doi.org/10.1086/209528>
- Steinmetz, H. (2018). Estimation and Comparison of Latent Means Across Cultures. In E. Davidov, P. Schmidt, J. Billiet, B. Meuleman (eds.), *Cross-Cultural Analysis: Methods and Applications* (s. 95–126). Routledge. <https://doi.org/10.4324/9781315537078-4>
- Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J. a Muthén, B. (2013). Facing off with Scylla and Charybdis: A Comparison of Scalar, Partial, and the Novel Possibility of Approximate Measurement Invariance. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00770>
- Vandenberg, R. J. a Lance, Ch. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational

- Research. *Organizational Research Methods*, 3(1), 4–70.
<https://doi.org/10.1177/109442810031002>
- Vassar, M. (2007). A Note on the Score Reliability for the Satisfaction With Life Scale: An RG Study. *Social Indicators Research*, 86(1), 47–57. <https://doi.org/10.1007/s11205-007-9113-7>
- Vittersø, J., Biswas-Diener, R. a Diener, E. (2005). The Divergent Meanings of Life Satisfaction: Item Response Modeling of the Satisfaction with Life Scale in Greenland and Norway. *Social Indicators Research*, 74(2), 327–348.
<https://doi.org/10.1007/s11205-004-4644-7>
- West, S. G., Taylor, A. B. a Wu, W. (2012). Model Fit and Model Selection in Structural Equation Modeling. In R. H. Hoyle (ed.). *Handbook of Structural Equation Modeling* (s. 209–231). The Guilford Press.
- Winter, S. D. a Depaoli, S. (2020). An Illustration of Bayesian Approximate Measurement Invariance with Longitudinal Data and a Small Sample Size. *International Journal of Behavioral Development*, 44(4), 371–382. <https://doi.org/10.1177/0165025419880610>
- Yuan, T., Honglei, Z., Xiao, X., Ge, W. a Xianting, C. (2021). Measuring Perceived Risk in Sharing Economy: A Classical Test Theory and Item Response Theory Approach. *International Journal of Hospitality Management*, 96, 102980.
<https://doi.org/10.1016/j.ijhm.2021.102980>
- Zercher, F., Schmidt, P., Ciecuch, J. a Davidov, E. (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.
<https://doi.org/10.3389/fpsyg.2015.00733>