



VLIV VERBÁLNÍCH KOTEV LIKERTOVY ŠKÁLY NA PSYCHOMETRICKÉ CHARAKTERISTIKY DOTAZNÍKŮ

Bc. Martin Tancoš

Katedra psychologie, FSS MU

446083@mail.muni.cz

Abstrakt

Tato studie se zabývá různými podobami verbálních kotev Likertovy škály a jejich vlivem na psychometrické charakteristiky dotazníků. Jednotlivé odpovědní formáty kotev byly vytvořeny vlivem dvou faktorů: různou extremitou kotev na krajních bodech a přítomností kotev na všech bodech škály, nebo jen na koncích. Vnitrosubjektový design s různými formáty měl překonat nedostatky předchozího výzkumu o tomto tématu. Při porovnávání byla zohledněna reliabilita, kritériální validita, konstruktová validita a distribuce jednotlivých položek. Bylo zjištěno, že reliabilita se mezi jednotlivými formáty neliší. V případě kritériální validity byly nalezeny drobné rozdíly, ale pouze u žen. Konstruktová validita byla zkoumána z hlediska vnitřní struktury a pomocí vztahů se skutečnou výškou. V obou případech nebyly zjištěny žádné rozdíly. Extrémní verbální kotvy však mírně snižovaly tendenci k extrémnímu odpovídání a tím i variabilitu škál. Lze tedy říci, že verbální kotvy mohou ovlivňovat rozdělení jednotlivých položek a celkového skóre, nikoli však charakteristiku měřeného rysu. V případě hrubých skóre je nutné ověřit, zda jsou skóre stejných škál s různými odpovědními formáty vzájemně srovnatelné. Na úrovni konstruktů je možné dojít k závěru, že bez ohledu na odpovědní formát měří škály daný konstrukt stejně validně.

Klíčová slova: Likertova škála, měření postojů, odpovědní formát, verbální kotvy

Abstract

This study focuses on different formats of Likert-type scales verbal labels and their influence on psychometric properties of questionnaires. Verbal label formats were defined by two factors: a different extremity of labels on the endpoints and whether the labels were on all points or just the endpoints. Within-subject design with different formats was to outdo the shortcomings of previous research on this topic. Reliability, criterion-related validity, construct validity and item distributions were considered. It was found that reliability is not different between the formats. In the case of criterion-related validity, minor differences were observed, however, only among women. Construct validity was examined in terms of inner structure and relationships with actual height. In both cases, no differences were ascertained. Nevertheless, extreme verbal labels slightly lowered extreme responding and hence the scale variability. Thus, we can claim that verbal labels are able to affect the item and score characteristics, but the scale measures the construct the same way. Evaluating the raw scores, it is necessary to check that the scores coming from different response formats are comparable. At the level of constructs, we can conclude that the scales measure the same construct with equal validity regardless of differences in verbal labels.

Key words: Likert-type scale, attitude measurement, response format, verbal labels

TEORETICKÁ VÝCHODISKA

Postoje, jakožto jeden z ústředních pojmů sociální psychologie, byly objektem zkoumání od počátku existence této vědy. Stejně jako je tomu u jiných psychologických konstruktů, postoje vzhledem k jejich latentní povaze nelze přímo pozorovat. Tato skutečnost příliš neulehčovala naplnění potřeby je nějakým způsobem zachytit, vysvětlit či kvantifikovat (Krosnick, Judd, & Wittenbrink, 2005). Z toho důvodu byla vyvinuta řada postupů jejich kvantifikace, mezi nimi Likertova škála (Likert, 1932).

Měření postojů poměrně záhy překročilo hranice sociální psychologie a postojů vůbec a v souvislosti s tímto rozmachem se Likertova škála nakonec stala jednou z nejvíce rozšířených metod v sociálních vědách vůbec (Carifio & Perla, 2008), což posiluje důvody, proč se metodologickými aspekty Likertovy škály zabývat. Domnívám se, že právě v souvislosti s masovým rozšířením této metody přestal být dáván důraz na možný vliv odpovědního formátu škály na získané odpovědi.

Furr (2011) popisuje odpovědní formát jako způsob, jakým jsou odpovědi na položky prezentovány. Jelikož respondenti s metodou nějakým způsobem interagují, mohou být jejich odpovědi zkresleny tím, jakým způsobem je metoda navržena (např. podoba instrukcí, podoba položek, odpovědní formát), a produkovat tak různě validní výsledky. V kontextu odpovědního formátu se tedy lze domnívat, že by různá konfigurace škály mohla produkovat různě validní odpovědi. Ústředním tématem tohoto výzkumu je pak jeden z konkrétních aspektů odpovědního formátu, a to verbální kotvy jednotlivých odpovědí Likertovy škály.

Navzdory nepřilíživému zájmu o toto téma vznikaly studie, které se vlivem verbálních kotev zabývaly. Bohužel je zde několik výhrad k jejich metodologii. Většina z nich např. používala mezisubjektový design (Dixon et al., 1984; Menold et al., 2014; Menold & Tausch, 2016; Wyatt & Meyers, 1987), s výjimkou studie Wengové (2004), která však ověřovala pouze test-retest reliabilitu, a Changa (1997), kde se vliv krajních bodů a extremity škály posuzoval ve dvou administracích, ačkoli na dvou oddělených vzorcích.

Tato studie se tedy bude zabývat, zda různá podoba verbálních kotev na Likertově škále ovlivní charakteristiky daných škál, a to z hlediska reliability, kriteriální validity, konstruktové validity a distribuce jednotlivých položek.

METODA

Výzkumný vzorek a sběr dat

Výzkumný soubor této studie obsahuje 1080 respondentů. Z tohoto vzorku je 78 % žen ($n = 843$) a 20 % mužů ($n = 215$), 2 % ($n = 22$) své pohlaví neuvedla. Účastníkům výzkumu bylo v době vyplňování v průměru 29,2 let ($SD = 12,2$, $Md = 24$, $Range = 12-77$). Přibližně 3 % ($n = 42$) pak věk v dotazníku neuvedla.

Sběr dat probíhal v elektronické podobě, přičemž potenciální respondenti byli osloveni příspěvky na sociálních sítích a diskusních fórech, jednalo se tedy o příležitostný výběr.

Respondenti nebyli informováni o skutečném záměru výzkumu, aby nedošlo ke zkreslení odpovídání v důsledku zapamatování odpovědí. Prezentovaným záměrem byla snaha zjistit, jak jsou některé psychologické rysy stabilní v čase, a zda nesouvisí se zdánlivě banálními proměnnými (počasím, náladou apod.).

Na základě měsíce narození byli respondenti pseudonáhodně zařazeni do jedné ze čtyř experimentálních skupin. Jakmile byli respondenti rozřazeni do skupin, byl jim následně zobrazen příslušný dotazník.

Výzkumné metody

Pro účel studie byly vybrány dvě škály, u nichž existuje předpoklad, že se jejich měřené konstrukty nemají tendenci měnit v krátkém časovém horizontu. První z nich je zkrácená škála výšky (Rečka, 2018), kde stabilitu konstruktu lze považovat za intuitivní. Druhou je pak škála autonomie z *Basic Needs Satisfaction in General Scale* (BNSG-S; Gagné, 2003), u níž také existují důkazy o její stabilitě v čase (např. Ježek, Macek, & Bouša, 2016).

Procedura

Ústředním aspektem výzkumného designu je manipulace s verbálními kotvami na Likertově škále. Aby případný efekt této manipulace nebyl ovlivněn i jinými aspekty odpovědního formátu, byla zvolena varianta pětibodové škály, jejíž délka se napříč jejími různými podobami neměnila.

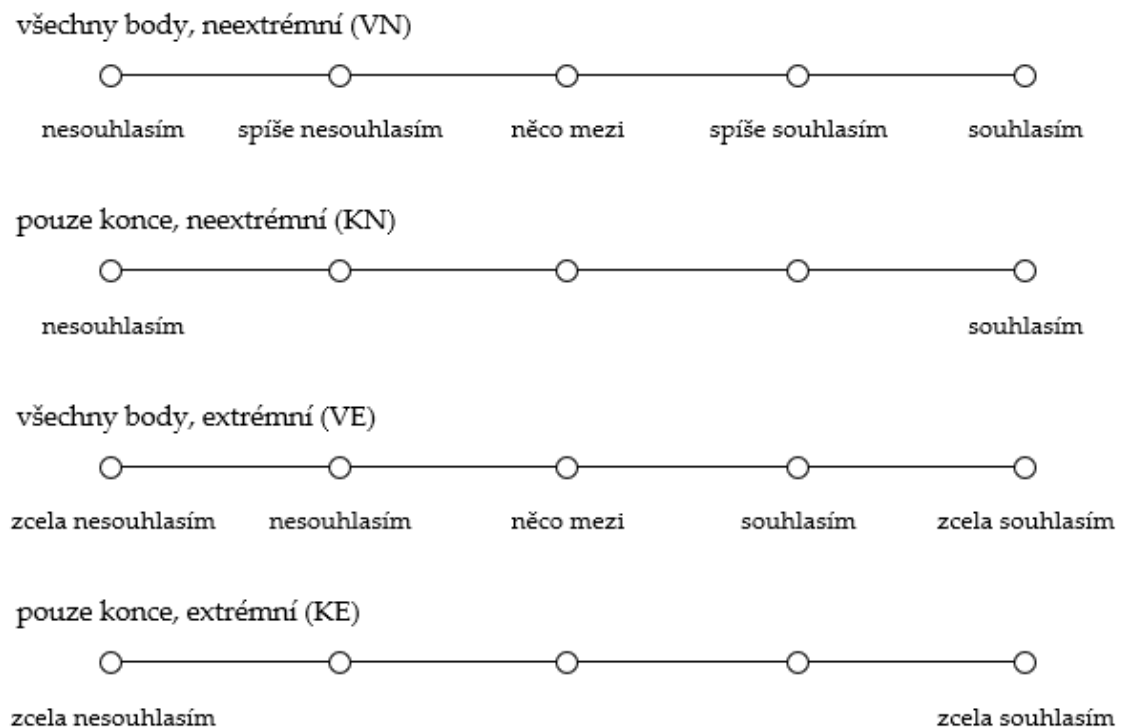
Zaměříme-li se na jednotlivé varianty škál, jejich podoba vznikla jako kombinace dvou faktorů. První z nich rozlišuje škály v závislosti na tom, zda jsou verbální kotvy přítomny na všech bodech, či jsou uvedeny pouze na krajních. Obsahem druhého faktoru jsou pak dvě varianty škál v závislosti na významové intenzitě kotev na jejich krajních bodech. Konkrétně

řeceno se jedná o varianty, zda na pravém konci škály bude mírnější fráze „*souhlasím*“, nebo možnost „*zcela souhlasím*“, které by mělo význam krajních bodů posunout do většího extrému než vyjádření mírné. Zkombinujeme-li oba dva faktory, vytvoříme tak čtyři kombinace Likertových škál, jejichž podoba je uvedena na *obrázku 1*.

Další zásadní charakteristikou tohoto výzkumu je, že se jedná o vnitrosubjektový experimentální design. Jak bylo řečeno výše, na začátku administrace byli respondenti rozděleni do čtyř skupin. Těm byly přiřazeny různé varianty verbálních kotev u stejných dotazníků. Poté jim byly v přibližných 14denních intervalech zasílány ještě třikrát nové instrukce k vyplnění dotazníku.

Zároveň se u každého měření střídalo pořadí dotazníku výšky a autonomie. K tomu všechny skupiny měly v každém měření náhodné pořadí položek, aby bylo zamezeno jejich lokální závislosti. Na konci každého měření byli respondenti dotázáni na svou tělesnou výšku.

Obrázek 1. Verbální kotvy Likertovy škály u jednotlivých měření.



pozn. zkratky jednotlivých formátů budou v této podobě používány v celé studii

Tabulka 1. Zastoupení respondentů v jednotlivých měřeních.

Skupina	T1	T2	T3	T4	T1–T4
1	289	207	174	167	144
2	261	173	157	140	117
3	279	170	159	157	118
4	251	174	151	147	125
Σ	1080	724	641	611	504
% z T1		67,0 %	59,4 %	56,6 %	46,7 %

pozn. T1–T4 – účast na všech čtyřech měřeních

Výzkum založený na opakovaných měřeních s sebou také přináší postupný úbytek vzorku s narůstajícím počtem měření, výjimkou není ani tato studie. V tabulce 1 jsou uvedeny velikosti vzorku za jednotlivé experimentální skupiny v různých měřeních.

Statistická analýza dat byla provedena v jazyce R verze 3.5.1. Datová matice společně se skriptem obsahujícím postupy analýzy dat jsou volně dostupné na *Open Science Framework* (<http://osf.io/twmb8/>).

VÝSLEDKY

Reliabilita

Pro porovnání jednotlivých odpovědních formátů mezi sebou byla jako první použita reliabilita vnitřní konzistence. Toto srovnání tedy bude testovat, zda nám některý z formátů verbálních kotev poskytne odlišnou míru chyby měření než jiný.

V tabulce 2 jsou zobrazeny koeficienty vnitřní konzistence pro škálu výšky pro každé pohlaví zvlášť, tabulka 3 pak ukazuje tyto koeficienty pro škálu autonomie. Koeficienty byly párově srovnávány mezi jednotlivými odpovědními formáty s Holmovou korekcí proti chybě I. typu (Holm, 1979). Tato srovnání jsou podrobně popsána v příloze č. 2. Žádné ze srovnání neukázalo žádný signifikantní rozdíl, koeficienty reliability lze tedy považovat za srovnatelné.

Tabulka 2. Reliability škály výšky v různých formátech.

	<i>ženy</i>				<i>muži</i>			
	ω	<i>SE</i>	95% <i>CI</i>		ω	<i>SE</i>	95% <i>CI</i>	
			<i>dolní</i>	<i>horní</i>			<i>dolní</i>	<i>horní</i>
VYS VN	0,859	0,010	0,837	0,877	0,792	0,028	0,714	0,835
VYS KN	0,828	0,014	0,799	0,853	0,707	0,038	0,607	0,765
VYS VE	0,860	0,011	0,836	0,878	0,757	0,042	0,652	0,823
VYS KE	0,825	0,013	0,795	0,847	0,708	0,050	0,583	0,787

Tabulka 3. Reliability škály autonomie v různých formátech.

formát	ω	<i>SE</i>	95% <i>CI</i>	
			<i>dolní</i>	<i>horní</i>
AUT VN	0,843	0,010	0,823	0,861
AUT KN	0,806	0,013	0,777	0,828
AUT VE	0,841	0,011	0,814	0,859
AUT KE	0,808	0,013	0,781	0,832

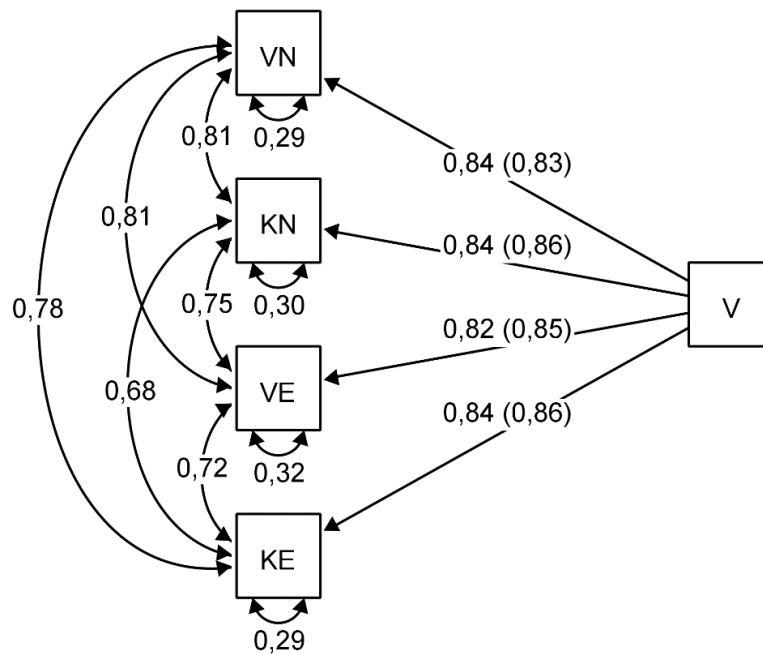
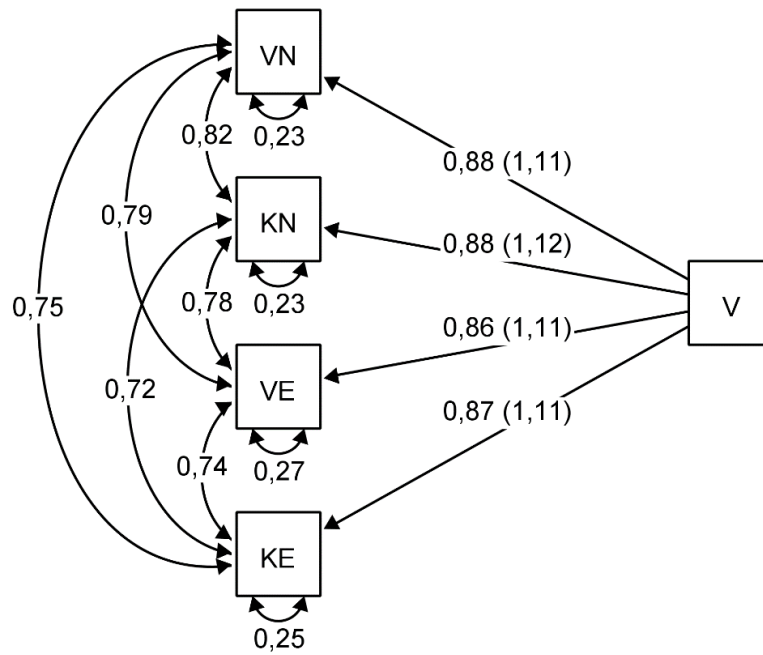
Kriteriální validita

Údaje o kriteriální validitě byly dostupné pouze u škály výšky. Pro testování hypotézy rozdílné kriteriální validity jsme použili úsekovou analýzu, kde byl použit průměr čtyř self-reportů výšky v centimetrech jako exogenní proměnnou, zatímco součtový skóre každého ze čtyř měření dotazníku výšky s různými odpovědními formáty byl proměnnou endogenní. Postupně byly fixovány regresní parametry na stejné hodnoty napříč podmínkami, zvláště pro muže a ženy.

Zafixování reziduálních rozptylů, které v tomto modelu představují kriteriální validitu, ve vzorku signifikantně zhoršilo shodu modelu s daty ($\Delta\chi^2(6) = 19,1$; $p = 0,004$; $\Delta TLI = 0,006$; $\Delta RMSEA = 0,038$; $\Delta SRMR = 0,007$). Podobně reportované ukazatele fitu všech modelů jsou k nahlédnutí v příloze č. 2.

Následná explorační analýza ukázala, že u žen měl neextrémní formát odpovědi vyšší kriteriální validitu než neextrémní ($\Delta\psi^2 = 0,03$; $z = 3,26$; $p = 0,001$), avšak mužů nebyl rozdíl signifikantní ($\Delta\psi^2 = 0,02$; $z = 1,27$; $p = 0,205$). Porovnání formátů se všemi kotvami a s kotvami pouze na krajích neukázalo signifikantní rozdíl jak u žen ($\Delta\psi^2 = 0,01$, $z = 0,65$, $p = 0,514$), tak u mužů ($\Delta\psi^2 = 0,01$, $z = 0,42$, $p = 0,677$).

Graf 1. Parametry modelu pro určení kriteriální validity jednotlivých formátů škály výšky rozdělených dle pohlaví.



pozn. horní diagram pro vzorek žen, dolní pro vzorek mužů; v závorce jsou uvedeny nestandardizované koeficienty

Konstruktová validita

Konstruktová validita se zabývá vazbou měřicího nástroje ke konstruktům, které by měl měřit. Zároveň se tyto předpoklady snaží ověřovat, přičemž způsobů ověřování konstruktové validity je mnoho a každý z nich přináší trochu odlišný úhel pohledu.

Jedním z nich je ověření vnitřní struktury metody, která vychází z teorie o daném konstrukt (Bohrnstedt, 1977; Furr, 2011). Vnitřní struktura v tomto kontextu by měla představovat to, že všechny varianty škál s různými odpovědními formáty budou mít k danému obecnému konstrukt stejné silný vztah, tedy nebude záležet na tom, jaký z nich bude použit.

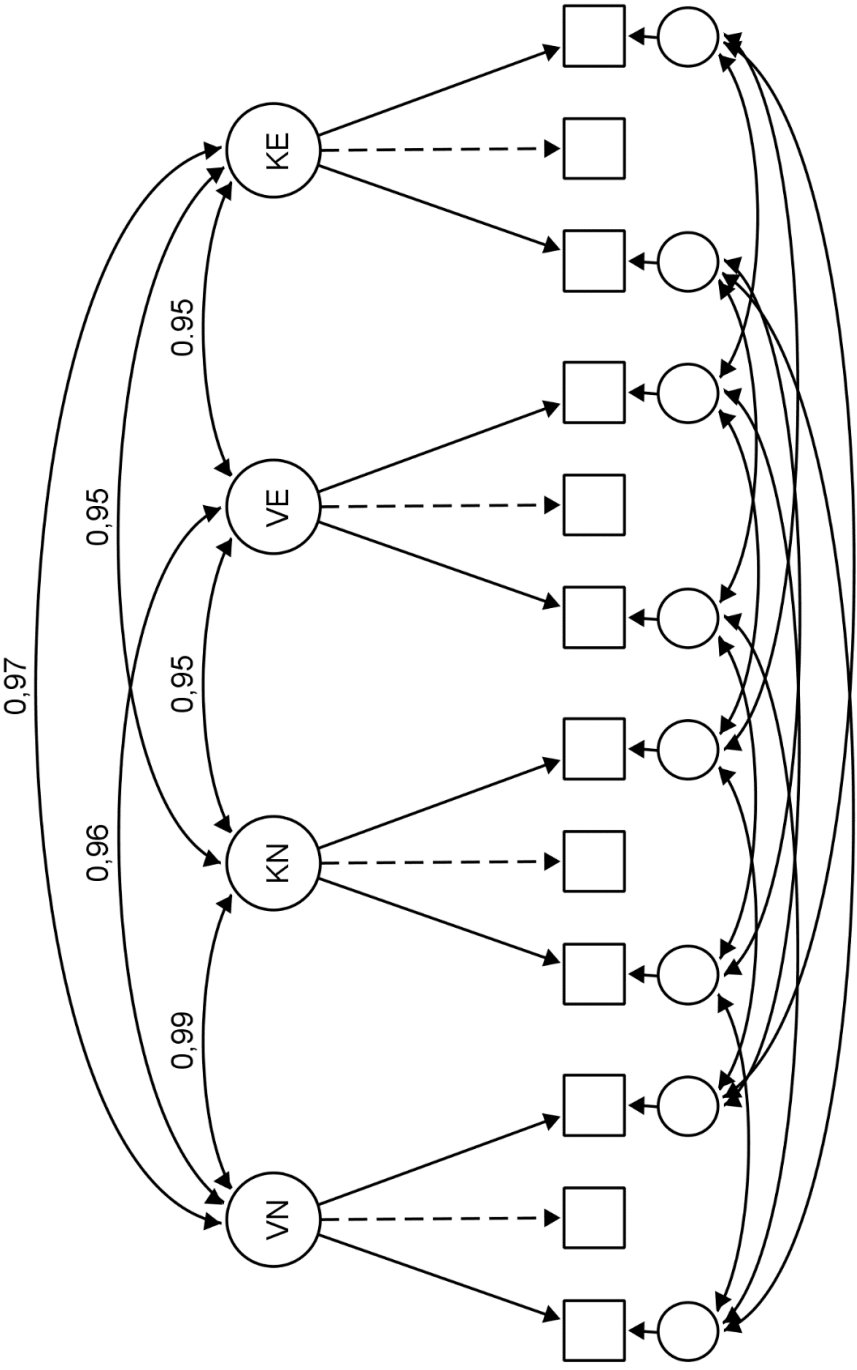
Tento předpoklad se dá ověřit např. modelem longitudinální invariance podle Liuové et al. (2017). Zde vystupují výsledky jednotlivých měření jako latentní proměnné, mezi kterými jsou definované korelační vztahy. Ty by měly reprezentovat míru shody mezi jednotlivými měřeními a zastupovat tak vztahy k obecnému konstrukt daných škál. Je však nutné také uvolnit reziduální korelace mezi stejnými položkami, jelikož lze teoreticky předpokládat, že stejné položky v různých měřeních budou mezi sebou také významně souviset.

K položkám samotným bylo přistupováno jako k ordinálním, což bylo zajištěno metodami odhadu k tomu učným (WLSMV). Tato metoda obnáší odhad pomocí polychorických korelací a prahů mezi jednotlivými možnostmi položek.

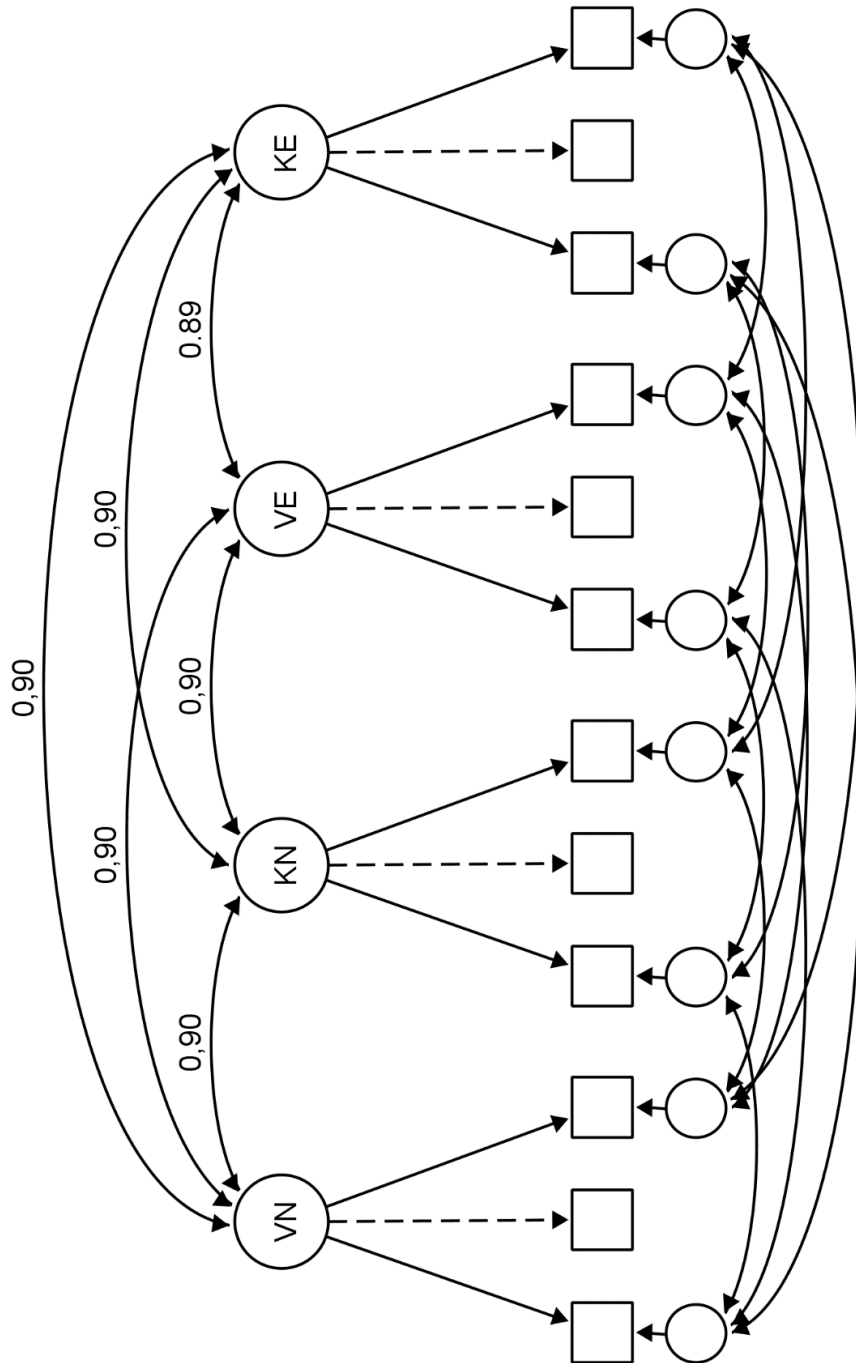
V tomto případě byly ve srovnávacím modelu fixovány hodnoty korelací mezi faktory na stejnou hodnotu. Jelikož škála výšky funguje odlišně pro obě pohlaví, tento model obsahuje velké množství parametrů a vzorek mužů je v tomto ohledu příliš malý, bylo nakonec rozhodnuto o tom, že do této analýzy bude vstupovat pouze vzorek žen. U škály autonomie vstupoval do analýzy celý vzorek, jelikož tyto problémy se zde nevyskytly.

Model s fixovanými korelacemi mezi faktory škály výšky nevykazoval signifikantně horší shodu s daty ($\Delta\chi^2(5) = 8,8$; $p = 0,118$; $\Delta TLI < 0,001$; $\Delta RMSEA < 0,001$; $\Delta SRMR = 0,001$), můžeme tedy konstatovat, že se tyto koeficienty vzájemně neliší. Ke stejnému závěru lze dojít i v případě škály autonomie ($\Delta\chi^2(5) = 0,4$; $p = 0,997$; $\Delta TLI = 0,001$; $\Delta RMSEA = 0,001$; $\Delta SRMR < 0,001$). Ukazatele fitu těchto modelů jsou také vypsány v příloze č. 2.

Graf 2. Hlavní parametry modelu pro odhad konstruktové validity vnitřní struktury škály výšky mezi jednotlivými formáty.



Graf 3. Hlavní parametry modelu pro odhad konstruktové validity vnitřní struktury škály autonomie mezi jednotlivými formáty.



Další způsob, jak nahlížet na ověření konstruktové validity, vznáší hypotézu, zda daný konstrukt koreluje s konstrukty, se kterými by měl mít na základě teoretických předpokladů souvislost (Furr, 2011). V tomto konkrétním případě je konstrukt tělesné výšky zastoupen kritériem respondenty uvedené výšky.

Toto lze také vnímat tak, že zde dochází k propojení kritériální a konstruktové validity, neboť budou odhadovány podobné regresní vztahy, avšak respondenty uvedená výška bude tentokrát predikovat latentní skóry jednotlivých odpovědních formátů očištěné od chyby měření. Opět kvůli absenci kritéria není možné tuto analýzu provést pro škálu autonomie.

Ve srovnávacím modelu byly tedy zafixovány reziduální rozptyly reprezentující míru konstruktové validity. V tomto případě se koeficienty od sebe neliší, neboť nedošlo k zhoršení shody s daty ($\Delta\chi^2(3) = 0,6$; $p = 0,896$; $\Delta TLI < 0,001$; $\Delta RMSEA < 0,001$; $\Delta SRMR < 0,001$). Pro podrobnosti modelů opět viz přílohu č. 2.

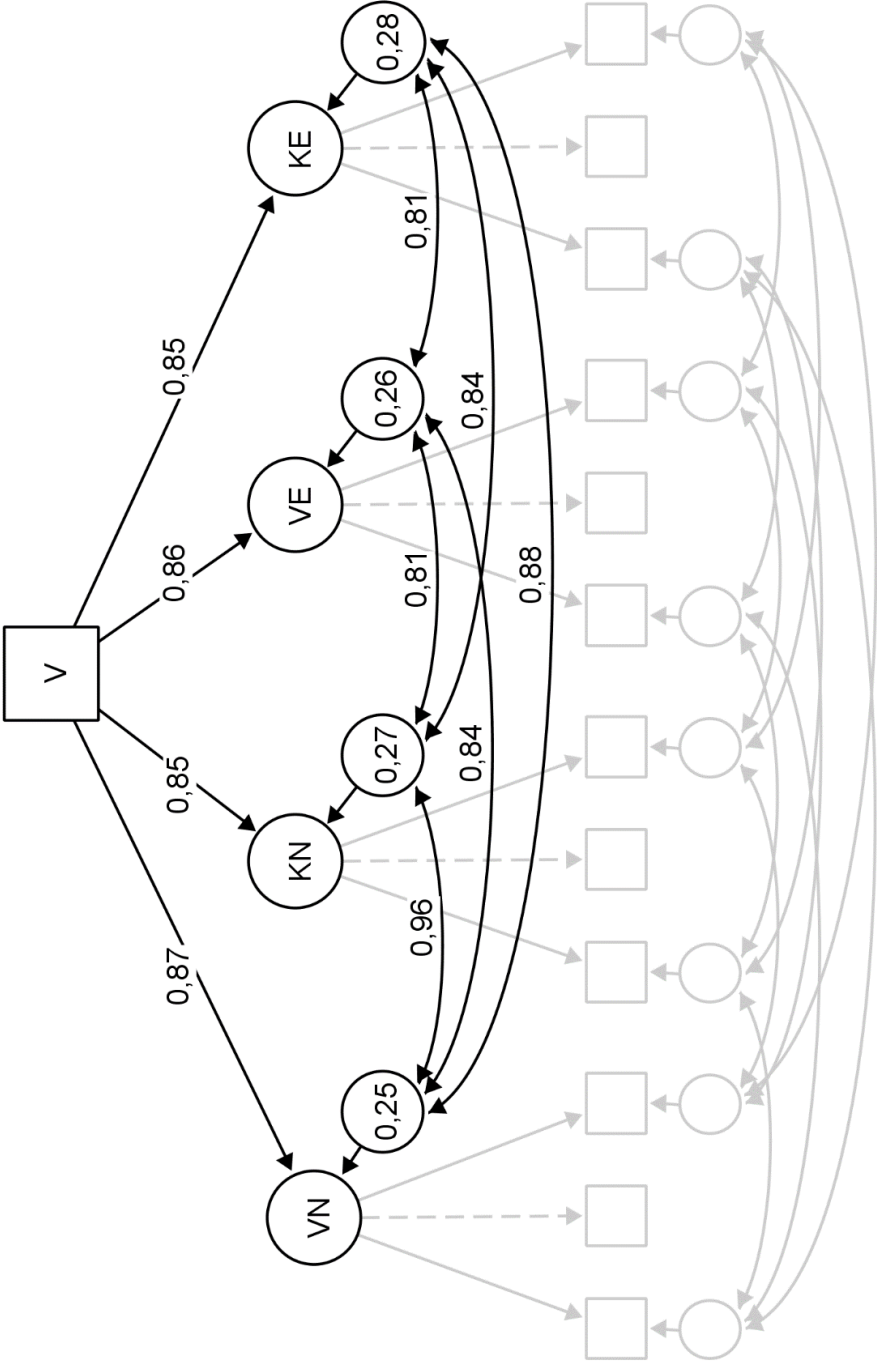
Distribuce položek

Tato část se bude zabývat tím, zda extremita škály má vliv na distribuci položek, tedy jestli se položky chovají stejně bez ohledu na jejich odpovědní formát. Tuto hypotézu lze testovat pomocí postupné metrické a skalární invariance jednotlivých položek mezi odpovědními formáty.

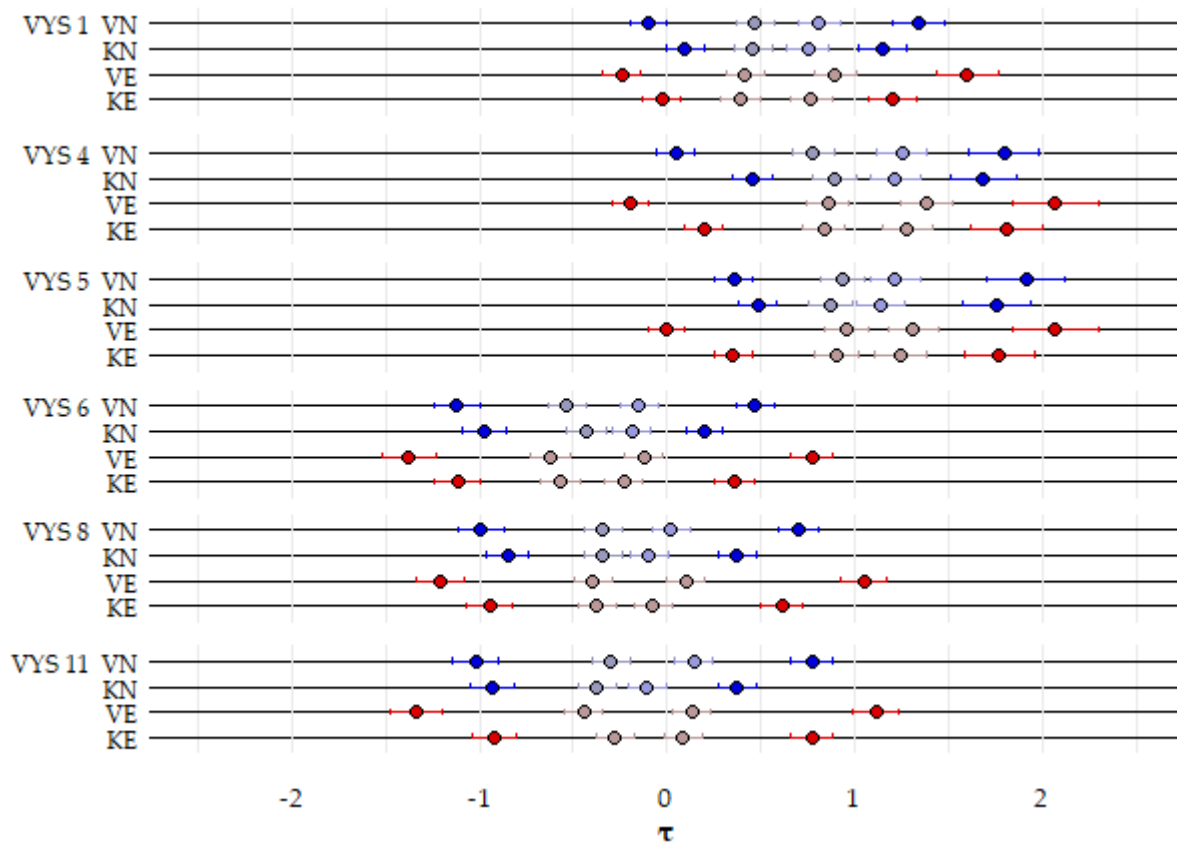
V případě obou metod byly zafixovány faktorové náboje jednotlivých položek za účelem testování metrické invariance. V obou případech tento krok nevedl k horší shodě s daty. Následný proces ověřování skalární invariance zahrnuje fixaci prahů jednotlivých položek na stejnou hodnotu. Jelikož hlavním zájmem jsou krajní odpovědi, byl tento proces rozdělen na dvě části. Nejdříve byly fixovány prahy vnitřní a následně krajní.

Fixace vnitřních prahů nepřinesla významné zhoršení shody modelu s daty u škály výšky ($\Delta\chi^2(66) = 48,8$; $p = 0,944$; $\Delta TLI = 0,001$; $\Delta RMSEA = 0,002$; $\Delta SRMR < 0,001$), avšak u škály autonomie měl model signifikantně horší shodu ($\Delta\chi^2(42) = 100,4$; $p < 0,001$; $\Delta TLI = 0,001$; $\Delta RMSEA < 0,001$; $\Delta SRMR < 0,001$), nicméně věcně zanedbatelnou. Následná fixace krajních prahů přinesla signifikantně horší shodu s daty jak škály výšky ($\Delta\chi^2(66) = 323,6$; $p < 0,001$; $\Delta TLI = 0,001$; $\Delta RMSEA = 0,001$; $\Delta SRMR < 0,001$), tak u škály autonomie ($\Delta\chi^2(42) = 343,9$; $p < 0,001$; $\Delta TLI = 0,011$; $\Delta RMSEA = 0,010$; $\Delta SRMR = 0,001$), ačkoli nebyla příliš výrazná.

Graf 4. Hlavní parametry modelu určující konstruktovou validitu s kritériem pro jednotlivé formáty škály výšky.

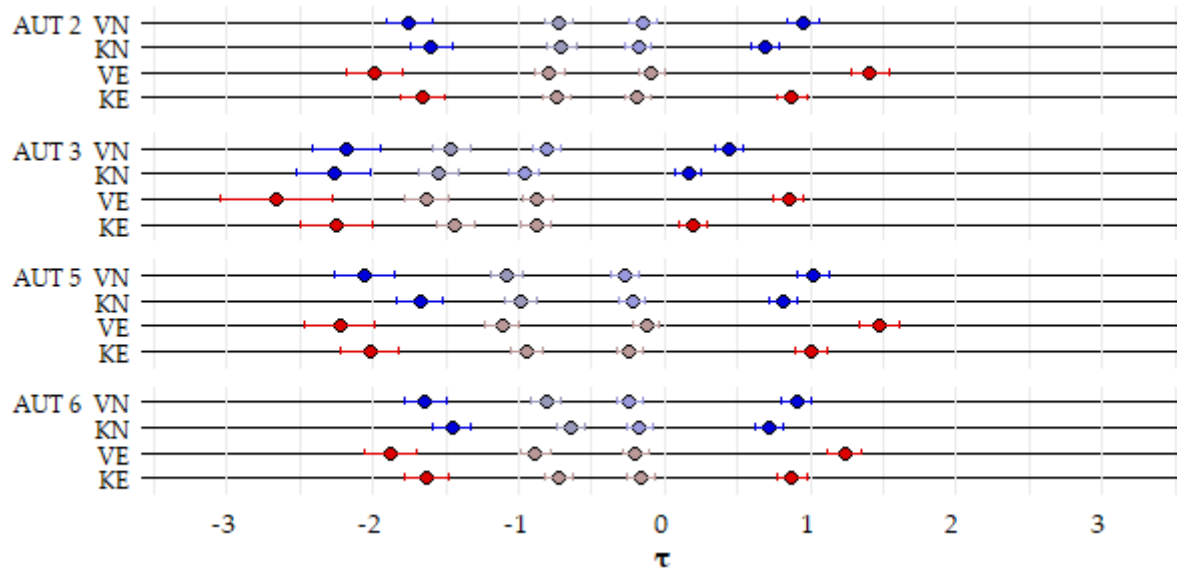


Graf 5. Hodnoty prahů vybraných položek škály výšky jednotlivých formátů.



pozn. prahy jsou na každém řádku uvedeny ve vzestupném pořadí ($\tau_1, \tau_2, \tau_3, \tau_4$)

Graf 6. Hodnoty prahů vybraných položek škály autonomie jednotlivých formátů.



pozn. prahy jsou na každém řádku uvedeny ve vzestupném pořadí ($\tau_1, \tau_2, \tau_3, \tau_4$)

Je však vhodné se podívat, jestli rozdíly mezi krajními prahy jednotlivých položek nevykazují nějakou pravidelnou strukturu. Na grafu 4 a 5 se znázorněnými standardizovanými prahy u vybraných položek lze vyzorovat, že prahy položek s extrémními kotvami jsou od sebe více vzdálené než v případě neextrémních kotev.

DISKUZE

Cílem této studie bylo prozkoumat, zda různé podoby verbálních kotev pětibodové Likertovy škály mohou ovlivnit psychometrické charakteristiky sebeposuzovacích škál. U verbálních kotev byla rozlišena extremita frází a přítomnost kotev na všech bodech škály, nebo pouze na koncích. Jednotlivé formáty byly porovnány z hlediska reliability, kriteriální validity, konstruktové validity a distribuce položek.

Co se týče reliability, byla sice zjištěna systematicky nižší vnitřní konzistence u obou škál v případě kotev pouze na koncích, nicméně tento rozdíl nebyl statisticky významný. Navzdory tomu, že je z toho hlediska nutné považovat jednotlivé míry reliability za ekvivalentní, je otázkou, zda by za jiných podmínek nedošlo k výraznějším odchýlkám. Závěr o shodných reliabilitách navazuje na tvrzení Wyatta a Meyerse (1987) s ohledem na extremitu škály, ale také na metaanalýzy Churchilla a Petera (1984) a Petersona (1994), které se týkaly porovnání všech a koncových kotev. Naproti tomu tato studie rozporuje závěr Menoldové et al. (2014), kde byly škály se všemi kotvami označeny za reliabilnější než kotvy pouze na krajních bodech.

Většina studií na toto téma sice porovnávalo reliabilitu jednotlivých formátů, výrazně opomíjeným tématem však zůstávala validita. V této studii byla nejprve porovnána kriteriální validita, a to pomocí predikce tělesné výšky ze součtových skóre škály výšky. V tomto případě se ukázalo, že extrémní škály byly méně validní než jejich neextrémní protějšky, nicméně byl tento rozdíl zjištěn pouze u žen, a to navíc s velmi malým efektem. Kriteriální validita v případě mužů byla naopak srovnatelná. Tento rozpor je nejspíše způsoben nekonzistentními rozdíly v centrální tendenci či variabilitě jednotlivých skóre mezi pohlavími, příp. nižší silou testu v případě menšího vzorku mužů. Závěr z tohoto srovnání je tedy poměrně nejednoznačný a je nejspíše nutný další výzkum, který by přinesl jasnější stanovisko. Každopádně pozorovaný efekt je velmi malý až zanedbatelný.

Srovnání na úrovni hrubých skóre je bohužel zatíženo chybou měření. Další srovnávání z hlediska konstruktové validity tedy probíhalo na úrovni latentních skóre. Konstruktová validita byla ověřena ze dvou různých aspektů, nejprve z hlediska vnitřní struktury jednotlivých

formátů. V obou použitých škálách byly vztahy mezi formáty prohlášeny za rovnocenné, tedy že se korelace mezi latentními rysy za různých podmínek neliší. Lze tedy konstatovat, že jednotlivé formáty mají k obecným rysům daných škál stejně silný vztah. Konstruktová validita byla dále posouzena z hlediska vztahu ke vnějšímu kritériu, šlo tedy do jisté míry o kombinaci obou předchozích přístupů. Ani v tomto případě nebyl nalezen rozdíl ve validitě mezi odpovědními formáty. Můžeme tedy usuzovat, že rozdíly v kritériální validitě hrubých skóre jsou spíše ovlivněny drobnými rozdíly v reliabilitě skóre či jejich variabilitě, než že by se lišila povaha měřeného konstruktů jako takového.

Mimo porovnání součtových či latentních skóre byly také zohledněny distribuce jednotlivých položek. Při podrobnějším srovnání byl odhalen poměrně systematický trend, kde byly prahy položek s extrémními kotvami více „roztažené“, což přímo souvisí s menší tendencí k extrémnímu odpovídání a s nižší variabilitou těchto položek. Lze zde tedy najít paralelu se studií Wyatta a Meyerse (1987), kde měly škály s extrémními kotvami nižší variabilitu.

Tato dílčí zjištění nám nakonec poskytují významné implikace pro validitu odpovědních procesů Likertovy škály a správnou interpretaci odpovědí z ní vzešlých. Prvním důležitým závěrem je, že jednotlivé položky mohou být mírně ovlivněny změnou verbálních kotev na škále, nicméně jako celek funguje celá škála prakticky srovnatelně ať po stránce reliability, tak i validity.

Dále je nutné zmínit, že hrubé skóre škál při změně verbálních kotev nejsou dle této studie zcela porovnatelné. Při měření daného rysu je tedy nutné zohlednit, že různé odpovědní formáty nemusí produkovat stejné výstupy, a v těchto situacích je třeba ověřit invarianci různých formátů jedné škály.

Poslední klíčové zjištění ukazuje, že na úrovni konstruktů měří škála při různých odpovědních formátech daný konstrukt stejně validně jako ostatní. V tomto ohledu lze tedy měření konstruktů danou škálou zobecnit napříč odpovědními formáty v rozsahu, ve kterém je zkoumala tato studie. Tento závěr vrhá v obecné rovině na koncept Likertovy škály dobré světlo, jelikož se ukazuje, že výběr odpovědního formátu s ohledem na verbální kotvy není určujícím faktorem vypovídající hodnoty škály. Tím je zajištěna velmi důležitá vlastnost přenositelnosti této škály mezi různými výzkumy a škálami v nich obsažených.

Limity studie a doporučení

Navzdory zajímavým zjištěním je však třeba zmínit i jejich možná omezení. Jak již bylo řečeno, výsledky této studie se dají v kontextu verbálních kotev zobecnit pouze do rozsahu, v jakém byly zkoumány v této studii. Je tedy do budoucna třeba zohlednit i jiné formáty verbálních kotev v dalším výzkumu.

Zobecnitelnost získaných zjištění však posiluje použití dvou prakticky nezávislých škál měřících zcela odlišné konstrukty. Lze tedy konstatovat, že by tyto efekty měly být nezávislé na výběru škály. Nicméně je nutné posoudit tyto vlivy na dalších škálách, neboť nezávislost škál ještě nemusí implikovat nezávislost jejich odpovědních procesů. Je totiž možné, že škály např. s odlišnou distribucí by mohly být ovlivněny verbálními kotvami jiným způsobem. Rovněž Wengová (2004) tvrdí, že reliabilita škál s homogenními položkami z hlediska jejich vztahu s měřeným konstruktem jsou odolnější proti změně odpovědního formátu. Škály s výrazně heterogenní strukturou by tedy mohly vykazovat jiné výsledky.

Dalším omezením na zobecnitelnosti této studie je délka škály. Tato studie ponechávala konstantní pětibodovou škálu, aby tento vliv do zjištěných efektů nezasahoval. Ačkoli z předchozího výzkumu nevychází obecná shoda na tom, jakým způsobem se mění reliabilita Likertovy škály v závislosti na její délce, existuje několik studií, které připouští, že reliabilita roste s délkou škály (např. Alwin, 1992; Churchill & Peter, 1984; Weng, 2004) nebo naopak klesá (např. McKelvie, 1978). Bylo by tedy zajímavé zkoumat interakci vlivů délky škály a verbálních kotev.

Naopak významným přínosem oproti předchozím výzkumům je použití vnitrosubjektového designu, který umožnil oddělení individuálních rozdílů mezi jednotlivými respondenty od vlivu verbálních kotev. Častý nedostatek dosavadních studií bylo totiž použití mezisubjektového designu právě kvůli nutnosti oddělit dva výše zmíněné efekty.

Další limity lze vztáhnout také ke zkoumanému vzorku. Je nutné říci, že vzorek se mohl lišit od zbytku populace větším zastoupením lidí, kteří mají bohaté zkušenosti s odpovídáním na dotazníky. Lidé s minimem těchto zkušeností by tedy nemuseli odpovídat na základě stejných vstupních informací. V naší většinové populaci však dle mého názoru není příliš mnoho lidí, kteří by s odpovídáním na dotazníky neměli zkušenost, a tudíž by se tento efekt v našem kontextu nemusel vůbec projevit, i kdyby v tomto ohledu nějaké rozdíly existovaly. Může také záležet na „obtížnosti“ výzkumných metod, tedy zda respondenti odpovídají spíše průměrnými, nebo extrémními odpověďmi. Je možné, že v případě extrémní obtížnosti dotazníku by závěry mohly být jiné.

Ačkoli by zmíněné závěry mohly spíše navádět k určité lhostejnosti a zbavení odpovědnosti při výběru odpovědního formátu, neboť dle závěrů studie různé formáty verbálních kotev produkují prakticky stejně validní výsledky, tak se naopak tato studie snaží znovuotevřít téma validity Likertovy škály s ohledem na její vlastnosti v různých kontextech. Pokud tedy chceme vznést nějaké předpoklady o vlastnostech Likertovy škály, tedy např. že na určitém aspektu toho, jak je navržena, nezáleží, je nutné tyto předpoklady neustále empiricky ověřovat.

LITERATURA

- Alwin, D. (1992). Information Transmission in the Survey Interview: Number of Response Categories and the Reliability of Attitude Measurement. *Sociological Methodology*, 22, 83–118. doi: [10.2307/270993](https://doi.org/10.2307/270993)
- Bohrnstedt, G. W. (1977). Reliability and Validity Assessment in Attitude Measurement. In G. F., Summers (Ed.), *Attitude Measurement* (pp. 80–99). London: Kershaw Publishing Company.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42(12), 1150–1152. doi: [10.1111/j.1365-2923.2008.03172.x](https://doi.org/10.1111/j.1365-2923.2008.03172.x)
- Chang, L. (1997). Dependability of Anchoring Labels of Likert-Type Scales. *Educational and Psychological Measurement*, 57(5), 800–807. doi: [10.1177/0013164497057005005](https://doi.org/10.1177/0013164497057005005)
- Churchill, G., & Peter, J. (1984). Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research*, 21(4), 360–375. doi: [10.2307/3151463](https://doi.org/10.2307/3151463)
- Dixon, P. N., Bobo, M., & Stevick, R. A. (1984). Response Differences and Preferences for All-Category-Defined and End-Defined Likert Formats. *Educational and Psychological Measurement*, 44(1), 61–66. doi: [10.1177/0013164484441006](https://doi.org/10.1177/0013164484441006)
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Los Angeles: Sage.
- Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion*, 27, 199–223. doi: [10.1023/A:1025007614869](https://doi.org/10.1023/A:1025007614869)
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70. doi: [10.2307/4615733](https://doi.org/10.2307/4615733)
- Ježek, S., Macek, P., & Bouša, O. (2016). Cesty k nezávislosti: Jak se vyvíjí autonomie s identitou. In L. Lacinová, S. Ježek, & P. Macek (Eds.), *Cesty do dospělosti: psychologické a sociální charakteristiky dnešních dvacátníků* (pp. 25-38). Brno: Masarykova univerzita.
- Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The Measurement of Attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 21–76). Mahwah, NJ: Psychology Press
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 5–55.

- Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. doi: [10.1037/met0000075](https://doi.org/10.1037/met0000075)
- McKelvie, S. J. (1978). Graphic rating scales - How many categories? *British Journal of Psychology*, 69(2), 185–202. doi: [10.1111/j.2044-8295.1978.tb01647.x](https://doi.org/10.1111/j.2044-8295.1978.tb01647.x)
- Menold, N., & Tausch, A. (2016). Measurement of Latent Variables With Different Rating Scales. *Sociological Methods & Research*, 45(4), 678–699. doi: [10.1177/0049124115583913](https://doi.org/10.1177/0049124115583913)
- Menold, N., Kaczmirek, L., Lenzner, T., & Neusar, A. (2014). How Do Respondents Attend to Verbal Labels in Rating Scales? *Field Methods*, 26(1), 21–39. doi: [10.1177/1525822x13508270](https://doi.org/10.1177/1525822x13508270)
- Peterson, R. (1994). A Meta-Analysis of Cronbach's Coefficient Alpha. *Journal of Consumer Research*, 21(2), 381–391. doi: [10.2307/2489828](https://doi.org/10.2307/2489828)
- Rečka, K. (2018). *Dotazník výšky a váhy* (Diplomová práce). Brno: Fakulta sociálních studií MU.
- Weng, L.-J. (2004). Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability. *Educational and Psychological Measurement*, 64(6), 956–972. doi: [10.1177/0013164404268674](https://doi.org/10.1177/0013164404268674)
- Wyatt, R. C., & Meyers, L. S. (1987). Psychometric Properties of Four 5-Point Likert Type Response Scales. *Educational and Psychological Measurement*, 47(1), 27–35. doi: [10.1177/0013164487471003](https://doi.org/10.1177/0013164487471003)

Příloha č. 1: Znění položek

VYS 1: „*Mám vhodnou výšku na hraní basketbalu nebo volejbalu.*“

VYS 2: „*Slychávám narážky na to, že jsem vysoký/á.*“

VYS 3: „*Lidem, kteří na koncertě stojí za mnou, většinou má postava dost brání ve výhledu.*“

VYS 4: „*Když chci někoho obejmout, většinou se musím sklonit.*“

VYS 5: „*Často si musím dávat pozor, abych se neuhodil/a hlavou např. o nízký strop nebo rám dveří.*“

VYS 6: „*Často potřebuji stoličku, abych dosáhl/a na něco, na co jiní lidé dosáhnou normálně.*“ (R)

VYS 7: „*Jednou z prvních věcí, které si na mně lidé všimnou, je to, jak moc jsem malý/á.*“ (R)

VYS 8: „*Často musím stát na špičkách, abych lépe viděl/a.*“ (R)

VYS 9: „*V autobuse mívám dostatek prostoru pro nohy.*“ (R)

VYS 10: „*Kvůli mé menší výšce lidé hádají, že jsem mladší, než ve skutečnosti jsem.*“ (R)

VYS 11: „*Když mluvím s jinými dospělými a chci se jim dívat do očí, častěji na ně spíš vzhlížím nahoru.*“ (R)

AUT 1: „*Můžu se sám/sama rozhodovat, jak chci žít svůj život.*“

AUT 2: „*Mám pocit, že mě pořád někdo někam tlačí.*“ (R)

AUT 3: „*Můžu volně vyjadřovat své nápady a názory.*“

AUT 4: „*V běžném životě musím často dělat, co mi druzí řeknou.*“ (R)

AUT 5: „*Lidé, se kterými se běžně setkávám, berou v potaz, jak se cítím.*“

AUT 6: „*V běžných každodenních situacích mám pocit, že se nemusím přetvařovat.*“

AUT 7: „*V běžném životě nemám moc příležitostí rozhodovat, jak chci co dělat.*“ (R)

Reverzní položky jsou označeny symbolem „(R)“.

Příloha č. 2: Tabulky modelů

Tabulka 4. Testy rozdílů mezi reliabilitami škály výšky u obou pohlaví.

ženy									
z					p				
	VN	KN	VE	KE		VN	KN	VE	KE
VN	0				VN	1	0,075	0,952	0,041
KN	1,78	0			KN	0,283	1	0,071	0,843
VE	-0,06	-1,81	0		VE	1,000	0,283	1	0,039
KE	2,04	0,20	2,07	0	KE	0,233	1,000	0,233	1

muži									
z					p				
	VN	KN	VE	KE		VN	KN	VE	KE
VN	0				VN	1	0,071	0,497	0,145
KN	1,81	0			KN	0,425	1	0,367	0,980
VE	0,68	-0,90	0		VE	1,000	1,000	1	0,450
KE	1,46	-0,03	0,76	0	KE	0,723	1,000	1,000	1

pozn. p-hodnoty pod diagonálou jsou korigovány Holmovou (1979) korekcí

Tabulka 5. Testy rozdílů mezi reliabilitami škály autonomie.

z					p				
	VN	KN	VE	KE		VN	KN	VE	KE
VN	0				VN	1	0,023	0,879	0,032
KN	2,28	0			KN	0,137	1	0,038	0,904
VE	0,15	-2,08	0		VE	1,000	0,158	1	0,051
KE	2,15	-0,12	1,95	0	KE	0,158	1,000	0,158	1

pozn. p-hodnoty na pod diagonálou jsou korigovány Holmovou (1979) korekcí

Tabulka 6. Testování rozdílů v kriteriální validitě škály výšky s oddělenými pohlavími.

Model	fix.	χ^2	df	$\Delta\chi^2_{SB}$	Δdf	p	TLI	RMSEA [90% CI]	SRMR
VYS 1a		0	0						
VYS 2a	b	1,7	6	1,7	6	0,946	1,000	0,000 [0,000; 0,000]	0,005
VYS 2b	b, ψ^2	21,2	12	19,1	6	0,004	0,994	0,038 [0,011; 0,062]	0,012

pozn. fix. – fixované parametry, χ^2_{SB} – chí-kvadrát korigovaný podle Satorry a Bentlera (2001), b – regresní koeficienty, ψ^2 – reziduální rozptyly

Tabulka 7. Testování konstruktové validity škály výšky na úrovni latentních skóru

Model	fix.	χ^2	df	$\Delta\chi^2_{SB}$	Δdf	p	TLI	RMSEA [90% CI]	SRMR
VYS 3a		4011,0	830				0,979	0,067 [0,065; 0,070]	0,096
VYS 3b	φ	4007,5	835	8,8	5	0,118	0,979	0,067 [0,065; 0,069]	0,097

pozn. fix. – fixované parametry, χ^2_{SB} – chí-kvadrát korigovaný podle Satorry a Bentlera (2001), φ – korelace mezi faktory

Tabulka 8. Testování konstruktové validity škály výšky na úrovni faktorových skóru

Model	fix.	χ^2	df	$\Delta\chi^2_{SB}$	Δdf	p	TLI	RMSEA [90% CI]	SRMR
AUT 3a		745,9	302				0,984	0,037 [0,034; 0,040]	0,041
AUT 3b	φ	728,0	307	0,4	5	0,997	0,985	0,036 [0,032; 0,039]	0,041

pozn. fix. – fixované parametry, χ^2_{SB} – chí-kvadrát korigovaný podle Satorry a Bentlera (2001), φ – korelace mezi faktory

Tabulka 9. Testování metrické a skalární invariance škály výšky mezi odpovědními formáty.

Model	fix.	χ^2	df	$\Delta\chi^2_{SB}$	Δdf	p	TLI	RMSEA [90% CI]	SRMR
VYS 3a		4011,0	830				0,979	0,067 [0,066; 0,069]	0,094
VYS 4a	λ	3585,2	860	30,2	30	0,455	0,983	0,061 [0,059; 0,062]	0,095
VYS 4b	λ, τ_{2-3}	3732,3	926	48,8	66	0,944	0,984	0,059 [0,057; 0,061]	0,095
VYS 4c	λ, τ_{1-4}	4083,9	992	323,6	66	< 0,001	0,983	0,060 [0,059; 0,062]	0,095

pozn. fix. – fixované parametry, χ^2_{SB} – chí-kvadrát korigovaný podle Satorry a Bentlera (2001), λ – faktorové náboje, τ_{2-3} – vnitřní prahy, τ_{1-4} – všechny prahy

Tabulka 10. Testování metrické a skalární invariance škály autonomie mezi odpovědními formáty.

Model	fix.	χ^2	df	$\Delta\chi^2_{SB}$	Δdf	p	TLI	RMSEA [90% CI]	SRMR
AUT 3a		745,9	302				0,984	0,037 [0,034; 0,040]	0,041
AUT 4a	λ	755,1	320	28,2	18	0,060	0,985	0,036 [0,032; 0,039]	0,048
AUT 4b	λ, τ_{2-3}	875,7	362	100,4	42	< 0,001	0,984	0,036 [0,033; 0,039]	0,048
AUT 4c	λ, τ_{1-4}	1302,9	404	343,9	42	< 0,001	0,975	0,046 [0,043; 0,048]	0,049

pozn. fix. – fixované parametry, χ^2_{SB} – chí-kvadrát korigovaný podle Satorry a Bentlera (2001), λ – faktorové náboje, τ_{2-3} – vnitřní prahy, τ_{1-4} – všechny prahy

Tabulka 11. Testování konstruktové validity škály výšky na úrovni odhadu skutečné výšky.

Model	fix.	χ^2	df	$\Delta\chi^2_{SB}$	Δdf	p	TLI	RMSEA [90% CI]	SRMR
VYS 5a	λ, τ_{2-3}	3809,9	966				0,982	0,059 [0,057; 0,061]	0,097
VYS 5b	$\lambda, \tau_{2-3}, \beta$	3815,7	969	0,6	3	0,896	0,982	0,059 [0,057; 0,061]	0,097

pozn. fix. – fixované parametry, χ^2_{SB} – chí-kvadrát korigovaný podle Satorry a Bentlera (2001), λ – faktorové náboje, τ_{2-3} – vnitřní prahy, β – standardizované regresní koeficienty