

**M A S A R Y K O V A
U N I V E R Z I T A**

FAKULTA SOCIÁLNÍCH STUDIÍ

**Dotazník
administrovaný videem:
psychometrické
vlastnosti a srovnání
s textovou podobou**

Bakalářská práce

ADAM STROJIL

Vedoucí práce: Mgr. Hynek Cígler, Ph.D.

Katedra psychologie
Program Psychologie

Brno 2023

MUNI
FSS

Anotace

Tato práce empiricky zkoumá možnost administrace internetových psychologických dotazníků pomocí krátkých videonahrávek (VD) a následně tuto podobu srovnává s její klasickou textovou variantou (TD). Žádné zásadní rozdíly nebyly nalezeny ani v případě reliabilit, ani v případě kritériální validity metod. V obou použitých škálách bylo mezi metodami dosaženo striktní (reziduální) invariance měření. VD nemusí nutně vést k delšímu času vyplňování, respondenti navíc hodnotili VD pozitivněji. Zdá se tak, že VD může být validní alternativní metoda administrace, nicméně je třeba k zobecňování výsledků přistupovat obezřetně kvůli suboptimální velikosti vzorku ($N = 321$).

Abstract

This thesis empirically examines the feasibility of administering web-based psychological questionnaires using short videos (VQ) and then compares this method to the traditional text-based version (TQ). No significant differences were found in either the reliability or criterion validity of the methods. In both employed scales, strict (residual) measurement invariance was established between the methods. VQ did not necessarily lead to longer completion times; in addition, respondents rated VQ more favourably. Thus, it appears that VQ may be a valid alternative method of administration, however, caution needs to be taken when generalizing the results due to the suboptimal sample size ($N = 321$).

Poděkování

Nejprve ze všeho, a také zejména, bych rád poděkoval svému vedoucímu, Hynkovi Cíglerovi, za konzultace, rady a asistenci, které práci pomohly dovést do podoby, v jaké je nyní odevzdávána. Děkuji i svým přátelům a spolužákům, kteří mi při psaní byli nenahraditelnou oporou. Zvláštní poděkování patří Verče za její ochotu a odvahu stát se jednou z tváří tohoto experimentálního dotazníku. Poděkování směřuje také panu Bartákovi na katedru žurnalistiky za propůjčení a přípravu nahrávacího studia. V neposlední řadě bych rád poděkoval svým rodičům, bez jejichž podpory bych jen stěží mohl studiu a této práci věnovat tolik času.

Obsah

1	Úvod	9
1.1	Způsoby administrace dotazníků	10
1.2	Internetové dotazníky	11
1.3	Sociální žádoucnost v dotaznících	15
1.4	Efekt tazatele	17
1.5	Humanizující elementy v dotaznících	19
1.6	Tato práce	23
2	Metoda	25
2.1	Vzorek.....	25
2.2	Měřicí nástroj.....	27
2.3	Kognitivní interview a předběžné testování	36
2.4	Analýza dat	37
3	Výsledky	38
3.1	Předčasná ukončení a vyřazené případy	38
3.2	Deskriptivní statistiky	40
3.3	Průměrné skóry a metody administrace	43
3.4	Kriteriální validita metod	45
3.5	Faktorová struktura škál a jejich reliabilita.....	48
3.6	Analýza invariance měření.....	50
3.7	Položka tělesné konstituce	52
3.8	Výstupní dotazník a zkušenosti s metodou	54
3.9	Časy vyplňování.....	56
4	Diskuse	60
4.1	Specifika administrace videem	60
4.2	Předčasná ukončení a neúčast	61
4.3	Dotazník výšky a kriteriální validita.....	62
4.4	Vlivy sociální žádoucnosti	62
4.5	Psychometrické vlastnosti škál a analýza invariance	63

4.6	Čas vyplňování	64
4.7	Spokojenost s metodou a pozornost.....	64
4.8	Limity.....	65
4.9	Shrnutí a všeobecná diskuse.....	65
	Použité zdroje	67
	PŘÍLOHA A: Předpoklady regresní analýzy	80
	PŘÍLOHA B: Průměrné časy odpovídání na položky	83

1 Úvod

Dotazníková šetření jsou jednou z nepoužívanějších metod výzkumu nejen v psychologii. Je výhodné, že zejména online dotazník umožňuje relativně snadno oslovit široké vzorky zkoumaných populací, zachovávat anonymitu respondentů a výsledky ověřovat s využitím nejrůznějších stále se zdokonalujících statistických a výpočetních nástrojů. Zatímco se první formy dotazníků objevily už poměrně dávno, se zřejmě prvním doloženým psaným dotazníkem již v první polovině 19. století, až technologický pokrok umožnil administraci velmi se přibližující té dnešní (Gault, 1907; Midena & Yeo, 2022). Tedy jako stejný soubor otázek poskytnutý ve stejný čas k vyplnění většímu počtu respondentů. Stejný pokrok změnil kdysi nejrozšířenější formu sebedministrace – metodu tužka-papír – na administraci s využitím výpočetních technologií. Dotazníky, které respondent vyplňuje skrze webové prohlížeče nebo aplikace se s rozvojem internetu rozšířily natolik, že jsou nyní naprostým standardem. Rozmach internetu a informačních technologií tak přinesl výzkumníkům levnou alternativu k získávání odpovědí na mnohem rozsáhlejších a různorodějších vzorcích, zatímco respondenti mohou dotazníky vyplnit téměř odkudkoliv (Couper, 2000; Wright, 2005). I tento trend však s sebou přinesl řadu nových potenciálních problémů, mezi nimi například obavy o výběr vzorků a validitu odpovědí.

Se zvyšující se dostupností internetu se rapidně zlepšuje i kvalita připojení a s ní se mění i internetový obsah. Konkrétně se stalo naprosto samozřejmé téměř neomezené sdílení a využívání multimédií (obrázky, hudba, videa). Podle predikce společnosti Cisco měl být celkový internetový provoz v roce 2022 tvořen videem až z 82 %, což je viditelný nárůst v porovnání se 72 % v roce 2017 (Cisco, 2019a). Také v rostoucím mobilním internetovém provozu se mělo jednat o výrazný nárůst, a to z 59 % v roce 2017 na 79 % v roce 2022 (Cisco, 2019b). Sledování videí na internetu se dnes už zdaleka neomezuje pouze na sociální síť YouTube. Značné popularity se těší platformy využívající pro svůj obsah krátká videa. Sociální síť TikTok, která je založena jen na tvorbě a sdílení krátkých videí, dosáhla už v roce 2021 jedné miliardy uživatelů (TikTok, 2021). Podobná krátká videa tak začalo nabízet i YouTube se svými „Shorts“, ale také služby Snapchat Spotlight nebo Instagram Reels, kterých je denně uživateli přehráno přes 140 miliard (META, 2022). Tento trend naznačuje, že lidé jsou na internetu videoobsahu vystavováni stále více. To vede k otázce, zda by dalším krokem ve vývoji internetových dotazníků nemohlo být nahrazení nebo alespoň doplnění jejich klasické textové podoby položkami administrovanými pomocí videozáznamů.

Administrace dotazníku skrze krátká videa by pak mohla být kombinací dnešních trendů, běžných textových dotazníků a osobních rozhovorů tváří v tvář. Výsledkem by potenciálně mohla být vyšší kvalita získaných dat například skrze vyšší motivaci

respondenta věnovat se jednotlivým položkám. Humanizující elementy v dotaznících, jako je lidský hlas nebo obličej, mohou nicméně kvalitu dat i ohrožovat. V tomto ohledu se nabízí například problematika sociální žádoucnosti. Je otázkou, zda jsou tyto vlivy natolik silné, aby způsobovaly rozdíly v porovnání s dotazníky textovými. Výzkum v oblasti využití videozáznamů pro administraci dotazníků je ovšem stále omezený a je zkoumán především v kontextu sociologických dotazníků. Studie o jejich využití v oblasti psychologie a aplikaci na konkrétní konstrukty z dostupných informací nejspíše neexistuje dokonce vůbec.

Tato bakalářská práce experimentálně ověří rozdíl mezi dvěma metodami administrace dotazníku – webovým dotazníkem v klasické textové podobě a webovým audiovizuálním dotazníkem. V audiovizuální verzi měřícího nástroje respondent odpovídá na položky, které administrátor pokládá prostřednictvím videozáznamu. Srovnání metod proběhne pomocí vícero způsobů. Půjde mimo jiné o rozdíly v čase vyplňování, deskriptivních statistikách nebo pomocí analýzy invariance. Předmětem zájmu bude ale také subjektivní spokojenost respondenta s danou metodou.

1.1 Způsoby administrace dotazníků

Pod pojmem dotazník se dnes nejspíše jako první vybaví buďto metoda administrace skrze webový prohlížeč, nebo klasická varianta tužka-papír, která byla z velké části variantou první nahrazena. Už třeba Gault (1907) na počátku 20. století nicméně poznamenal, že dotazník nemusí být nutně prezentovaný psanou formou. Je stejně tak možné získávat odpovědi na připravené otázky skrze osobní rozhovor. Takovýto způsob administrace je zmiňován i v definici pojmu dotazník (*questionnaire*) ve slovníku Americké psychologické asociace (APA, n.d.).

Zmíněné způsoby administrace, tedy tužka-papír, interview a nejnovější webové dotazníky, jsou ovšem jen naprostým základem mnohem širšího spektra metod průzkumu, které prošly od svého vzniku dlouhým vývojem (Smith & Kim, 2015). Metody sběru dat se měnily a různě kombinovaly ruku v ruce s vývojem a zvyšující se dostupností technologií. Zejména z počátku byl u vyplňování dotazníku přítomen administrátor, který četl předpřipravené otázky a odpovědi respondenta zaznamenával do záznamového archu. Tato metoda je zvaná PAPI (*Paper and Pencil Interview*) a často je její součástí i dotazník, který vyplňuje sám respondent, tedy SAQ (*Self-administred Questionnaire*). Metodu PAPI samozřejmě později z velké části nahradila ekvivalentní metoda CAPI (*Computer-assisted Personal Interview*), když papír začaly nahrazovat výpočetní technologie (Leisher, 2014; Tourangeau & Smith, 1996). S rozšířením telefonního spojení bylo možné začít odpovědi získávat také dodnes používanou metodou

telefonního interview (TI; *Telephone Interview*). Evan & Miller (1969) byli pravděpodobně první, kteří experimentálně využili pro sběr dotazníkových dat počítač, a to dokonce jako formu sebedministrace. Započali tak to, co se s dalším vývojem počítačů stalo metodou CASI (*Computer-assisted Self-interview*), také jinak zvanou CSAQ (*Computer Self-administered Questionnaire*). Zde respondent čte a zodpovídá položky přímo v počítači, ideálně s minimální nebo nulovou pomocí administrátora. I přesto však bývá administrátor přítomen. Ať už z důvodu navázání kontaktu s respondentem nebo kvůli vysvětlení ovládání zařízení. Metoda CASI se už velmi blíží dnešním internetovým dotazníkům (OQ; *Online Questionnaire*), jejichž rozšíření s vývojem internetu samozřejmě nevyhnutelně následovalo. Nástup CASI pak umožnil tvorbu svých nejrůznějších variant. Jedná se například o ACASI (*Audio CASI*), kdy respondent odpovídá na předem nahrané zvukové otázky, což bylo využíváno zejména v kombinaci s telefonním průzkumem (T-ACASI; *Telephone ACASI*; jinak také IVR; *Interactive Voice Response*). V neposlední řadě se jedná i o zvláštní podskupinu, tedy metodu AVCASI (*Audio-video CASI*), kdy jsou dotazníky administrovány formou videí i se zvukem.

Každý ze zmíněných přístupů v sobě nese výhody, ale také nevýhody, a každá metoda administrace má svou vlastní strukturu chyby, která plyne jak z neochoty zúčastnit se průzkumu (*nonresponse*), tak samozřejmě i ze samotného měření (measurement; Biemer, 2010; Smith & Kim, 2015). S nástupem počítačově administrovaných dotazníků například vyvstaly obavy ze zkreslení výsledků kvůli zhoršeným schopnostem různých skupin uživatelů pochopit a zvládnout jejich ovládání (Couper & Rowe, 1996). Z výše popsaného vývoje je ovšem také patrné, že se metody značně měnily s ohledem na přítomnost administrátora, respektive jeho roli. Zatímco u některých způsobů administrace odpovídají respondenti přímo tazateli, jindy v jeho přítomnosti pouze vyplňují dotazník, nebo přítomnost administrátora není dokonce vyžadována vůbec. Zvláštní případ tvoří ty automatizované metody, kdy administrátor přítomen reálně není a odpovědi mohou zůstat plně anonymní, ale respondent je vystaven nějakým humanizujícím elementům, jako je hlas administrátora (ACASI, T-ACASI) nebo i jeho obraz (AVCASI). Možnost automatického sběru odpovědí a snadnějšího šíření dotazníků prostřednictvím internetu umožnila znatelný rozmach sebedministrace. Výzkumníci tak byli motivováni více se zabývat efektem, jaký má přítomnost tazatele v různých metodách administrace na kvalitu dat, kterou zejména u citlivých otázek ohrožoval efekt sociální žádoucnosti (Tourangeau & Smith, 1996).

1.2 Internetové dotazníky

Zatímco zpočátku si provádění internetových průzkumů žádalo znalosti programovacích jazyků a nebylo tak mnohdy výhodné (Wright, 2005), dnes je situace naprosto

odlišná. Vývoj internetu a příchod nejrůznějších služeb pro administraci dotazníků, jako jsou Qualtrics, SurveyMonkey, Google Forms nebo Microsoft Forms, přinesl pohodlnou cestu sběru dat, často s naprosto nulovými náklady. Administrace dotazníků skrze internet je svým způsobem specifická ve vícero ohledech. Výzkumníci zabývající se metodologií průzkumů považují za potenciální zdroje chyb a hlavní problémy spojené s kvalitou dat ty plynoucí z pokrytí (*coverage*), výběru vzorku (*sampling*), nezodpovídání (*nonresponse*) a chyb měření (Couper, 2000).

1.2.1 Pokrytí, výběr vzorku a nezodpovězení

Pokrytí, které znamená možnost oslovit zájmovou populaci, respektive vzorek z ní, představovalo mnohem markantnější problém dříve, když byly na internetu zastoupeny spíše specifické komunity. Dřívější autoři, kteří vyjadřovali obavy z nízké dostupnosti internetu, tak pracovali například s takovými čísly, jako 35% zastoupení uživatelů internetu mezi obyvateli USA (Couper, 2000) nebo 58% dostupností internetu v Německu (de Leeuw, 2012). Dnešní situace je však ve vyspělých zemích zcela opačná. Lze říci, že na internetu spíše nejsou (než jsou) zastoupeny jen specifické komunity. Pro srovnání, dostupnost internetu konkrétně v německých domácnostech byla v roce 2022 už 91,4 % (Eurostat, 2023a). Tento nárůst je dokreslen i 90% podílem obyvatel Evropské unie, kteří v roce 2022 uvedli, že v posledních třech měsících použili internet (Eurostat, 2023b). Podotkněme, že internet stále nemusí být zcela optimální pro provádění výzkumu ve všech skupinách obyvatel, například mezi seniory. Problém s výběrem vzorku v internetových dotaznících se pak pokrytí úzce týká, nebo z velké části spíše týkal, protože nízké pokrytí komplikovalo náhodný výběr reprezentativního vzorku, a tím i následnou inferenci (Couper, 2000).

Může se zdát, že rozhodnutí potenciálního respondenta nezúčastnit se průzkumu, tedy úplné nezodpovězení (*unit-nonresponse*¹), mohlo být také spojeno s obavami z používání nové technologie (Couper, 2000) a dnes by se už nemuselo tolik projevovat. Míra nezodpovězení je v případě online dotazníků s otevřenou účastí špatně odhaditelná, protože není snadné zjistit, kolik uživatelů sdílený odkaz na webu vidělo. Nedávná metaanalýza 114 studií nicméně potvrdila nižší míru odpovídání na internetové dotazníky v porovnání s ostatními metodami administrace, a to o 12 % (Daikeler et al., 2020). S podobnými výsledky přišla i metaanalýza dřívější, kdy se jednalo o 11 % (Manfreda et al., 2008). Daikelerová et al. (2020) zmiňují jako jednu z možných příčin právě to, že internetové dotazníky jsou v dnešní době internetu nadužívány kvůli svým zanedbatelným nákladům, a to bez ohledu na jejich důležitost nebo kvalitu. Dalším

¹ *Unit-nonresponse* nastává v případě, kdy se oslovená osoba průzkumu vůbec nezúčastní. Na druhou stranu, *item-nonresponse* nastává, když respondent nezodpoví pouze určité položky.

vysvětlením může být, že respondenti, kterým je dotazník administrován jiným způsobem, a to například osobně, telefonicky, nebo jim je alespoň zaslán poštou, ho ve výsledku vnímají jako legitimnější a účast ve výzkumu jako důležitější. Tato vysvětlení spolu nakonec mohou samozřejmě souviset.

Přesné zobecnění na populaci ale není vždy hlavním záměrem, a tak zmíněné problémy ani problémem být nevyhnutelně nemusí. Například některé psychologické dotazníky vyplňované respondenty na webu byly svými výsledky srovnatelné s laboratorními vzorky i v otázce psychometrických vlastností, a to hned na přelomu století (pro přehled: Krantz & Dalal, 2000). Nereprezentativnost výběru vzorku na internetu, která tak začala působit starosti zejména sociologům, v některých případech psychologie nemusela vůbec trápit. Dokonce právě naopak. Protože, jak poznamenávají Krantz a Dalal (2000), velká část klasických psychologických studií a laboratorních experimentů na externí validitu zcela rezignovala, a tak internetová administrace dotazníků nemohla v tomto směru nijak více uškodit. To se ukázalo jako opodstatněné s nástupem a výzkumem dopadu takových služeb, jako je Amazon Mechanical Turk², kdy se internet zdá být pro výběr vzorku pro psychologická měření poměrně výhodnou alternativou poskytující diverzifikovanější vzorky respondentů (E. M. Buchanan & Scofield, 2018; Casler et al., 2013; Mason & Suri, 2012). To platí zejména v porovnání s takovým tradičním postupem, jako je rekrutování respondentů z řad (amerických) univerzitních studentů. Ekvivalence metod v některých psychologických testech však neznamená ekvivalenci metod administrace napříč všemi měřícími nástroji (T. Buchanan et al., 2005). Je také samozřejmé, že i v případě psychologických měřících nástrojů musí být vzorek dostatečně reprezentativní, hlavně pokud má sloužit pro diagnostické účely.

1.2.2 Kvalita dat a chyba měření

Chyba měření může být v tomto kontextu definována jednoduše jako rozdíl v odpovědi udané respondentem a opravdovou, respektive správnou hodnotou (Couper, 2000). Aby respondent odpověděl správně, prochází podle modelu CASM (*Cognitive Aspects of Survey Methodology*) několika kognitivními procesy (Tourangeau, 1984, 2018). Ty sestávají z porozumění otázce, vybavení si relevantních informací, integrace těchto informací k vytvoření odhadu nebo úsudku a nakonec reportování daného úsudku. Je pochopitelné, že respondent musí během každého z těchto kognitivních procesů vynaložit určité úsilí a být nakonec ochoten své úsudky sdílet.

² Virtuální tržiště umožňující outsourcovat úkoly, například vyplnění dotazníku, mezi registrované pracovníky po celém světě.

Motivace respondenta, ale také způsob administrace (tužka-papír/počítač) by měly být brány v potaz, protože mohou ovlivňovat statistiky jednotlivých položek, jak také uvádí jeden z komentářů ve Standardech pro pedagogické a psychologické testování (standard 4.10; AERA, APA, & NCME, 2014, p. 89). Má se za to, že přítomnost administrátora je v příslušných metodách výhodná třeba právě proto, že je tazatel schopen respondenta motivovat v jeho snahách zodpovědět otázky a dostatečně se na ně soustředit (Couper, 2000). V sebeadministrovaných dotaznících, kterými ty internetové jsou, žádné podobné interakce ani vztah tazatel-respondent neexistuje a respondentovo úsilí tak záleží jen na něm samotném. Pokud jsou nároky na respondenta příliš vysoké a motivace vyvinout úsilí k prožití všech zmíněných kognitivních procesů nízká, přejde podle Krosnicka (1991) respondent k poskytování pouhých uspokojivých odpovědí (*satisficing*). Ty mohou být výsledkem zmíněných čtyřech kognitivních procesů, ale respondent označí nebo napíše odpověď, aniž by nad otázkou a správností odpovědi hlouběji přemýšlel. V těžších formách může respondent některé kroky zcela vynechat. V tomto případě jde zpravidla o dva procesy: vybavení si informací a formování úsudku. Celá ztráta motivace se pak může projevit v ukazatelích kvality dat. Těmi může být vícero odpovědí „nevím“, případně vyšší výběr středních hodnot nebo také úplné přeskočení otázky (*item-nonresponse*). Jde také o nižší diferenciaci odpovědí (*non-differentiation*³), kdy respondent zdaleka nevyužívá celou škálu nabízených odpovědí.

V souladu se zmíněnými procesy popisují Heerwegh a Loosveldt (2008) rozdíl v kvalitě dat mezi webovým dotazníkem a administrací tváří v tvář. Webový dotazník byl zodpovězen rychleji, protože osobní administrace otázek trvá déle a tazatel se může doptávat nebo otázky upřesňovat. I když by podezřele krátký čas bylo možné považovat za indikaci „satisficingu“, v porovnání metod je použití problematické. Co se porovnatelných ukazatelů kvality dat týče, internetový průzkum vedl k 2,6krát vyšší průměrné míře odpovědí „nevím“. Respondenti v internetové metodě také statisticky významně méně diferenciovali své odpovědi, přičemž se navíc výrazně více uchylovali k označení střední možnosti nabízených odpovědí. Míra přeskočených položek byla také významně vyšší u internetových respondentů. Podobné zhoršení kvality dat se ukázalo i při přechodu z osobní administrace tváří v tvář k telefonním interview, při kterém je interakce tváří v tvář odstraněna také (Holbrook et al., 2003).

Na druhou stranu, odstraněním tazatele a přechodem k internetové sebeadminitraci měly být některé závažné zdroje zkreslení dat také redukovány. Jedná se například o sociální žádoucnost (1.3), kdy respondent své odpovědi upravuje tak, aby

³ Někdy také *straight-lining*, protože respondent označuje v každé následující položce odpověď na stejném místě, jako byla jeho předchozí, takže zvolené odpovědi seřazené pod sebou tvoří jeden sloupec (ale občas i jiný pravidelný motiv).

působil lépe, což přítomnost administrátora podněcuje. Eliminována měla být také chyba vzniklá z příbuzného jevu, a to efektu tazatele (1.4), kdy některé charakteristiky tazatele působily na specifické skupiny respondentů, a ty tak opět poskytovaly zkreslené odpovědi. Je také možné, že respondenti jsou v případě internetové sebedministrace pod menším časovým tlakem a mohou tak své odpovědi více promýšlet, pokud tedy sami chtějí.

Standardy pro pedagogické a psychologické testování také uvádějí, že pokud tvůrce testu usoudí, že se podmínky administrace mohou mezi participanty lišit, měl by tento rozdíl popsat a zdokumentovat (standard 4.5; AERA, APA, & NCME, 2014, p. 87). Tato doporučení se sice týkají zejména potřeb diagnostiky, nicméně stále poukazují na jeden z problematických aspektů internetových dotazníků, kde by něco podobného šlo jen velmi obtížně. V případě osobní administrace na podmínky, ve kterých je dotazník zodpovídan, dohlíží administrátor. V případě sebedministrace internetového dotazníku je však celý tento proces plně v režii respondenta. Respondent si tak kromě času vyplnění volí také místo a zařízení, které se může lišit rozlišením nebo kvalitou připojení k internetu. Rušivé okolí, technické problémy nebo problémy se soukromím během vyplňování mohou hrát roli v případné frustraci a ztrátě motivace. Zdá se však, že například odpovědi získané na mobilních telefonech svou kvalitou horší spíše nejsou (Tourangeau et al., 2018).

1.3 Sociální žádoucnost v dotaznících

Sociální žádoucnost, také zvaná sociální desirabilita (*social desirability*), je fenomén, který se s největší pravděpodobností může v určité míře objevovat bez ohledu na metodu administrace dotazníku. Jedná se o typ zkreslení odpovědí (*response bias*), který spočívá v tendenci respondenta poskytovat takové odpovědi, aby vypadal dobře (Paulhus, 1991). Sociální žádoucnost tak může zkreslovat výsledky dotazníkového šetření, protože respondent neodpovídá na položky v tom smyslu, jak byly zamýšleny. Jedná se tedy o jev, který může potenciálně ohrožovat validitu dotazníků (Nederhof, 1985).

Paulhus (1991) identifikoval dva faktory sociální žádoucnosti. Zaprvé jde o pozitivní sebeklamání (*self-deceptive positivity*), kdy se respondent sám vidí v příliš pozitivním světle. Tento typ sociální žádoucnosti je spíše spojen s osobnostními rysy respondenta a změny v metodách administrace dotazníků tak budou v řešení tohoto jevu neefektivní. Zadruhé se však jedná o strategické sebezprezentace (*impression management*). V tomto pojetí mluvíme o sociální žádoucnosti v tradičnějším slova smyslu. Svě odpovědi respondent zkresluje záměrně a skrze tyto úpravy se snaží zlepšit dojem, jakým působí na své okolí. Může tak činit například kvůli obavám z

negativního hodnocení nebo kvůli nedostatečně vnímané anonymitě dotazníku. V každém případě právě zde vystupují do popředí sociální konvence. Holtgraves (2004) zjistil, že respondentům zabraly odpovědi nejdélejší čas v podmínkách, které vyvolávaly největší obavy ze sociální žádoucnosti, a to jak v případě osobnostních rysů, tak chování. Z toho lze usoudit, že zvláště sebe prezentace bude ovlivňovat až poslední z kognitivních procesů při odpovídání (1.2.2), tedy reportování odpovědi. Respondenti totiž projdou všemi procesy a žádoucnost své odpovědi následně zvažují a odpověď upravují⁴, až v posledním kroku, a to nejspíše vědomě. Pokud by byl proces automatický a respondent by některé procesy přeskakoval, čas odpovědi by byl naopak kratší.

Jedná se právě o druhou z variant sociální žádoucnosti, strategickou sebe prezentaci, která je naopak k ovlivnění výběrem metody administrace značně náchylná. Na rozdíl od pozitivního sebeklamání je totiž typicky ovlivnitelná situačními vlivy a přechodnými motivy (Paulhus, 1991). Je tak pravděpodobné, že v případě dotazníku administrovaného tazatelem, by mohl být respondent více náchylný k tomu své odpovědi kontrolovat a upravovat, aby byly v souladu se sociálně žádoucími názory nebo chováním. Pro ilustraci Tourangeau a Smith (1996) například empiricky ověřili, že odpovědi respondentů na citlivá témata se s ohledem na způsob administrace lišily. Dotazníky formou sebeadministrace (CASI, ACASI) vedly k vyšším reportovaným počtům sexuálních partnerů u žen v porovnání s metodou, kdy byly otázky administrovány tazatelem (CAPI). Opačný trend byl zaznamenán u mužů, kteří v případě sebeadministrace reportovali partnerů stejně nebo méně, čímž se původní významný rozdíl zmenšil. Tato zjištění jsou v souladu s přesvědčením o tom, jak sociální normy v tomto citlivém kontextu rozdílně ovlivňují muže a ženy ve vnímání toho, co je pro každé pohlaví sociálně přijatelné (Mitchell et al., 2018).

Výsledek zmíněné studie také demonstruje případ, kdy se sociální žádoucnost stává problémem nejen ve chvíli získávání absolutních odpovědí, při kterých existuje odpověď správná. V tomto případě by se tak jednalo o správný počet sexuálních partnerů. V kontextu normativních odpovědí, kdy je respondentův výsledek porovnáván s ostatními respondenty ve zkoumané populaci, je žádoucnost problémem právě tehdy, když na různé respondenty působí rozdílně (Gordon & Gross, 1978; McCrae & Costa, 1983). V kontextu studie Tourangeau a Smitha (1996) se tedy ženy zařazovaly v počtu sexuálních partnerů níže než muži, zejména v podmínkách metody, při které otázky pokládal tazatel (CAPI). Validita odpovědí by tedy byla ohrožena, kdyby byly závěry vyvozovány bez ohledu na použitou metodu a její možnou interakci s příslušnou skupinou (pohlavím).

⁴ Poslední kognitivní proces, reportování úsudku (Tourangeau, 1984), je podle jiného modelu (Sudman et al., 1996) právě úprava odpovědi (*response editing*).

Je na druhou stranu nutné podotknout, že právě interakce mezi pohlavím a metodou administrace v této studii, ač zjevná, sama o sobě statisticky signifikantní nebyla. Signifikantní se interakce stala ve chvíli, kdy do ní byla zahrnuta i forma otázek, které položkám ohledně počtu partnerů předcházely, a měly za cíl ovlivnit respondentovu sebe prezentaci. Výsledky jsou i přesto konzistentní s poznatky o sebe prezentaci a sociální žádoucnosti. Self-reportové metody poskytují vyšší anonymitu a v tomto duchu se nesou i ostatní výsledky studie. Větší podíl participantů byl ochoten přiznat užívání drog nebo některé sexuální praktiky právě v případě sebeadministrace.

Nepřekvapí, že obdobné výsledky vykazuje i sebeadministrace dotazníku skrze internet. To například platí v případě srovnání s telefonním interview, kdy byli stejní respondenti ve stejných otázkách ochotnější odpovědět (projevili vyšší *disclosure*) v případě internetové administrace (Milton et al., 2017). Tento efekt metody administrace však záležel na vnímané citlivosti otázek. Nejčastěji tak existoval statisticky významný rozdíl v odpovědích u položek, které respondenti hodnotili jako velmi citlivé (5 z 8 citlivých položek), zatímco nulový rozdíl v odpovědích způsobovaly položky hodnocené jako neutrální (0 ze 4 neutrálních položek). Zhang et al. (2017) provedli srovnání metod se zaměřením na pozitivní (např.: self-efficacy) i negativní (např.: depresivita) psychologické škály. V porovnání s internetovou administrací se ve 12 z 13 škálách ukázal vyšší průměrný skóre v těch pozitivních a nižší v těch negativních. To jak v případě telefonního interview, tak osobního rozhovoru. Jedinou škálu neovlivněnou metodou administrace považují autoři za nejvíc neutrální v otázkách sociální žádoucnosti. To znamená, že efekt sociální žádoucnosti se tak nejspíše neprojevuje stejně ve všech dotaznících a u všech položek, ale záleží na citlivosti jejich obsahu (pro přehled: Tourangeau & Yan, 2007). Navíc, i vnímaná citlivost položky respondentem se může lišit s ohledem na metodu administrace (Kreuter et al., 2008).

1.4 Efekt tazatele

Odpovědi může zkreslovat i efekt tazatele (*interviewer effect*) nebo chyba způsobená tazatelem (*interviewer error*). Tuto problematiku se nabízelo zkoumat zejména v kontextu osobních rozhovorů. Při výzkumech využívajících této metody se ukazovalo, že tazatelé, kterých bylo více, získávali odpovědi, jejichž rozptyl se napříč jednotlivými tazateli lišil (West & Blom, 2017; West & Olson, 2010). Odpovědi respondentů dotazovaných jedním tazatelem jsou si tak podobnější než odpovědi ostatních respondentů. Jinými slovy, chyba způsobená tazatelem znamená, že data získaná specifickým tazatelem nebo specifickou skupinou tazatelů se liší od dat získaných jiným tazatelem nebo skupinou tazatelů, kteří administrují stejný dotazník stejné populaci (Davis et al., 2010). To znamená, že některé charakteristiky tazatele zřejmě respondenty jistým

způsobem ovlivňují. Kish (1962) tento vliv statisticky vyjádřil pomocí koeficientu vnitřní korelace (ICC, *intra-class-correlation coefficient*), kterou lze v tomto specifickém kontextu také pojmenovat jinými slovy jako *intra-interviewer* korelaci (West & Olson, 2010). Je zřejmé, že zmiňovanou chybu měření lze statisticky kontrolovat víceúrovňovými modely, které rozdělení respondentů podle tazatelů zohledňují (Davis et al., 2010). Pro odhad efektu tazatele tak studie přidělování respondentů k administrátorům randomizují (*interpenetrated design*; Biemer & Stokes, 1985). Je však nutné podotknout, že využití dostatečně vysokého počtu administrátorů není mnohdy z praktických důvodů možné. Problematickým by se tak efekt tazatele mohl stát zejména ve chvíli, kdy by jediný tazatel administroval dotazník nepoměrně velké části respondentů (Holbrook et al., 2019).

Efekt tazatele je v rámci metodologie průzkumů komplexní problematikou týkající se i takových náležitostí, jako je oslovení a získávání potenciálních respondentů, získávání si jejich přízně a dalších činností tazatele, které ve výsledku přispívají například k chybě vzniklé systematickým odmítáním odpovídat (unit-nonresponse error; pro přehled: West & Blom, 2017). Efekt tazatele a variabilita, kterou svou přítomností vnáší, jsou ale nevyhnutelně spojeny přímo s administrací otázek a procesy odpovídání na ně. Jak poznamenávají Davis et al. (2010), významnost efektu tazatele se může lišit i s použitou metodou administrace. Jako konkrétní příklad uvádějí rozdíl mezi telefonním rozhovorem a rozhovorem tváří v tvář. V druhém případě je totiž respondent vystaven kromě hlasových charakteristik administrátora i charakteristikám vizuálním. V tomto ohledu se dá podotknout, že tyto vlivy jsou v naprostém souladu s poznatkem Paulhuse (1991), a to že sebezprezentace je ovlivněna situací a přechodnými motivy. To mimo jiné implikuje, že se koncepty sociální žádoucnosti a efektu tazatele do jisté míry překrývají, nebo přinejmenším doplňují, a jsou si tak v některých případech velice blízké, nebo dokonce identické. Efekt tazatele se tak zdá být zvláštním případem sociální žádoucnosti, ale nemusí tomu tak být vždy.

Mezi nejdůležitější charakteristiky tazatele, které mohou ovlivňovat respondentovy odpovědi, a tím přispívat k chybě měření, lze zařadit třeba administrátorovo chování. Nápadné jsou však především fyzické charakteristiky administrátora, jako je rasa nebo etnikum, věk, pohlaví nebo vzhled. Jedno z možných vysvětlení efektu tazatele se pojí se sociálně atribučním modelem (*social attribution model*; Johnson et al., 2000). V tomto případě respondent vnímá charakteristiky tazatele a přiřazuje k nim své domněnky o jeho postojích a názorech. Následně pak upravuje své odpovědi takovým stylem, aby byly s těmito „atribucemi“ kongruentní a s tazatelem si tak vytvořil a udržel pozitivní vztah. Podobnost se sociální žádoucností je zde natolik zjevná, že není ani autory ponechána bez povšimnutí (Holbrook et al., 2019). Podle Johnsona et al. (2000) se druhé možné vysvětlení nachází v modelu sociální vzdálenosti (*social distance model*; Williams, 1964). Tento model také počítá s atribucemi vycházejícími

z charakteristik tazatele. Vyzdvihuje však, že respondenti mohou mít problém důvěřovat, komunikovat a sdílet informace s lidmi, kteří se jim v nějakém ohledu nepodobají a jsou jim tak „sociálně vzdálení“. Čím menší je shoda ve vnímané sociální identitě tazatele a respondenta, tím větší je tato vzdálenost, a tím i vyšší a vlivnější efekt tazatele. V tomto ohledu se nabízí i možná aktivace stereotypů.

Administrátorova rasa je jedním z nejnápadnějších fyzických znaků. Navíc, pokud je společnost někde rozdělená, jedná se o problematiku rasovou, v rámci které je nedůvěra mezi skupinami charakteristická (Dovidio et al., 2002, 2011). V těchto případech není ani příliš možností k vzájemnému pozitivnímu kontaktu, potenciálně oslabujícím tuto nedůvěru a vzniklé stereotypy (např.: Allport, 1954; Pettigrew, 1998; Pettigrew & Tropp, 2008). V kontextu toho se nabízí, že by efekt tazatele mohl být obzvláště zjevný právě při kontaktu respondenta a tazatele odlišné rasy nebo etnika. Přehledové studie se v tomto kontextu shodují ve svých zjištěních na jediném. Výsledky studií na vliv tazatelovy rasy nejsou jednoznačné a jsou moderovány rasou respondenta, relevancí rasového obsahu otázek a citlivostí jejich obsahu (Davis et al., 2010; Holbrook et al., 2019; West & Blom, 2017). Například West a Blom (2017) uvádějí, že v 17 z 18 studií v rasově citlivých otázkách poskytovali respondenti spíše takové odpovědi, aby neurazili rasu tazatele. V případě otázek, které citlivé nebyly, se efekt projevil už pouze v 9 z 16 studií, ale přesto existoval.

I jiný zjevný znak administrátora, jeho pohlaví, produkuje smíšené výsledky (West & Blom, 2017). V 10 studiích byl vliv pohlaví tazatele nulový. V 9 studiích získaly kvalitnější data ženy, kdežto ve 4 studiích muži. West a Blom (2017) podotýkají, že i zde může být efekt ovlivněn jak úctou vůči danému pohlaví, tak aktivací stereotypů. Zdá se například, že mužský tazatel může podněcovat u respondentů vyšší tendenci k souhlasu (*acquiescence*; Liu & Wang, 2016), zatímco tazatelky mohou obdržet více sociálně žádoucích odpovědí (Liu & Stainback, 2013). Efekt byl patrný také v případě interakce pohlaví, kdy mužští respondenti reportovali méně sexuálních partnerů tazatelkám než tazatelům (Wilson et al., 2002).

1.5 Humanizující elementy v dotaznících

Předchozí kapitoly popisovaly rozdíly, výhody a nevýhody seadministrovaných (internetových) dotazníků a dotazníků administrovaných tazatelem osobně, nebo skrze telefonní hovor. Technologický pokrok ovšem umožnil kombinovat prvky obou metod. V rámci průzkumů tak vznikly například metody automatizovaného sběru odpovědí administrované audionahrávkami (ACASI) nebo videonahrávkami (AVCASI; 1.1), potenciálně vhodné i pro respondenty s omezenou schopností číst (O'Reilly et al., 1994).

Jak ale lidské znaky ovlivňují respondenta? Stačí pouhá integrace lidských znaků k vyvolání efektu tazatele nebo posílení sociální žádoucnosti? Nebo naopak domnělá lidská přítomnost respondenta lépe zaujme a data jsou tak kvalitnější?

Některé indicie ve prospěch metod sebeadministrace, které používají lidské znaky, vyplynuly docela nezáměrně. Efekt administrace hlasovými nahrávkami tak podotýkají i Tourangeau a Smith (1996) v již zmíněné studii (1.3). Zatímco v porovnání osobním rozhovorem metoda CASI (pouze počítač) zvýšila průměrný počet sexuálních partnerů, které reportovaly ženy za posledních 5 let 1,16krát, metoda ACASI administrovaná audionahrávkami až 1,88krát. Větší vliv metody využívající zvukové nahrávky byl pozorován i v několika dalších sledovaných proměnných, jako například ochotě přiznat užívání drog (ACASI: 2,84krát; CASI: 1,37krát). Pokud se ovšem studie na tuto otázku zaměřila konkrétně, podobný efekt se už replikovat nepodařilo (Couper et al., 2003). Roli zřejmě nehraje ani vnímané soukromí, které bylo experimentálně manipulováno.

Už v 90. letech začali výzkumníci v oblasti interakce člověka s počítačem experimentovat se zahrnutím lidských elementů do uživatelského rozhraní, aby tak učinili tuto interakci přirozenější a příjemnější (Sproull et al., 1996; Walker et al., 1994). Vycházeli tak z vlivných poznatků ze sociální psychologie. Šlo například o vyšší výkonnost (u snadných úkolů) v přítomnosti druhých (sociální facilitace; Zajonc, 1965) nebo roli lidských obličejů v sociální interakci (např.: Ekman, 1982). Ve studiích se nakonec ukázalo, že v porovnání s textem byly odpovědi respondentů delší a přesnější, pokud jim byly otázky pokládány animovaným 3D modelem lidského obličeje (Walker et al., 1994). Respondenti se také prezentovali v lepším světle a cítili se méně uvolnění, což dle autorů mohlo znamenat, že se více soustředí (Sproull et al., 1996). Zdálo se také, že i pouhý ženský, popřípadě mužský hlas v počítači je schopný aktivovat stereotypy o daném pohlaví (Nass et al., 1997).

Tourangeau et al. (2003) se pokusili přenést tyto poznatky do praxe internetových dotazníků. V sérii experimentů měnili autoři míru „personalizace“ dotazníku s položkami ohledně genderových rolí, nežádoucího chování nebo sociální žádoucnosti – BIDR (Paulhus, 1984) i Marlow-Crowne (Crowne & Marlowe, 1964). Webový dotazník obsahoval buďto obrázek výzkumníka muže, ženy nebo neutrální logo studie. V jedné z podmínek navíc dotazník oslovoval respondenta osobněji, oproti druhé neutrální podmínce. Tato studie nicméně prokázala jen velmi slabý efekt. Objevila se například slabá tendence respondentů odpovídat na otázky více pro-feministicky v případě, kdy byli vystaveni obrázku ženy. Respondenti také hodnotili personalizovanější verze dotazníků více jako rozhovor s kamarádem a méně jako odpovídání stroji. I sami autoři nicméně podotýkají, že jejich manipulace nemusela být dostatečná. Je zjevné, že pouhý

obrázek výzkumníka je velmi odlišný od přímé administrace otázek virtuálním tazatelem, jako je popsáno výše, nebo dokonce videozáznamem reálného člověka.

Videozáznam administrátora však v laboratorní studii použili Krysan & Couper (2003), a to k prozkoumání efektu tazatele způsobeného jeho rasou. V tomto případě nalezla studie pro efekt tazatele a interakci s metodou podporu. Autoři uvádějí, že bílí respondenti tázaní afroamerickým virtuálním administrátorem měli vyšší tendenci k rasovým stereotypům nebo popírání rasové diskriminace, oproti podmínce s bílým virtuálním tazatelem. Tento jev ale nenastal v případě osobního rozhovoru s černošským tazatelem. Byla tak zřejmě vidět interakce metody i rasy, kdy rasový stimul na videu (afroamerický tazatel) mohl svou pouhou přítomností, ač neosobní, aktivovat negativní rasové postoje (už např.: Devine, 1989), které však v případě osobního rozhovoru potlačila sociální žádoucnost.

1.5.1 Administrace dotazníků videem oproti textové podobě

V následujících odstavcích je shrnuto několik studií, které jsou dílčí inspirací pro tuto práci. Dopředu je dobré upozornit na jejich metodologickou různorodost a nekonklusivní výsledky, které mohou být dány i dobou jejich vzniku a dostupnými technologiemi. Také je potřeba upozornit, že jsou popsány všechny mně známé dosavadní studie na toto téma, kterých je tak dohromady pouze pět.

Fuchs a Funke (2007) nejspíše jako první provedli srovnání administrace skrze videozáznam a textového internetového dotazníku ($N = 459$). Obě podoby dotazníku zobrazovaly položky jednu po druhé, aby byly podmínky co nejjednodušší. V otázkách sociální žádoucnosti ovšem studie nepřinesla signifikantní výsledky. Jediným poměrně překvapivým efektem byla nižší vnímaná sociální přítomnost v podmínce videa. V případě administrace videem také více respondentů studii nedokončilo (*break-off*), nebo ji ani nespustilo (*unit-nonresponse*). To ale mohlo být z velké části způsobené technickými problémy, protože ne všechna zařízení byla dříve schopna videa přehrát. Také sociální žádoucnost byla měřena jen čtyřmi dichotomickými položkami.

Gerich (2008) laboratorně porovnal administraci dotazníku osobně, textem (tužka-papír), videem (AVCASI), audiozáznamem (ACASI). Videozáznam byl v tomto případě doplněn i o titulky. Studie ($N = 189$) sledovala sociální žádoucnost (část Marlow-Crownovy škály), odklon od norem (riskantní chování) a sebekontrolu. Jejich kombinace, tedy nižší žádoucnost, nižší sebekontrola a vyšší odklon od norem měl znamenat vyšší ochotu svěřit se (*self-disclosure*). Touto operacionalizací měla administrace s použitím videa *self-disclosure* dokonce nejvyšší. To tedy přinejmenším přináší určitou podporu pro tvrzení, že vyšší sociální žádoucnost u administrace videem nemusí

být problémem. Autoři ale poznamenávají, že existovaly určité indicie o efektu tazatele (jeho pohlaví).

Fuchs (2009) se zaměřil přímo na efekt tazatelova pohlaví v internetovém dotazníku administrovaném videem a srovnal tento efekt s textovou verzí ($N = 880$). Muži například připustili zkušenost s pohlavní nemocí spíše muži, zatímco ženy spíše ženě. Stejný vzorec se objevil u položky týkající se nevhodného sexuálního chování. Trend se otočil v případě přiznání panictví, s čímž se muži svěřovali méně jinému muži. Z výsledků se zdá, že respondenti své odpovědi upravují s ohledem na sociální žádoucnost problému. Výsledek textové verze pak koresponduje s tímto efektem sociální žádoucnosti. Je například méně přípustné přiznat pohlavní chorobu opačnému pohlaví (2 %), takže výsledek textové verze (9 %) se spíše blíží podmínce komunikace se stejným pohlavím (7 %). Zejména je však patrné, že efekt tazatelova pohlaví se projevuje i v případě administrace skrze videozáznam, a to v podobném duchu jako při osobní administraci, nicméně jde zřejmě o velmi citlivá témata, která se v psychologickém testování objevují spíše výjimečně.

V aktuálnější studii pak Haan et al. (2017) srovnávali webovou administraci textem, videem, ale i telefonické a osobní interview ($N = 536$) v rámci sociologických dotazníků. Sociální žádoucnost byla měřena jako rozdíl v počtu odpovědí na otázky potenciálně citlivé (genderová rovnost, příjem, náboženství). Rozdíl mezi administrací videem a textem nebyl nijak významný. S ohledem na kvalitu dat byla sledována tendence k pouze uspokojivému odpovídání (*satisficing*). Jednalo se o efekt primárnosti (volba prvních možností odpovědi), efekt novosti (*recency*; volba posledních možností) a počet odpovědí nevím. Ani v těchto ukazatelích se však metody významně nelišily. Autoři tak podotýkají, že ani dynamické lidské elementy nejsou dostatečné k tomu, aby respondentovu pozornost více zaujaly, nebo ovlivňovaly odpovědi sociální žádoucností. V sociální žádoucnosti se nicméně administrace videem lišila od osobního a telefonního interview.

Conrad et al. (2022) provedli k mé dosavadní znalosti poslední, a tak i nejaktuálnější studii porovnávající metody internetové administrace ($N = 1066$). V tomto případě šlo o administraci pomocí videí, textu, ale také živého interview skrze videohovor. Stejně jako Haan et al. (2017), využívají i zde autoři různé položky ze sociologických dotazníků. „Satisficing“ byl sledován prevalencí zaokrouhlených číselných dopovědí udaných respondentem dělitelných beze zbytku pětkou. Dále také tendencí nediferencovat odpovědi, tedy zvolením odpovědi na stejné pozici ve všech položkách dané části dotazníku. Ochotu svěřit se (*disclosure*) studie sledovala jako vyšší počet stigmatizujících nebo potenciálně ztrapňujících odpovědí na senzitivní otázky. Čas strávený vyplňováním měl být ukazatelem zátěže kladené na respondenta. Autoři se také zaměřili na respondentovu zkušenost s použitou metodou. Administrace videozáznamem vedla

k nižší tendenci zaokrouhlovat číselné odpovědi, a to jak v porovnání s textovou verzí, tak videohovorem, kde zaokrouhlování probíhalo nejvíce. To podporuje například domněnku, že lidé jsou vystaveni menšímu tlaku odpovědět rychle v případě sebadministrace. Především však autoři poznamenávají, že respondenti mohou být přítomností virtuálního tazatele motivováni k vyšší snaze vybavit si informace potřebné k odpovědi. Videozáznam také vedl k vyšší ochotě svěřit se u některých položek, ale nikoli k významně nižší tendenci nediferencovat odpovědi v porovnání s textovou verzí. Rozdíl ve spokojenosti mezi dotazníkem s videem a textovou verzí nebyl významný, a to i přes to, že textová verze zabrala respondentům nejkratší dobu ($M = 7,8$ minut), zatímco verze s videem nejdelší ($M = 12,43$ minut).

1.6 Tato práce

Primárním cílem této práce je mezisubjektové experimentální srovnání internetového dotazníku, který bude respondentovi administrován v textové podobě, a internetového dotazníku, který bude administrován formou krátkých videí. Tato práce tak volně navazuje na přechodí pokusy o srovnání forem administrace (1.5.1). V kontextu těchto výzkumů je nutné podotknout, že studií není mnoho a jsou metodologicky poměrně různorodé. Navíc od doby provedení některých z nich uplynulo i více než 10 let. Vzhledem k rychlému technologickému pokroku, jakým internet, multimédia a všechny typy elektronických zařízení v poslední době prošly, se jedná o příliš dlouhou dobu na vyvozování závěrů o dnešních respondentech. Těch pár studií, které lze považovat za aktuálnější (Conrad et al., 2022; Haan et al., 2017), se zase zabývá výhradně sociologickými dotazníky a nezkoumá škály zaměřené na konkrétní konstrukty, které často využívá psychologie.

Tato práce využije také psychologické škály, u kterých jsou známé a zjistitelné jejich psychometrické vlastnosti. Ekvivalenci metod administrace tak bude možné ověřit i pomocí strukturního modelování (SEM; *structural equation modeling*) a z něj vycházející analýzy invariance měření (*measurement invariance*). Tyto metody budou s největší pravděpodobností pro srovnání „videodotazníku“ (dále také VD) a internetového textového dotazníku (dále také TD) použity vůbec poprvé. Jako jeden z klasických zástupců psychologických škál byla vybrána Rosenbergova škála sebehodnocení (Rosenberg, 1965). Využití dotazníku výšky (Rečka, 2018; Tancoš, 2019), jakožto druhé ze škál, pak umožní srovnat získané skóry s objektivním kritériem, jakým je výška udaná respondenty, a tím vyjádřit a porovnat kriteriální validitu metod.

Z některých studií vyplývá, že se i v případě videí může projevovat sociální žádoucnost a efekt tazatele (Fuchs, 2009; Krysan & Couper, 2003). Zřejmě se jedná

zejména o otázky na citlivá témata, jako jsou například sexuální zkušenosti, nemoci nebo rasové postoje. Aby bylo možné efekt tazatele lépe sledovat, bude dotazník administrován v každé ze dvou experimentálních podmínek (VD) jiným administrátorem. Části respondentů tak budou položky administrovány ženou (dále také VD_ž) a jiné části zase mužem (dále také VD_M). Bude tak možné pozorovat i rozdíly, jaké může produkovat interakce pohlaví, kterou ilustruje například studie Fuchse (2009). I když je možné, že se efekt projeví i v samotných škálách, byla pro tento účel do dotazníku zahrnuta položka ohledně fyzické konstituce respondenta. Problematiku tělesné váhy lze považovat za potenciálně ovlivnitelnou sociální žádoucností, aniž by však respondenty vystavovala tématům na hranici výzkumné etiky.

Zároveň budou sledovány i jiné ukazatele, které by od sebe mohly metody administrace odlišovat. Jedná se o čas strávený vyplňováním, čas na každé položce, počet nedokončených průchodů (*break-off*), subjektivně vnímanou, ale i objektivní soustředěnost respondenta na dotazník, jeho spokojenost s danou metodou a technické problémy, na které mohl narazit. Sledován bude i možný vliv pohlaví, ale také respondentův postoj ke čtení, který může vést k preferenci videa oproti textu.

Kvůli velmi limitovanému počtu studií je obtížné stanovit jakékoli silné hypotézy. Tato práce tak má převážně explorativní charakter. Na základě indicií z předešlých výzkumů je hypotézy přesto možné odhadnout, nicméně jde spíše o očekávání spojená s touto prací⁵.

H1a: Bude dosaženo skalární invariance (silné faktorové invariance) mezi dotazníkem administrovaným videem (VD) a textovou verzí (TD) dotazníku výšky u žen.

H1b: Bude dosaženo skalární invariance (silné faktorové invariance) mezi dotazníkem administrovaným videem (VD) a textovou verzí (TD) dotazníku výšky u mužů.

H2: Bude dosaženo skalární invariance (silné faktorové invariance) mezi dotazníkem administrovaným videem (VD) a textovou verzí (TD) pro Rosenbergovu škálu sebehodnocení.

H3: VD bude oproti TD respondenty považován za zábavnější.

H4: Subjektivní pozornost věnovaná dotazníku bude vyšší ve VD oproti TD.

H5: VD bude mít vyšší počet nedokončených průchodů oproti TD.

H6: VD budou lépe hodnotit respondenti, kteří neradi čtou.

H7: Doba strávená vyplňováním bude vyšší ve VD oproti TD.

H8a: Rozložení ženami reportované figury se bude mezi VD a TD lišit.

H8b: Rozložení ženami reportované figury se bude mezi VD_M a VD_ž lišit.

⁵ Např. Rubin & Donkin (2022) doporučují rozlišovat podobné predikce od hypotéz s hlubším teoretickým podkladem.

2 Metoda

2.1 Vzorek

2.1.1 Odhad velikosti vzorku

Pravděpodobně nejnáročnější metodou na velikost vzorku, která má být pro analýzu dat v této studii využita, je analýza invariance měření. Odhad velikosti vzorku komplikuje fakt, že síla testů v rámci strukturního modelování závisí nejen na velikosti vzorku, ale také na komunalitách, případně počtu položek na faktor, a to jak v případě exploratorní faktorové analýzy, tak jednoskupinové konfirmační faktorové analýzy (Meade & Bauer, 2007). Protože analýza invariance je zde vlastně víceskupinovou konfirmační faktorovou analýzou, stejné problémy se přenáší i do tohoto odhadu. Na základě simulací, které autoři provedli, se zdá, že vzorky o $n = 100$ (na skupinu) nejsou pro potřeby analýzy invariance zcela uspokojivé, zatímco u vzorků o $n = 200$ (na skupinu) už záleží na dalších psychometrických vlastnostech. Protože byly pro potřeby této studie záměrně vybrány škály s jednoduchou faktorovou strukturou a obecně dobrými psychometrickými vlastnostmi, za dostatečnou velikost vzorku lze považovat už zhruba $N = 400$ (ve všech podmínkách dohromady), ideálně však alespoň $N = 500$. Je krajně nepravděpodobné, že by ostatní analytické postupy použité v této práci vyžadovaly vzorek větší.

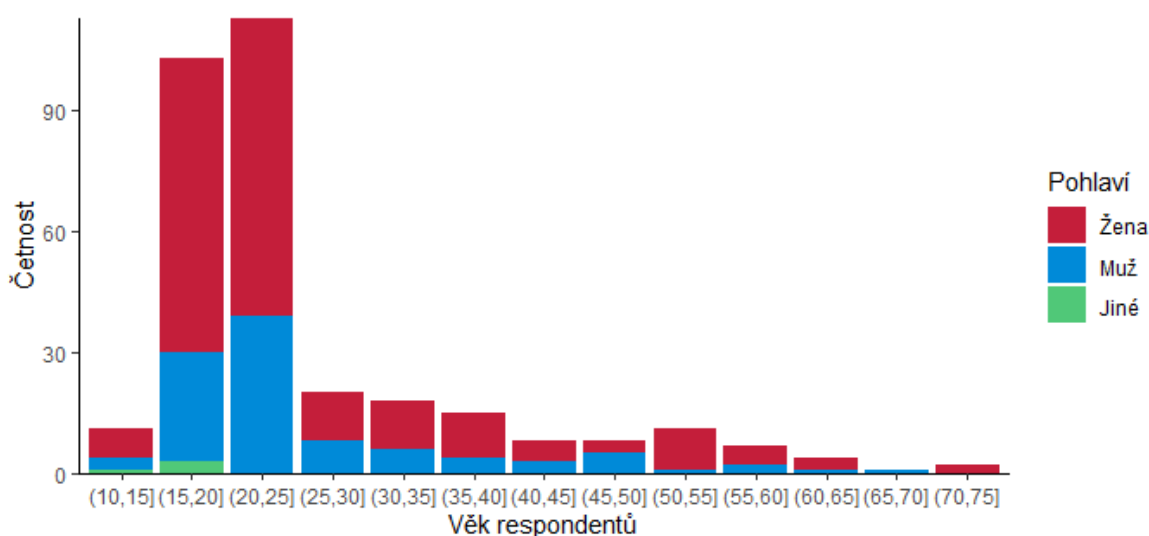
2.1.2 Základní charakteristiky výzkumného souboru

Dotazník byl sdílen primárně skrze sociální síť. Jednalo se převážně o veřejné skupiny na síti Facebook, ale také sdílení na síti Instagram nebo Reddit. Sběr dat probíhal od 11. 3. do 7. 4. 2023 a dotazník zaznamenal za tuto dobu celkem 389 průchodů. Po odstranění zejména těch průchodů, které byly ukončené před rozdělením do experimentální/kontrolní skupiny nebo byly pochybné kvality (blíže v 3.1), bylo do analýz zahrnuto $N = 321$ respondentů, z nichž 67,6 % ($n = 217$) tvořily ženy a 31,2 % ($n = 100$) muži. Další 4 respondenti (1,2 %) se označili jako „Jiné“. Drtivou většinu respondentů tvořili mladí lidé mezi 15 a 25 lety (70,7 %), což šlo předpokládat z povahy šíření dotazníku (Graf 1). Dolní hranicí věku pro vyplnění bylo 15 let. Průměrný věk respondentů činil 26,31 let ($SD = 12$; $Mdn = 22$), u žen $M = 26,21$ ($SD = 12,31$; $Mdn = 22$) a u mužů $M = 26,91$ ($SD = 11,4$; $Mdn = 23$). Značné množství, tedy 248 respondentů (77,3 %), vyplnilo dotazník na mobilním zařízení, zatímco jen 73 (22,7 %) na stolních počítačích.

2.1.3 Rozdělení do experimentálních podmínek

Ze vzorku $N = 321$ respondentů bylo $n = 157$ (48,9 %) přiděleno k experimentální podmínce, dotazníku administrovanému skrze video (VD), z toho $n = 80$ k mužskému administrátorovi (VD_M; 51,0 %) a $n = 77$ k ženské administrátorce (VD_ž; 49,0 %; Tabulka 1). Věk respondentů se mezi jednotlivými metodami administrace významně nelišil.

Graf 1: Rozložení věku a pohlaví respondentů



Tabulka 1: Rozdělení respondentů do experimentálních podmínek podle pohlaví

Verze	Pohlaví respondenta						Celkem	
	Žena		Muž		Jiné			
	n	Rel. Četnost	n	Rel. Četnost	n	Rel. Četnost	n	Rel. Četnost
TD	115	53,0%	47	47,0%	2	50,0%	164	51,1%
VD _M	51	23,5%	29	29,0%	0	0%	80	24,9%
VD _ž	51	23,5%	24	24,0%	2	50,0%	77	24,0%
Σ	217	100%	100	100%	4	100%	321	100%

2.2 Měřicí nástroj

2.2.1 Funkcionalita dotazníku

Internetový dotazník využitý v tomto výzkumu byl vytvořen v programu Qualtrics a do značné míry přizpůsoben skrze HTML, JavaScript a CSS. Kromě běžných textových položek obsahuje dotazník položky, které předčítá administrátor formou krátkých videí (Obr. 1). Protože se jedná o studii zaměřenou na rozdíly v metodách administrace, zvláštní pozornost byla kladena na co nejjednodušší podmínky pro všechny respondenty. Využití videozáznamů pro administraci například přináší poměrně specifické problémy ve standardizaci zobrazení a ovládání pro různá zařízení a internetové prohlížeče. Šlo také velmi dobře předpokládat, že alespoň určitá, a zřejmě také podstatná část respondentů bude tento dotazník vyplňovat na mobilních zařízeních. Zejména mladší respondenti, ale zdaleka nejen ti, mohou být navíc zvyklí na některé aspekty přehrávačů krátkých videí (Instagram Reels, YouTube Shorts, TikTok apod.). Velké odlišnosti by mohly být matoucí, a tak zbytečně navyšovat procento nedokončených průchodů nebo narušovat soustředěnost na položky.

Aby bylo zobrazení videa co nejjednodušší napříč zařízeními, byly zcela odebrány ovládací prvky videa, jako ovládání hlasitosti nebo posuvník. Stejně tak bylo zakázáno přehrávání v režimu celé obrazovky, zvláště kvůli tomu automatickému na mobilních zařízeních, nebo funkce obraz v obraze (*PiP*). Byla také zajištěna plná responzivita videí, aby vždy co nejlépe vyplnila obrazovku použitého zařízení a administrátor byl bez obtíží viditelný. To také znamená, že byly položky zobrazovány vždy po jedné (jak ve VD, tak i v TD; pro srovnání viz Obr. 2). Pro co nejintuitivnější ovládání byla přidána vlastní grafika překrývající video, která instruuje respondenta k jeho přehrání, a která je zobrazena vždy, když je video pozastaveno, přehrávání skončilo, nebo ještě nezačalo. Vzhledem k délce videí, která někdy nepřekročila ani tři sekundy, byla videa také defaultně nastavena na automatické přehrání, což mělo formát co nejvíce přiblížit chování výše zmíněných platforem a ušetřit respondenta zbytečné přítěže⁶.

Diagram (Obr. 3) zobrazuje kompletní průchod dotazníkem. Velká část položek dotazníku, včetně videí, umožňuje shromažďovat data o respondentově chování v dotazníku, ale také metadata o použitém zařízení, například pro kontrolu možného vlivu typu zařízení na předčasné ukončení z důvodu nekompatibility. Ve všech podmínkách je sledován čas strávený na položce (*Timer*), aby bylo možné porovnat reakční časy a dobu vyplňování. Experimentální podmínky (VD) navíc na straně uživatele umožňují sledovat i počet přehrání každého z videí, konkrétně začátek (*Start*), pozastavení

⁶ Chování této funkce se může drobně lišit s konfigurací zařízení respondenta (operační systém, prohlížeč apod.). V některých případech je například nutné spustit první video manuálně.

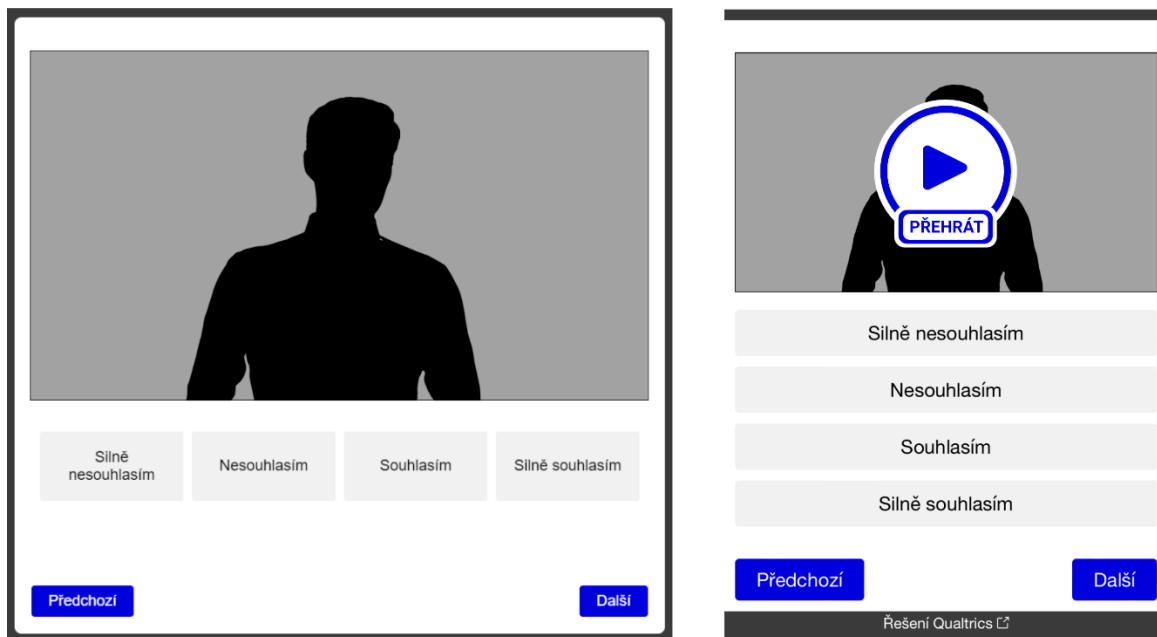
(*Pause*) a konec (*Ended*), aby bylo možné pozorovat, zda respondenti potřebovali (nebo chtěli) otázku zopakovat vícekrát. Dotazník také na straně uživatele sleduje, zda respondent v době experimentu přepíná do jiných oken nebo aplikací (*Počet přepnutí okna A/B*), což umožňuje objektivně sledovat soustředěnost na danou metodu.

2.2.2 Začátek dotazníku a rozdělení do skupin

Před spuštěním dotazníku je respondentům zobrazen informovaný souhlas s účastí ve výzkumu, který popisuje i účel výzkumu a nastiňuje obsah použitých položek a škál. Mimo informace o dobrovolné účasti, anonymitě a odhadovaného času trvání je zahrnuto i upozornění na možný audiovizuální obsah a to, že položky vyžadují pro své zodpovězení nejen obraz, ale i poslech zvuku. Tato výstraha se tak zobrazuje všem respondentům navzdory tomu, že polovina z nich je přiřazena ke kontrolní skupině s klasickým textovým dotazníkem (TD). Tento krok má zabránit otevření dotazníku respondentem v situacích nedovolujících poslech zvuku, ale také snížit nebezpečí výběrové chyby v případě, že by se lidé ochotní vyplnit textový dotazník systematicky lišili od lidí ochotných vyplnit dotazník s videi. V případě souhlasu jsou pak respondenti vyzváni k zodpovězení některých základních údajů, jako je věk, pohlaví, ale také výška v centimetrech.

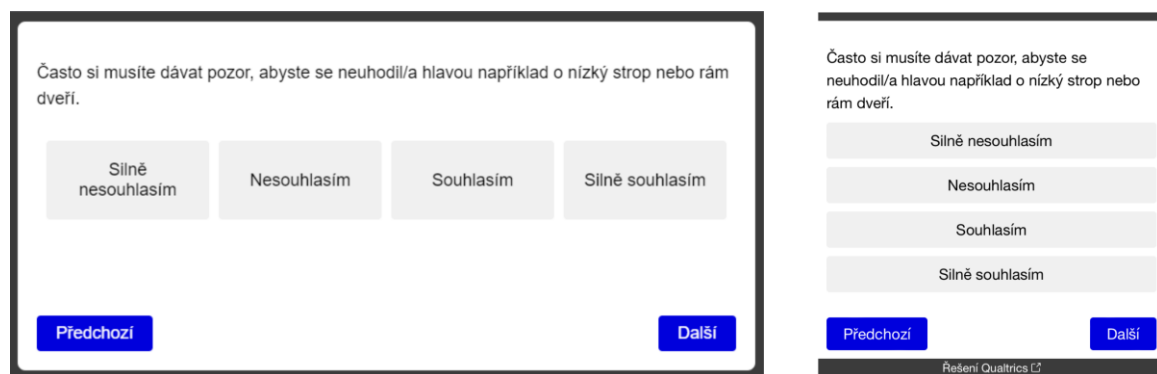
Následně dochází k náhodnému rozdělení respondentů do tří skupin. Dvou experimentálních podmínek: dotazníku formou videí administrovaným mužem (VD_M), ženou (VD_Ž) a jedné kontrolní skupiny (textový dotazník; TD). Respondenti jsou rozděleni v poměru 1:1:2 (VD_M:VD_Ž:TD). Ve všech podmínkách byli respondenti v první poloze po rozdělení instruováni ohledně ovládání dotazníku ve formátu odpovídajícímu přidělené podmínce. Respondenti ve VD tak zhlédli instruktážní video prezentované příslušným administrátorem, zatímco respondenti přiřazení k TD obdrželi návod textový. Všichni respondenti byli upozorněni na automatické posouvání dotazníku dále po výběru odpovědi, ale také na možnost vracet se nebo přeskočit položky, které si nepřejí zodpovědět. Na začátku každé experimentální podmínky (VD) je respondentům krátce vysvětleno, že položky jsou administrovány formou videí. Respondenti jsou navíc upozorněni na možnost videa přehrávat opakovaně. V případě VD se také kromě instrukcí administrátoři představují jménem, aby byl kontakt více osobní.

Obr. 1: Položka dotazníku v experimentální podmínce (VD)



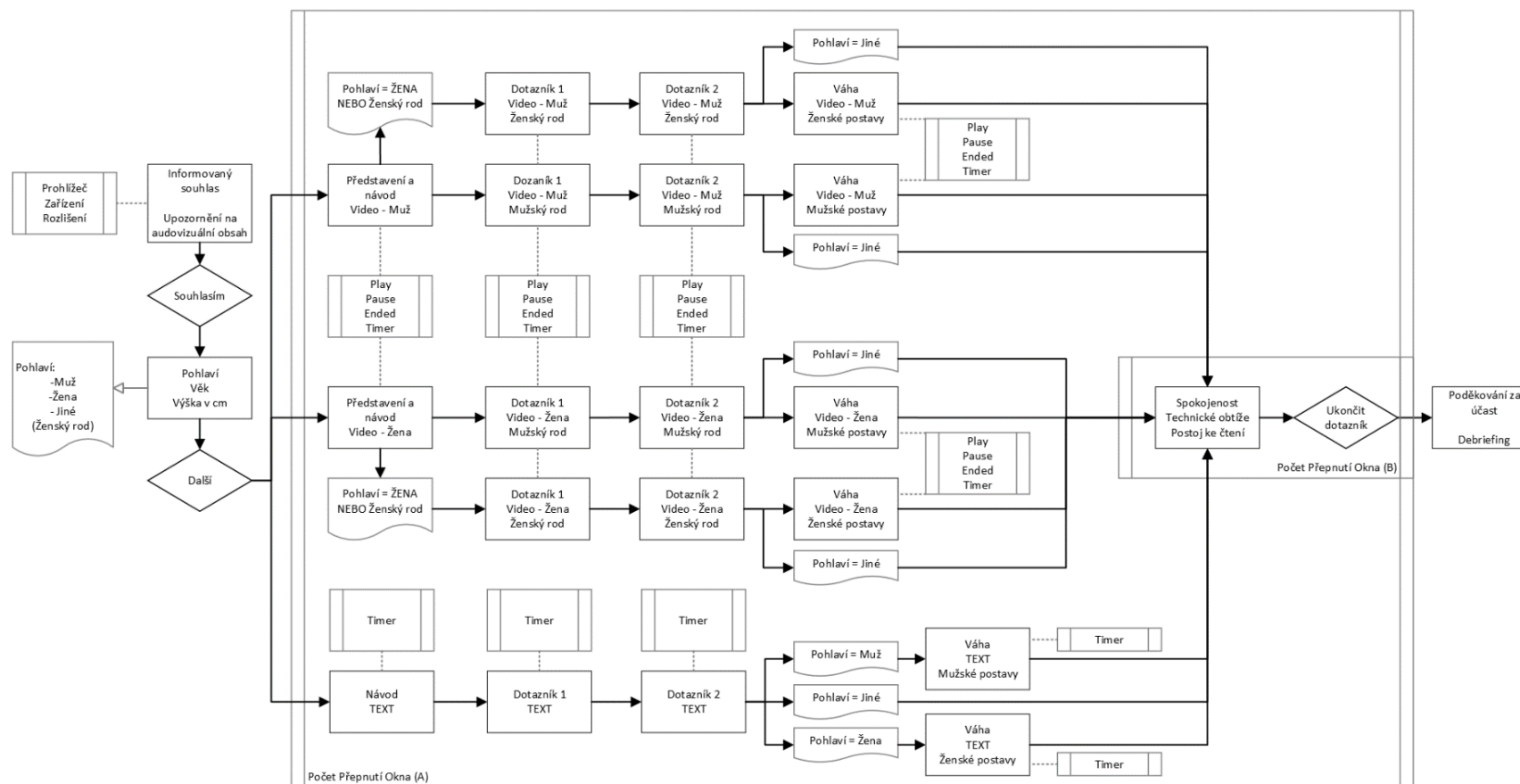
Pozn.: Nalevo: zobrazení pro PC – přibližná pozice administrátora na video; Napravo: mobilní zobrazení – grafika pozastaveného přehrávání.

Obr. 2: Položka dotazníku v kontrolní podmínce (TD)



Pozn.: Nalevo: zobrazení pro PC; Napravo: mobilní zobrazení

Obr. 3: Diagram průchodu dotazníkem



Pozn.: Dvojitý okraj = metadata; Dolní hrana vlnovkou = volba/zvolené pohlaví; Kosočtverec = akce vyžadovaná pro pokračování;
 Dotazník 1 = Dotazník výšky; Dotazník 2 = Rosenbergova škála sebehodnocení; Váha = Položka tělesné konstituce

2.2.3 Úprava znění položek a rodové varianty

Položky použitých škál prošly úpravou, aby bylo jejich znění přizpůsobeno administraci skrze video. Tato úprava zaprvé spočívala v převedení položek z původní první osoby, která je pro sebeposuzovací škály typická, do podoby ve druhé osobě. Problém ilustrujme na položce: „*Často musím stát na špičkách, abych lépe viděl/a.*“. Tato podoba je optimálním řešením pro textovou verzi dotazníku, protože respondent čte tvrzení o sobě a vybírá míru, s jakou s ním souhlasí. Je ovšem zjevné, že v případě, kdy položku prezentuje administrátor, respondent by vyslechl tuto položku ve formě tvrzení o administrátorovi, nikoli o sobě. Tuto variantu tak pro potřeby VD považuji za nevhodnou a příliš matoucí. Tato konkrétní položka je tak transformována do podoby: „*Často musíte stát na špičkách, abyste lépe viděl/a.*“, čímž je pro respondenta jasné, že předmětem tohoto tvrzení je on sám a jeho výška, nikoli výška administrátora. Stejná podoba položek ve druhé osobě je pro jednotnost použita i v textové podmínce (TD), i když se zde může jednat o suboptimální variantu, oproti variantě v osobě první.

Druhou úpravou, která má optimalizovat položky pro potřeby prezentace skrze video, je rozdělení gramatických rodů položek. Zatímco třeba v anglickém jazyce by podobné rozdělení nebylo nutné, v českém jazyce je tento problém nevyhnutelný. V případě textové verze položky: „*Často musíte stát na špičkách, abyste lépe viděl/a.*“, je část složeného slovesného tvaru, tedy sloveso „*vidět*“, ve svém ženském rodě příčestí činného zapsáno pomocí písmene „*a*“ za lomítkem. V mluvené řeči však ke stejnému řešení přistoupit pochopitelně nelze. Stejně tak rušivě by působilo i vyjádření obou gramatických rodů najednou („*...abyste lépe viděl nebo viděla.*“, případně „*...abyste lépe viděl, viděla.*“). Z tohoto důvodu mají experimentální podmínky (VD) dvě varianty, v ženském a mužském rodě zvlášť. Gramatický rod je respondentovi přiřazen na základě výběru odpovědi (Muž/Žena) na položku „*Jste:*“ (viz také diagram v Obr. 3). V případě volby „*Nechci odpovídat/Jiné*“ je respondent upozorněn na výchozí administraci dotazníku v mužském rodě, a je mu nabídnuta alternativa v rodě ženském. Při natáčení videí byl kladen zvláštní důraz na to, aby byly varianty stejné položky v obou rodech prezentovány administrátorem srovnatelně (intonace, výraz tváře). Toto rozdělení do gramatických rodů bylo využito pouze v experimentálních podmínkách (VD). V kontrolní podmínce (TD) byla použita varianta ve druhé osobě, jak je popsáno výše, ale s kombinací obou rodů, jak je pro textové dotazníky zvykem (např.: „*Často musíte stát na špičkách, abyste lépe viděl/a.*“).

2.2.4 Dotazník výšky

Dotazník výšky je jednou ze škál vyvinutých Rečkou (2018), která je schopna odhadnout fyzickou výšku respondenta pomocí podobných položek, jaké využívají běžné

psychologické škály. Dotazník výšky měří spíše subjektivně vnímanou výšku, která zřejmě částečně reflektuje i referenční skupinu, kterou respondent využívá ke srovnání. Z tohoto důvodu tak škála funguje rozdílně pro muže a pro ženy, což bude v analýze zohledněno. Možné odpovědi na každou z položek jsou pak mírou souhlasu respondenta na čtyřbodové Likertově škále od „Silně nesouhlasím“ po „Silně souhlasím“. Pro účely této práce bude adaptována zkrácená verze dotazníku výšky, která sestává z 11 položek (Tabulka 2) s nejlepšími psychometrickými vlastnostmi ve vztahu k měřenému konstrukt (Tancoš, 2019). Položky škály se respondentům zobrazují v randomizovaném pořadí. V analýze budou mimo jiné získané skóry porovnatelné s výškou udanou respondenty, a tím bude možné určit, která z metod administrace vede k přesnějším odhadům, pokud však takový rozdíl nastane.

2.2.5 Rosenbergova škála sebehodnocení

Rosenbergova škála sebehodnocení (RSES; Rosenberg's self-esteem scale; Rosenberg, 1965) je jednou z nejpoužívanějších škál pro měření sebehodnocení v psychologickém výzkumu (Blascovich & Tomaka, 1991; Jordan, 2020). Jedná se o reliabilní, validní a snadno administrovatelný nástroj, který je považován za jednodimenzionální, i když mnohdy produkuje spíše dva faktory (Jordan, 2020). Respondenti i zde odpovídají na čtyřstupňové Likertově škále od „Silně nesouhlasím“ po „Silně souhlasím“ na celkem 10 položkách (Tabulka 3). Položky škály se i zde zobrazují v randomizovaném pořadí. Pro tuto práci byl adaptován český překlad od Blatného a Osecké (1994). Při své analýze došli autoři ke dvěma možným řešením. Buďto dvoufaktorovému, nebo třífaktorovému. Dvoufaktorové řešení pak odpovídá pozitivně a negativně formulovaným položkám a podle autorů mimo jiné úzce souvisí i s poznatky Paulhuse (1989) o faktorech sociální žádoucnosti. Je poměrně intuitivní, že škála jako taková, respektive sebehodnocení, které měří, se sociální žádoucností souvisí a koreluje (Fleming & Courtney, 1984; Jordan, 2020). Pokud by tak sociální žádoucnost ovlivňovala výsledky v experimentální podmínce, je možné, že se právě projeví i zde.

Tabulka 2: Položky dotazníku výšky ve druhé osobě

POLOŽKA	TEXTOVÝ DOTAZNÍK (TD)	ROD	DOTAZNÍK VIDEEM (VD)
1	Máte vhodnou výšku na hraní basketbalu nebo volejbalu.		Máte vhodnou výšku na hraní basketbalu nebo volejbalu.
2*	Když mluvíte s jinými dospělými a chcete se jim dívat do očí, častěji na ně spíš vzhlížíte nahoru.		Když mluvíte s jinými dospělými a chcete se jim dívat do očí, častěji na ně spíš vzhlížíte nahoru.
3	Lidem, kteří na koncertě stojí za vámi, většinou vaše postava dost brání ve výhledu.		Lidem, kteří na koncertě stojí za vámi, většinou vaše postava dost brání ve výhledu.
4*	Často musíte stát na špičkách, abyste lépe viděl/a.	M	Často musíte stát na špičkách, abyste lépe viděl.
		Ž	Často musíte stát na špičkách, abyste lépe viděla.
5	Když chcete někoho obejmout, většinou se musíte sklonit.		Když chcete někoho obejmout, většinou se musíte sklonit.
6*	Často potřebujete stoličku, abyste dosáhl/a na něco, na co jiní lidé dosáhnou normálně.	M	Často potřebujete stoličku, abyste dosáhl na něco, na co jiní lidé dosáhnou normálně.
		Ž	Často potřebujete stoličku, abyste dosáhla na něco, na co jiní lidé dosáhnou normálně.
7*	Jednou z prvních věcí, které si na vás lidé všimnou, je to, jak moc jste malý/á.	M	Jednou z prvních věcí, které si na vás lidé všimnou, je to, jak moc jste malý.
		Ž	Jednou z prvních věcí, které si na vás lidé všimnou, je to, jak moc jste malá.
8	Často si musíte dávat pozor, abyste se neuhodil/a hlavou například o nízký strop nebo rám dveří.	M	Často si musíte dávat pozor, abyste se neuhodil hlavou například o nízký strop nebo rám dveří.
		Ž	Často si musíte dávat pozor, abyste se neuhodila hlavou například o nízký strop nebo rám dveří.
9*	V autobuse míváte dostatek prostoru pro nohy.		V autobuse míváte dostatek prostoru pro nohy.
10	Slýcháváte narážky na to, že jste vysoký/á.	M	Slýcháváte narážky na to, že jste vysoký.
		Ž	Slýcháváte narážky na to, že jste vysoká.
11*	Kvůli vaší menší výšce vám lidé hádají, že jste mladší, než ve skutečnosti jste.		Kvůli vaší menší výšce vám lidé hádají, že jste mladší, než ve skutečnosti jste.

*Reverzní položka

METODA

Tabulka 3: Položky Rosenbergovy škály sebehodnocení

POLOŽKA	TEXTOVÝ DOTAZNÍK (TD)	ROD	DOTAZNÍK VIDEEM (VD)
1	Máte pocit, že si uchováváte svoji osobní důstojnost nejméně v takové míře, jako většina ostatních lidí.		Máte pocit, že si uchováváte svoji osobní důstojnost nejméně v takové míře, jako většina ostatních lidí.
2*	Jste vždy náchylný/á považovat se za neúspěšného člověka.	M	Jste vždy náchylný považovat se za neúspěšného člověka.
		Ž	Jste vždy náchylná považovat se za neúspěšného člověka.
3	Myslíte si, že máte řadu dobrých vlastností.		Myslíte si, že máte řadu dobrých vlastností.
4	Jste schopný/á dělat mnoho věcí stejně dobře jako ostatní.	M	Jste schopný dělat mnoho věcí stejně dobře jako ostatní.
		Ž	Jste schopná dělat mnoho věcí stejně dobře jako ostatní.
5*	Zdá se vám, že nemůžete být u sebe na nic zvláštního hrdý/á.	M	Zdá se vám, že nemůžete být u sebe na nic zvláštního hrdý.
		Ž	Zdá se vám, že nemůžete být u sebe na nic zvláštního hrdá.
6	Máte k sobě dobrý vztah.		Máte k sobě dobrý vztah.
7*	Chtěl/a byste si sám/sama sebe více vážit.	M	Chtěl byste si sám sebe více vážit.
		Ž	Chtěla byste si sama sebe více vážit.
8	Jste se sebou vcelku spokojený/á.	M	Jste se sebou vcelku spokojený.
		Ž	Jste se sebou vcelku spokojená.
9*	Občas jasně pociťujete svoji neužitečnost.		Občas jasně pociťujete svoji neužitečnost.
10*	Někdy si myslíte, že jste naprosto neschopný/á.	M	Někdy si myslíte, že jste naprosto neschopný.
		Ž	Někdy si myslíte, že jste naprosto neschopná.

**Reverzní položka*

2.2.6 Tělesná konstituce – váha

Součástí dotazníku je také položka zaměřená na váhu, respektive postavu respondenta. Některé předešlé výzkumy ukázaly, že i v dotazníku administrovaném formou videí ovlivňuje ochotu svěřit se také interakce pohlaví administrátora a respondenta (Fuchs, 2009). Protože není záměrem tohoto výzkumu vystavovat respondenty intimním nebo eticky sporným otázkám, je tělesná konstituce použita jako alternativa, která může být za určitých okolností sociální žádoucností ovlivněna. Je zřejmé, že ve společnosti existuje tlak směrem k ideální postavě a ideálu krásy, který se s ní pojí. Self-reportové metody tak mohou produkovat zkreslené údaje o respondentově váze, které se liší i s pohlavím respondenta (Betz et al., 1994; Gil & Mora, 2011). V takovém případě by bylo eventuelně možné sledovat rozdílné fungování položky v interakci pohlaví respondenta s pohlavím administrátora, kdy se mohou projevit například určité sociálně žádoucí tendence svoji postavu idealizovat. Pro tuto položku byly použity ilustrace postav ze studie Harrise et al. (2007). Respondenti vybírají z 10 obrázků lidských postav tu, kterou považují za nejpodobnější své postavě (Obr. 4). Respondentům se zobrazily vždy takové postavy, které korespondovaly se zvoleným pohlavím v položce „Jste:“ (Muž/Žena). Respondentům, kteří zvolili možnost „Nechci odpovídat/Jiné“, nebyla položka zobrazena (viz také diagram v Obr. 3). Znění položky bylo pro všechny verze dotazníku stejné:

„Dole je zobrazeno deset různých lidských postav. Vyberte z nich takovou, která je nejpodobnější vaší postavě.“

Obr. 4: Obrázková škála váhy (Harris et al., 2007)



Pozn.: Zobrazení v dotazníku; Nalevo: mužská varianta, Napravo: ženská varianta

2.2.7 Výstupní dotazník

Respondenti ze všech podmínek končili své odpovídání na krátkém textovém výstupním dotazníku s 5 položkami. Tento dotazník měl primárně zjistit subjektivní zkušenosti respondentů s dotazníkem. Čtyři položky měly stejný formát jako většina dotazníku, tedy s odpověďmi na čtyřstupňové Likertově škále od „silně nesouhlasím“ po „silně souhlasím“. Položka „*Dotazník mě bavil.*“ se snaží zjistit zábavnost metody pro respondenta a umožnit následné srovnání. Položka „*Na dotazník jsem se plně soustředil/a.*“ je zaměřena na měření subjektivně vnímané pozornosti věnované dotazníku. Kvůli specifikům tohoto experimentálního dotazníku je zařazena i položka „*Měl/a jsem technické problémy.*“, která má sledovat, zda dotazník formou videa nekladl příliš vysoké nároky, ať už na respondentovo zařízení, nebo respondentovu technickou způsobilost, navzdory snaze o co nejintuitivnější ovládání. Je však třeba brát v potaz, že závažné technické problémy by nejspíše vedly k předčasnému ukončení, a na položku tak bude nutné pohlížet s rezervou. Položka „*Ve svém volném čase čtu nerad/a.*“ slouží k ověření možné preference dotazníků administrovaných videem osobami, které nerady čtou. Poslední z položek dotazníku je volná odpověď, která umožňuje respondentům připojit svůj komentář nebo připomínku. Po tomto dotazníku následoval krátký debriefing, kde byli respondenti blíže seznámeni s účelem výzkumu. Zejména respondentům z kontrolní skupiny (TD) bylo vysvětleno, proč neviděli žádné video, na které byly předtím upozorněni.

2.3 Kognitivní interview a předběžné testování

Za normálních okolností slouží kognitivní interview, prováděné zpravidla na malém počtu respondentů, k ověření pochopitelnosti a srozumitelnosti obsahu vytvořených položek. Protože tento výzkum používá položky z již odzkoušených škál, nebylo interview v tomto smyslu tolik potřebné. Kognitivní interview ovšem proběhla z důvodu ověření znění položek ve druhé osobě (viz 2.2.3), zejména však k ověření srozumitelnosti položek v pravém slova smyslu. Na vzorku šesti respondentů bylo ověřeno, že veškeré zvukové stopy náležející k daným videonahrávkám jsou dostatečně hlasité, a že je administrátorům ve všech verzích zřetelně rozumět. U žádné z videonahrávek (celkem 68; 34 na administrátora) nebyl respondenty shledán žádný problém.

Funkčnost dotazníku, zejména VD, byla před spuštěním kontrolována na několika zařízeních (mobilní zařízení, tablety, PC) s různými kombinacemi operačních systémů (Windows, Android, iOS), prohlížečů (Google Chrome, Safari, MS Edge) a připojení (Wi-Fi, ethernet, mobilní síť). Testování odhalilo určité problémy s rychlostí přenosu dat ze serveru hostujícího videa, které by mohly vést k nadměrnému předčasnému ukončování dotazníku u respondentů v experimentálních podmínkách (VD), a to

navzdory malým velikostem videí (řádově stovky kB v kvalitě 720p). Video tak byla přesunuta na jiný hostingový server, což učinilo načítání a přehrávání videí relativně bezproblémové i v suboptimálních podmínkách (mobilní připojení; přenosové rychlosti < 20 Mb/s). Zveřejnění a ostrý sběr dat byl odstartován až po ověření fungování všech variant dotazníku v uzavřeném okruhu respondentů ($n = 10$), kteří potvrdili, že na jejich zařízeních nebyl s dotazníkem, ani jeho ovládáním žádný problém.

2.4 Analýza dat

Pro manipulaci s daty a následnou analýzu bude použit programovací jazyk R, verze 4.2.3 (R Core Team, 2023) ve vývojovém prostředí R Studio. Pro manipulaci s daty bude použit zejména balíček *tidyverse* (Wickham et al., 2019). Balíček *psych* (Revelle, 2023) bude použit zejména pro získání deskriptivních statistik. Jako doplněk regresních analýz budou využity balíčky *car* (Fox & Weisberg, 2019) nebo *lm.beta* (Behrendt, 2022) pro získání standardizovaných regresní koeficientů (β). Pro srovnání korelačních koeficientů bude použit balíček *cocor* (Diedenhofen & Musch, 2015). Pro potřeby strukturního modelování bude využit primárně balíček *lavaan* (Rosseel, 2012), ale také *semTools* (Jorgensen et al., 2023), který bude sloužit i k výpočtu koeficientů omega (ω). Většina grafů použitá v této práci využívá balíček *ggplot2* (Wickham, 2016; součást balíčku *tidyverse*), nebo balíčky z něj vycházející.

Všechny výsledky interpretované jako statisticky významné budou mít p-hodnotu pod hladinou významnosti $\alpha = 0,05$, pokud nebude uvedeno jinak.

Data použitá pro analýzu a analytický skript v jazyce R jsou dostupné online z webových stránek Open Science Framework na adrese: <https://osf.io/s5q8d/>

3 Výsledky

3.1 Předčasná ukončení a vyřazené případy

Vzhledem k experimentálnímu designu dotazníku se dalo očekávat, že předčasně ukončených případů by mohlo být značné množství. Pro ekologickou validitu administrace dotazníku skrze video je důležité ověřit, zda toto ukončování není způsobeno právě přítomností videí, například kvůli technickým problémům.

3.1.1 Globálně vyřazené případy a úplná neúčast

Z celkových 389 zaznamenaných průchodů dotazníkem bylo poměrně značné množství odfiltrováno jako nevyhovující ($n = 68$; 17,5 %). Tyto průchody naplnily alespoň jednu ze stanovených podmínek, na základě kterých byly vyřazeny ty pokusy o vyplnění, kde:

1. Respondent nedošel ani k rozdělení do experimentální podmínky (VD/TD).
2. Počet chybějících hodnot v dotazníku výšky a RSES byl vyšší než 12.
3. Odpovědi v dotazníku výšky mají pouze stejné hodnoty navzdory přítomnosti reverzních položek (*straight-lining*).
A zároveň: Odpovědi ve škále sebehodnocení mají pouze stejné hodnoty navzdory přítomnosti reverzních položek.
4. Čas strávený vyplňováním dotazníku byl nižší než 30 sekund.

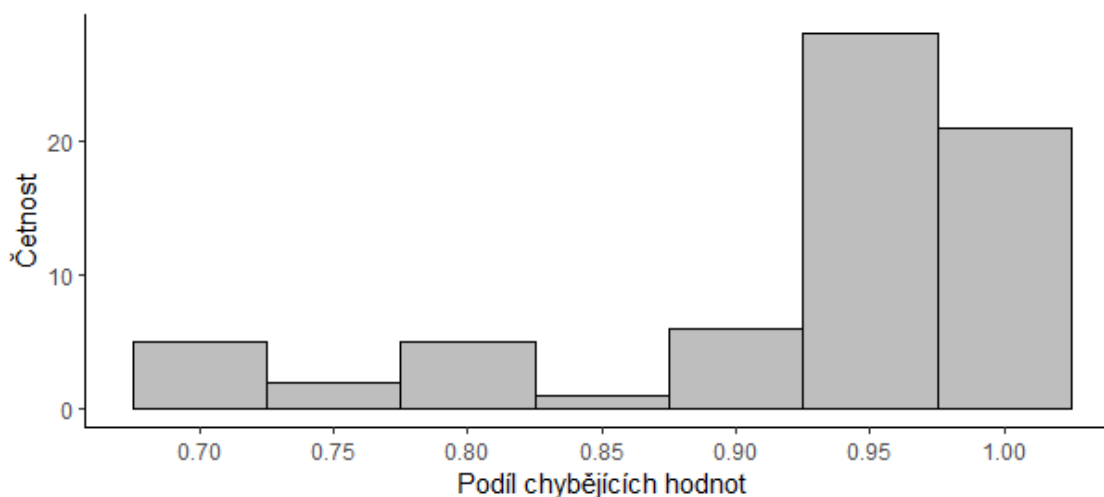
Uvedená kritéria jsou však do velké míry redundantní, protože případy označené jako nevyhovující obsahují v klíčových škálách nejvýše 7 označených odpovědí (z 21 možných). V drtivé většině případů však není označena odpověď téměř žádná, protože se jedná o předčasná ukončení (*break-off*). Podíl chybějících hodnot ve 28 položkách dotazníku se u vyřazených případů pohybuje od $Min = 68,9 \%$ do $Max = 100 \%$ ($M = 91,6 \%$; $Mdn = 92,9 \%$; Graf 2).

U celkem 21 pokusů (5,4 %) bylo odpovídání ukončeno před randomizovaným rozdělením do experimentálních podmínek. Všichni z těchto respondentů nicméně dotazník opustili na jeho samotném začátku, kde byli mimo jiné upozorněni na přítomnost audiovizuálního obsahu. Není však možné určit, zda se rozhodli odpovídání ukončit z tohoto důvodu s přihlédnutím k tomu, že samotné pozvánky k vyplnění dotazníku tento obsah jasně deklarovaly. Převažují mobilní zařízení ($n = 15$), i když v případě respondentů, kteří strávili v dotazníku alespoň 10 sekund ($n = 12$) a pravděpodobněji

tak účast ve výzkumu zvažili, je počet ukončených průchodů na mobilních zařízeních shodný s počtem ukončených průchodů na PC. Vzhledem k velkému podílu mobilních zařízení by tak byly případy s PC dominantní.

Ve zbylých neplauzibilních průchodech, kde k rozdělení do experimentálních podmínek došlo, bylo 26 případů (6,7 %) přiřazených buďto k mužskému administrátorovi (VD_M; $n = 11$; 2,8 %), nebo k administrátorce ženě (VD_ž; $n = 15$; 3,9 %). U těchto respondentů je možné předpokládat, že jejich odpovídání mohl narušit problém s videoobsahem, a to buď technické povahy, nebo samotná přítomnost videí za předpokladu, že respondenti přešli předchozí upozornění. Alespoň v 19 případech však metadata implikují, že přehrávání úvodního videa přinejmenším začalo. Nelze však říci v jaké kvalitě. V případě kontrolní skupiny (TD) se jedná také o 21 případů (5,4 %), což v porovnání s VD spíše zpochybňuje, že by problémy způsobovala přítomnost videí. Mobilní zařízení tvoří nepřekvapivě v obou podmínkách většinu vyřazených případů. V případě VD se jedná dokonce o 21 případů (80,8 %), nicméně také u TD jde o 14 případů (66,7 %). Znatelně vyšší podíl vyřazených respondentů s mobilním zařízením u VD by mohl naznačovat problémy s kompatibilitou. Nezdá se však, že by něco podobného nastalo.

Graf 2: Rozložení podílů chybějících hodnot (NA) globálně vyřazených případů



3.1.2 Předčasná ukončení

K případům, kde došlo k ukončení už na samotném začátku, je potřeba dodat také případy předčasného ukončení, kdy respondenti vyplnili první škálu (dotazník výšky), ale druhou škálu (RSES) buďto nedokončili, nebo dokončili s velkým množstvím

chybějících hodnot (> 5). Těchto respondentů bylo celkem 9, a i zde náleželi jak k VD ($n = 5$), tak také k TD ($n = 4$).

Respondentů, kteří dotazník nedokončili (*break-off*) tak, že byli přiděleni do příslušné skupiny, ale své odpovídání ukončili před nebo během odpovídání na položku tělesné konstituce, která byla poslední položkou experimentální/kontrolní podmínky, bylo celkem $n = 59$, z toho 32 ve VD ($z n = 183$) a 27 v TD ($z n = 185$). To znamená, že i proporčně se jedná o více ukončení ve VD (17,5 %) ve srovnání s TD (14,60 %), ale nikoli významně (jednostranný Fisherův exaktní test: $p = 0,270$; $OR = 0,81$). Hypotéza H_5 (VD bude mít vyšší počet nedokončených průchodů oproti TD.) nemůže být podpořena.

3.1.3 Zbylé neplauzibilní hodnoty

Označeny byly ty případy, kdy respondent zvolil pro všechny položky jedné, nebo druhé škály stejnou odpověď. Zejména kvůli přítomnosti reverzních položek je nepravděpodobné, že by všechny odpovědi měly stejnou hodnotu. Jednalo se tak zřejmě o případy „straight-lining“, nicméně vždy jen v jedné ze škál, a tak nebyly vyloučeny list-wise (viz 3.1.1, podmínka č. 2). Tito respondenti jsou vyloučeni pouze z analýz, kde by tento styl odpovídání narušoval jejich výsledky (analýzy dané škály, časy vyplňování apod.).

U jednoho respondenta (muže) byla jeho výška změněna na chybějící hodnotu, protože udal nepravděpodobně nízkou výšku, a to na dolní hranici přípustných hodnot ověřovaných dotazníkem ($h = 130$ cm).

3.2 Deskriptivní statistiky

Průměrný testový skóre dotazníku výšky byl spočítán pro ty respondenty, kteří se v něm nedopustili „straight-lining“, zodpověděli alespoň 9 z 11 položek a jako své pohlaví neuvedli „Jiné“. Průměrný testový skóre Rosenbergovy škály sebehodnocení byl spočítán pro ty respondenty, kteří se v něm nedopustili „straight-lining“, a kteří zodpověděli alespoň 9 z 10 položek. Tabulka 4 zobrazuje deskriptivní statistiky pro klíčové položky a škály dotazníku. Graf 3 pak zobrazuje rozložení průměrného testového skóru RSES.

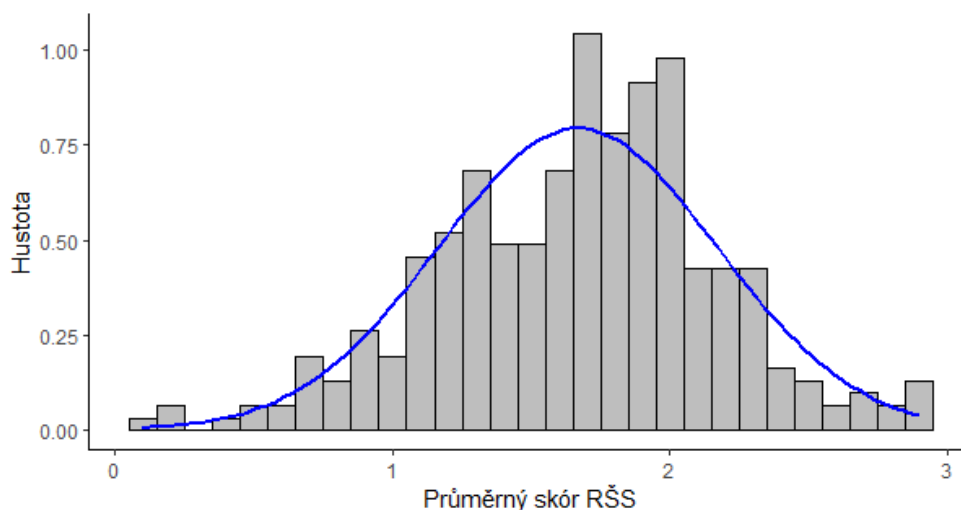
V případě skóre dotazníku výšky se pak podle očekávání průměry žen ($M = 1,31$; $SD = 0,55$) a mužů ($M = 1,85$; $SD = 0,49$) liší, a to signifikantně ($t(310) = -8,19$; $p < 0,001$; Cohenovo $d = -1,00$; Graf 4). V případě reálné výšky udané

respondenty se průměrná výška žen ($M = 167,40$; $SD = 6,96$) od průměrné výšky mužů ($M = 182,32$; $SD = 0,49$) liší, a to dle očekávání také statisticky významně ($t(310) = -17,74$; $p < 0,001$; Cohenovo $d = -2,16$; Graf 5). Obrázková položka tělesné konstituce měla odlišnou podobu pro každé pohlaví (viz 2.2.6), a tak jsou její statistiky napříč pohlavími porovnatelné jen stěží.

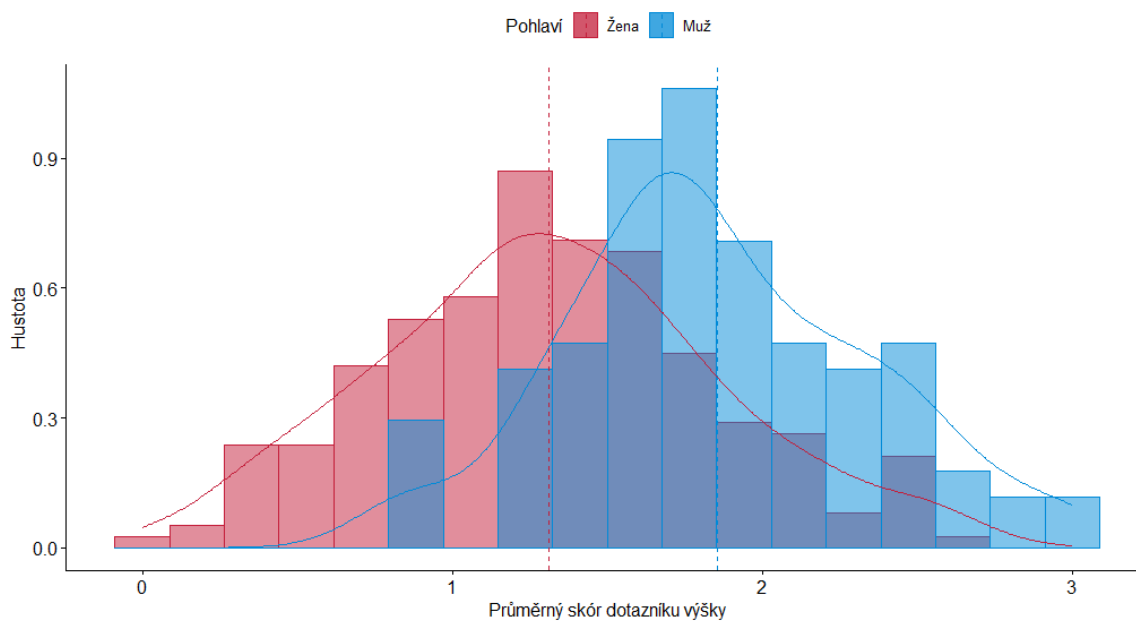
Tabulka 4: Deskriptivní statistiky

Škála/položka	Pohlaví	N	M	SD	Mdn	Min	Max
Dotazník výšky	-	312	1,48	0,59	1,45	0,00	3,00
RSES	-	307	1,67	0,50	1,70	0,10	2,90
Dotazník výšky	Ženy	215	1,31	0,55	1,27	0,00	2,73
	Muži	97	1,85	0,49	1,73	0,82	3,00
Výška/cm	Ženy	214	167,40	6,96	167,00	148,00	187,00
	Muži	98	182,32	6,74	183,00	162,00	200,00
Tělesná konstituce	Ženy	209	2,98	1,43	3,00	1,00	10,00
	Muži	96	3,58	1,83	3,00	1,00	10,00

Graf 3: Rozložení průměrných testových skóre Rosenbergovy škály sebehodnocení

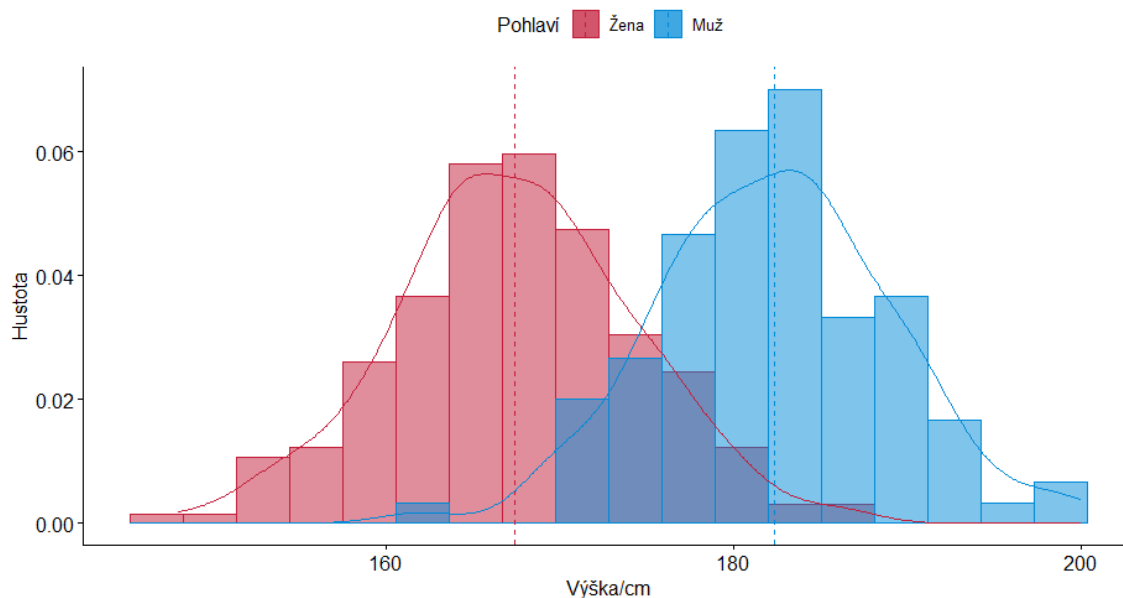


Graf 4: Rozložení průměrných testových skóreů dotazníku výšky podle pohlaví



Pozn.: Čárkované vertikální přímky značí průměry

Graf 5: Rozložení výšky udané respondentem v centimetrech podle pohlaví



Pozn.: Čárkované vertikální přímky značí průměry

3.3 Průměrné skóry a metody administrace

3.3.1 Výška udaná respondenty v cm

Výška respondentů v centimetrech byla získávána v textové podobě ještě před rozdělením do experimentálních podmínek, a tak se neměla ve skupinách výrazně lišit. V případě žen jsou průměry srovnatelné ($M_{TD} = 167,20$; $M_{VD} = 167,63$). V případě mužů však překvapivě vidíme vyšší průměr v případě TD ($M = 183,81$; $SD = 6,56$) oproti VD ($M = 180,94$; $SD = 6,68$), a to statisticky významně ($t(96) = 2,14$; $p = 0,035$; Cohenovo $d = 0,43$). Podobně pak i v případě mediánu ($Mdn_{TD} = 184$; $Mdn_{VD} = 180$). Tento jev je zřejmě výsledkem výběrové chyby a menší velikosti vzorku ($n = 98$). V následujících analýzách bude tento problém zohledněn.

3.3.2 Rozdíly ve způsobech administrace

V případě průměrných skóru škál se u žen průměrný skór dotazníku výšky v případě administrace skrze video (VD; $M = 1,30$; $SD = 0,57$) od kontrolní textové podmínky (TD; $M = 1,31$; $SD = 0,54$) signifikantně neliší ($t(213) = -0,13$; $p = 0,896$; Cohenovo $d = -0,02$), dokonce je téměř stejný. Co se týče mužů, vzhledem k jejich rozdílné reálné výšce lze očekávat, že se jejich průměrné skóry dotazníku výšky budou lišit, a to s nižším průměrem ve VD ($M = 1,76$; $SD = 0,5$) oproti TD ($M = 1,94$; $SD = 0,46$), což podporují i výsledky jednostranného t-testu ($t(95) = -1,91$; $p = 0,029$; Cohenovo $d = -0,39$).

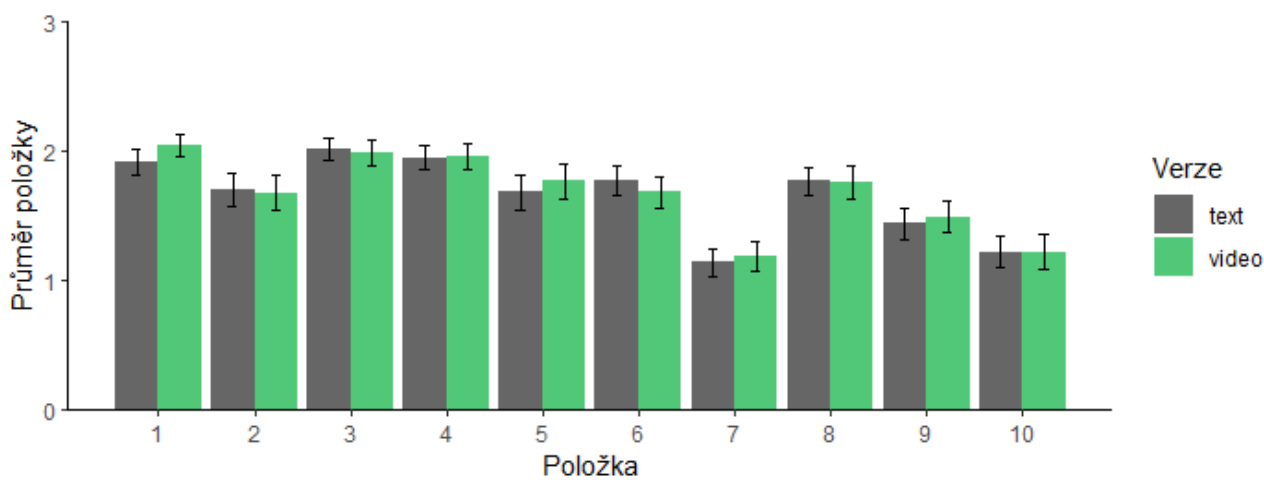
V případě Rosenbergovy škály sebehodnocení se průměrný skór respondentů ve VD ($M = 1,68$; $SD = 0,54$) oproti průměrnému skóru v TD ($M = 1,66$; $SD = 0,47$) významně neliší ($t(305) = 0,29$; $p = 0,770$; Cohenovo $d = 0,03$). Naopak, průměry jsou i zde srovnatelné.

Rosenbergova škála sebehodnocení je potenciálně ovlivnitelná sociální žádoucností. Pokud by sociální žádoucnost hrála v metodách administrace roli, mohla by se projevit na úrovni položek škály s citlivějším obsahem. I zde jsou však rozdíly minimální (Tabulka 5), s pouze marginálně signifikantním rozdílem u položky číslo 1: *Máte pocit, že si uchováváte svoji osobní důstojnost nejméně v takové míře, jako většina ostatních lidí*, s celkově slabým efektem ($r_{r-b} = 0,10$ (95% CI [-0,02; 0,23])).

Tabulka 5: Srovnání skóre položek RSES mezi metodami administrace

Pol.	TD				VD				Wilcoxon r-s	
	N	M	SD	Mdn	N	M	SD	Mdn	W	p
1	160	1,91	0,62	2	146	2,04	0,52	2	12905	0,051
2	162	1,7	0,86	2	148	1,68	0,83	2	11713	0,710
3	160	2,01	0,56	2	149	1,99	0,6	2	11885	0,955
4	160	1,95	0,59	2	149	1,96	0,62	2	12195	0,659
5	160	1,68	0,84	2	147	1,77	0,83	2	12530	0,285
6	160	1,78	0,74	2	147	1,68	0,76	2	10813	0,170
7	159	1,14	0,69	1	149	1,18	0,71	1	12201	0,613
8	159	1,77	0,69	2	147	1,76	0,81	2	11691	0,995
9	160	1,44	0,78	1	148	1,49	0,79	2	12231	0,590
10	162	1,22	0,78	1	149	1,22	0,83	1	11897	0,813

Graf 6: Průměry položek škály sebehodnocení



Pozn.: Chybové úsečky značí 95% intervaly spolehlivosti

3.4 Kriteriační validita metod

3.4.1 Korelace

Statistické srovnání Pearsonových korelačních koeficientů je problematické kvůli zeshikmení výběrového rozdělení. Jedním z možných postupů je Fisherova z-transformace v podobě:

$$F(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \tanh^{-1}(r), \quad (1)$$

kde r je příslušný Pearsonův korelační koeficient, a kde se výběrové rozdělení $F(r)$ už blíží normálnímu.

Pro ženy v případě VD měla korelace mezi výškou respondentů a průměrným testovým skórem dotazníku výšky hodnotu $r(97) = 0,86$ (95% CI [0,80; 0,91]). V případě kontrolní podmínky (TD) pak $r(111) = 0,85$ (95% CI [0,79; 0,90]). Tento rozdíl tak statisticky významně odlišný není ($z = 0,35$; $p = 0,728$).

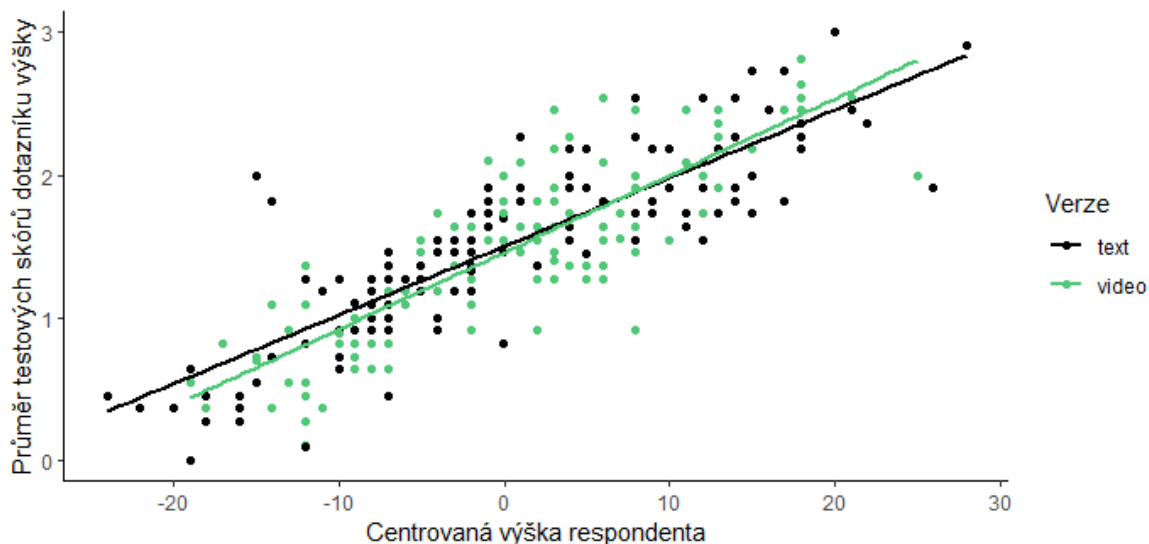
Pro muže měl v případě VD korelační koeficient hodnotu $r(48) = 0,84$ (95% CI [0,73; 0,91]), zatímco v případě TD $r(43) = 0,78$ (95% CI [0,63; 0,87]). Vzhledem k šířce konfidenčních intervalů se vztahy neliší ani zde ($z = 0,83$; $p = 0,405$). Z těchto srovnání tak nelze říci, že by se varianty administrace lišily ve vztahu ke kritériu, jakým byla výška udaná respondenty.

3.4.2 Hierarchická lineární regrese

Za účelem srovnání vlivu metod administrace na průměrné skóry získané z dotazníku výšky byla provedena hierarchická lineární regrese se třemi bloky ($n = 307$; Tabulka 6). V prvním bloku byl průměrný skór výšky predikován experimentální podmínkou a pohlavím. Podle očekávání a předchozích výsledků je jen pohlaví prediktorem signifikantním ($\beta = 0,43$; $p < 0,001$). Co se vlivu metody administrace týče (TD/VD), ten statisticky významný není ($\beta = -0,05$; $p = 0,317$). Ve druhém bloku byla zahrnuta výška respondentů v centimetrech, která byla centrována pro redukci multikolinearity. Tato proměnná je také podle očekávání významným prediktorem ($\beta = 1,08$; $p < 0,001$). Po statistické kontrole reálné výšky se výrazněji změnil jen regresní koeficient pro mužské respondenty, a to na statisticky významně záporný ($\beta = -0,34$; $p < 0,001$).

Ve třetím bloku byly zahrnuty interakční členy. Statisticky signifikantní se stala interakce metody administrace a centrované výšky respondenta ($\beta = 0,12$; $p = 0,019$). S větší výškou respondentů tak má trochu překvapivě administrace skrze video vést k vyššímu průměru v dotazníku výšky. Lze však jednoduše vidět, že směrnici přímků ovlivňují vlivné případy (Graf 7)⁷. Z tohoto důvodu byly nejvlivnější případy vyřazeny ($n = 4$; Cookova vzdálenost $> 0,04$) a poslední krok regrese byl zopakován ($n = 303$; Tabulka 6, blok 3b). V tomto případě už zmíněný interakční člen za signifikantní označen být nemůže ($\beta = 0,09$; $p = 0,061$). Zdá se tak, že pokud má metoda administrace nějaký vliv na skóry dotazníku výšky, je tento vliv spíše zanedbatelný.

Graf 7: Interakce výšky respondenta a metody administrace



Pozn.: Graf zobrazuje stav před odebráním vlivných případů (model 3a)

⁷ Pro detailnější diagnostiku předpokladů lineární regrese viz Přílohu A.

Tabulka 6: Hierarchický regresní model

Závislá proměnná: Průměr testového skóru dotazníku výšky (n = 307)																				
Proměnná	Blok 1					Blok 2					Blok 3a					Blok 3b (n = 303)				
	B	SE	β	t	P	B	SE	β	t	p	B	SE	β	t	p	B	SE	β	t	p
(Průsečík)	1,34	0,05		28,77	<0,001*	1,62	0,03		59,47	<0,001*	1,62	0,03		51,00	<0,001*	1,62	0,03		55,76	<0,001*
Pohlaví (Muž)	0,55	0,07	0,43	8,40	<0,001*	-0,43	0,05	-0,34	-8,43	<0,001*	-0,31	0,09	-0,24	-3,63	<0,001*	-0,38	0,08	-0,30	-4,85	<0,001*
Verze (Video)	-0,06	0,06	-0,05	-1,00	0,317	-0,02	0,03	-0,02	-0,62	0,533	0,01	0,05	0,01	0,20	0,845	0,02	0,04	0,01	0,40	0,689
Výška						0,06	0,00	1,08	27,09	<0,001*	0,06	0,00	1,05	18,14	<0,001*	0,07	0,00	1,08	20,66	<0,001*
Výška:Pohlaví(Muž)											-0,01	0,01	-0,11	-1,95	0,052	-0,01	0,01	-0,06	-1,16	0,248
Výška:Verze(Video)											0,01	0,00	0,12	2,36	0,019*	0,01	0,00	0,09	1,88	0,061
Pohlaví(Muž):V(Video)											-0,12	0,10	-0,08	-1,17	0,245	-0,08	0,09	-0,05	-0,84	0,400
R^2	0,189					0,763					0,770					0,811				
ΔR^2	0,189					0,574					0,007					-				
F	35,40					749,88					3,17					-				
df	304					1; 303					3; 300					-				
p	<0,001					<0,001					0,025					-				

3.5 Faktorová struktura škál a jejich reliabilita

3.5.1 Faktorová struktura

Pro Rosenbergovu škálu sebehodnocení byl využit dvoufaktorový model s jedním faktorem pro pozitivně formulované a druhým pro negativně formulované položky. Pro dotazník výšky byl také použit dvoufaktorový model s faktory nadprůměrné a podprůměrné výšky s tím rozdílem, že položka č. 9 (*V autobuse míváte dostatek prostoru pro nohy.*) spadá do faktoru nadprůměrné výšky. Souhlas s ní totiž „znamená spíše absenci nadprůměrné výšky, nikoli nutně přítomnost podprůměrné výšky“, jak uvádí už Rečka (2018). Tyto modely zřejmě nejsou těmi nejideálnějšími (Tabulka 7), nicméně ověřování faktorové struktury RSES nebo dotazníku výšky není cílem této práce (pro tyto účely: Blatný & Osecká, 1994; Urbánek, 2000; respektive Rečka, 2018). Použité modely slouží primárně k výpočtu koeficientu ω (3.5.2) a ověření invariance měření (3.6). Za povšimnutí ovšem stojí, že se tyto jednoduché faktorové struktury viditelně lépe shodují s daty obdrženy z VD.

Tabulka 7: Ukazatele shody modelů s daty pro jednotlivé škály a metody a administrace

Škála	χ^2	df	p	TLI	RMSEA	RMSEA 95% CI		SRMR
						Dolní	Horní	
RSES TD	149,76	34	<0,001	0,904	0,145	0,122	0,170	0,092
RSES VD	71,84	34	<0,001	0,982	0,086	0,058	0,114	0,049
VYS TD ženy	135,46	43	<0,001	0,897	0,138	0,112	0,164	0,088
VYS VD ženy	51,91	43	0,165	0,993	0,046	0,000	0,085	0,048
VYS TD muži	90,03	43	<0,001	0,901	0,158	0,112	0,203	0,135
VYS VD muži	44,09	43	0,425	0,998	0,022	0,000	0,098	0,081

3.5.2 Reliabilita

Běžně používaným ukazatelem vnitřní konzistence je Cronbachova α (Cronbach, 1951). Použití Cronbachovy α je však problematické kvůli striktním předpokladům jako tau-ekvivalence nebo spojitě položky škály s normálním rozložením (McNeish, 2018). Proto bude jako vhodnější koeficient reliability použita McDonaldova ω (McDonald, 1999) vycházející ze strukturního modelování.

Statistické srovnání koeficientů ω není možné přímo, nicméně pomocí bootstrappingu lze odhadnout výběrové rozdělení koeficientu ω , které se blíží normálnímu, a jeho standardní chybu SE . Srovnání lze pak provést pomocí z-testu:

$$Z = \frac{|\omega_{TD} - \omega_{VD}|}{\sqrt{SE_{TD}^2 + SE_{VD}^2}}, \quad (2)$$

kde Z je testová statistika, ω_i je koeficient omega pro danou škálu a metodu administrace a SE_i je jeho standardní chyba. Koeficienty α , celkové koeficienty ω a koeficienty ω pro jednotlivé faktory zobrazuje Tabulka 8. V případě mužů je i zde potřeba upozornit na menší velikost vzorku ($n = 97$). Ve zbylých dvou škálách je reliabilita ve VD mírně vyšší. Žádný rozdíl v reliabilitách však není signifikantní, s jediným marginálně signifikantním rozdílem na faktoru pozitivních položek (f1) pro RSES, s vyšší hodnotou ve VD (Tabulka 9).

Tabulka 8: Koeficienty alfa a omega pro škály, jejich faktory a metody administrace

Škála	α	ω Celková	ω f1	ω f2
RSES TD	0,90	0,897	0,806	0,811
RSES VD	0,85	0,925	0,877	0,855
VYS TD ženy	0,90	0,911	0,854	0,854
VYS VD ženy	0,86	0,924	0,886	0,833
VYS TD muži	0,87	0,942	0,915	0,771
VYS VD muži	0,86	0,924	0,827	0,889

Tabulka 9: Srovnání koeficientů omega mezi VD a TD

Škála	ω Celková				ω Faktor 1 (pozitivní/nadprům. výška)				ω Faktor 2 (negativní/podprům. výška)			
	$\Delta\omega$	SE	Z	p	$\Delta\omega$	SE	Z	p	$\Delta\omega$	SE	Z	p
RSES	0,028	0,02	1,36	0,174	0,071	0,04	1,85	0,065	0,043	0,05	0,84	0,402
VYS ženy	0,013	0,02	0,56	0,575	0,032	0,04	0,83	0,406	0,021	0,05	0,43	0,664
VYS muži	0,018	0,05	0,39	0,699	0,088	0,07	1,24	0,216	0,118	0,10	1,21	0,228

Pozn.: SE = bootstrapová standardní chyba; 500–2000 iterací (viz limity: 4.8)

3.6 Analýza invariance měření

Analýza invariance testuje psychometrickou ekvivalenci konstruktů napříč skupinami nebo různými podmínkami (Putnick & Bornstein, 2016). Slouží tak i pro ověření stability faktorové struktury a psychometrické ekvivalence mezi metodami administrace (Taris et al., 1998). V tomto případě bude použita metoda víceskupinové konfirmační faktorové analýzy (*MG-CFA*) pro ordinální data s WLSMV estimátorem a théta parametrizací, která umožňuje posuzovat také ekvivalenci reziduálních rozptylů mezi skupinami (Kline, 2016; Muthén & Asparouhov, 2002). Jednotlivé modely pak byly parametrizovány podle doporučení Wu & Estabrook (2016).

Konfigurální invariance využívá stejnou faktorovou strukturu (3.5.1) pro obě metody administrace. V rámci ordinálních proměnných jsou pak pro metrickou invarianci nejprve zvlášť fixovány prahy, až potom navíc faktorové náboje. Pro ověření skalární invariance jsou navíc fixovány průsečíky a ve striktní invarianci také reziduální rozptyly. Kromě testů exaktní shody založených na $\Delta\chi^2$, které nemusí být ideálním kritériem pro zamítnutí invariance (např.: Meade et al., 2008; Putnick & Bornstein, 2016), je nutné sledovat změny v jiných absolutních a inkrementálních ukazatelích. Možné hodnoty ukazatelů pro zamítnutí invariance jsou $\Delta RMSEA > 0,015$, $\Delta CFI < -0,01$, $\Delta SRMR > 0,03$, kdy zejména u menších vzorků by za hlavní kritérium měl být považován CFI (Chen, 2007). Zde použitý ukazatel TLI pak využívá ekvivalentní hranici jako příbuzný CFI.

V případě dotazníku výšky pro ženy (Tabulka 10) bylo dosaženo striktní (reziduální) invariance, a to i přes signifikantní $\Delta\chi^2$. U žádného modelu RMSEA ani TLI nevykazují zhoršení, ale dokonce implikují zlepšení. $\Delta SRMR$ je vždy zanedbatelná. Pro RSES stejné platí v případě skalární (silné faktorové) invariance (Tabulka 11), a to i navzdory signifikantnímu testu exaktní shody v případě metrické. U striktní invariance došlo k nárůstu RMSEA a snížení TLI, vždy však v přípustných hodnotách. Nárůst SRMR už je mírně vyšší, nicméně se zde model dá ještě považovat za hraničně přípustný. Zdá se tak, že striktní invariance bylo dosaženo v obou případech. Tyto výsledky tak plně podporují hypotézy H1a a H2 o dosažení skalární (silné faktorové) invariance. Ověření invariance dotazníku výšky pro mužské respondenty není kvůli malé velikosti vzorku v odpovídající kvalitě možné. Hypotéza H1b tak zůstane nepodpořena.

Protože by způsob administrace mohl ovlivňovat i rozložení latentního rysu, byly nad rámec běžného testování invariance doplněny modely s fixovanými populačními parametry: průměry, rozptyly (Var) a kovariancemi (CoVar) latentní proměnné. Ani v jednom z těchto modelů nedošlo k zásadnímu zhoršení, jen v případě RSES: CoVar se opět pohybujeme na hranici přípustnosti.

Tabulka 10: Analýza invariance měření mezi metodami administrace (VD/TD) pro dotazník výšky (ženy)

Model	χ^2	df	TLI	RMSEA	RMSEA 95% CI		SRMR	$\Delta\chi^2$	Δdf	p	ΔTLI	$\Delta RMSEA$	$\Delta SRMR$
					Dolní	Horní							
Vys_ž: Konfig	194,07	86	0,950	0,109	0,088	0,129	0,069						
Vys_ž: Prahy	203,48	97	0,956	0,102	0,082	0,121	0,069	10,19	11	0,514	0,006	-0,007	0,000
Vys_ž: Metr	204,79	106	0,963	0,094	0,074	0,113	0,070	11,75	9	0,228	0,007	-0,008	0,001
Vys_ž: Skal	210,10	115	0,967	0,088	0,069	0,107	0,071	10,13	9	0,340	0,004	-0,005	0,000
Vys_ž: Strikt	227,42	126	0,968	0,087	0,069	0,105	0,078	26,19	11	0,006	0,001	-0,001	0,008
Vys_ž: Prům	203,45	128	0,977	0,074	0,055	0,093	0,078	0,67	2	0,716	0,009	-0,013	0,000
Vys_ž: Var	207,24	130	0,976	0,075	0,055	0,093	0,086	5,56	2	0,062	0,000	0,000	0,008
Vys_ž: CoVar	211,59	131	0,975	0,076	0,057	0,094	0,098	5,25	1	0,022	-0,001	0,001	0,011

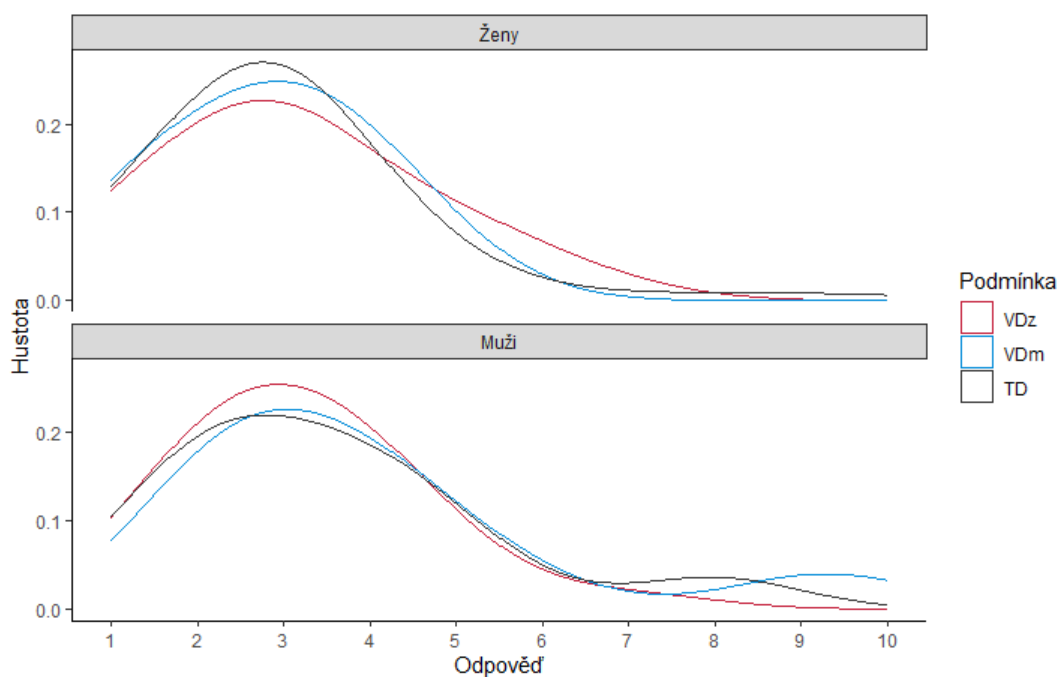
Tabulka 11: Analýza invariance měření mezi metodami administrace (VD/TD) pro Rosenbergovu škálu sebehodnocení

Model	χ^2	df	TLI	RMSEA	RMSEA 95% CI		SRMR	$\Delta\chi^2$	Δdf	p	ΔTLI	$\Delta RMSEA$	$\Delta SRMR$
					Dolní	Horní							
RSES: Konfig	231,52	68	0,951	0,124	0,107	0,142	0,071						
RSES: Prahy	240,40	78	0,957	0,116	0,099	0,133	0,071	14,81	10	0,139	0,007	-0,009	0,000
RSES: Metr	240,83	86	0,963	0,108	0,092	0,124	0,073	19,86	8	0,011	0,006	-0,008	0,002
RSES: Skal	244,19	94	0,967	0,101	0,086	0,117	0,073	7,65	8	0,469	0,004	-0,006	0,000
RSES: Strikt	295,91	104	0,962	0,109	0,094	0,124	0,090	57,93	10	<0,001	-0,005	0,008	0,017
RSES: Prům	255,76	106	0,971	0,095	0,080	0,110	0,090	0,57	2	0,753	0,009	-0,014	0,000
RSES: Var	247,03	108	0,974	0,091	0,076	0,106	0,095	5,66	2	0,059	0,003	-0,004	0,005
RSES: CoVar	263,44	109	0,971	0,095	0,081	0,110	0,112	11,82	1	0,001	-0,003	0,004	0,016

3.7 Položka tělesné konstituce

V případě obrázkové položky tělesné konstituce bylo očekáváno, že se rozložení skóre bude lišit, a to zejména pro ženy vzhledem k sociálně žádoucí povaze této problematiky. Podle očekávání se rozložení mezi VD a TD signifikantně neliší v případě mužů (Fisherův exaktní test: $p = 0,328$), nicméně ani v případě žen ($p = 0,427$). V případě srovnání s přihlédnutím na pohlaví administrátora se rozložení odpovědí žen mezi VD_M a VD_Z signifikantně neliší ($p = 0,334$; Graf 8). Hypotézy H8a (*Rozložení ženami reportované figury se bude mezi VD a TD lišit.*) a H8b (*Rozložení ženami reportované figury se bude mezi VD_M a VD_Z lišit.*) podpořeny být nemohou, i přes mírně vyšší tendenci žen reportovat postavy B a C (Obr. 5) v případě TD (Graf 9).

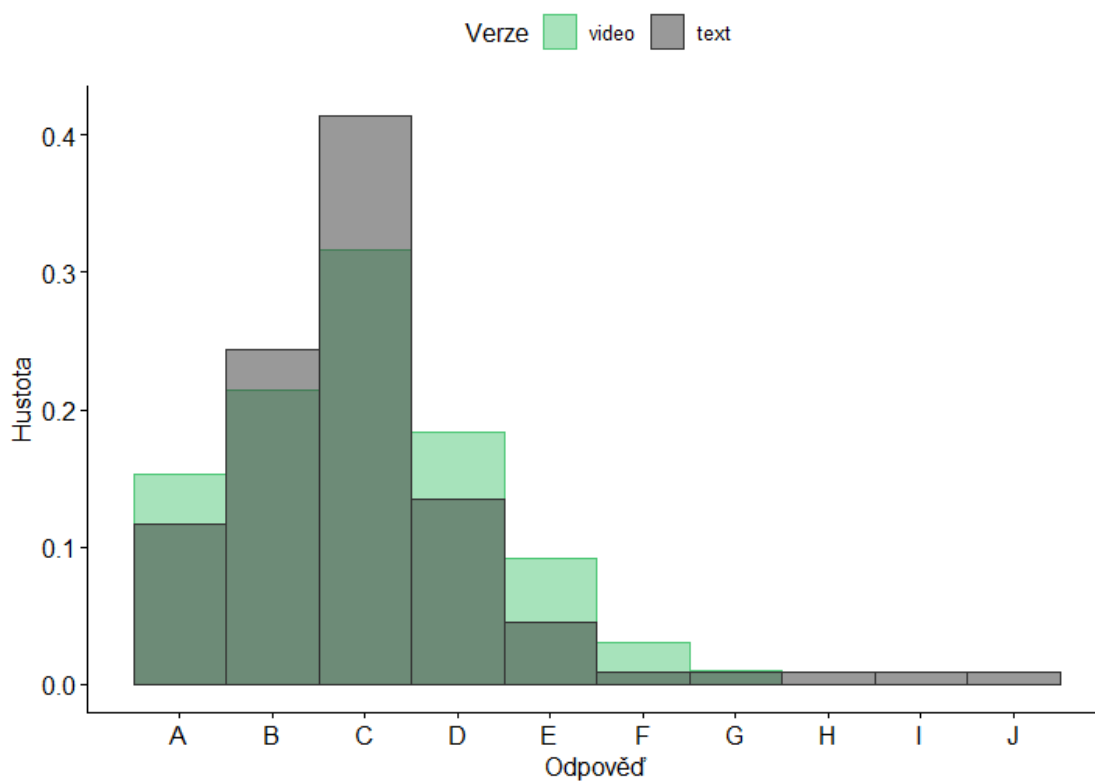
Graf 8: Srovnání rozložení skóre tělesné konstituce mezi podmínkami a pohlavími



Obr. 5: Možné odpovědi položky tělesné konstituce pro ženy



Graf 9: Srovnání rozložení skóreů tělesné konstituce mezi verzemi pro ženy



3.8 Výstupní dotazník a zkušenosti s metodou

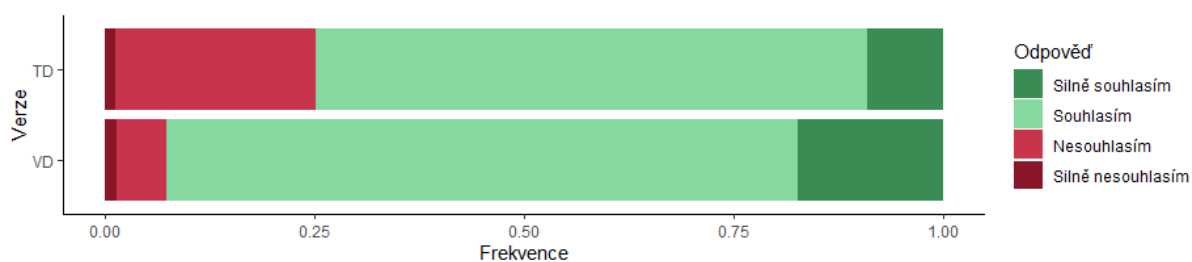
3.8.1 Spokojenost s metodou

Položka „Dotazník mě bavil.“ mohla nabývat 4 hodnot od „Silně nesouhlasím“ po „Silně souhlasím“. Nesouhlasné hodnoty tak značí, že respondenta dotazník nebavil, zatímco souhlasné značí opak. VD ($M = 2,09$; $SD = 0,53$; $Mdn = 2$; $n = 150$) tak byla hodnocena jako zábavnější forma administrace oproti TD ($M = 1,83$; $SD = 0,59$; $Mdn = 2$; $n = 155$), a to statisticky signifikantně (jednostranný Wilcoxon rank-sum: $W = 14204$; $p < 0,001$; $r_{r-b} = 0,22$ (95% CI [0,10; 0,34])). Rozložení reakcí podle forem administrace zobrazuje Graf 10 a Tabulka 12. Z nich je naprosto patrné, že VD produkovala nejen vícero souhlasných odpovědí, značících zábavnost dané metody, ale také vícero odpovědí „Silně souhlasím“ v porovnání s textovou variantou (TD).

Tabulka 12: Rozložení skóre položky spokojenosti

Verze	Silně nesouhlasím		Nesouhlasím		Souhlasím		Silně souhlasím		Σ n
	n	Rel. četnost	n	Rel. četnost	n	Rel. četnost	n	Rel. četnost	
TD	2	1,3%	37	23,9%	102	65,8%	14	9,0%	155
VD	2	1,3%	9	6,0%	113	75,3%	26	17,3%	150
Σ	4	1,3%	46	15,1%	215	70,5%	40	13,1%	305

Graf 10: Rozložení skóre položky „Dotazník mě bavil“

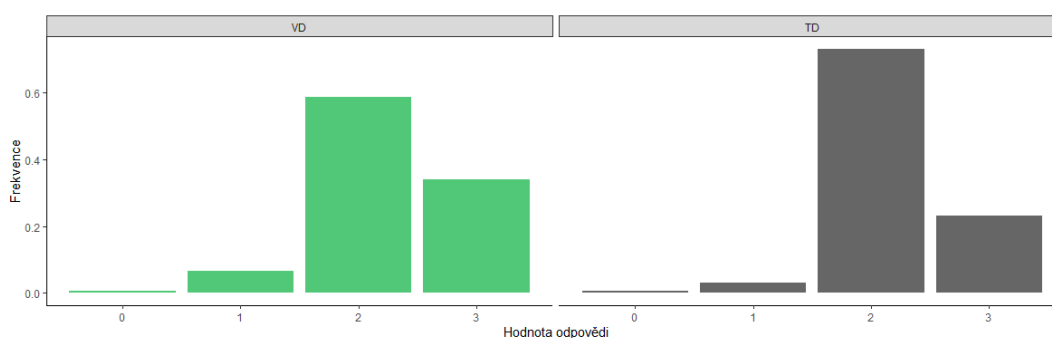


Vyšší hodnota odpovědi „Ve svém volném čase čtu nerad/a.“ znamenala vyšší odpor ke čtení. Vztah mezi těmito hodnotami a vnímanou zábavností metody pro respondenty ve VD ($n = 150$) je slabý, dokonce mírně záporný (Kedallovo $\tau = -0,08$; $z = -1,11$; $p = 0,265$). Experimentální hypotéza H6 (VD budou lépe hodnotit respondenti, kteří neradi čtou.) tak zůstává nepodpořena.

3.8.2 Pozornost věnovaná dotazníku

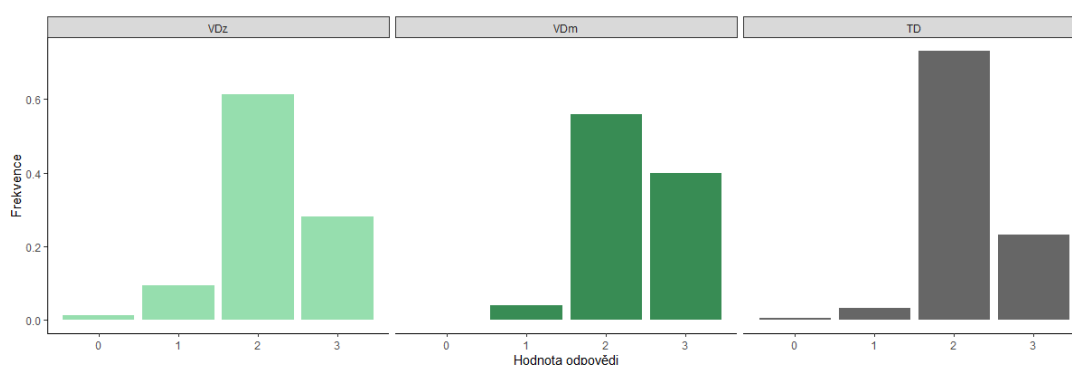
Položka „Na dotazník jsem se plně soustředil/a.“ mohla nabývat 4 hodnot od „Silně nesouhlasím“ po „Silně souhlasím“. Subjektivní pozornost věnovaná dotazníku ve VD ($M = 2,26$; $SD = 0,61$; $Mdn = 2$; $n = 150$) je oproti TD ($M = 2,19$; $SD = 0,51$; $Mdn = 2$; $n = 156$) svým průměrem vyšší, nicméně hodnoty jsou vyšší pouze marginálně (jednostranný Wilcoxon rank-sum test: $W = 12618$; $p = 0,077$; $r_{r-b} = 0,08$ (95% CI [-0,03; 1,00])). To znamená, že H4 (*Subjektivní pozornost věnovaná dotazníku bude vyšší ve VD oproti TD.*) není podpořena. Rozložení skóre se nicméně liší, a to statisticky významně (Fisher's exact test: $p = 0,035$). Jde vidět, že větší procento respondentů označilo „Souhlasím“ v případě TD (73,1 %) oproti VD (58,7 %; Graf 11). Nicméně „Silně souhlasit“ měli větší tendenci respondenti ve VD (34,0 %) oproti TD (23,1 %). Podotkneme, že tato tendence byla také mírně vyšší v případě srovnání na úrovni administrátorů, konkrétně v případě VD_M (40,0 %; Graf 12).

Graf 11: Rozložení skóre subjektivně vnímané pozornosti podle verze



Pozn.: Vyšší hodnota znamená vyšší reportovanou pozornost.

Graf 12: Rozložení skóre subjektivně vnímané pozornosti podle verze a administrátora



Pozn.: Vyšší hodnota znamená vyšší reportovanou pozornost.

Objektivní pozornost byla operacionalizována jako počet přepnutí okna dotazníku (do jiných aplikací/oken). Za předpokladu správné funkcionality se zdá, že okno opustilo jen minimum respondentů (Tabulka 13). Rozložení těchto respondentů se pak ani signifikantně neliší (Fisherův exaktní test: $p = 0,837$).

Tabulka 13: Opouštění okna dotazníku respondenty mezi metodami administrace

Verze	Přepnutí okna respondentem					
	0x	1x	2x	3x	5x	6x
VD	140	11	4	1	0	1
TD	151	8	3	1	1	0

3.9 Časy vyplňování

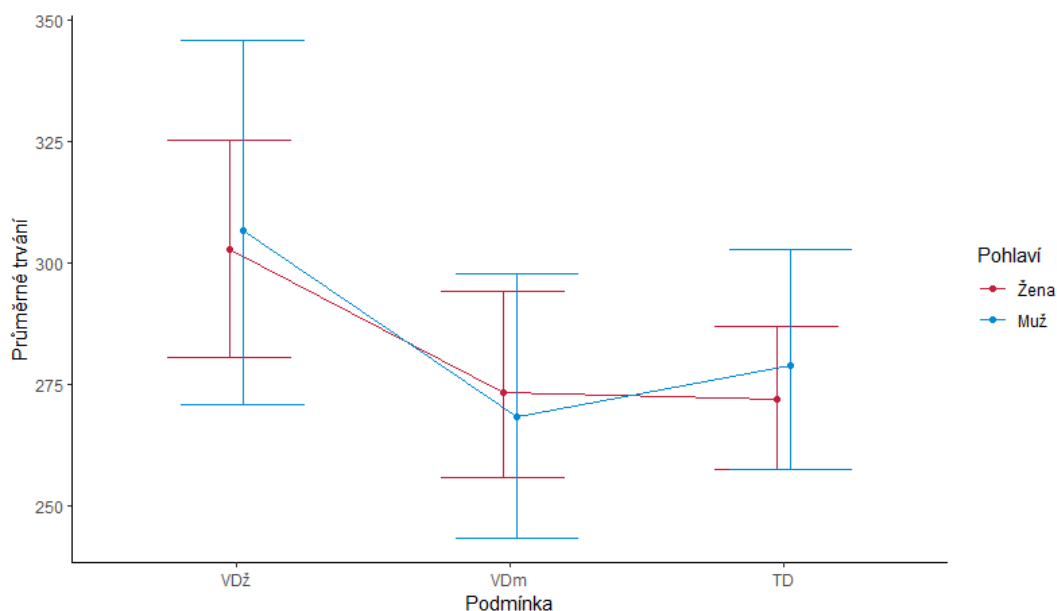
3.9.1 Celkový čas vyplňování

Jednou z možných nevýhod administrace skrze video může být delší čas, který respondenti stráví u dotazníku a který tak může působit jako nadbytečná zátěž. Tento dotazník zabral respondentům, kteří se svým časem spadali mezi 10. a 90. percentil a zároveň jeho průchod dokončili celý ($n = 260$), průměrně $M = 281,82$ sekund ($SD = 71,43$; $Mdn = 264$). Pro srovnání skupin byly brány v potaz časy mezi 10. a 90. percentilem vždy pro každou skupinu zvlášť, aby nedošlo k možnému zkreslení. Respondenti ve VD ($n = 122$) strávili vyplněním dotazníku průměrně $M = 290,61$ sekund ($SD = 73,83$; $Mdn = 269$). Oproti tomu respondentům v TD ($n = 124$) průměrně vyplnění zabralo $M = 273,54$ sekund ($SD = 71,54$; $Mdn = 258,5$). To znamená, že respondentům ve VD zabralo vyplnění průměrně statisticky signifikantně déle (jednostranný t-test: $t(243) = 1.84$; $p = 0,033$; Cohenovo $d = 0,23$). Výsledek jednostranného Wilcoxon rank-sum testu také indikuje, že časy ve VD jsou signifikantně vyšší ($W = 8694$; $p = 0,021$), ale velikost účinku je zde malá ($r_{r-b} = 0,15$ (95% CI [0,01; 0,29])).

Doba trvání videí, které respondent v tomto dotazníku mohl zhlédnout (celkem 24), se mezi administrátory lišila, což je dáno různou rychlostí mluvy (viz Tabulka 14). Zároveň drobný, ale spíše zanedbatelný rozdíl produkovaly i rozdílné gramatické varianty. Je vidět, že ač jsou rozdíly v časech na úrovni položek mnohdy řádově v setinách sekund, s přibývajícím počtem položek má administrátor na trvání už vliv viditelný. S přihlédnutím k tomuto pak respondentům ve VDž ($n = 77$) zabralo vyplnění nejdéle ze všech variant ($M = 305,30$; $SD = 74,98$; $Mdn = 289$). Zatímco skupině ve VD_M ($n = 80$),

zabralo vyplnění nejen kratší dobu v porovnání s VDž, ale dokonce dobu nejkratší ($M = 271,47$; $SD = 64,41$; $Mdn = 256$), i když jen těsně (Graf 13).

Graf 13: Průměrný čas vyplňování dotazníku podle podmínky a pohlaví respondenta



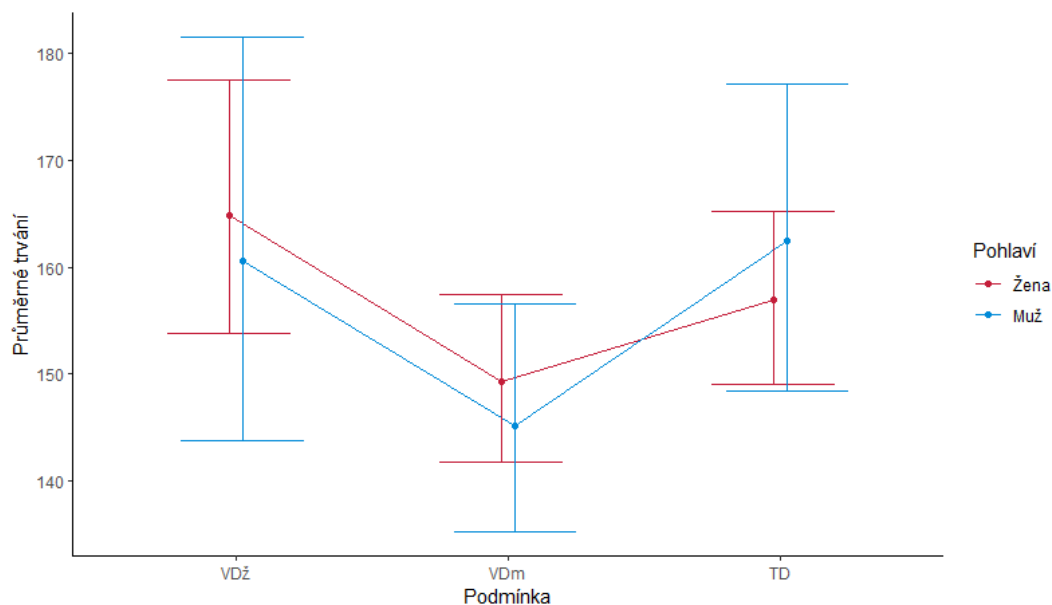
Pozn.: Chybové úsečky značí 95% bootstrapové konfidenční intervaly

3.9.2 Čas strávený v experimentálních podmínkách

Předmětem zájmu by však mohly být pouze experimentální podmínky a kontrolní podmínka bez započítávání časů, které rozdělily předcházejely, případně následovaly ve výstupním dotazníku, aby se striktně oddělil vliv videí. Také nebude započítáván čas strávený na úvodním videu. Tito respondenti tak museli projít (ne nutně zodpovědět) jen položky škál (dotazník výšky, RSES) a tělesné konstituce. V tomto případě byl průměrný čas strávený respondenty v TD ($n = 126$) roven $M = 158,57$ sekund ($SD = 41,45$; $Mdn = 150,10$). Respondenti ve VDž i v tomto případě zodpovídali položky déle ($M = 163,68$; $SD = 38,82$; $Mdn = 152,34$). Respondenti ve VDM strávili vyplněním dobu nejkratší ($M = 147,85$; $SD = 26,77$; $Mdn = 143,41$), což je oproti průměrnému času v TD rozdíl statisticky významný (Welchův t-test: $t(168) = -2,12$; $p = 0,036$; Coheno $d = -0,29$). Graf 14 zobrazuje tento rozdíl v průměrech i podle pohlaví respondenta.

I když v některých případech může administrace skrze video vést k delším časům vyplňování, s přihlédnutím k rozdílnému trvání napříč administrátory tomu tak být nemusí. Hypotéza *H7 (Doba strávená vyplňováním bude vyšší ve VD oproti TD.)*, tak jak byla formulována, může být podpořena, nicméně pouze částečně.⁸

Graf 14: Prům. čas vyplňování v experimentálních podmínkách podle pohlaví respondenta



Pozn.: Chybové úsečky značí 95% bootstrapové konfidenční intervaly

⁸ Pro vizualizaci průměrných časů vyplňování jednotlivých položek, viz Přílohu B.

Tabulka 14: Trvání videí jednotlivých položek VD v různých variantách

Položka	Trvání/s			
	VD _M		VD _ž	
	Ž. rod	M. rod	Ž. rod	M. rod
Úvod	19,07	19,07	23,02	23,02
V1	3,05	3,05	3,14	3,14
V2	5,05	5,05	5,19	5,19
V3	4,19	4,19	4,23	4,23
V4*	2,22	2,23	3,02	3,00
V5	3,08	3,08	3,06	3,06
V6*	5,00	4,22	5,23	5,15
V7*	4,11	4,05	4,18	4,20
V8*	5,13	5,05	5,14	5,22
V9	2,20	2,20	2,22	2,22
V10*	2,20	2,16	2,16	2,15
V11	5,00	5,00	5,11	5,11
R1	6,01	6,01	7,06	7,06
R2*	3,04	3,01	3,09	3,05
R3	2,21	2,21	2,16	2,16
R4*	3,12	3,10	3,06	3,06
R5*	3,10	3,07	3,13	3,07
R6	2,00	2,00	1,19	1,19
R7*	2,09	2,06	2,11	2,11
R8*	2,02	2,00	2,05	2,04
R9	2,13	2,13	3,02	3,02
R10*	2,18	2,15	2,17	2,23
Váha	6,02	6,02	7,02	7,02
Σ	94,22	93,11	101,76	101,70

Pozn.: V = dotazník výšky; R = RSES; Úvod = instruktážní video; *odlišné znění ž. a m. rodu

4 Diskuse

Když Gault (1907) na počátku 20. století poznamenal, že dotazníky nemusejí být jen textové, samozřejmě tak odkazoval na osobní administraci a kontakt respondenta a administrátora tváří v tvář. Co si však ještě dlouho potom dokázali výzkumníci představit jen stěží, bylo, že administraci „tváří v tvář“ bude možné provádět plošně, ve stejný čas a na neomezeném množství respondentů. Jen už ne osobně v pravém slova smyslu. Zatímco rozvoj technologií a internetu umožňuje měnit i jeho obsah, který stále více tvoří video (Cisco, 2019a, 2019b), psychologické dotazníky se sice přesunuly z listů papíru na obrazovky počítačů a záhy na to i mobilních zařízení, jejich elementární podoba, tedy ta textová, však zůstává stále z velké části nezměněna. Vzhledem k oblíbenosti videoobsahu a aplikací, které jej využívají (META, 2022; TikTok, 2021), v kombinaci s často převážně mladými respondenty v psychologických dotaznících, se nabízelo alespoň zvážit možnost určité inovace. Tato práce tak experimentálně prozkoumala využití online administrace psychologického dotazníku skrze video a srovnala tuto metodu s tradiční administrací v textové formě.

4.1 Specifika administrace videem

Pro účely této práce byl upraven program Qualtrics, jehož funkcionalita, ale také funkcionalita ostatních běžných nástrojů pro online průzkumy, není za normálních okolností dostatečně přizpůsobena pro přehrávání krátkých videí a získávání odpovědí na jejich základě. Pro správnou integraci této funkce tak bylo potřeba celou podobu dotazníku navrhnout víceméně od samotného základu.

Při konstrukci podobného dotazníku je nutné myslet zejména na uživatelskou zkušenost respondentů, kteří se s videoobsahem setkávají převážně v dobře optimalizovaných přehrávačích. Vyšší nároky na respondenta, který v tomto případě musí být schopen a ochoten při vyplňování poslouchat zvuk, sami o sobě jako dodatečná zátěž postačují. Nutnost manuálně spouštět každé z desítek přítomných videí, čekání na jejich načtení i přesto, že trvají jen několik sekund, špatně nastavená hlasitost, nesrozumitelnost mluvy administrátorů, matoucí přítomnost ovládacích prvků nebo rozdílné chování dotazníku lišící se s použitým typem zařízení a prohlížeče. To je pouhý nástin komplikací, na které výzkumník ve snaze administrovat dotazník pomocí videa narazí, respektive by mohli narazit jeho respondenti. Ani výchozí podoba sebesposuzovacích škál nemusí být zcela vhodným řešením vzhledem k jejich klasickému znění v první osobě, ale také vzhledem k jejich podobě gramatické, která v textové verzi zahrnuje v českém jazyce obě rodové varianty.

Je tak nutné připustit, že administrace dotazníku pomocí videí je mnohem náročnější jak po technické, tak po časové stránce. Například jakýkoli problém se zněním položek v textové verzi je možné kdykoli napravit jejich přepsáním. Chybné znění položky administrované skrze video vyžaduje po jejím nahrání buďto složitou post-produkční úpravu, znovunatočení dané videopoložky nebo i celé sekvence videí.

4.2 Předčasná ukončení a neúčast

I přes všechny potenciální technické potíže je však možné na základě výsledků konstatovat, že *break-off* nebyl v případě administrace skrze video jakkoli zásadně vyšší. Už vůbec ne pro ohrožení ekologické validity této metody. Z velké části se tak nejspíše povedlo tyto technické problémy vyřešit nebo alespoň minimalizovat, protože případy s nejvážnějšími technickými problémy by k předčasnému ukončení jistě vedly, a to v mnohem větším měřítku.

Výsledky také implikují, že administrace skrze video by mohla vést k dalšímu navýšení počtu vyplnění dotazníků na mobilních zařízeních (mobilní telefony, tablety), které se v této práci blíží 80 %, což je více, než bývá zvykem (např.: Gummer et al., 2023). Také ze získaných metadat se zdá, že rozhodnutí nezúčastnit se odpovídání (*unit-nonresponse*) může být pravděpodobnější pro respondenty na počítačích, i když to s jistotou tvrdit nelze. Velké procento vyplnění na mobilních zařízeních je dáno zejména stylem šíření a sdílení tohoto dotazníku. Na druhou stranu, svou roli může hrát i běžný způsob využití daného typu zařízení. Vzhledem k hojně využívanému streamování hudby a videí v mobilních aplikacích (Spotify, Apple Music, TikTok, Instagram) je možné, že respondenti mají více možností k poslechu zvuku na mobilních zařízeních, už jen kvůli běžné bezprostřední dostupnosti sluchátek.

Je stejně tak nutné podotknout, že tempo, se kterým jsou získáváni respondenti může být v případě administrace videem pomalejší a sběr dat tak zabere delší čas. Protože respondent musí být pro vyplnění dotazníku schopen a ochoten poslouchat zvuk doprovázející video, ne všichni z nich mohou ve chvíli procházení sociálních sítí pozvánku k výzkumu přijmout. Také může být výhodné zvážit alternativní způsoby šíření dotazníku než zřejmě nejběžněji používanou sociální síť Facebook. Je možné, že uživatelé sociálních sítí, na nichž je videoobsah využíván běžněji, jako například Instagram, budou ochotnější a připravenější se takového dotazníku zúčastnit. Tyto domněnky podporuje například problém se získáváním respondentů pro dotazník použitý v této práci, nicméně relevantnější data pro ověření toho tvrzení chybí.

4.3 Dotazník výšky a kriteriální validita

Zkrácená verze dotazníku výšky (Rečka, 2018; Tancoš, 2019) se v obou verzích administrace chovala podle očekávání. Její průměrný testový skór dokonce indikoval výběrovou chybu u reálné výšky mužských respondentů. Kriteriální validita jednotlivých metod administrace byla ověřena pomocí korelace reálné výšky udané respondenty a průměrným skórem dotazníku výšky. Těsné vztahy těchto proměnných se ani v tomto případě nelišily. Vliv metody byl prozkoumán také skrze regresní analýzu s průměrem testového skóru jako závislou proměnnou. Vliv metod administrace byl buďto zcela minimální, nebo v jednom případě pouze marginálně významný. Vliv metody na kriteriální validitu nelze označit za jakkoli významný a v tomto ohledu nelze ani tvrdit, že by administrace skrze video měla být jakýmkoli hendikepem.

4.4 Vlivy sociální žádoucnosti

Rosenbergova škála sebehodnocení (Blatný & Osecká, 1994; Rosenberg, 1965) se svými průměrnými skóry mezi metodami nijak nelišila. Škála byla také v rámci explorační podrobená srovnání skóru na úrovni položek, kvůli svému možnému navázání na sociální žádoucnost (Blatný & Osecká, 1994; Fleming & Courtney, 1984; Jordan, 2020). S některými položkami této škály by mohlo být těžší souhlasit či nesouhlasit v případě, že by u respondenta přítomnost tazatele posilovala sociálně žádoucí vlivy. Tato teorie se však ukázala jako neopodstatněná, kdy se skóry jednotlivých položek napříč verzemi významně nelišily, pouze s marginálním rozdílem u položky č.1 (*Máte pocit, že si uchováváte svoji osobní důstojnost nejméně v takové míře, jako většina ostatních lidí.*). Obsah této položky se nezdá jako nejcitlivější ve srovnání s ostatními (např.: „*Někdy si myslíte, že jste naprosto neschopný/á.*“ apod.). Vzhledem k tomu, že jde o zřejmě nejkomplexnější a poměrně dlouhou položku, nabízí se otázka, zda by u delších položek mohl hrát roli rozdíl v porozumění mezi textovou a mluvenou administrací. Také vzhledem k slabému efektu však nelze něco podobného tvrdit.

Výlučně pro srovnání vlivu sociální žádoucnosti byla zařazena obrázková položka týkající se tělesné konstituce respondentů (Harris et al., 2007). U této položky bylo předpokládáno její rozdílné působení na muže a ženy (např.: Betz et al., 1994; Gil & Mora, 2011) mezi textovou podmínkou a dotazníkem administrovaným videem. Také bylo prozkoumáno možné rozdílné fungování pro ženy v případě administrace mužem, nebo ženou s možnou interakcí jejich pohlaví. V případě žen byla v textové podmínce viditelná lehce vyšší tendence k reportování zejména dvou figur. Analýza nicméně ani jednu z těchto hypotéz nepodpořila. Nelze tak tvrdit, že by odpovědi na tuto položku ovlivňovala sociální žádoucnost, natož způsobovala rozdíl mezi metodami administrace.

Je nutné připustit, že manipulace nemusela být dostatečná a váha není dostatečně citlivým tématem. Zdá se, že pokud by má metoda administrace videem vůbec nějaký vliv na sociální žádoucnost, půjde spíše o velice citlivá témata, jako je sexuální chování, pohlavní nemoci, omamné látky nebo rasismus, jak ukázali například Fuchs (2009) nebo Wilson et al. (2002). Protože stejné se ostatně týká i osobní administrace (metody PAPI, CAPI apod.; Tourangeau & Yan, 2007), bylo by spíše překvapivé, pokud by sociální žádoucnost způsobovala problémy v případě administrace virtuálním tazatelem.

4.5 Psychometrické vlastnosti škál a analýza invariance

Při srovnání metod administrace byly viditelné rozdíly v reliabilitách škál, a to jak v případě vnitřní konzistence vyjádřené Cronbachovou alfou (Cronbach, 1951), tak v případě McDonalduvy omegy (McDonald, 1999). U koeficientu omega došlo ke statistickému srovnání mezi metodami administrace, a to pro celkové koeficienty, ale také pro koeficienty jednotlivých faktorů škál s pomocí bootstrapové standartní chyby. I když se zdálo, že dotazník administrovaný videem může vést k vyšším hodnotám reliability, ani v jednom případě nebyl rozdíl významný. Dotazník administrovaný videem sice nemůže být prohlášen za statisticky významně reliabilnější, jednoznačně však v tomto srovnání nebyl tím horším, většinou právě naopak.

Jako faktorová struktura škál byla využita zřejmě ta nejjednodušší možná, tedy dvoufaktorová s pozitivně a negativně formulovanými položkami. Jen v případě dotazníku výšky došlo k úpravě podle poznatků Rečky (2018), kterou podpořily i modifikační indexy. Je zajímavé, že tato struktura lépe vystihovala data získaná administrací skrze video, jak šlo vidět z ukazatelů shody modelů s daty. Z výsledků provedené analýzy invariance měření se nicméně zdá, že metody administrace invariantní jsou. V obou škálách bylo dosaženo silné faktorové invariance (skalární invariance), jak bylo také předpokládáno. To znamená, že průměry získané z obou metod jsou mezi sebou porovnatelné a rozdíly v získaných skórech existují vlivem rozdílu v měřených konstruktech, nikoli vlivem rozdílu v metodách administrace. V obou případech se nicméně povedlo docílit také reziduální invariance (striktní invariance), ač v případě RSES byly hodnoty hraniční. Striktní invariance je podmínkou pro plnou faktorovou invarianci (Meredith, 1964). Ač některými výzkumníky opomíjena (Kline, 2016), až dosažení této úrovně implikuje, že jednotlivé položky nejsou napříč skupinami systematicky ovlivněny intervenujícími proměnnými, se kterými by v modelu nebylo počítáno (Lubke et al., 2003; A. D. Wu et al., 2007). Zdá se také, že metoda administrace nemá nijak zásadní vliv na rozložení latentních rysů, jak ukázaly dodatečné modely s fixovanými populačními parametry.

Je potřeba zdůraznit, že počet respondentů byl v této práci oproti odhadu nižší a zdaleka ne ideální. To platí zejména pro dotazník výšky, který je rozdělen podle pohlaví. V případě mužů, kde bychom mluvili jen zhruba o 40 respondentech na skupinu, bylo od analýzy invariance upuštěno zcela. Psychometrické vlastnosti administrace pomocí videozáznamů byly k mé dosavadní znalosti prozkoumány vůbec poprvé. Ve výsledku si tato metoda administrace vedla v pilotáži až překvapivě nadějně. Je otázkou, jak a zda se podobné výsledky podaří v budoucnu replikovat, ideálně na mnohem větším počtu respondentů.

4.6 Čas vyplňování

Vyšší čas vyplňování může být považován za přítěž kladenou na respondenta, která může vést k jeho únavě, frustraci a potenciálně také k úplnému ukončení odpovídání. U dotazníku administrovaného videem bylo předpokládáno, že čas vyplnění bude delší, protože celková doba zhlédnutí videa měla zabrat delší čas než doba potřebná k přečtení položky, nemluvě o možných technických překážkách. Tento jev například ve své studii popisují Conrad et al. (2022). Celková doba strávená v dotazníku se opravdu ukázala jako delší v případě videa. Pokud však izolujeme pouze experimentální a kontrolní podmínky a přihlídneme k rozdílnému tempu mluvy administrátorů, tak toto zjištění neplatí.

V případě mužského tazatele, jehož administrace zabrala dobu nejkratší, byla doba odpovídání dokonce významně kratší než u textové verze. Vzhledem k tomu, že čtení by mělo být rychlejší metodou přenosu informací se zdá, že respondenti volili svou odpověď rychleji. To by mohlo nasvědčovat pouhému uspokojivému odpovídání (*satisficing*; Krosnick, 1991), což však tvrdit nelze s přihlédnutím ke srovnatelným výsledkům škál administrovaných videem, například v případě reliabilit. Nabízí se tak, že respondenti mohli být v případě videí více motivováni zaměřit svou pozornost na položky, zvýšit kognitivní snahu zpracovat jejich obsah (1.2.2; např.: Tourangeau, 1984, 2018), a dojít k rozhodnutí už během jejich poslechu. To může být určitou podporou tvrzení, že přítomnost administrátora může podporovat snahu věnovat dotazníku pozornost, jak podotýká například Couper, (2000). Plynule přečtená položka také může mít určité výhody ve srozumitelnosti oproti textové administraci, kdy má respondent číst mnohdy komplexně znějící tvrzení vůbec poprvé.

4.7 Spokojenost s metodou a pozornost

Hypotéza, že dotazník administrovaný skrze video budou respondenti vnímat jako zábavnější, se potvrdila. Tento efekt je pochopitelný s přihlédnutím k obrovskému počtu internetových dotazníků, které mají stále stejnou formu, zato mnohdy diskutabilní

kvalitu a obsah. Odpovídání na dotazník touto formou tak pro mnohé může být vítanou změnou, která vede k lepšímu zážitku. Administrátor, běžně skrytý za textem a krátkým poděkováním, je nyní také alespoň virtuálně přítomný a dotazník tak může simulovat opravdový rozhovor. Jednalo by se tedy o podobný efekt, jakého docílili už Tourangeau et al. (2003), když v internetovém dotazníku kombinovali osobnější oslovení respondentů s fotkami výzkumníků. Jedním z důvodů může být také snižující se tendence a obliba mladých lidí číst (např.: Clark & Teravainen-Goff, 2020). Tato hypotéza byla také ověřena, nicméně tak, jak byla operacionalizována, nepřinesla pro toto tvrzení žádnou podporu.

Subjektivně vnímaná pozornost věnovaná dotazníku s videem byla vyšší pouze marginálně, i když rozložení skóre napovídá, že více respondentů se u dotazníku skrze video cítilo jako „silně“ zaujatých oproti textové verzi. Pokus o objektivní měření pozornosti nepřinesl žádná zásadní zjištění. Sledování přepínání okna dotazníku se i přesto zdá jako možný způsob sledování soustředěnosti respondentů a lze jej využít v jakémkoli internetovém dotazníku. Podotkneme, že celý tento dotazník průměrně zabral respondentům necelých 5 minut. Je otázkou, zda by k jiným závěrům dospěl dotazník, jehož vyplnění by zabralo dobu výrazně delší. Zda by administrace videem pozornost respondentů udržela spíše než textová verze, nebo právě naopak.

4.8 Limity

Největší limitací této práce je bezpochyby malá velikost vzorku. Zejména u mužských respondentů je tak potřeba k získaným výsledkům přistupovat s opatrností. Tento problém tak například přispěl k výběrové chybě u reálné výšky mužských respondentů, která se navzdory randomizaci mezi podmínkami lišila. Ovlivňuje však také strukturní modely, kde se naráží na nízké zastoupení respondentů zejména v krajních hodnotách některých položek („Silně souhlasím“/„Silně nesouhlasím“). To následně může mít vliv, ač zřejmě nijak zásadní, také na bootstrapové odhady výběrového rozdělení koeficientů omega. V případě dotazníku výšky pro muže byla použita funkce z 2000 iterací schopna výpočet dokončit jen zhruba ve čtvrtině případů. Obezřetnost při jakémkoli zobecňování je však na místě ve všech výsledcích, a to nejen v případě mužských respondentů.

4.9 Shrnutí a všeobecná diskuse

Tato práce svým způsobem potvrdila, co naznačovaly už výsledky limitovaného počtu studií, které jsou k dispozici (Conrad et al., 2022; Fuchs, 2009; Fuchs & Funke, 2007; Gerich, 2008; Haan et al., 2017). Administrace skrze video je v porovnání s textovou verzí do velké míry ekvivalentní a nevede k jiným výsledkům, což se v tomto případě

ukázalo i na psychometrické úrovni. Vlivy sociální žádoucnosti se v rámci běžných škál a otázek ukázaly jako zanedbatelné. Metoda administrace dotazníku skrze video nebyla, co se týče výsledků, horší variantou. To lze vidět na srovnatelné kriteriální validitě, reliabilitě i invariantní povaze metod v obou měřených konstruktech. Některé slibné indicie ve prospěch administrace videem, jako je lepší shoda modelů s daty nebo mírně vyšší reliability, si žádají další prozkoumání, nejlépe na větších vzorcích. Také pozitivnější zkušenost respondentů implikuje, že metoda může být při správném provedení výhodnou alternativou k nadužívaným textovým dotazníkům. Delší čas vyplňování, možná nejvýznamnější potenciální nevýhoda videem administrovaných dotazníků, se také ukázala jako opodstatněná jen v některých případech. Za některých okolností překvapivě vedla dokonce k časům kratším.

Měly by se psychologické škály do budoucna raději nahrávat? To jistě ne. O revoluci v psychologickém měření mluvit nelze. Jde však o překvapivě nadějný pokus podobu dotazníků pozměnit, přizpůsobit moderním technologiím, trendům a zejména dnešním respondentům. Podobná metoda administrace také nabízí možnost manipulace velkým množstvím proměnných, sluchových i vizuálních, a tak poskytuje prostor k dalším experimentům. Použitelnost a ekologická validita této metody bude nicméně vždy záležet na jejím technickém provedení.

Použité zdroje

- Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- APA. (n.d.). *APA Dictionary of Psychology*. <https://dictionary.apa.org/questionnaire>
- Behrendt, S. (2022). *lm.beta: Add Standardized Regression Coefficients to lm-Objects* (1.6.2).
- Betz, N. E., Mintz, L., & Speakmon, G. (1994). Gender differences in the accuracy of self-reported weight. *Sex Roles, 30*(7–8), 543–552.
<https://doi.org/10.1007/BF01420801/METRICS>
- Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly, 74*(5), 817–848.
<https://doi.org/10.1093/POQ/NFQ058>
- Biemer, P. P., & Stokes, S. L. (1985). Optimal Design of Interviewer Variance Experiments in Complex Surveys. *Journal of the American Statistical Association, 80*(389), 158. <https://doi.org/10.2307/2288066>
- Blascovich, J., & Tomaka, J. (1991). Measures of Self-Esteem. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 115–160). Academic Press. <https://doi.org/10.1016/B978-0-12-590241-0.50008-3>
- Blatný, M., & Osecká, L. (1994). Rosenbergova škála sebehodnocení: struktura globálního vztahu k sobě. *Československá Psychologie, 38*(6), 481–488.
- Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods, 50*(6), 2586–2596. <https://doi.org/10.3758/S13428-018-1035-6>

- Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, 37(1), 148–154. <https://doi.org/10.3758/BF03206409/METRICS>
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156–2160. <https://doi.org/10.1016/J.CHB.2013.05.009>
- 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>
- Cisco. (2019a). *Cisco Visual Networking Index: Forecast and Trends, 2017-2022*.
- Cisco. (2019b). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2017-2022*.
- Clark, C., & Teravainen-Goff, A. (2020). Children and Young People's Reading in 2019: Findings from Our Annual Literacy Survey. National Literacy Trust Research Report. *National Literacy Trust*. <https://eric.ed.gov/?id=ED607777>
- Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2022). Video in Survey Interviews: Effects on Data Quality and Respondent Experience. *Methods, Data, Analyses*, 0(0), 35. <https://doi.org/10.12758/MDA.2022.13>
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464–494. <https://doi.org/10.1086/318641>
- Couper, M. P., & Rowe, B. (1996). EVALUATION OF A COMPUTER-ASSISTED SELF-INTERVIEW COMPONENT IN A COMPUTER-ASSISTED PERSONAL INTERVIEW SURVEY. *Public Opinion Quarterly*, 60(1), 89–105. <https://doi.org/10.1086/297740>

- Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly*, 67(3), 385–395. <https://doi.org/10.1086/376948>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555/METRICS>
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.
- Daikeler, J., Bosnjak, M., & Manfreda, K. L. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology*, 8(3), 513–539. <https://doi.org/10.1093/JSSAM/SMZ008>
- Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, 25(1), 14. <https://doi.org/10.1093/HER/CYP046>
- de Leeuw, E. D. (2012). Counting and Measuring Online. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 114(1), 68–78. <https://doi.org/10.1177/0759106312437290>
- Devine, P. G. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, 56(1), 5–18. <https://doi.org/10.1037/0022-3514.56.1.5>
- Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, 10(4), e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Dovidio, J. F., Eller, A., & Hewstone, M. (2011). Improving intergroup relations through direct, extended, and other forms of indirect contact. *Group Processes & Intergroup Relations*, 14(2), 147–160. <https://doi.org/10.1177/1368430210390555>

- Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology, 8*(2), 88–102. <https://doi.org/10.1037/1099-9809.8.2.88>
- Ekman, P. (Ed.). (1982). *Emotion in the Human Face* (2nd ed.). Cambridge University Press.
- Eurostat. (2023a). *Households with access to the internet at home*. https://ec.europa.eu/eurostat/databrowser/view/isoc_r_iacc_h/default/table?lang=en
- Eurostat. (2023b). *Internet use by individuals*. <https://ec.europa.eu/eurostat/databrowser/view/tin00028/default/table?lang=en>
- Evan, W. M., & Miller, J. R. (1969). Differential effects on response bias of computer vs. conventional administration of a social science questionnaire: An exploratory methodological experiment. *Behavioral Science, 14*(3), 216–227. <https://doi.org/10.1002/BS.3830140306>
- Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology, 46*(2), 404–421. <https://doi.org/10.1037/0022-3514.46.2.404>
- Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third Edition). Sage.
- Fuchs, M. (2009). Gender-of-Interviewer effects in a video-enhanced Web survey. *Social Psychology, 40*(1), 37–42. <https://doi.org/10.1027/1864-9335.40.1.37>
- Fuchs, M., & Funke, F. (2007). Video Web Survey - Results of an experimental comparison with a text-based Web survey. In M. Trotman (Ed.), *The challenges of a changing world. Proceedings of the fifth international conference of the association for survey computing* (pp. 63–80). ASC.
- Gault, R. H. (1907). A History of the Questionnaire Method of Research in Psychology. *Pedagogical Seminary, 14*(3), 366–383. <https://doi.org/10.1080/08919402.1907.10532551>

- Gerich, J. (2008). Real or Virtual? Response Behavior in Video-Enhanced Self-Administered Computer Interviews. *Field Methods*, 20(4), 356–376.
<https://doi.org/10.1177/1525822X08320057>
- Gil, J., & Mora, T. (2011). The determinants of misreporting weight and height: The role of social norms. *Economics & Human Biology*, 9(1), 78–91.
<https://doi.org/10.1016/J.EHB.2010.05.016>
- Gordon, M. E., & Gross, R. H. (1978). A critique of methods for operationalizing the concept of fakeability. *Educational and Psychological Measurement*, 38(3), 771–782. <https://doi.org/10.1177/001316447803800316>
- Gummer, T., Höhne, J. K., Rettig, T., Roßmann, J., & Kummerow, M. (2023). Is there a growing use of mobile devices in web surveys? Evidence from 128 web surveys in Germany. *Quality and Quantity*.
<https://doi.org/10.1007/s11135-022-01601-8>
- Haan, M., Ongena, Y. P., Vannieuwenhuyze, J. T. A., & De Glopper, K. (2017). Response Behavior in a Video-Web Survey: A Mode Comparison Study. *Journal of Survey Statistics and Methodology*, 5(1), 48–69.
<https://doi.org/10.1093/JSSAM/SMW023>
- Harris, C. V., Bradlyn, A. S., Coffman, J., Gunel, E., & Cottrell, L. (2007). BMI-based body size guides for women and men: development and validation of a novel pictorial method to assess weight-related concepts. *International Journal of Obesity* 2008, 32(2), 336–342. <https://doi.org/10.1038/sj.ijo.0803704>
- Heerwegh, D., & Loosveldt, G. (2008). Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly*, 72(5), 836–846. <https://doi.org/10.1093/POQ/NFN045>
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, 67(1), 79–125. <https://doi.org/10.1086/346010>

- Holbrook, A. L., Johnson, T. P., & Krysan, M. (2019). Race- and Ethnicity-of-Interviewer Effects. In Lavrakas Paul, Traugott Michael, Kennedy Courtney, Holbrook Allyson, de Leeuw Edith, & West Brady (Eds.), *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment* (pp. 197–224). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781119083771.CH11>
- Holtgraves, T. (2004). Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin*, 30(2), 161–172. <https://doi.org/10.1177/0146167203259930>
- Johnson, T. P., Fendrich, M., Shaligram, C., Garcy, A., & Gillespie, S. (2000). An evaluation of the effects of interviewer characteristics in an RDD telephone survey of drug use. *Journal of Drug Issues*, 30(1), 77–102.
<https://doi.org/10.1177/002204260003000105>
- Jordan, C. H. (2020). Rosenberg Self-Esteem Scale. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 4518–4520). Springer, Cham. https://doi.org/10.1007/978-3-319-24612-3_1155
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2023). *sem-Tools: Useful tools for structural equation modeling*. (0.5-6.917).
- Kish, L. (1962). Studies of Interviewer Variance for Attitudinal Variables. *Journal of the American Statistical Association*, 57(297), 92–115.
<https://doi.org/10.1080/01621459.1962.10482153>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th edition). Guilford Press.
- Krantz, J. H., & Dalal, R. (2000). Validity of Web-Based Psychological Research. In Birnbaum Michael H. (Ed.), *Psychological Experiments on the Internet* (pp. 35–60). Academic Press. <https://doi.org/10.1016/B978-012099980-4/50003-4>
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly*, 72(5), 847–865. <https://doi.org/10.1093/POQ/NFN063>

- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.
<https://doi.org/10.1002/ACP.2350050305>
- Krysan, M., & Couper, M. P. (2003). Race in the Live and the Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys. *Social Psychology Quarterly*, 66(4), 364–383. <https://doi.org/10.2307/1519835>
- Leisher, C. (2014). A Comparison of Tablet-Based and Paper-Based Survey Data Collection in Conservation Projects. *Social Sciences*, 3(2), 264–271.
<https://doi.org/10.3390/SOCSCI3020264>
- Liu, M., & Stainback, K. (2013). Interviewer Gender Effects on Survey Responses to Marriage-Related Questions. *Public Opinion Quarterly*, 77(2), 606–618.
<https://doi.org/10.1093/POQ/NFT019>
- Liu, M., & Wang, Y. (2016). Interviewer Gender Effect on Acquiescent Response Style in 11 Asian Countries and Societies. *Field Methods*, 28(4), 327–344.
<https://doi.org/10.1177/1525822X15623755>
- Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology*, 56(2), 231–248. <https://doi.org/10.1348/000711003770480020>
- Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50(1), 79–104.
<https://doi.org/10.1177/147078530805000107>
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882–888.
<https://doi.org/10.1037/0022-006X.51.6.882>

- McDonald, R. P. (1999). *Test Theory*. Psychology Press.
<https://doi.org/10.4324/9781410601087>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/MET0000144>
- Meade, A. W., & Bauer, D. J. (2007). Power and Precision in Confirmatory Factor Analytic Tests of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 611–635.
<https://doi.org/10.1080/10705510701575461>
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology*, 93(3), 568–592. <https://doi.org/10.1037/0021-9010.93.3.568>
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29(2), 177–185.
<https://doi.org/10.1007/BF02289699/METRICS>
- META. (2022, October 26). *Third Quarter 2022 Results Conference Call*.
https://s21.q4cdn.com/399680738/files/doc_financials/2022/q3/Meta-Q3-2022-Earnings-Call-Transcript.pdf
- Midena, D., & Yeo, R. (2022). Towards a history of the questionnaire. *Intellectual History Review*, 32(3), 503–529. <https://doi.org/10.1080/17496977.2022.2097576>
- Milton, A. C., Ellis, L. A., Davenport, T. A., Burns, J. M., & Hickie, I. B. (2017). Comparison of Self-Reported Telephone Interviewing and Web-Based Survey Responses: Findings from the Second Australian Young and Well National Survey. *JMIR Mental Health*, 4(3). <https://doi.org/10.2196/MENTAL.8222>
- Mitchell, K. R., Mercer, C. H., Prah, P., Clifton, S., Tanton, C., Wellings, K., & Copas, A. (2018). Why Do Men Report More Opposite-Sex Sexual Partners Than Women? Analysis of the Gender Discrepancy in a British National Probability Survey. *The Journal of Sex Research*, 56(1), 1–8.
<https://doi.org/10.1080/00224499.2018.1481193>

- Muthén, B., & Asparouhov, T. (2002). Latent Variable Analysis with Categorical Outcomes: Multiple-Group and Growth Modeling in Mplus. *Mplus Web Notes*, 4(5). <http://www.statmodel.com/download/webnotes/CatMGLong.pdf>
- Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27(10), 864–876. <https://doi.org/10.1111/J.1559-1816.1997.TB00275.X>
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. <https://doi.org/10.1002/EJSP.2420150303>
- O'Reilly, J. M., Hubbard, M. L., Lessler, J. T., Biemer, P. P., & Turner, C. F. (1994). AUDIO AND VIDEO COMPUTER-ASSISTED SELF INTERVIEWING: Preliminary Tests of New Technologies for Data Collection. *Journal of Official Statistics*, 10(2), 197.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>
- Paulhus, D. L. (1989). Socially Desirable Responding: Some New Solutions to Old Problems. In D. M. Buss & N. Cantor (Eds.), *Personality Psychology* (pp. 201–209). Springer. https://doi.org/10.1007/978-1-4684-0634-4_15
- Paulhus, D. L. (1991). Measurement and Control of Response Bias. In Robinson J. P., Shaver P. R., & Wrightsman L. S. (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17–59). Elsevier. <https://doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65–85. <https://doi.org/10.1146/annurev.psych.49.1.65>
- Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, 38(6), 922–934. <https://doi.org/10.1002/EJSP.504>

- Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review, 41*, 71–90. <https://doi.org/10.1016/J.DR.2016.06.004>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.2.3). R Foundation for Statistical Computing.
- Rečka, K. (2018). *Dotazník výšky a váhy* [Diplomová práce, Masarykova univerzita]. <https://is.muni.cz/th/ug7c2/>
- Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research* (2.3.3).
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2). <https://doi.org/10.18637/jss.v048.i02>
- Rubin, M., & Donkin, C. (2022). Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2022.2113771>
- Smith, T. W., & Kim, J. (2015). A Review of Survey Data-Collection Modes: With a Focus on Computerizations. *Sociological Theory and Methods, 30*(2), 185–200. <https://doi.org/10.11218/OJJAMS.30.185>
- Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*(2), 97–124. https://doi.org/10.1207/S15327051HCI1102_1
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass.
- Tancoš, M. (2019). *Vliv verbálních kotev Likertovy škály na psychometrické charakteristiky dotazníků* [Diplomová práce, Masarykova univerzita]. <https://is.muni.cz/th/uk8cb/>
- Taris, T. W., Meijer, Z. Y., Bok, I. A., & Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different

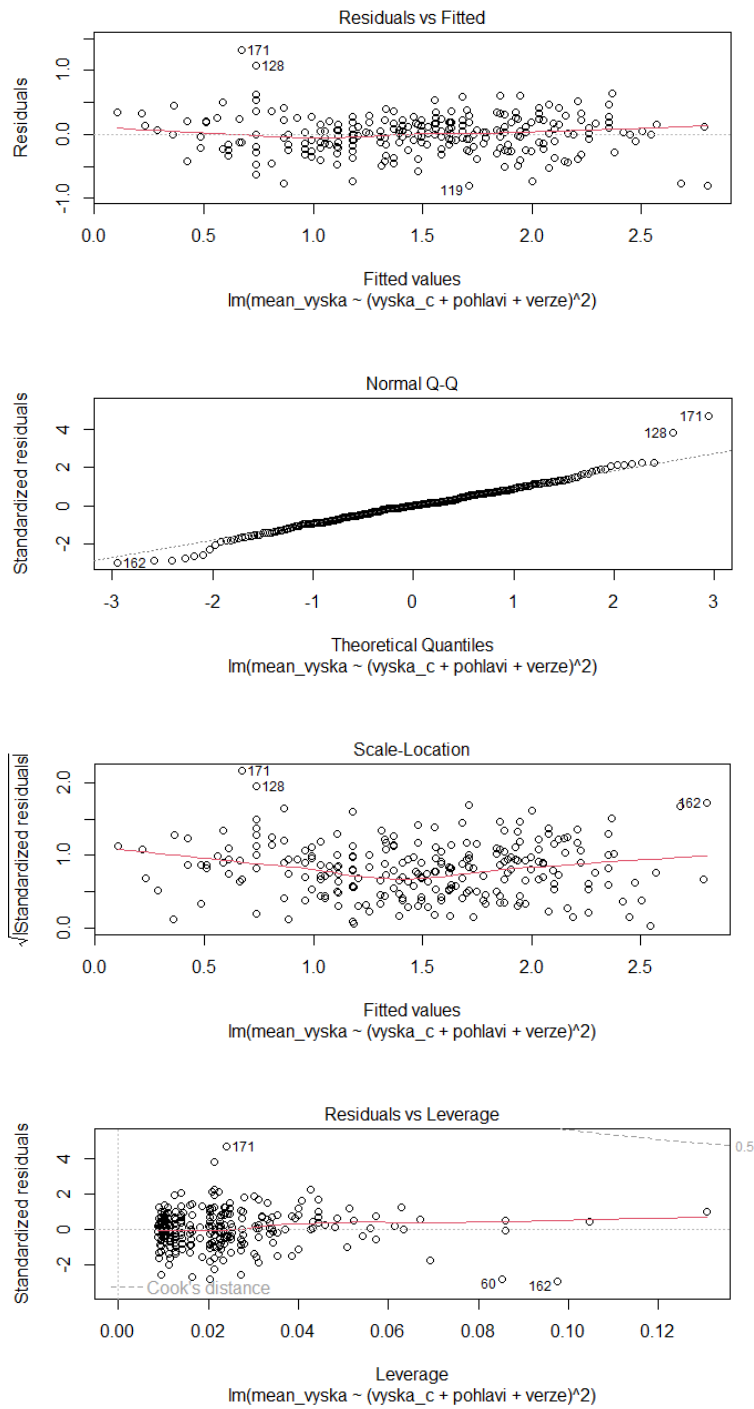
- situations: A general approach. *Journal of Psychology: Interdisciplinary and Applied*, 132(3), 301–316. <https://doi.org/10.1080/00223989809599169>
- TikTok. (2021, September 27). *Thanks a billion!* <https://newsroom.tiktok.com/en-us/1-billion-people-on-tiktok>
- Tourangeau, R. (1984). Cognitive science and survey methods. In T. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive Aspects of Survey Design: Building a Bridge between Disciplines*. National Academy Press.
- Tourangeau, R. (2018). The survey response process from a cognitive viewpoint. *Quality Assurance in Education*, 26(2), 169–181. <https://doi.org/10.1108/QAE-06-2017-0034>
- Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior*, 19(1), 1–24. [https://doi.org/10.1016/S0747-5632\(02\)00032-8](https://doi.org/10.1016/S0747-5632(02)00032-8)
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60(2), 275–304. <https://doi.org/10.1086/297751>
- Tourangeau, R., Sun, H., Yan, T., Maitland, A., Rivero, G., & Williams, D. (2018). Web Surveys by Smartphones and Tablets: Effects on Data Quality. *Social Science Computer Review*, 36(5), 542–556. <https://doi.org/10.1177/0894439317719438>
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
- Urbánek, T. (2000). *Strukturální modelování v psychologii*. AV ČR, Psychologický ústav.
- Walker, J. H., Sproull, L., & Subramani, R. (1994). Using a human face in an interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, 85–91. <https://doi.org/10.1145/191666.191708>

- West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5(2), 175–211.
<https://doi.org/10.1093/JSSAM/SMW024>
- West, B. T., & Olson, K. (2010). How Much of Interviewer Variance is Really Nonresponse Error Variance? *Public Opinion Quarterly*, 74(5), 1004–1026.
<https://doi.org/10.1093/POQ/NFQ061>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
<https://doi.org/10.21105/joss.01686>
- Williams, J. A. (1964). Interviewer-Respondent Interaction: A Study of Bias in the Information Interview. *Sociometry*, 27(3), 338–352.
<https://doi.org/10.2307/2785623>
- Wilson, S. R., Brown, N. L., Mejia, C., & Lavori, P. W. (2002). Effects of Interviewer Characteristics on Reported Sexual Behavior of California Latino Couples. *Interviewer Effects Hispanic Journal of Behavioral Sciences*, 24(1), 38–62.
<https://doi.org/10.1177/0739986302024001003>
- Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3). <https://doi.org/10.1111/j.1083-6101.2005.tb00259.x>
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration with TIMSS Data. *Practical Assessment, Research, and Evaluation*, 12, 3. <https://doi.org/10.7275/mhqa-cd89>

- Wu, H., & Estabrook, R. (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika*, *81*(4), 1014–1045. <https://doi.org/10.1007/S11336-016-9506-0>
- Zajonc, R. B. (1965). Social facilitation. *Science*, *149*(3681), 269–274. <https://doi.org/10.1126/SCIENCE.149.3681.269>
- Zhang, X. C., Kuchinke, L., Woud, M. L., Velten, J., & Margraf, J. (2017). Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behavior*, *71*, 172–180. <https://doi.org/10.1016/J.CHB.2017.02.006>

PŘÍLOHA A: Předpoklady regresní analýzy

Diagnostika bloku 3a lineární regrese s označenými vlivnými případy

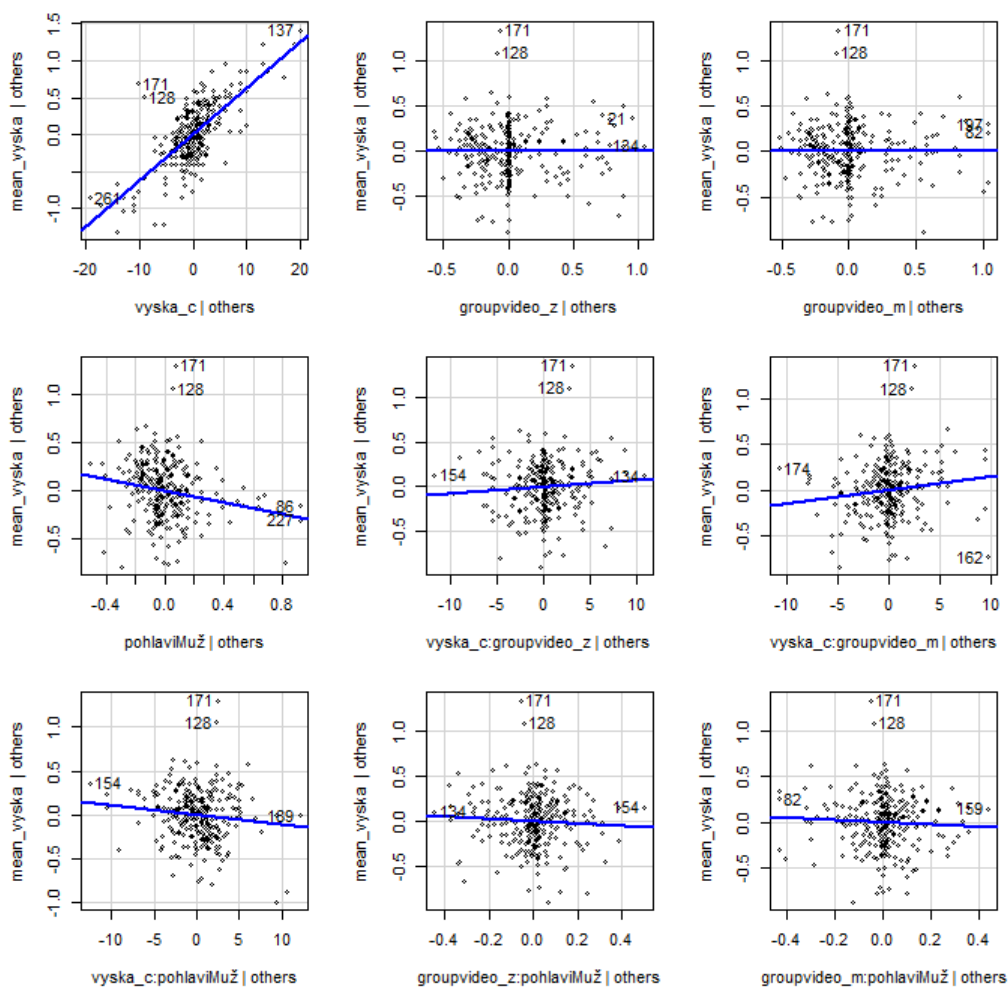


Kolinearita prediktorů

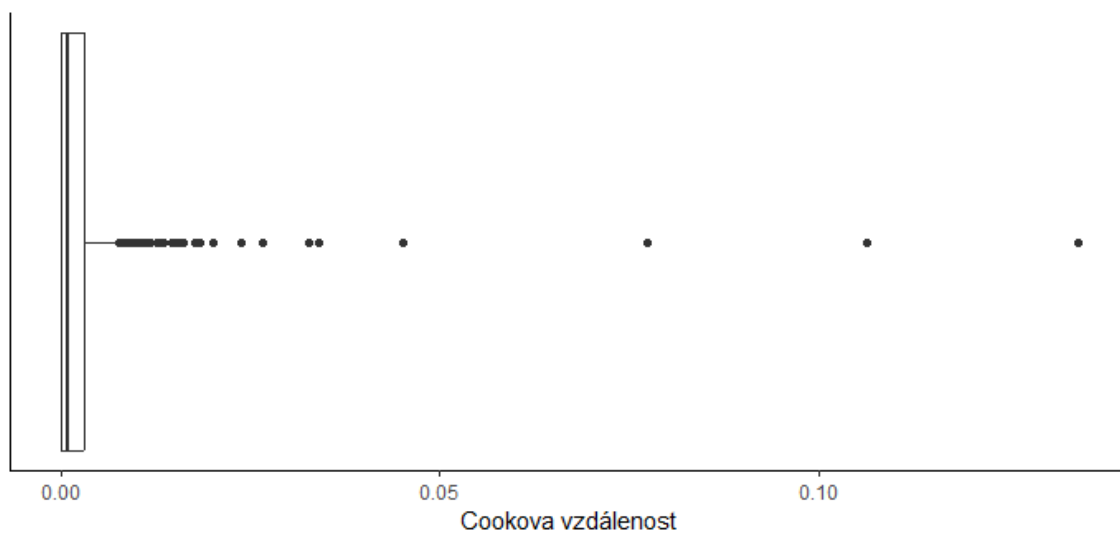
Proměnná	VIF
Pohlaví	5,82
Verze	1,92
Výška	4,36
Výška:Verze	3,56
Výška:Pohlaví	3,95
Verze:Pohlaví	5,42

Vizualizace parciálních korelací

Added-Variable Plots

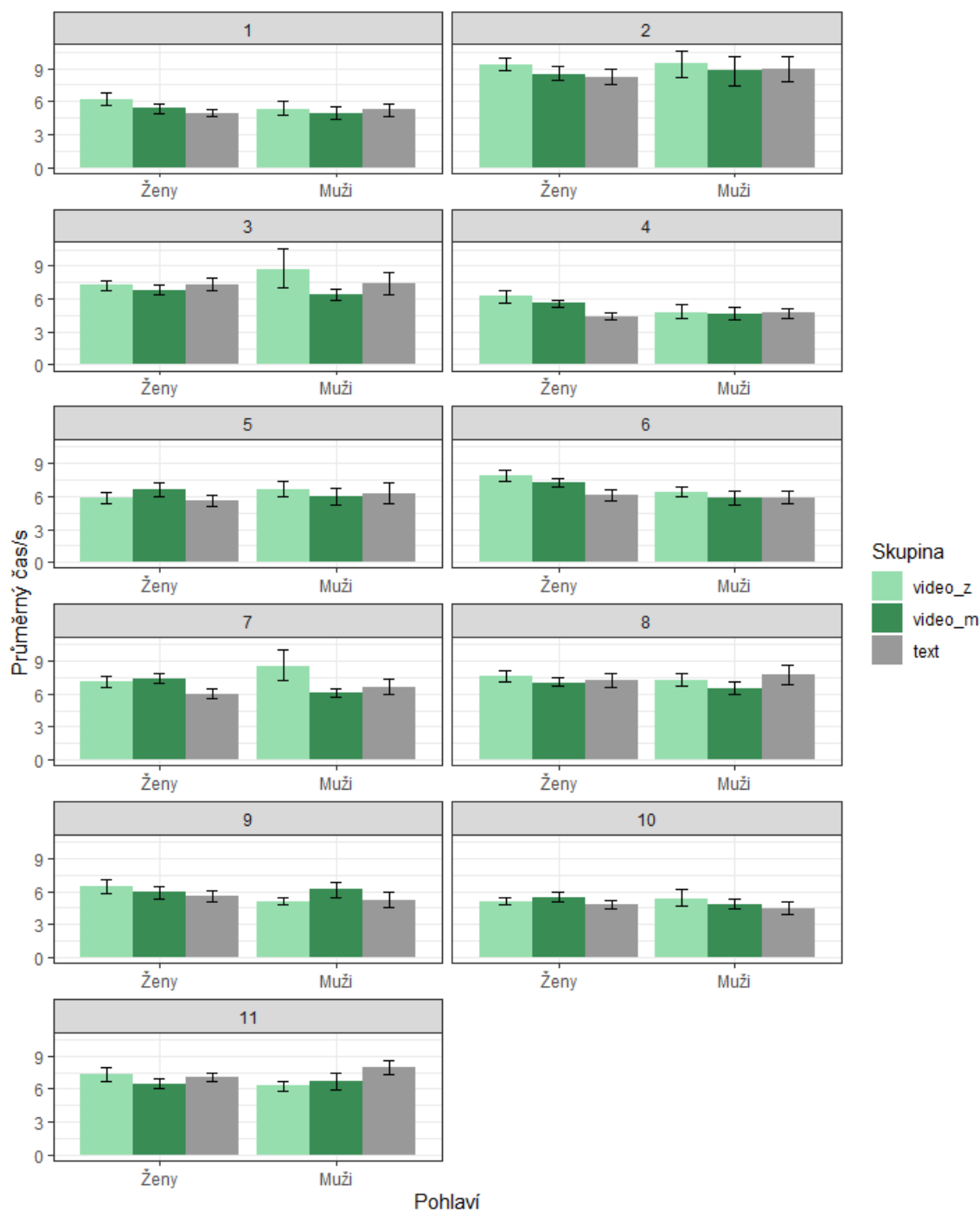


Boxplot Cookovy vzdálenosti modelu 3a hierarchické regrese



PŘÍLOHA B: Průměrné časy odpovídání na položky

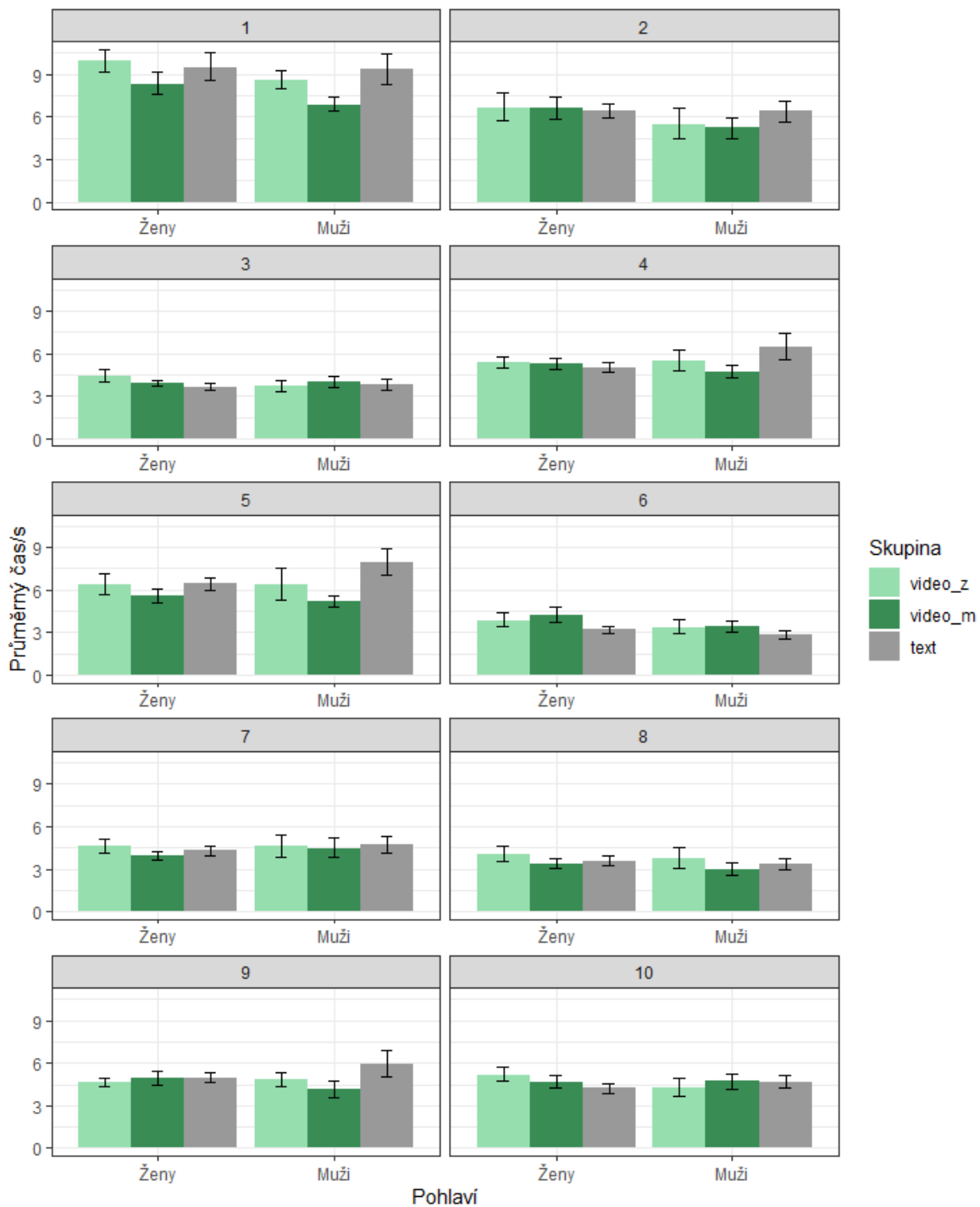
Průměrné časy odpovídání na položky dotazníku výšky podle podmínky a pohlaví



Chybové úsečky značí 95% bootstrapové konfidenční intervaly; video_z/_m = administrátor žena/muž

PŘÍLOHA B: PRŮMĚRNÉ ČASY ODPOVÍDÁNÍ NA POLOŽKY

Průměrné časy odpovídání na položky RSES podle podmínky a pohlaví



Chybové úsečky značí 95% bootstrapové konfidenční intervaly; video_z /_m = administrátor žena/muž