



Metoda ukotvujících vinět a její využití v pedagogickém výzkumu

Hana Voňková



PEDAGOGICKÁ FAKULTA
Univerzita Karlova

Metoda ukotvujících vinět a její využití v pedagogickém výzkumu

Hana Voňková

Dedikace:

Kniha je výstupem projektu GA ČR: Vztahy mezi dovednostmi, vzděláváním a výsledky na trhu práce: longitudinální studie (číslo P402/12/G130).

Recenzovali:

Doc. Mgr. et Mgr. Kateřina Vlčková, Ph.D.

Ing. Bc. Stanislav Michek, Ph.D.

Jazyková korektura:

František Rezek

Obsah

Předmluva	11
Úvod	13
1. Teoretické základy metody ukotvujících vinět	19
1.1. Metodologické problémy při porovnávání skupin	19
1.1.1. Zkreslení (<i>bias</i>) a jeho taxonomie	20
1.1.2. Zkreslení a metoda ukotvujících vinět	22
1.2. Základní myšlenka metody ukotvujících vinět	24
1.2.1. Porovnání (sebe)hodnocení dvou hypotetických respondentů	26
1.2.2. Porovnání rozdělení (sebe)hodnocení ve dvou hypotetických zemích	29
2. Aplikace v sociálních vědách	33
2.1. Klíčové aplikace v sociálních vědách	34
2.1.1. Vnímaná politická účinnost (<i>political efficacy</i>)	34
2.1.2. Ostrost zraku (<i>visual acuity</i>)	36
2.1.3. Zdravotní problémy při výkonu práce (<i>work disability</i>)	38
2.2. Dosavadní využití metody ukotvujících vinět v pedagogickém výzkumu	41
2.2.1. Rodičovské hodnocení spolupráce s učiteli, ředitelem a ostatními rodiči	43
2.2.2. Žákovské sebehodnocení motivace, osobnostních charakteristik, znalostí, nekázně a životní spokojenosti	44
2.2.3. Žákovské hodnocení učitelova řízení třídy a učitelovy podpory žáků	53

3.	Statistické přístupy metody ukotvujících vinět	57
3.1.	Neparametrický (jednoduchý) přístup	57
3.1.1.	Základy neparametrického přístupu pro začínající výzkumníky	57
3.1.2.	Korigované sebehodnocení – výčet všech relativních pozic sebehodnocení vůči hodnocení vinět	64
3.1.3.	Intervalové hodnoty korigovaného sebehodnocení	68
3.2.	Parametrický model	73
3.2.1.	Specifikace modelu	73
3.2.2.	Odhad parametrů modelu metodou maximální věrohodnosti	76
3.2.3.	Identifikace parametrů modelu	80
3.2.4.	Rozšíření standardního modelu	82
4.	Zpracování dat ve statistických softwarech	89
4.1.	Neparametrický přístup	89
4.1.1.	Pořadí vinět	92
4.1.2.	Výpočet korigovaného sebehodnocení	97
4.1.3.	Výběr vinět	103
4.1.4.	Grafické znázornění korigovaného sebehodnocení	104
4.2.	Parametrický model	108
5.	Formulace vinět a jejich umístění v dotazníku	115
5.1.	Předpoklady metody ukotvujících vinět	115
5.1.1.	Předpoklad konzistence odpovědí (<i>response consistency</i>)	115
5.1.2.	Předpoklad ekvivalence vinět (<i>vignette equivalence</i>)	124
5.2.	Diskriminační síla, počet a pořadí vinět	126
5.2.1.	Diskriminační síla vinět	126
5.2.2.	Počet vinět a počet respondentů	127
5.2.3.	Pořadí (sebe)hodnocení a vinět	129

5.3.	K obsahu vinět	130
6.	Nová aplikace parametrického modelu: porovnání učitelovy podpory žáků v zemích z výzkumu PISA 2012	133
6.1.	Cíle výzkumu	133
6.2.	Data z výzkumu PISA 2012	134
6.2.1.	Důležitost zařazení ukotvujících vinět do žákovského dotazníku PISA 2012	134
6.2.2.	Deskriptivní statistiky hodnotících otázek a ukotvujících vinět	137
6.3.	Model CHOPIT a jeho rozšíření pro analýzu dat	147
6.4.	Heterogenita ve stylu odpovídání a její dopad na hodnocení učitelovy podpory žáků	149
6.4.1.	Heterogenita ve stylu užívání jednotlivých kategorií škály	149
6.4.2.	Vztah mezi stylem užívání různých kategorií škály	161
6.4.3.	Korigované versus nekorigované hodnocení učitelovy podpory na úrovni zemí	163
6.5.	Porovnání výsledků analýzy s dalšími studiemi	171
7.	Alternativní postupy korekce	175
7.1.	Tendence k volbě určitých škálových kategorií	178
7.2.	Technika přehánění (<i>overclaiming technique</i>)	184
7.3.	Hodnocení nucenou volbou (<i>forced choice assessment</i>)	188
7.4.	Test posouzení situace (<i>situational judgement test</i>)	190
8.	Závěr	193
	Literatura	197

1.	Příloha: Příklady ukotvujících vinět ve výzkumech zdraví a vnímané politické účinnosti (<i>political efficacy</i>)	211
2.	Příloha: Příklady ukotvujících vinět v pedagogickém výzkumu	223
3.	Příloha: Nová aplikace parametrického modelu – deskriptivní statistiky	233
4.	Příloha: Hodnocení nucenou volbou (<i>forced choice assessment</i>) – příklad otázek	245
5.	Příloha: Test posouzení situace (<i>situational judgement test</i>) – příklad otázek	247
	Seznam tabulek	250
	Seznam obrázků	253
	Summary	255
	Rejstřík	259

Věnováno mamince. Chybíš nám.

Poděkování

Děkuji oběma recenzentům, docentce Kateřině Vlčkové a doktoru Stanislavu Michkovi, za jejich konstruktivní a velmi detailní připomínky, které přispěly k zásadnímu zkvalitnění textu. Autorům jiných publikací bych přála podobně kvalifikované recenzenty. Dále děkuji kolegovi doktoru Michalu Zvírotskému za jeho předmluvu, která se vztahuje nejen k odbornému pohledu na metodologii pedagogického výzkumu, ale také k mému vědeckému zrání pod vedením docenta Petra Byčkovského. Můj dík patří i doktoru Jaroslavu Šárochovi, se kterým jsem diskutovala úskalí překládání mnohých termínů z anglického do českého jazyka, což nepochybně přispělo ke kultivaci předloženého textu. Za jazykovou korekturu děkuji panu Františku Rezkovi. V neposlední řadě děkuji svému manželovi magistru Davidu Voňkovi za (nekritickou) podporu a (kritickou) diskuzi nad předkládanou monografií.

Předmluva

Právě před dvaceti lety, v prvním vydání dnes již klasické vysokoškolské učebnice a zároveň vědecké monografie *Moderní pedagogika*, konstatoval Jan Průcha neutěšený stav pedagogického výzkumu v naší zemi. Poukázal mj. na absenci mladých vědeckých pracovníků, nedostatek výzkumných pracovišť, nekoordinovanost pedagogického výzkumu a jeho nedostatečnou internacionalizaci. O šest let později provedla Eliška Walterová expertní šetření vývoje a stavu české pedagogiky, v němž vedle řady přetrvávajících slabin oboru identifikovala několik povzbudivých trendů – mezi nimi rozvoj metodologie pedagogického výzkumu, obohacování metodologického inventáře a kultivaci výzkumných metod. V monotematickém čísle *Pedagogické orientace* z roku 2016 (2016/3) hovoří jeho editoři Petr Knecht, Kateřina Lojdrová a Jana Majerčíková dokonce o renesanci pedagogické metodologie v České republice. V naší pedagogické literatuře poslední dobou přibývá pojednání o nových metodologických přístupech, často včetně jejich hlubší epistemologické reflexe; četné metody jsou úspěšně přejímány ze zahraničí anebo upravovány pro specifické potřeby našeho pedagogického bádání.

Příkladem takového uvedení nové metody do českého pedagogického výzkumu je počín Hany Voňkové, který systematicky představuje předkládaná monografie. Autorka v původním článku publikovaném v časopise *Orbis Scholae* (2012) zavedla pro tuto metodu české pojmenování metoda ukotvujících vinět (anglicky *anchoring vignettes method*). Hlavním úkolem vinět je totiž *ukotvit* odpovědi na sebehodnotící otázky tak, aby je bylo možno porovnávat. Metoda ukotvujících vinět tak může pomoci odstranit jednu z častých slabin kvantitativního výzkumu, totiž omezenou validitu při použití posuzovacích škál. Hana Voňková metodu ukotvujících vinět nejen představila české pedagogické obci, ale publikovala též několik empirických studií, v nichž tento přístup použila. V neposlední řadě se také věnuje jeho další kultivaci, založila například výzkumnou skupinu Educational Research Anchors Group (ERA Group), která se zabývá aplikacemi metody ukotvujících vinět v pedagogickém výzkumu a sdružuje badatele, kteří viněty ve svých výzkumech používají.

Pedagogické metodologii velice svědčí matematické a technické myšlení, jak svým dílem ukázal například zapálený pedagogický výzkumník a průkopník

teorie testování Petr Byčkovský. Dokazuje to ostatně i tato publikace Hany Voňkové, která na odkaz svého velkého učitele směle navázala. Stejně jako on našla inspiraci na zahraničních univerzitách, které často navštěvuje, stejně jako on se rozhodla pěstovat především kvantitativní pedagogickou metodologii a teorii didaktických testů. S Petrem Byčkovským spojuje Hanku i přístup k výuce, jenž se vyznačuje vysokou náročností a důrazem na aktivitu studentů. A také spravedlivost a nezdolnost.

Jedním z podstatných trendů v současném pedagogickém výzkumu je sblížení kvantitativních a kvalitativních přístupů, rozvoj smíšené metodologie, v níž přicházejí ke slovu nejen tradiční kritéria validity, reliability a zobecnitelnosti, ale též kritéria nová, jež by se dala zastřešit výrazem hodnověrnost (*trustworthiness*). Ať již pedagogický výzkum vychází z jakýchkoli paradigmatických pozic, musí být vždy především pečlivě a poctivě proveden, výzkumník si musí být vědom své odpovědnosti k vědecké komunitě i k adresátům svého výzkumu. Poctivé badatelské práci by se měli učit již studenti magisterských programů – musí k tomu mít však dostatek inspirujících příkladů dobré praxe. Předkládaná publikace může k takovému účelu dobře posloužit – je zcela původní vědeckou monografií (a to nejen v českém, ale i v mezinárodním kontextu) a její čtivý text jistě uvítají začínající i pokročilí výzkumníci nejen na pedagogických fakultách. Kniha výborně poslouží každému, kdo již získal základní orientaci v metodologii společenskovědního výzkumu a zajímá se o využití metody ukotvujících vinět.

Michal Zvírotsky

Úvod

Na jednom z mezinárodních setkání k výzkumu PISA zaznamenali Kyllonen a Bertling (2014a) vystoupení zástupce jedné země. Ten měl dobrou a špatnou zprávu. Dobrou zprávou bylo, že jeho země opět dosáhla vynikajících výsledků studentů v testovaných oblastech. Špatnou zprávou bylo, že jeho země je jednou z nejhorších v přístupu studentů ke škole. Z údivu a obav zástupce této země je znatelné jeho celkem pochopitelné očekávání, že přístup ke škole a výsledky v testovaných oblastech by měly být kladně korelovány. Dále je znatelná vážnost, jaká je připisována zjištěním v dotazníkových šetřeních.

Otázkou však je, zdali data získaná na základě odpovědí na (sebe)hodnoticí dotazníkové položky (např. *Škola mě nepřipravila příliš dobře na dospělý život*) s posuzovací škálou jako je Likertova škála (*rozhodně souhlasím, souhlasím, nesouhlasím a rozhodně nesouhlasím*) můžeme využít pro přímé porovnávání zemí. Je možné, že studenti této země mají ve skutečnosti dokonce lepší postoj ke škole ve srovnání s ostatními. Odpovědi žáků na otázky týkající se postojů mohou být ovlivněny rozdílným užíváním škálových kategorií ve srovnání s ostatními zeměmi. Co žáci v jedné zemi hodnotí jako vynikající, může být žáky v jiné zemi hodnoceno pouze jako dobré.

Obecně lze říci, že (sebe)hodnoticí otázky se v pedagogickém výzkumu využívají velmi často, a to nejen při porovnávání zemí, ale také při porovnání různých skupin nebo jednotlivců. Můžeme například porovnávat postoje, hodnoty, subjektivní normy a (ne)kognitivní dovednosti. K rozšíření využití (sebe)hodnoticích otázek ve výzkumných šetřeních přispívá jednoduchost jejich administrace a také časová a finanční úspornost. Přímé porovnání respondentů na základě jejich odpovědí na (sebe)hodnoticí otázky však, jak již bylo naznačeno, může vést k chybnému závěru o objektivním stavu. Předkládané slovní charakteristiky u kategoriálních posuzovacích škál mohou různí posuzovatelé chápat odlišným způsobem (např. Chráska, 2007). Do odpovědi na otázku se může vedle objektivního stavu promítat i rozdílnost užívání škály odrážející rozdílné kulturní zázemí respondentů, rozdílný hodnotový systém apod.

Řešení problému heterogenity ve stylu odpovídání (*heterogeneity in reporting behavior*, v literatuře týkající se ukotvujících vinět se též užívají termíny

odlišné fungování položek, *differential item functioning* se zkratkou DIF, pocházející z teorie odpovědí na položku, či *response scale differences*) ve výzkumných šetřeních nabízí metoda ukotvujících vinět (*anchoring vignette method*). Jejím cílem je pomocí přímého změření rozdílnosti využívání škály různými respondenty očistit (přizpůsobit, korigovat) sebehodnocení respondentů tak, aby bylo porovnatelné. Základní myšlenkou metody ukotvujících vinět je to, že respondenti mimo (sebe)hodnotící otázky hodnotí v dané oblasti i hypotetickou osobu popsanou v krátkém příběhu – ukotvující viněti. Vzhledem k tomu, že všichni respondenti hodnotí tentýž příběh, může být heterogenita hodnocení vinět interpretována jako rozdílnost užívání škály. Na základě takto odhadnuté rozdílnosti se následně provede korekce (sebe)hodnocení respondentů v dané oblasti.

Ve výzkumu PISA prozatím nebyly použity ukotvující viněti týkající se přístupu ke škole, nicméně v roce 2012 byly mimo jiné do žakovského dotazníku PISA (OECD, 2014b) zahrnuty celkem tři ukotvující viněti týkající se učitelovy podpory žáků. Samotné hodnocení učitelovy podpory bylo měřeno například pomocí otázky *Když je třeba, učitel se nám více věnuje se čtyřbodovou Likertovou škálou rozhodně souhlasím, souhlasím, nesouhlasím a rozhodně nesouhlasím*. Jedna z ukotvujících vinět týkající se učitelovy podpory je následující: *Pan učitel Novák dává domácí úkoly z matematiky jednou za týden. Opravené je žákům vždy vrací ještě před zkoušením nebo písemkou. Žáci měli na stejné škále jako pro hodnotící otázku odpovědět, zda panu učiteli Novákovi záleží na tom, jak se žáci učí. Zatímco jeden žák může rozhodně souhlasit s tím, že panu učiteli Novákovi záleží na tom, jak se žáci učí, druhý žák může nesouhlasit*. Ačkoli žáci hodnotí stejnou situaci popsanou v ukotvující viněti, jejich styl odpovídání na otázky je jiný. Lze ho například interpretovat tak, že první žák má ve srovnání s druhým žákem nižší standardy pro hodnocení. Důležité je, že na základě hodnocení ukotvujících vinět identifikujeme odlišné užívání škály. To následně použijeme k očištění původních hodnotících otázek týkajících se podpory žáků ze strany jejich *skutečného* učitele. Analýza učitelovy podpory v různých zemích účastnících se PISA 2012 je provedena v této monografii.

Předložená monografie pojednává o metodě ukotvujících vinět a jejím využití v pedagogickém výzkumu. Je členěna do sedmi klíčových kapitol. Pořadí studia jednotlivých kapitol může záviset na typu čtenáře. Čtenáři, který chce nejprve

získat přehled o teoretických základech metody a jejího vztahu k obecným problémům řešeným v metodologii nejen pedagogického výzkumu, doporučujeme začít se studiem první kapitoly. Čtenáři, který chce nejprve získat přehled o konkrétních příkladech aplikací metody ve výzkumu v pedagogice a dalších sociálních vědách, doporučujeme začít se studiem druhé kapitoly. Čtenáři, který chce nejprve získat přehled o statistických modelech metody, doporučujeme začít se studiem třetí kapitoly.

Podrobnější obsah jednotlivých kapitol je následující. V první kapitole jsou popsány teoretické základy metody ukotvujících vinět. Nejprve jsou diskutovány metodologické problémy při porovnávání skupin, detailněji je popsán pojem zkreslení (*bias*) a jeho taxonomie. Následuje vysvětlení, jak by mohla metoda ukotvujících vinět přispět k řešení uvedených metodologických problémů. Kapitola je zakončena popisem základní myšlenky metody ukotvujících vinět. První způsob prezentuje základní myšlenku na základě porovnání (sebe)hodnocení dvou hypotetických respondentů a objasňuje tím základní principy statistického neparametrického přístupu metody ukotvujících vinět. Druhý způsob je založen na porovnání rozdělení (sebe)hodnocení ve dvou hypotetických zemích a objasňuje tím základní principy parametrického modelu metody ukotvujících vinět.

Druhá kapitola shrnuje klíčové aplikace v sociálních vědách a dále i všechny autorce známé aplikace v pedagogice. Podrobně jsou představeny studie zaměřující se na měření vnímané politické účinnosti (King et al., 2004), ostrosti zraku (King et al., 2004) a zdravotních problémů při výkonu práce (Kapteyn, Smith, & van Soest, 2007) s využitím sebesposuzovacích škál. Tyto studie jsou jedny z klíčových. V King a kol. (2004) jsou jedny z prvních aplikací metody ukotvujících vinět vůbec. Zároveň jsou zde popsány formální základy metody. Co se týká pedagogického výzkumu, jsou aplikace metody rozděleny do třech skupin: a) rodičovské hodnocení týkající se spolupráce s učiteli, ředitelem a ostatními rodiči, b) žákovské sebehodnocení motivace, osobnostních charakteristik, znalostí, nekázně a životní spokojenosti, c) žákovské hodnocení učitelova řízení třídy a učitelovy podpory žáků.

Třetí kapitola prezentuje statistické přístupy metody ukotvujících vinět. Prvním přístupem je neparametrický (jednoduchý) přístup (*nonparametric [simple] approach*), který je založen na principu relativního porovnání (sebe)hodnocení

a hodnocení vinět(y). Jeho pochopení je snadné, a proto bývá nazýváno jednoduchým přístupem. Nejprve jsou popsány základy neparametrického přístupu pro začínající výzkumníky, následované zobecněním popsaných úvah pro úplné pochopení tohoto principu. Druhým přístupem je parametrický CHOPIT model (*compound hierarchical ordered probit model*). U tohoto modelu je uvedena jeho statistická specifikace, postup odhadu parametrů modelu a také rozšíření modelu popsaná v literatuře.

Čtvrtá kapitola na konkrétním příkladu s volně dostupnými daty ukazuje, jak lze neparametrický přístup a parametrický model odhadnout v softwarových balíčcích. V softwaru R jsou implementovány oba přístupy, v softwaru STATA je implementován pouze parametrický model. Pro oba přístupy je uveden *vzorový* kód, s jehož pomocí by čtenář měl být schopen odhadnout všechny zásadní statistiky pro vlastní data. Pro neparametrický přístup se jedná o statistiky důležité pro určení pořadí vinět a jejich volbu pro výpočet korigovaného sebehodnocení, dále je uveden postup výpočtu korigovaného sebehodnocení s jeho následným grafickým znázorněním. Pro parametrický model je pro oba softwary ukázán kód pro odhad parametrů, na ukázkovém příkladu jsou parametry odhadnuty a interpretovány.

Pátá kapitola se zaměřuje na formulaci vinět a jejich umístění v dotazníku. Jsou uvedena doporučení týkající se dodržení klíčových předpokladů metody – konzistence odpovědí (*response consistency*) a ekvivalence vinět (*vignette equivalence*). Předpoklad konzistence odpovědí odkazuje na to, že respondenti používají škálu v (sebe)hodnocících otázkách stejným způsobem jako při hodnocení vinět. Předpoklad ekvivalence vinět odkazuje na to, že všichni respondenti interpretují viněty stejným způsobem. Předpoklady metody je nutné splnit pro korektní fungování metody. Dále se v této kapitole zabýváme diskriminační silou vinět (určující do jaké míry pomáhají viněty korigovat rozdílné užívání škály v sebehodnocících otázkách), počtem vinět a respondentů, pořadím (sebe)hodnocení a vinět a zakončujeme několika poznámkami k obsahu vinět. Metoda ukotvujících vinět s jejími statistickými modely je kvantitativní metodou, nicméně formulace vinět a konstrukce dotazníku s vinětami představuje výzvu i pro kvalitativní výzkum.

Šestá kapitola představuje novou aplikaci parametrického modelu pro porovnání učitelovy podpory žáků v zemích účastnících se výzkumu PISA 2012.

Nejprve je analyzována heterogenita ve stylu odpovídání žáků v různých zemích, dále je provedena korekce žakovského hodnocení učitelovy podpory o identifikovanou heterogenitu, porovnání korigovaného a nekorigovaného hodnocení žáků a nakonec je analyzován vztah (ne)korigovaného studentského hodnocení učitelovy podpory s výsledkem žáků v testu. V analýze se pracuje s daty od více než 300 000 žáků z 68 zemí. Standardní model CHOPIT musel být rozšířen, neboť do analýzy vstupuje více než jedno hodnocení učitelovy podpory od každého žáka. Země jsou mimo jiné i s využitím grafického znázornění v mapě světa rozděleny dle tendence užívání škály identifikované pomocí vinět.

Primárním cílem metody ukotvujících vinět je korekce (sebe)hodnocení různých skupin respondentů o jejich odlišné užívání škály. Ke korekci odlišného užívání škály by však mohly pomoci i čtyři přístupy popsané v poslední klíčové sedmé kapitole. První z nich je identifikace tendence k volbě určitých škálových kategorií, jako je například obecná tendence respondenta k souhlasu bez ohledu na to, o jaký hodnocený pojem se jedná. Následně se odpovědi respondenta o tuto identifikovanou tendenci korigují. Druhý přístup je technika přehánění (*overclaiming technique*), jejímž principem je identifikovat tendenci k přehánění či ukazování svých znalostí v mnohem lepším světle. Třetí přístup využívá hodnocení nucenou volbou (*forced choice assessment*), tedy otázky, v nichž je respondent obvykle žádán o volbu jednoho z předložených tvrzení. Například může být respondent požádán zvolit z určité nabídky jeden způsob práce v matematice, který ho nejlépe vystihuje. Poslední čtvrtý přístup využívá test posouzení situace (*situational judgement test*). Respondentům je předložen popis nějaké situace a v následných otázkách se vyjadřuje, jak by se v dané situaci zachoval. Oproti tradičním otázkám v osobnostních dotaznících na chování jedince se ukazuje být tento typ otázek méně citlivý na kulturní rozdíly.

I přes existující a dobře zdokumentovaný problém uvedené heterogenity ve stylu odpovídání na položky je aplikací metody ukotvujících vinět v pedagogice na celosvětové úrovni velmi málo. Pokud je nám známo, monografie k tomuto tématu prozatím chybí. Tato monografie je vůbec prvním počinem o metodě ukotvujících vinět a jejím využití v pedagogickém výzkumu. Mimo pedagogický výzkum může nalézt též uplatnění ve výzkumu v dalších sociálních vědách jako je psychologie, sociologie a politologie. Statistické

modely metody ukotvujících vinět mohou být inspirativním tématem a výzvou též pro statistiky.

1. Teoretické základy metody ukotvujících vinět

1.1. Metodologické problémy při porovnávání skupin

Ve výzkumech porovnávajících měřený koncept v různých kulturách, zemích, socio-ekonomických či jiných skupinách a u různých individuí vzniká otázka, zda výsledek měření (testové skóry, odpovědi na dotazníkové položky apod.) může být v těchto populacích interpretován stejným způsobem. Dva nejčastější termíny, které se v této souvislosti používají, jsou zkreslení (*bias*) a ekvivalence (*equivalence*). Van de Vijver a Tanzer (2004) uvádí příklady, definici a taxonomii termínů včetně jejich možných zdrojů. Pro pochopení těchto termínů je vhodné začít příkladem. Předpokládejme, že je při mezinárodním srovnání zahrnuta v testu z geografie následující otázka: Jaké je hlavní město Polska? Procento správných odpovědí bude mimo jiné záviset na intelektových schopnostech žáka, kvalitě výuky zeměpisu a také na vzdálenosti od Polska. Pokud je naším cílem porovnat geografické znalosti žáků v různých zemích jako je Austrálie a Německo, není tato otázka vhodná, neboť lze předpokládat, že žáci ze střední Evropy se srovnatelnou znalostí geografie jako žáci z Austrálie budou na tuto otázku odpovídat s větší pravděpodobností správně. O zkreslení (*bias*) tedy hovoříme v případě, když změřené rozdíly (procento žáků v jednotlivých zemích, které ví, že hlavním městem Polska je Varšava) neodpovídají reálným rozdílům v měřeném konstruktu (znalost geografie). Závěry činěné na základě takového měření jsou neplatné, zavádějící. Ekvivalenci (*equivalence*) lze definovat jako protiklad ke zkreslení (*bias*). Ekvivalence je základ validního porovnání skupin. Je nutno poznamenat, že o zkreslení a ekvivalenci má smysl hovořit pouze při porovnávání *konkrétních* skupin (kultur, zemí atd.). Naš měřicí nástroj nemusí vykazovat zkreslení při porovnání českých a slovenských žáků, ale naopak může vykazat zkreslení při porovnání českých a australských žáků.

1.1.1. Zkreslení (*bias*) a jeho taxonomie

Vhodným vodítkem pro analýzu zkreslení (*bias*) je taxonomie uvedená ve studii van de Vijver a Tanzer (2004). V této taxonomii jsou vymezeny tři základní typy zkreslení – zkreslení způsobené konstruktem (*construct bias*), zkreslení způsobené metodou (*method bias*) a zkreslení způsobené obsahem položky (*item bias*).

Zkreslení způsobené konstruktem (*construct bias*)

Zkreslení způsobené konstruktem (*construct bias*) se objevuje v případě, kdy není měřený konstrukt identický napříč různými kulturami či skupinami. Van de Vijver a Tanzer (2004) s odkazem na Super (1983) uvádí jako příklad testy inteligence. Obecné testy inteligence kladou často důraz na uvažování, znalost a paměť. Sociální aspekty inteligence jsou zdůrazňovány méně často. Tyto aspekty jsou však mimo západní kultury více zdůrazňovány. Dalším příkladem může být měření chování asociovaného s *bytím* dobrým synem či dobrou dcerou (koncept je označován jako *filial piety*). Například v Číně mohou být povinnosti dětí vůči rodičům odlišné ve srovnání se západními kulturami. Rozdíly mohou být způsobeny také vzděláním a příjmem. Kagitcibasi (1996) došel k závěru, že v tureckých rodinách u vzdělanějších rodičů ztrácí na významu pomoc dětí s domácími pracemi. Ochota postarat se o své staré rodiče finančně v západních kulturách jen částečně souvisí s konceptem *filial piety*. Ze všech těchto důvodů může tedy koncept *filial piety* podléhat zkreslení.

Zkreslení způsobené metodou (*method bias*)

Pro zkreslení způsobené metodou (*method bias*) vymezujeme tři základní typy:

- a) **zkreslení způsobené výběrem** (*sample bias*) – nesrovnatelnost vzorku v proměnné, která sice není cílem měření, ale může ho ovlivnit. V případě kognitivních testů mohou být takovými rušivými proměnnými rozdílnost ve vzdělávacím systému, novost testové situace, motivace k psaní testu nebo podmínky získávání participantů do studie.
- b) **zkreslení způsobené administrací** (*administration bias*) – může být vyvoláno rozdílnostmi v podmínkách při sběru dat. Například využití

techniky jako je kamera či diktafon může být pro nějaké respondenty zcela novou věcí, což může vést k větším rozdílům v měřeném konceptu. Nemusí se však jen jednat o využitou techniku při sběru dat. Podstatnou roli též může hrát individuální versus skupinová administrace dat, velikost prostoru mezi testovanými (např. velikost třídy), nejasné instrukce pro respondenty či administrátory dat, přítomnost testujícího či pozorovatele nebo komunikační problémy mezi testovaným/dotazovaným a testujícím / dotazujícím se.

- c) **zkreslení způsobené nástrojem** (*instrument bias*) – van de Vijver a Tanzer (2004) s odkazem na Deregowski a Serpell (1971) uvádí příklad tohoto zkreslení při řazení miniaturních modelů zvířat a motorových vozidel versus fotografií těchto modelů u dětí ze Skotska a Zambie. Při použití miniaturních modelů nebyly nalezeny žádné rozdíly mezi těmito skupinami dětí, zatímco při použití fotografií byly výsledky skotských dětí lepší. Postupy volené při odpovídání respondentů mohou také sehrát zásadní roli. Serpell (1979) ukazuje rozdíly mezi britskými a zambijskými dětmi při volbě postupu reprodukce předložených obrazců. V případě, že byla zvolena metoda tužka – papír, dosahovali lepších výsledků britští žáci. V případě, že byl zvolen železný drát, dosahovali lepších výsledků zambijské děti (dělání modelů z drátu je oblíbenou činností zambijských chlapců). Dalším příkladem je užití dotazníkového šetření v práci Hui a Triandis (1989), kteří ukázali, že Hispánci mají větší tendenci užívat extrémních bodů na pětibodové hodnoticí škále ve srovnání s bílými Američany. Pokud však byla užitá desetibodová hodnoticí škála, k tomuto jevu nedocházelo.

Zkreslení způsobené obsahem položky (*item bias*)

Zkreslení způsobené obsahem položky (*item bias*) oproti zkreslení způsobeným konstruktem či metodou odkazuje na úroveň položky. V literatuře bývá užíván též termín rozdílné fungování položky (*differential item functioning*, zkratka DIF) (Holland a Wainer, 1993). Zkreslení způsobené obsahem položky se může týkat sebehodnocení, které respondent uvádí při dotazníkovém šetření. Například odpovědi na jednu z položek měřících určitý koncept mohou být v určité kultuře oproti jiným kulturám ovlivněny tím, co je v ní sociálně žádoucí. Takováto otázka pak při srovnání kultur poskytuje zavádějící odpovědi

a měla by být vypuštěna. Ze statistického pohledu existuje celá řada publikací zabývajících se analýzou DIF (např. Camili & Shepard, 1994; Osterlind & Everson, 2009). Položka vykazuje DIF, pokud osoby z různých skupin se stejným skórem (skór je operacionalizovaný jako například celkový skór v testu či celková hodnota indexu při měření nějakého konstrukt) nemají stejné očekávané skóre v rámci této položky. Pro osoby z různých kultur se stejným celkovým skórem by měly být položky srovnatelně náročné. Jako typické proměnné, které se při analýze zkreslení způsobené obsahem položky berou v úvahu, jsou rasa, země, či pohlaví (např. Hunter & Schmidt, 2000; Sheppard et al., 2006).

1.1.2. Zkreslení a metoda ukotvujících vinět

V případě výzkumů, kde se k měření jevů využívají (sebe)hodnotící otázky (např. jako součást dotazníků), lze říci, že metoda ukotvujících vinět pomáhá řešit problém se všemi uvedenými zkresleními. V první řadě však lze jmenovat zkreslení způsobené metodou a s ním spojený efekt heterogenity ve stylu odpovídání na otázky.

Heterogenita ve stylu odpovídání na otázky (*heterogeneity in reporting behavior*) se vyskytuje například v případě, kdy respondenti se stejnou úrovní měřeného jevu (například respondenti se stejnou úrovní sebekázně) mohou hodnotit svou úroveň odlišně (například jeden respondent hodnotí svou sebekázeň jako výbornou, jiný respondent jen jako dobrou). Heterogenita ve stylu odpovídání se může u jednotlivých respondentů lišit napříč studovanými jevy, při hodnocení některých jevů může být respondent přísnější ve srovnání s ostatními respondenty, u některých jevů může být zase méně přísný ve srovnání s ostatními respondenty. Pokud hovoříme o heterogenitě ve stylu užívání škály, jedná se vždy o *styl*, který uplatňuje respondent v rámci dané měřené oblasti. Heterogenita ve stylu odpovídání na otázky může mít různé zdroje. Tuto heterogenitu rozlišíme hierarchicky – na úrovni odlišných kultur a zemí, dále na úrovni skupin v jednotlivých zemích a na úrovni jednotlivců.

1. **Různé kultury a země:** Známý příklad heterogenity ve stylu užívání škály v různých zemích a kulturách je ukázán ve studii Chen a kol. (1995). Autoři analyzovali odpovědi žáků středních škol z Japonska

(město Sendai, N = 944), Tchaj-peje (N = 1357), Kanady (Edmonton a Kalgary, N = 687) a USA (Virginie, N = 2174) na padesát sedm dotazníkových položek se sedmibodovou Likertovou škálou. Dotazníkové položky se týkaly hodnoty vzdělání (4 položky), hodnoty sociálního a fyzického vývoje (8 položek), akademického sebepojetí (9 položek), sociálního a fyzického sebepojetí (6 položek), postoje k matematice (4 položky), spokojenosti se školním výkonem (5 položek) a obavy ze školy (3 položky). Dalších 18 položek se týkalo různých konceptů, jako je spokojenost sám se sebou a důvěra ve své schopnosti. Hlavním závěrem bylo, že japonští a čínští žáci užívají ve srovnání s americkými a kanadskými žáky častěji středové kategorie škály. Američtí žáci užívají nejčastěji ze všech srovnávaných skupin krajní, extrémní body škály. V rámci každé kulturní skupiny bylo navíc užívání škály spojeno s podporou individualismu či kolektivismu. Podpora individualismu byla pozitivně korelována s užitím extrémních kategorií škály a naopak podpora kolektivismu byla pozitivně korelována s užitím středových kategorií. Dalším příkladem může být Buckley (2009), který provedl analýzu stylů odpovídání na otázky v žákovském dotazníku PISA 2006 (více viz část 7.1).

- Různé skupiny v jednotlivých zemích:** Heterogenita ve stylu odpovídání na položky se vyskytuje nejen při srovnávání různých kultur a zemí, ale také při srovnávání různých skupin respondentů, jako jsou školy a socio-ekonomické skupiny. Příkladem může být studie West a kol. (2016), kde je porovnávána sebekontrola a vůle amerických žáků z tzv. *charter schools* a žáků z běžných státních škol. Ve výzkumu jsou záměrně vybrány *charter schools* s pozitivními dopady na výsledky žáků měřenými pomocí státních testů z matematiky a angličtiny. Na úrovni jednotlivců v rámci školy je vztah mezi těmito nekognitivními dovednostmi a výsledky v testech pozitivní. Tento vztah však mizí na úrovni srovnání typů škol. *Charter schools* vykazují výrazně nižší úroveň sebekontroly (měřené pomocí sebehodnocení studentů), což je v rozporu s tím, co výzkumníci očekávali. Vysvětlením může právě být referenční zkreslení (*reference bias*) vyplývající z rozdílného školního klimatu obou typů škol.
- Různí jednotlivci:** Problém s heterogenitou ve stylu užívání škály se může vyskytnout již při porovnávání jednotlivců. V podstatě pokud

porovnáváme dva jedince, můžeme na tento problém narazit a získat zkreslená data. Příkladem může být efekt velké ryby v malém rybníce (*Big Fish in a Little Pond Effect*), kdy se jedinci porovnávají se svými vrstevníky a dochází k tomu, že stejně schopní jedinci mají vyšší sebehodnocení v méně schopné skupině než ve více schopné skupině (více k tomuto efektu Marsh, 1987; Marsh & Parker, 1984).

Dotazníkový nástroj může vykazovat i zkreslení způsobené obsahem položky. Jedna z položek v rámci celé baterie otázek měřících určitý jev může být ovlivněna rozdílným užíváním škály. V tomto případě můžeme využít metodu ukotvujících vinět pro očištění dané položky od tohoto zkreslení. Principiálně se jedná o stejnou myšlenku jako při zmiňovaném zkreslení způsobeném metodou, kde mohou odpovědi na všechny položky měřeného konceptu podléhat efektu heterogenity ve stylu užívání škály. Očištění od rozdílného užívání škály můžeme provést pomocí metody ukotvujících vinět jak pro jednotlivé položky, tak pro celou baterii položek.

Metoda ukotvujících vinět však může přispět i k řešení zkreslení způsobeného konstruktem. Některé měřené pojmy (například politická svoboda) je vhodné vysvětlit respondentům pomocí příkladu či nějakého příběhu, neboť daný pojem je pro ně příliš komplikovaný, těžko uchopitelný a především nemusí být identický napříč různými kulturami. Při aplikaci metody ukotvujících vinět pracujeme s ukotvujícími vinětami – krátkými příběhy, v nichž popisujeme určitou úroveň měřeného konceptu pro hypotetickou osobu. Tento příběh může pomoci respondentům pochopit, na co je vlastně dotazník zaměřen, co chce výzkumník měřit. Zde je však nutné poznamenat, že v případě, kdy má respondent odpovídat na otázky tak, že se nemůže zpětně vracet k předchozím otázkám, musí být ukotvující viněty na začátku dotazníku, aby tuto funkci plnily. Tato úvaha již však není přímočará a vyžaduje pochopení alespoň základního principu metody ukotvujících vinět (část 1.2) a také analýzu pořadí otázek v dotaznících (část 5.2).

1.2. Základní myšlenka metody ukotvujících vinět

Primárním cílem metody ukotvujících vinět je učinit původně neporovnatelné (sebe)hodnocení respondentů (např. v dotazníkových šetřeních) zatížené

heterogenitou ve stylu odpovídání porovnatelné. Snažíme se tedy učinit neporovnatelné porovnatelným. Modelově vycházíme z toho, že (sebe)hodnocení respondentů obsahuje dvě složky – jedna složka odrážející reálnou, skutečnou úroveň měřeného konceptu, druhá složka odrážející styl užívání škály. (Sebe)hodnocení respondenta je tedy jedno *číslo*, které však v sobě obsahuje dvě složky. Vzhledem k tomu, že se zaměříme především na první složku, tj. chceme získat skutečnou úroveň měřeného konceptu, je nutné tyto dvě části oddělit. Bez další informace to nelze. Jako další zdroj informace v metodě ukotvujících vinět slouží právě ukotvující viněty – krátké příběhy popisující určitou úroveň měřeného konceptu pro hypotetickou osobu. V jedné z prvních aplikací metody ukotvujících vinět (King et al., 2004) je měřen koncept vnímané politické účinnosti (*political efficacy*). Respondentům je předložena následující otázka: *Do jaké míry máte možnost ovlivňovat vládu v záležitostech, které se Vás týkají?* se škálou *vůbec, málo, středně, velmi, neomezeně*. Příklad jedné z ukotvujících vinět je: *[Alison] nemá dostatek pitné vody. Společně se svými sousedy podporuje v následujících volbách opozičního kandidáta, který slibuje, že se touto záležitostí bude zabývat. Zdá se, že mnoho lidí v její oblasti má stejně jako ona pocit, že opoziční kandidát porazí současného zastupitele. Do jaké míry má [Alison] možnost ovlivňovat vládu v záležitostech, které se jí týkají?* (více podrobností k měření tohoto konceptu lze najít v části 1).

Respondentovi je ve výzkumném šetření předložena ke zhodnocení nejen (sebe)hodnotící otázka, ale také ukotvující viněta. Vzhledem k tomu, že respondenti hodnotí *tentýž* příběh, lze rozdíly v jejich hodnocení ukotvující viněty připsat rozdílům ve stylu užívání škály. Takto identifikované rozdíly ve stylu užívání škály lze pak využít pro korekci (sebe)hodnocení a tedy oddělit od sebe reálnou úroveň měřeného konceptu u daného respondenta (první složka) a styl užívání škály (druhá složka).

Metoda ukotvujících vinět je metodou kvantitativního výzkumu, její modely jsou popisovány pomocí statistického aparátu. Existují dva základní modely metody ukotvujících vinět – neparametrický přístup a parametrický model CHOPIT. Pro první seznámení čtenáře s touto metodou a pochopení základní myšlenky metody je však možné statistický aparát vynechat. Základní myšlenka bývá popisována dvěma způsoby. První způsob je pomocí porovnání (sebe)hodnocení dvou hypotetických respondentů. Tento způsob objasňuje

základní princip neparametrického přístupu. Druhý způsob je pomocí porovnání rozdělení (sebe)hodnocení ve dvou hypotetických zemích. Tento způsob zase objasňuje základní princip parametrického modelu CHOPIT. Ve své podstatě se však u obou způsobů jedná o totéž – učinit odpovědi respondentů porovnatelné, očistit je od odlišného užívání škály.

Oba způsoby jsou v české literatuře popsány v článku Voňková (2012a), který mimo jiné zavedl i český termín metoda ukotvujících vinět jako překlad anglického termínu *anchoring vignette method*. Zde uvedeme podobný popis obou způsobů s tím, že myšlenku nově objasníme na příkladu, kde žáci hodnotí, jakou míru podpory dostávají od svého učitele. V kapitole 6 na toto navazujeme a prezentujeme novou aplikaci parametrického modelu metody ukotvujících vinět týkající se právě podpory učitele. Dále v případě představení myšlenky pomocí porovnání rozdělení (sebe)hodnocení ve dvou hypotetických zemích (druhý způsob) provedeme rozšíření o rozbor všech možností, které mohou při porovnávání hypotetických zemí v případě existence heterogenity v užívání škály nastat.

1.2.1. Porovnání (sebe)hodnocení dvou hypotetických respondentů

Představme si, že porovnáваме učitelovu podporu dvou žáků 1 a 2 (žáci mají různé učitele) na základě jejich odpovědí na následující dotazníkovou položku: *Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícím tvrzením? Když je třeba, učitel se nám více věnuje.* Tato dotazníková položka byla např. využita v mezinárodním srovnávacím výzkumu PISA 2012 jako jedna ze čtyř položek měřících učitelovu podporu žáků (všechny položky jsou uvedeny v příloze 2). Žáci uvádějí svou odpověď na čtyřbodové Likertově škále *rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím*. Dále si představme, že tito dva žáci hodnotí navíc i tři ukotvující viněty popisující různou úroveň podpory žáků hypotetickými učiteli. Příkladem jedné z vinět může být ukotvující viněta využitá ve zmíněném výzkumu PISA 2012, která popisuje hypotetickou učitelku Novotnou: *Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrací ještě před zkoušením nebo písemkou.* Žáci odpovídají, jak *paní učitelce Novotné záleží na tom, jak se žáci učí* na stejné škále jako v případě hodnotící otázky: *rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně*

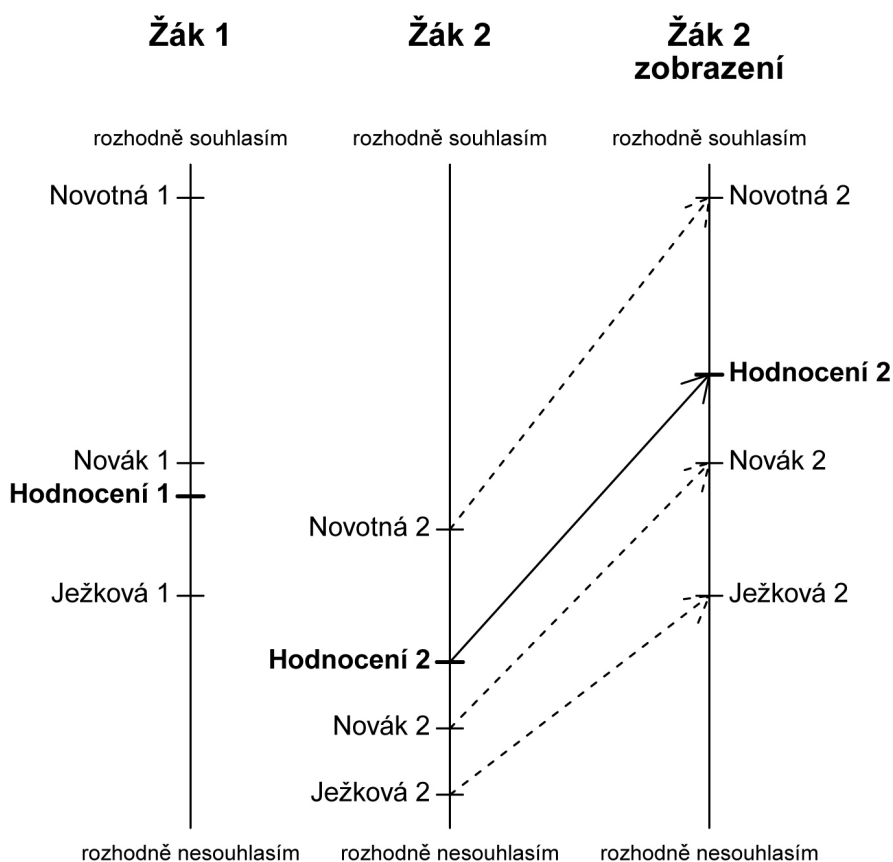
nesouhlasím. Tato viněta popisuje nejvyšší úroveň učitelovy podpory žáků. Další dvě viněty popisují hypotetické učitele Nováka (střední úroveň učitelovy podpory) a Ježkovou (nejnižší úroveň učitelovy podpory). Znění všech vinět je uvedeno v příloze 2.

Neparametrický přístup je založen na relativním porovnání hodnocení svého učitele s hodnocením situace popsané v ukotvujících vinětách. Na obrázku 1 je znázorněno, jak žáci 1 a 2 hodnotí svého učitele a také hypotetické učitele popsané ve vinětách. Pro výstižnější popis situace je při ilustraci neparametrického přístupu volena spojitá škála. V našem případě lze pracovat se škálou spojitou s krajními body *rozhodně souhlasím* a *rozhodně nesouhlasím*.

Žák 1 umísťuje hodnocení podpory svého učitele přibližně doprostřed škály. Žák 2 hodnotí podporu žáků u svého učitele v dolní části škály, tj. u pólu *nesouhlasím*. Na základě přímého porovnání těchto hodnocení bychom dospěli k závěru, že učitel žáka 1 poskytuje vyšší podporu při učení než učitel žáka 2.

Hodnocení hypotetických učitelů je však ze strany žáků odlišné. Žák 1 hodnotí podporu hypotetické učitelky Ježkové jako horší v porovnání se svým učitelem a podporu hypotetických učitelů Nováka a Novotné naopak jako lepší (levá část obrázku 1). Žák 2 hodnotí podporu dvou hypotetických učitelů Ježkové a Nováka jako horší v porovnání se svým učitelem a podporu hypotetické učitelky Novotné jako lepší (prostřední část obrázku 1).

Porovnáme-li hodnocení ukotvujících vinět žáků 1 a 2, je zjevné, že žák 2 hodnotí podporu hypotetických učitelů vždy hůře než žák 1. Toto lze označit tak, že žák 2 je při hodnocení pesimističtější, či že má vyšší standardy pro hodnocení ve srovnání s žákem 1.



Obrázek 1. Základní myšlenka metody ukotvujících vinět: ilustrace neparametrického přístupu pro dva respondenty

Pozn. Upraveno pro ilustraci hodnocení učitelovy podpory žáků podle Voňková (2012a).

Nyní zobrazme hodnocení žáka 2 na škálu žáka 1 (pravá část obrázku 1). Hodnocení třech ukotvujících vinět žáka 2 zobrazíme do úrovně hodnocení žáka 1 (přerušované šipky z prostřední části obrázku 1 do jeho pravé části). Hodnocení žáka 2 pak umístíme mezi jeho zobrazené hodnocení vinět (plná šipka na obrázku 1. Na základě zobrazení hodnocení žáka 2 na škále žáka 1 můžeme konstatovat, že podpora učitele žáka 2 je vyšší než podpora učitele žáka 1. (Analogickým způsobem by také šlo vyjádřit hodnocení žáka 1 na škále žáka 2. Závěr by byl totožný.)

Na základě přímého porovnání hodnocení žáků (před korekcí o rozdílné užívání škály) jsme dospěli k závěru, že učitel žáka 1 poskytuje vyšší podporu při učení než učitel žáka 2. Po korekci o rozdílné užívání však dospíváme k opačnému závěru. Bez korekce by naše výsledky byly chybné.

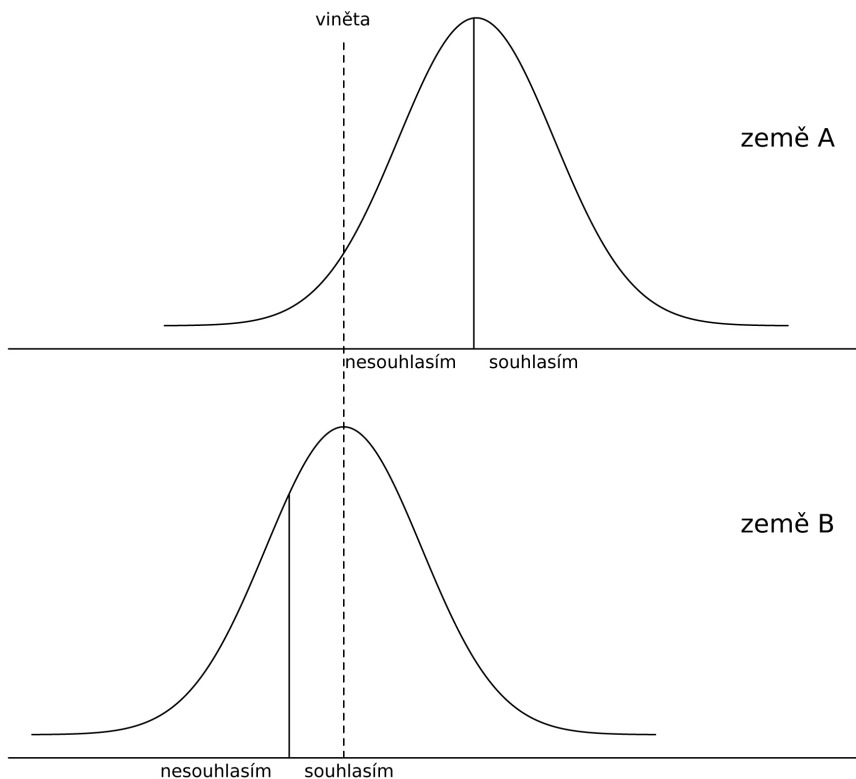
1.2.2. Porovnání rozdělení (sebe)hodnocení ve dvou hypotetických zemích

Základní myšlenka parametrického modelu metody ukotvujících vinět bývá často ilustrována na příkladu porovnání zdravotního stavu obyvatel ve dvou hypotetických zemích (např. Kapteyn, Smith a van Soest, 2007; Voňková, 2013). V tomto textu však stejně jako při představení myšlenky metody ukotvujících vinět při porovnání dvou různých respondentů opět využijeme položku z výzkumu PISA 2012 týkající se učitelovy podpory žáků: *Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícím tvrzením: Když je třeba, učitel se nám více věnuje.* Zde pro jasnější ilustraci a snadné popsání všech relevantních porovnání zemí zjednodušíme škálu odpovědí na dvoubodovou – *souhlasím a nesouhlasím*.

Pokud žáci ve dvou hypotetických zemích užívají škálové kategorie odlišným způsobem, dochází ke zkreslení porovnání učitelovy podpory žáků v těchto zemích. Na obrázku 2 je zakreslena reálná (objektivní) distribuce latentní proměnné (nepozorovaná proměnná, v našem příkladu spojitá) učitelova podpora žáků pro zemi A a zemi B. Rozdělení pro zemi A je posunuto vpravo, což znamená, že v zemi A je reálně vyšší úroveň učitelova řízení třídy než v zemi B (čím vyšší úroveň řízení třídy, tím více vpravo se distribuce nachází).

Tyto distribuce latentní proměnné však výzkumníkům nejsou dostupné. Místo nich jsou výzkumníkům k dispozici odpovědi žáků na dvoubodové škále *souhlasím a nesouhlasím* (místo spojité latentní proměnné pozorujeme diskrétní proměnnou; pokud latentní proměnná překročí určitou mez, pozorujeme změnu volby škálových kategorií). Jak je ukázáno na obrázku 2, v zemi A uvádí odpověď *souhlasím* polovina žáků, zatímco v zemi B je to více než polovina žáků. Na základě porovnání těchto odpovědí žáků bychom došli k závěru, že v zemi A je nižší úroveň učitelovy podpory žáků než v zemi B. To je však opačný závěr než při porovnání objektivních úrovní řízení třídy.

Chybný závěr je zapříčiněn rozdílným užíváním škály. Žáci v zemi A mají vyšší standard pro užívání škály (jsou přísnější, více pesimističtí) než žáci v zemi B. Rozdílné užívání škály ve dvou zemích je na obrázku 2 ilustrováno pomocí hodnocení určité úrovně učitelovy podpory žáků zobrazené přerušovanou horizontální čarou. Zatímco žáci v zemi A se vyjadřují nesouhlasně, že danému učiteli záleží na tom, jak se žáci učí, žáci v zemi B se naopak vyjadřují souhlasně (volí škálovou kategorii *souhlasím*).



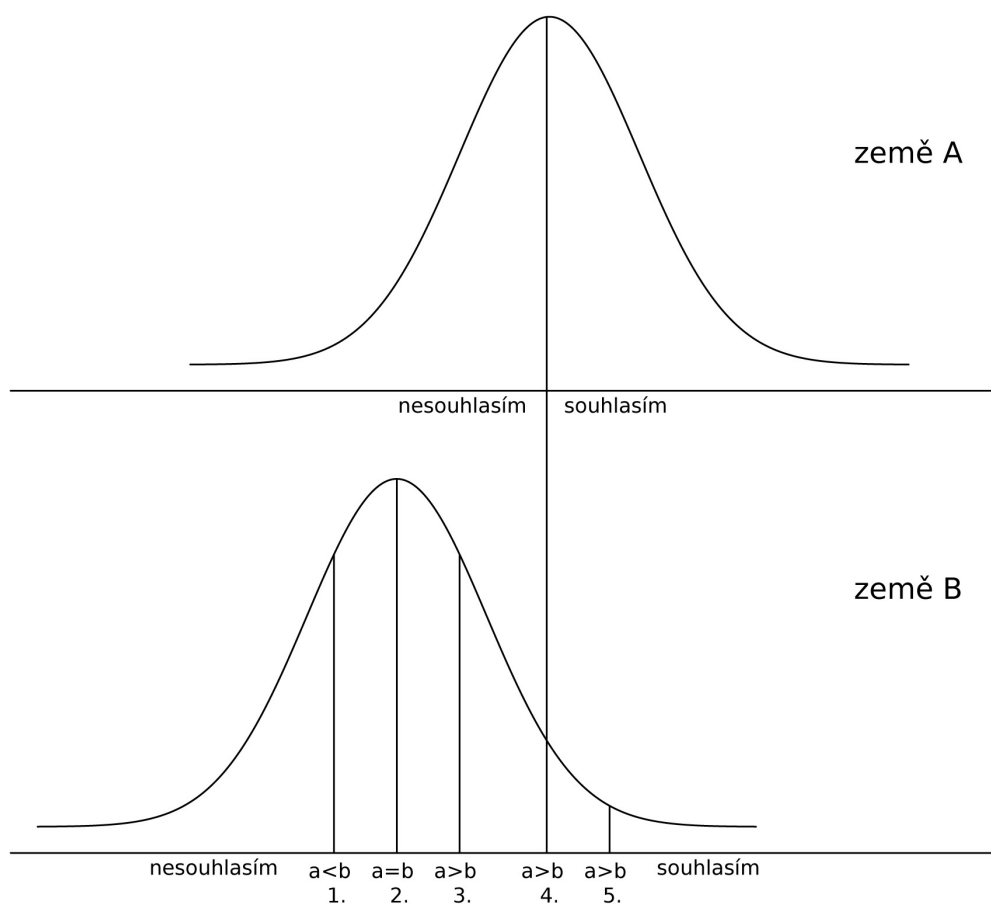
Obrázek 2. Základní myšlenka metody ukotvujících vinět: ilustrace parametrického modelu pro dvě země

Metoda ukotvujících vinět pomáhá tento problém vyřešit pomocí identifikace heterogenity v užívání škály a následné korekce žákovských odpovědí o tuto identifikovanou heterogenitu. Žáci v obou zemích hodnotí nejen úroveň podpory od jejich učitele, ale také úroveň podpory žáků u hypotetického učitele,

jež je popsána v krátkém příběhu, ukotvující viněť. K uvedení odpovědi žáků je využita stejná škála jako u hodnocení jejich reálného učitele. Všichni žáci hodnotí jednu a tutěž viněť, a proto může být rozdílné hodnocení viněty interpretováno jako rozdílné užívání škály. Tato úroveň učitelovy podpory je na obrázku 2 zachycena přerušovanou horizontální čarou. Zatímco žáci v zemi A nesouhlasí s tím, že hypotetickému učiteli záleží na tom, jak se žáci učí, žáci v zemi B vyjadřují naopak souhlas. Heterogenitu ve stylu užívání škály identifikovanou pomocí viněť využijeme ke korekci žakovského hodnocení tak, že zvolíme užívání škály jedné země za fixní a na ní vyjádříme rozdělení hodnocení druhé země, tj. jakoby druhá země využívala škálu stejným způsobem jako země první. Porovnáním korigovaného hodnocení o heterogenitu ve stylu užívání škály dospíváme ke správnému závěru, že učitelovo řízení třídy má vyšší úroveň v zemi A.

Obecně v případě, že objektivní úroveň učitelovy podpory žáků je lepší v zemi A než v zemi B ($A > B$), můžeme pomocí porovnání hodnocení žáků v těchto zemích (hodnocení žáků označme a a b) dospět k pěti odlišným závěrům, které jsou zobrazeny na obrázku 3:

1. $a < b$: hodnocení žáků není v souladu s objektivní úrovní řízení třídy, žáci v zemi A mají vyšší standardy (jsou přísnější, více pesimističtí) než žáci v zemi B
2. $a = b$: hodnocení žáků v obou zemích je srovnatelné, což neodpovídá objektivní úrovni, žáci v zemi A mají vyšší standardy pro hodnocení
3. $a > b$: hodnocení žáků při prostém porovnání odpovídá tomu, že objektivní úroveň je vyšší v zemi A, nicméně absolutní rozdíl mezi zeměmi má být vyšší (rozdíl je podhodnocen), žáci v zemi A mají vyšší standardy pro hodnocení
4. $a > b$: hodnocení žáků odpovídá objektivní úrovni, rozdíl mezi zeměmi je správný, žáci v zemi A užívají škálu stejným způsobem jako v zemi B
5. $a > b$: hodnocení žáků při prostém porovnání odpovídá tomu, že objektivní úroveň je vyšší v zemi A, nicméně absolutní rozdíl mezi zeměmi má být menší (rozdíl je nadhodnocen), žáci v zemi A mají nižší standardy pro hodnocení



Obrázek 3. Základní myšlenka metody ukotvujících vinět: ilustrace parametrického modelu s výčtem všech možností porovnání skutečných a reportovaných distribucí

2. Aplikace v sociálních vědách

Metoda ukotvujících vinět je v určitých oblastech sociálních věd často využívána. Jednou z takových oblastí je měření zdraví (Salomon, Tandon, & Murray, 2004; Bago d'Uva, O'Donell, & van Doorslaer, 2008; Grol-Prokopczyk, Freese, & Hauser, 2011; Voňková & Hullegie, 2011; Peracchi & Rossetti, 2012; Kok et al., 2012; Voňková, 2013), dále spokojenosti se zdravotním systémem (Murray et al., 2003; Sirven, Santos-Eggimann, & Spagnoli, 2012), životní spokojenosti (Kapteyn, Smith, & van Soest, 2010; Angelini et al., 2012a; Angelini et al., 2012c), spokojenosti se sociálními kontakty (Bonsang & van Soest, 2012), spokojenosti s prací (Kristensen & Johansson, 2008) a zdravotních omezení při výkonu práce (Angelini, Cavapozzi, & Paccagnella, 2012b; Kapteyn et al., 2007).

Jedna z prvních oblastí, kde byla metoda aplikována, je vnímaná politická účinnost (*political efficacy*). Aplikace je popsána ve studii King a kol. (2004). Tuto aplikaci podrobněji popisujeme v následující části. King et al. (2004) zde také poprvé formálně popisuje základy této metody a shrnuje předcházející metodologické přístupy. V české literatuře jsou tyto přístupy uvedeny ve Voňková (2012b).¹

¹ Jedná se například o tzv. sebekotvící škálu (*self-anchoring scale*), v nichž respondenti identifikují koncové body škály pomocí svých vlastních příkladů, jako jsou konkrétní osoby či události. Následně mají mezi tyto body vložit vlastní hodnocení předloženého jevu. Je zjevné, že příklady uváděné respondenty se mohou lišit, a tudíž tento přístup není robustní pro srovnávání výpovědí velmi odlišných respondentů. Někdy se též používají tzv. zavedené kotvící otázky (*designated anchors*), což jsou otázky týkající se měřeného konceptu, a o kterých jsou výzkumníci přesvědčeni, že nejsou zatíženy heterogenitou v interpretaci škály u jednotlivých respondentů. Zůstává však otázkou, zdali tyto otázky skutečně nejsou touto heterogenitou zatíženy. Dále výzkumníci používají porovnání odpovědi na položku s (váženým) průměrem odpovědi na zbylé položky měřící daný koncept, přičemž předpokládají, že průměr není zatížen heterogenitou, některé otázky jsou pod a jiné zase nad „zlatým standardem“. Tato úvaha však může vést k systematickým chybám, může například docházet k tomu, že většina otázek heterogenitou prostě zatížena je. Metoda ukotvujících vinět se snaží tyto problémy redukovat a pomocí přímého změření heterogenity (sebe)hodnocení respondentů očistit.

Mimo aplikace v oblasti vnímané politické účinnosti uvádíme v dalších částech i podrobnější popis dalších dvou oblastí, měření ostrosti zraku pomocí sebehodnocení respondentů (King et al., 2004) a měření zdravotních problémů při výkonu práce na základě výpovědi respondentů (Kapteyn et al., 2007), které jsou v literatuře k metodě ukotvujících vinět dobře známy. U těchto třech příkladů mimo jiné uvádíme i příklady znění ukotvujících vinět. Další příklady formulací vinět pro oblast zdraví, které byly užity ve výzkumu Survey of Health, Ageing and Retirement in Europe (SHARE-ERIC, 2016), jsou uvedeny v příloze 1. Gary King na svých webových stránkách uvádí též příklady ukotvujících vinět z různých oblastí (King, 2017).

2.1. Klíčové aplikace v sociálních vědách

2.1.1. Vnímaná politická účinnost (*political efficacy*)

Metodu ukotvujících vinět použil poprvé v sociálních vědách King a kol. (2004). Popsal základní strukturu statistického modelu a jeho předpokladů a aplikoval ji ve výzkumu možností obyvatel Číny a Mexika ovlivňovat politická rozhodnutí a také ve výzkumu ostrosti zraku v Číně a na Slovensku. Výzkumu o možnostech ovlivňování politických rozhodnutí se účastnilo 430 respondentů z Číny a 551 respondentů z Mexika. Sběr dat byl ukončen v červnu 2002 za podpory Světové zdravotnické organizace. Respondentům byla v dotazníkovém šetření předložena následující sebehodnotící otázka:

*Do jaké míry máte možnost ovlivňovat vládu v záležitostech, které se Vás týkají?
vůbec, málo, středně, velmi, neomezeně.*

Mimo této otázky byli respondenti vyzváni zhodnotit na stejné škále jako sebehodnotící otázku pět ukotvujících vinět:

[Alison] nemá dostatek pitné vody. Společně se svými sousedy podporuje v následujících volbách opozičního kandidáta, který slibuje, že se touto záležitostí bude zabývat. Zdá se, že mnoho lidí v její oblasti má stejně jako ona pocit, že opoziční kandidát porazí současného zastupitele.

*Do jaké míry má [Alison] možnost ovlivňovat vládu v záležitostech, které se jí týkají?
vůbec, málo, středně, velmi a neomezeně*

[Imelda] nemá dostatek pitné vody. Společně se svými sousedy se snaží na tento problém upozornit sběrem podpisů pod petici. Plánují ji předložit všem politickým stranám před nadcházejícími volbami.

Do jaké míry má [Imelda] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Jane] nemá dostatek pitné vody, protože úřady provádějí změny v územním plánu za účelem jeho průmyslového rozvoje. V kampani k nadcházejícím volbám opoziční strana slíbila, že se touto záležitostí bude zabývat, ale Jane má pocit, že je zbytečné volit opozici, protože současní zastupitelé jistě vyhrají.

Do jaké míry má [Jane] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Toshiro] nemá dostatek pitné vody. Je zde skupina místních vedoucích představitelů, kteří s problémem mohou něco dělat, ale ti říkají, že v současné době je nejdůležitější průmyslový rozvoj a nikoli pitná voda.

Do jaké míry má [Toshiro] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Moses] nemá dostatek pitné vody. Ráda by to změnila, ale nemůže volit, a má pocit, že nikdo z vlády tuto záležitost neřeší. V tichosti trpí a doufá, že se něco v budoucnu stane.

Do jaké míry má [Moses] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

Při porovnání samotného sebehodnocení došli autoři k závěru, že respondenti v Číně uváděli signifikantně větší možnosti ovlivňovat politická rozhodnutí ve srovnání s respondenty v Mexiku. Jak autoři uvedli, výsledek byl neočekávaný, politická situace v těchto zemích mu neodpovídala. Porovnání sebehodnocení bylo provedeno pomocí histogramu, ale i probitového modelu pro ordinální proměnnou (*ordered probit model*), ve kterém byly jako vysvětlující proměnné mimo *dummy* proměnné za Čínu použity i věk, pohlaví a vzdělání respondentů

(výsledky i po kontrole o tyto proměnné vykazovaly signifikantně lepší výsledek v Číně).

Nicméně analýza odpovědí na viněty poukázala na zásadní problém. Předložené hypotetické příběhy byly respondenty z Číny hodnoceny velmi mírně ve srovnání s respondenty z Mexika (systematicky častěji volili respondenti z Číny vyšší škálové kategorie *středně, velmi a neomezeně*) a poukázaly tak na rozdílné standardy při hodnocení daného konceptu. Sebehodnocení bylo následně korigováno o rozdílné užívání škály pomocí neparametrického přístupu i parametrického modelu metody ukotvujících vinět (specifikace modelů v kapitole 3). V parametrickém modelu byly využity kontrolní proměnné stejné jako v probitovém modelu pro ordinální proměnnou. Pomocí obou přístupů dospěli autoři k závěru, že vnímaná politická účinnost je signifikantně vyšší v Mexiku než v Číně, což bylo v souladu s očekávaným výsledkem. Je zásadní zdůraznit, že bez korekce o rozdílné užívání škály bychom došli k neočekávaným výsledkům. Pokud bychom vzali hodnocení politické situace experty jako zlatý standard měření, lze vyslovit ještě silnější tvrzení, že porovnání možností ovlivňovat politické rozhodnutí u respondentů v Číně a Mexiku pouze na základě sebehodnocení je chybné.

2.1.2. Ostrost zraku (*visual acuity*)

Další výzkum, na který se autoři King a kol. (2004) zaměřili, se týkal ostrosti zraku respondentů v Číně a na Slovensku. Data byla sebrána v rámci šetření Světové zdravotnické organizace. V Číně byla sebrána data od 9484 respondentů v únoru roku 2001. Na Slovensku se výzkumu účastnilo 1183 respondentů a sběr dat se uskutečnil v prosinci roku 2000. Všem respondentům byla položena následující sebehodnotící otázka:

*Jak velké obtíže jste měl/a v posledních třiceti dnech s viděním a rozeznáváním známých osob na druhé straně ulice (tj. ve vzdálenosti přibližně 20 metrů)?
žádné, mírné, střední, vážné, extrémní / neviděl/a jsem.*

Do výzkumu bylo zařazeno osm ukotvujících vinět. Příklad jedné z vinět je následující:

Angela potrebuje brýle při čtení novin (a při navlékání jehly). Zřetelně rozeznává lidské obličej a rozpoznává detaily na obrázcích ze vzdálenosti 10 metrů. Nemá problémy vidět při tlumeném světle.

Po každé z vinět následovala otázka ohledně obtíží s viděním a rozeznáváním známých osob se stejnou škálou, jako je v sebehodnotící otázce. Na viněty odpovídala náhodně vybraná polovina respondentů. Do výzkumu byla také zařazena *objektivní míra ostrosti zraku*, tzv. *Snellen Eye Chart test* (test s využitím Snellenovy tabule pro určení ostrosti zraku). Jedná se o test s otáčejícím se písmenem E, které je na každé další řádce tabule zobrazeno v menší velikosti. Respondenti měli pak určit, jakým směrem je E skloněno. Test s využitím Snellenovy tabule je znám ve více variantách. Někdy se používají různá písmena, která jsou na jednotlivých řádcích tabule postupně menší. Výhodou otáčejícího se E je, že testovaný nemusí umět písmena a určuje pouze sklon písmene E. Ačkoli i tento test může obsahovat chybu měření, tato chyba by neměla být závislá na rozdílnosti v užívání škály v různých kulturách či zemích, a tudíž by měla poskytnout relativně dobrý standard pro srovnání se sebehodnocením a následně i korigovaným sebehodnocením pomocí vinět. Test ostrosti zraku byl proveden s náhodně vybranou polovinou respondentů. Výsledky testu se Snellenovou tabulí ukázaly, že respondenti z Číny mají podstatně horší zrak než respondenti ze Slovenska. Tento výsledek byl vzhledem k horší zdravotní péči a horší dostupnosti brýlí v Číně očekáván.

Při porovnání sebehodnocení bez korekce o rozdílné užívání škály pomocí běžného probitového modelu pro ordinální proměnnou nebyl tento zásadní rozdíl detekován, rozdíl mezi zeměmi byl statisticky nesignifikantní. Oproti tomu byl výsledek po korekci o rozdílné užívání škály pomocí parametrického modelu metody ukotvujících vinět (detaily modelů v kapitole 3) v souladu s výsledkem očního testu. Srovnání výsledků testu s výsledky probitového modelu pro ordinální proměnnou a modelu metody ukotvujících vinět umožňuje srovnat užitečnost obou modelů.

Jak autoři King a kol. (2004) dále poznamenávají, bylo by obecně užití testů řešením k problému heterogenního užívání škály při porovnávání respondentů. Nicméně jejich užití především ve velkých sběrech dat je mnohem dražší a dodržení kvality měření bývá mnohem náročnější než u běžně administrovaných dotazníkových položek. Běžní tazatelé nejsou většinou

školení pro provádění medicínských testů. A i kdyby testy prováděli lékařští experti, bylo by pravděpodobně měření ve zcela různých podmínkách, v nichž respondenti žijí, ovlivněno velkou chybou. Autoři mimo jiné i provedli několik *objektivních* testů pro různé koncepty a došli k závěru, že chyba v některých verzích těchto testů dokonce výrazně převážila chybu z měření pomocí nekorigovaného sebehodnocení. Navíc, a to je zásadní, pro mnoho konceptů, jako je vnímaná politická účinnost (*political efficacy*), nejsou objektivní míry dostupné.

2.1.3. Zdravotní problémy při výkonu práce (*work disability*)

Další velmi zdařilá aplikace metody ukotvujících vinět se týkala toho, že respondenty udávaná míra zdravotních omezení při výkonu práce je v mnoha západních zemích výrazně odlišná, zatímco *objektivní* zdravotní stav je v těchto zemích podobný. Kapteyn a kol. (2007) se konkrétně zaměřili na Nizozemsko a Spojené státy americké. V Nizozemsku byla data sbírána v roce 2003 od 1977 respondentů tvořících reprezentativní výběr dospělé nizozemské populace (byl využit internetový panel CentERpanel, který je provozován oddělením CentERdata sídlícím na Univerzitě v Tilburgu). Ve Spojených státech byl pro sběr dat využit RAND MS Internet panel, jehož participanti byli získáni oslovením respondentů starších 40 let z Monthly Survey (MS) prováděném Michigan's Survey Research Center. Celkově se výzkumu zúčastnilo 672 amerických respondentů, sběr dat byl uskutečněn v první polovině roku 2004. Pro doplnění relativně malého vzorku amerických respondentů byla využita též data z roku 1998 o sebehodnocení zdravotních omezení při výkonu práce od 15 740 respondentů mladších 75 let z Health and Retirement Study (HRS). Otázka k hodnocení míry vlastních zdravotních problémů při výkonu práce byla:

Máte nějaké zdravotní poškození nebo nějaký zdravotní problém, které vás omezují ve vykonávání určitého typu či množství placené práce?

Otázka měla dvoubodovou škálu *ano* a *ne* pro americké respondenty a jednu část nizozemských respondentů. Druhá část nizozemských respondentů odpovídala na pětibodové škále: 1) *vůbec ne*, 2) *ano, jsem mírně limitován*, 3) *ano, jsem středně limitován*, 4) *ano, jsem vážně limitován*, 5) *ano, jsem extrémně limitován, nemohu pracovat*.

Dále bylo administrováno pět vinět v každé ze tří domén zdravotních problémů při výkonu práce – bolest, poruchy emocí a onemocnění srdce. Volba domén byla učiněna na základě předchozích analýz, které indikovaly, že tyto tři domény patří k nejčastějším příčinám pracovní neschopnosti. Příklad viněty popisující nejnižší úroveň bolesti je:

[Katie] příležitostně cítí při práci bolest v zádech, ale v posledních několika měsících tento pocit neměla. Pokud bolest pocítí, typicky trvá jen několik dní.

Příklad viněty popisující nejvyšší úroveň bolesti je:

[Mark] má bolest v zádech a nohách, která je přítomná téměř pořád. Bolest se zhorší, když pracuje. Ačkoli mu léky pomáhají, při pohybu a držení a zvedání věcí v práci se necítí dobře.

Za vinětami následovala otázka, zdali si respondenti myslí, že daná osoba je omezena ve výkonu své práce s pětibodovou škálou *vůbec, mírně, středně, vážně, extrémně*. Sebehodnoticí otázka byla u části nizozemských respondentů pouze dvoubodová a u některých pětibodová (totožná se škálou ve vinětách). Bylo tudíž učiněno porovnání mezi těmito škálami, hranice mezi odpověďmi *ano* a *ne* na dvoubodové škále se nacházela mezi hranicí mezi odpověďmi *ne* a *mírně* a hranicí mezi *mírně* a *středně*, a tudíž hranice na dvoubodové škále byla modelována jako vážený průměr uvedených hranic na pětibodové škále.

Pro zjištění *objektivní* úrovně zdravotního stavu sloužily otázky:

Řekl ti někdy tvůj lékař, že máš cukrovku? Artritidu? Hypertenzi? Srdeční problém? Poruchy emocí? Že jsi prodělal mozkovou mrtvici?

Výsledky výzkumu jsou velmi zajímavé. Respondenty udávaná míra zdravotních problémů při výkonu práce je v Nizozemsku přibližně dvakrát větší než ve Spojených státech (ve věkové skupině 25–34 je podíl reportovaných problémů v Nizozemsku 17,3 % versus jen 7,4 % ve Spojených státech, ve věkové kohortě 25–44 je to 22,8 % a 11,3 %, v kohortě 45–54 je to 36,8 % a 17,6 %, v poslední sledované věkové kohortě sice není podíl reportovaných problémů dvojnásobný, ale absolutní rozdíl je stále velký – 37,1 % a 25,9 %). Indikátory *objektivního* zdravotního stavu však poukazují na protichůdný závěr. Nizozemci mají dle konstatování jejich lékaře podstatně menší výskyt cukrovky, artritidy a prodělané mozkové mrtvice ve srovnání s Američany. Podíl výskytu

hypertenze a srdečních problémů je vyrovnaný. Jediný indikátor, který je vyšší v Nizozemsku než v Americe, je podíl výskytu poruch emocí, což je ale nejvíce *subjektivní doména*.

V centru zájmu autorů však není konstatovat, že Nizozemci jsou zdravější než Američani, nýbrž to, že pomocí reportovaných zdravotních problémů nelze vysvětlit hodnoty dostupných objektivních indikátorů při porovnání Nizozemska a Spojených států. Hodnocení patnácti vinět pomáhá vysvětlit tento kontrast. Především u vinět k doménám bolesti a poruch emocí se projevila velká heterogenita. Američtí respondenti udávají mnohem častěji, že hypotetická osoba ve vinětě nemá žádné bolesti (např. u viněty popisující nejnižší úroveň bolesti je to 38,4 % u amerických respondentů versus 24,9 % u nizozemských respondentů; u viněty popisující druhý nejnižší stupeň bolesti je to 29,3 % u amerických respondentů versus 10,5 % u nizozemských respondentů). Volí mnohem méně často ve srovnání s nizozemskými respondenty kategorie *mírné* a *střední* (např. u výše uvedené viněty popisující nejnižší stupeň bolesti je volba kategorie *mírné* 49,1 % u amerických respondentů versus 63,3 % u nizozemských respondentů). Nejvyšší kategorie škály *extrémní* je častěji volena americkými respondenty (u viněty popisující nejvyšší úroveň bolesti volí kategorii *extrémní* 23,4 % amerických respondentů versus jen 14,9 % nizozemských respondentů). Podobná tendence byla též pozorována při hodnocení vinět popisujících poruchy emocí. Nejméně viditelný rozdíl v hodnocení vinět byl nalezen u domény popisujících srdeční problémy (doména popisující nejméně subjektivní zdravotní stav). Obecně lze říci, že američtí respondenti mají tendenci volit krajní body škály *žádné* a *extrémní* bolesti, zatímco nizozemští respondenti mají naopak sklon k volbě středových kategorií. U problémů popisujících mírné projevy nemoci jsou tedy nizozemští respondenti přísnější, mají vyšší standard pro hodnocení. Naopak u problémů popisujících závažné projevy nemoci jsou nizozemští respondenti mírnější, mají nižší standard pro hodnocení.

Pro očištění vlivu užití škály při reportování zdravotních problémů při výkonu práce byl odhadnut model s DIF (*differential item functioning*). Ten byl porovnán s modelem, který toto neumožňoval, tedy nebylo vzato v úvahu potenciální rozdílné užívání škály (model bez DIF). Predikované zdravotní problémy při výkonu práce v modelu bez DIF pro věkovou kohortu 51–64 let byly 22,7 % pro Spojené státy a 35,8 % pro Nizozemsko. Nizozemsko tedy

reportuje o více než polovinu zdravotních problémů více než Spojené státy v této věkové kohortě. Model s DIF však zásadně tento procentní rozdíl zmenšil. Pokud by nizozemští respondenti užívali škálu stejným způsobem jako američtí respondenti (používání americké škály je bráno jako fixní), snížilo by se procento reportovaných problémů na 28,3 %, rozdíl se tedy zmenší na přibližně čtvrtinu.

Zkoumány byly nejen rozdíly mezi zeměmi, ale i rozdíly mezi jednotlivými skupinami respondentů definovanými dle vzdělání, věku a pohlaví v rámci jednotlivých států. Pro americké respondenty například platí, že jejich reportované problémy při výkonu práce (predikce pomocí modelu bez DIF) signifikantně klesají s úrovní dosaženého vzdělání, rostou s věkem a nejsou signifikantně odlišné pro muže a ženy. Efekty vzdělání a věku jsou vyšší ve Spojených státech než v Nizozemsku. Po očištění o rozdíly v užívání škály (predikce pomocí modelu s DIF) je efekt úrovně vzdělání a věku u amerických respondentů mnohem větší. V Nizozemsku nebyl nalezen vztah mezi užíváním škály a vzděláním. Ve Spojených státech jsou ženy méně přísné než muži (hranice mezi *žádné* a *mírné* je vyšší pro ženy než muže). V Nizozemsku nebyl opět nalezen vztah mezi užíváním škály a pohlavím. Vztah mezi pohlavím a zdravotními problémy při výkonu práce je v obou zemích po očištění vlivu užívání škály nesignifikantní. Lze shrnout, že v případě Spojených států jsou ženy, lidé s nižším vzděláním a starší lidé méně přísní při hodnocení zdravotních problémů při výkonu práce. Dále obecně lze konstatovat, že nizozemští respondenti jsou mnohem homogennější při hodnocení zdravotních problémů při výkonu práce než američtí respondenti.

2.2. Dosavadní využití metody ukotvujících vinět v pedagogickém výzkumu

V této části shrneme všechny nám známé aplikace metody ukotvujících vinět v pedagogickém výzkumu. Aplikace rozdělíme dle typu respondentů (rodič a žák) a dle toho, zda respondent hodnotí sám sebe či nikoli (sebehodnocení a hodnocení jiných): a) rodičovské hodnocení týkající se spolupráce s učiteli, ředitelem a ostatními rodiči, b) žákovské sebehodnocení motivace, osobnostních charakteristik, znalostí a dovedností v určitých oblastech, nekázně a životní spokojenosti, c) žákovské hodnocení týkající se učitelova řízení třídy

a učitelovy podpory žáků.

Pedagogický výzkum se z velké části vztahuje k populaci žáků. Tato populace je oproti populaci rodičů (dospělých) novou populací, která začíná být zkoumána pomocí metody ukotvujících vinět. Přináší tím mnoho metodologických výzev, jako je vhodná formulace ukotvujících vinět pro respondenty nízkého věku, hledání nových zdrojů heterogenity ve stylu užívání škály u těchto respondentů a s tím související hledání odpovědí na paradoxy při porovnávání (skupin) těchto respondentů (např. negativní korelace mezi motivací učit se a výsledky v testech při srovnávání zemí v He & van de Vijver, 2016).

Doposud se většina aplikací metody ukotvujících vinět mimo pedagogický výzkum týkala sebehodnocení. V některých nových aplikacích metody v pedagogickém výzkumu se však setkáváme nejen se sebehodnocením, ale také s hodnocením jiných osob (např. žák hodnotí podporu žáků ze strany jejich učitele). Z hlediska splnění předpokladů metody ukotvujících vinět, jako je konzistence odpovědí (více o tomto předpokladu je pojednáno v části 5.1.1), může toto hrát důležitou roli. Každopádně je nutné předpoklady metody ověřovat a řešit jejich případné porušení již při konstrukci dotazníku a sběru dat.

Přehled některých studií v pedagogickém výzkumu je stručně uveden i v přehledovém článku Voňková, Papajoanu a Bendl (2016). Zde připojujeme další aplikace, které v době vydání článku nebyly ještě publikovány či nebyly známy jejich první výsledky (výzkum byl pouze rozpracován). Navíc mimo detailnějšího popisu připojujeme i příklady ukotvujících vinět užitých v těchto studiích. Dále článek Voňková (2012a) popisuje možnosti využití metody ukotvujících vinět v pedagogickém výzkumu, které v době vydání tohoto článku byly aktuální.

Mimo uvedených studií je nutno dodat, že na Pedagogické fakultě Univerzity Karlovy vzniklo pod vedením autorky této knihy několik diplomových prací, které se zabývají aplikací metody ukotvujících vinět. Vlasta Beranová a Kateřina Králová aplikovaly metodu v oblasti hodnocení vlastních znalostí anglického jazyka (Beranová, 2015; Králová, 2015). Jedná se o návazný výzkum na Voňková a Šrámková (2013), který se zaměřoval na porovnávání znalostí anglického jazyka u jednotlivých žáků středních škol. Vlasta Beranová pak porovnávala žáky na víceletých gymnáziích a základních školách, Kateřina

Králová studenty pedagogických a ekonomických oborů. Jan Hrabák porovnával ve své diplomové práci hodnocení vlastních znalostí a dovedností v oblasti informačních a komunikačních technologií u studentů středních škol se zaměřením a bez zaměření na IT (Hrabák, 2014). Se svou školitelkou publikoval výsledky výzkumu v odborných časopisech (Voňková & Hrabák, 2015a; Voňková & Hrabák, 2015b). Petra Vlčková se zaměřila na měření životní spokojenosti učitelů (Vlčková, 2016).

V České republice byla založena výzkumná skupina ERA Group (Educational Research Anchors Group), která se zaměřuje na využití metody ukotvujících vinět v pedagogickém výzkumu a demonstraci jejího potenciálu pro zvyšování validity měření jevů. Výzkumná skupina má webové stránky, které doporučujeme prostudovat výzkumníkům v pedagogice (Educational Research Anchors Group, 2017).

2.2.1. Rodičovské hodnocení spolupráce s učiteli, ředitelem a ostatními rodiči

Buckley a Schneider (2007) se zabývají porovnáním hodnocení školy rodiči žáků navštěvujících tzv. *charter schools* (americké veřejné školy s relativně velkou autonomií při jejich řízení a tvorbě kurikula) a tradiční veřejnou školu ve Washingtonu D. C. Rodiče žáků *charter schools* hodnotí školu svých dětí lépe a jsou spokojenější s jejími různými charakteristikami ve srovnání s rodiči žáků navštěvujících tradiční školy. Otázkou však je, do jaké míry zde hraje roli tzv. efekt růžových brýlí (Erikson, 1982). Ve svém výzkumu se Buckley a Schneider (2007) dotázali rodičů nejen na to, jak jsou spokojeni se školou svých dětí, ale předložili jim též viněty. Konkrétně rodiče hodnotili tři dimenze školy svých dětí. Jednalo se o míru spolupráce: a) rodičů a ředitele, b) rodiče a učitele jeho dítěte, c) mezi rodiči. Ke každé dimenzi byla formulována jedna sebehodnoticí otázka a tři ukotvující viněty. Například sebehodnoticí otázka na míru spolupráce rodičů a ředitele je následující:

*Jak byste zhodnotili spolupráci mezi ředitelem a rodiči na škole Vašeho dítěte?
A, ..., F.*

Poznamenejme, že tuto škálu rodiče dobře znají, na amerických školách se

běžně využívá. Příkladem ukotvující viněty k této otázce je:

Představte si následující situaci: Ve třídě je dítě, které pravidelně vyrušuje při vyučování a je učitelem pravidelně posíláno k řediteli.

Ředitel této školy pošle rodičům dopis o problémovém chování jejich dítěte, ale situaci již dále nijak neřeší.

Jak byste zhodnotili spolupráci mezi ředitelem a rodiči na této škole?

A, ..., F.

Závěry studie jsou zajímavé. Volba *charter school* sama o sobě nevede k systematické preferenci pozitivních odpovědí ani v jedné dimenzi. Navíc se ukázalo, že demografické charakteristiky (např. rasa, pohlaví a vzdělání) rodičů dětí v *charter schools* predikují posun odpovědí opačným směrem, tj. že jsou mnohem náročnější a hodnotí školu kritičtěji.

Buckley (2008) se následně pro zmiňované měřené oblasti (spolupráce rodičů s ředitelem, učitelem a ostatními rodiči) zaměřoval na metodologické otázky při sestavování dotazníku týkající se pořadí sebehodnoticích otázek a vinět a také řazení samotných vinět. Poznamenejme, že znění všech vinět k jednotlivým oblastem lze najít v příloze 2.

2.2.2. Žákovské sebehodnocení motivace, osobnostních charakteristik, znalostí, nekázně a životní spokojenosti

V této části shrneme výsledky všech studií s aplikacemi metody ukotvujících vinět, ve kterých bylo využito žákovské sebehodnocení. První typ aplikací se týká hodnocení motivace k učení (He & van de Vijver, 2016) a osobnostních charakteristik svědomitosti a otevřenosti (Primi et al., 2016). Druhý typ aplikace se týká hodnocení vlastních znalostí a dovedností, konkrétně hodnocení vlastních ICT znalostí a dovedností (Voňková & Hrabák, 2015a; Voňková & Hrabák, 2015b) a hodnocení vlastních dovedností v anglickém jazyce (Voňková & Šrámková, 2013). Dále je metoda aplikována též pro hodnocení nekázně (Voňková, Bendl, & Papajoanu, 2017). Mimo to byla v rámci šetření CLoSE (Czech Longitudinal Study in Education, <http://czechlongitudinal.blogspot.cz/>) sebrána data k životní spokojenosti

u reprezentativního vzorku českých žáků 5. ročníků. Aplikace týkající se nekázně žáků a první výsledky analýzy životní spokojenosti žáků s využitím metody ukotvujících vinět shrneme ve třetím typu aplikací.

Sebehodnocení motivace k učení a osobnostních charakteristik

He a van de Vijver (2016) studovali vztah mezi motivací žáků učit se matematiku a výsledkem v testu matematické gramotnosti s využitím dat PISA 2012. Snažili se objasnit paradoxní negativní vztah mezi těmito proměnnými na úrovni zemí.

Vnitřní motivace učit se matematiku je v PISA 2012 měřena následujícími čtyřmi otázkami: 1) *Baví mě číst články nebo knihy s matematickou tematikou.* 2) *Na hodiny matematiky se těším.* 3) *Učím se matematiku, protože mě baví.* 4) *Věci, které se učíme v matematice, mě zajímají.* Vnější motivace je měřena též čtyřmi otázkami: 1) *V matematice má smysl se snažit, protože se mi znalosti a dovednosti budou hodit v práci, kterou chci v budoucnu vykonávat.* 2) *Učit se matematiku se mi vyplatí, protože si tak zlepším své vyhlídky na získání dobrého zaměstnání.* 3) *Matematika je pro mě důležitý předmět, protože ji budu potřebovat ve svém dalším studiu.* 4) *V matematice se naučím mnoho věcí, které mi pomohou získat dobré zaměstnání.* Žáci odpovídali na čtyřbodové Likertově škále *rozhodně souhlasím, souhlasím, nesouhlasím a rozhodně nesouhlasím.*

Korelace mezi vnitřní motivací učit se matematiku a výsledkem v testu matematické gramotnosti (udané pěti přijatelnými hodnotami (*plausible values*)² v datovém souboru PISA 2012) se na úrovni studentů v jednotlivých zemích pohybovala od $-0,13$ do $0,42$, mediánová hodnota byla $0,17$. Co se týká vnější motivace, korelace se pohybovala od $-0,11$ do $0,43$, mediánová hodnota byla $0,13$. Na úrovni jednotlivců v jednotlivých zemích lze tedy říci, že je vztah mezi motivací a výsledkem v testu pozitivní, ačkoli velikost účinku (*effect size*) se napříč jednotlivými zeměmi velmi liší (He & van de Vijver, 2016, citováno dle d' Ailly, 2003; Watkins, 2000). Na úrovni zemí jsou však korelace negativní. Medián korelací vnitřní motivace s pěti přijatelnými hodnotami je $-0,53$, pro vnější motivaci je roven $-0,52$.

² Přijatelné hodnoty (*plausible values*) jsou v kontextu výzkumu PISA (Programme for International Student Assessment) náhodné hodnoty z rozdělení měřené gramotnosti.

Uvedené korelace byly porovnány s korelacemi po očištění o rozdílné užívání škály pomocí metody ukotvujících vinět. Všechny uvedené korelace se zvýšily. Na individuální úrovni se korelace zvýšily z 0,17 (před očištěním) na 0,26 (po očištění) pro vnitřní motivaci a dále z 0,13 na 0,20 pro vnější motivaci. Na úrovni zemí bylo zvýšení znatelnější, a to z -0,53 na -0,11 pro vnitřní a dále z -0,52 na -0,04 pro vnější motivaci. Ve studii jsou též analyzovány další přístupy jako korekce tendence k volbě extrémních kategorií a technika přehánění. O těchto metodách si čtenář může přečíst v kapitole 7. Všechny techniky se ukázaly jako vhodné (posunuly záporné korelace směrem ke kladným hodnotám), neúčinnější se ukázala metoda ukotvujících vinět.

Další studií s využitím metody ukotvujících vinět uvádí Primi a kol. (2016). Autoři zjišťovali vlastnosti škál pro měření svědomitosti a otevřenosti vůči novým věcem u studentů. Jednalo se o vnitřní konzistenci (*internal consistency*, závislost mezi jednotlivými položkami měřeného konceptu měřená pomocí Cronbachova alfa), diskriminační validitu (*discriminant validity*, korelace mezi různými škálami) a kriteriální validitu (*criterion validity*, predikce výsledků v testech z matematiky a jazyka). Pracovali s korigovanými škálami, škály očistili pomocí ukotvujících vinět týkajících se oblasti popsané v sebehodnotící otázce (např. hodnocení vlastní svědomitosti očistili pomocí ukotvujících vinět popisujících svědomitost), ale také vinět týkajících se jiné oblasti (k očištění hodnocení vlastní svědomitosti použili viněty popisující otevřenost).

Pracovali s reprezentativním vzorkem studentů z Rio de Janeira navštěvujících 12. ročník. Počet zapojených škol je 216, mediánový počet studentů z každé školy je 50, minimální počet 4, maximální počet 150. Průměrný věk studentů je 18,2 (SD = 1,1). Vzhledem k velkému počtu studentů, kteří ze škol odcházejí a pak se vracejí, je více než čtvrtina studentů ve věku 19 až 24 let. Celkový počet pozorování je 8582.

Použitý dotazník je brazilský dotazník SENNA 1.0 obsahující otázky týkající se hodnocení vlastních dovedností s úzkým vztahem k taxonomii *Big Five* (John, Nauman, & Soto, 2008). Při sběru dat se tedy zaměřili nejen na svědomitost a otevřenost, ale také na neuroticismus (*neuroticism*), extraverci (*extraversion*) a na vstřícnost (*agreeableness*). V článku však data k těmto dalším třem oblastem nejsou diskutována. Mimo sebehodnotících otázek sbírali autoři i data k ukotvujícím vinětám. Příklad ukotvujících vinět týkajících se svědomitosti je

následující (další viněty jsou uvedeny v příloze 2):

Manuela má dobrý smysl pro pořádek, ale někdy je v její místnosti několik dní nepořádek. Má sklon dokončovat školní úkoly těsně před termínem odevzdání.

Jak je Manuela podle tvého názoru svědomitá?

1 = vůbec, 2 = trochu, 3 = středně, 4 = hodně, 5 = zcela

Při zpracování dat byl využit neparametrický přístup, přičemž při práci s intervalovými hodnotami bylo zvoleno minimum v rámci daného intervalu jako ta hodnota, která interval reprezentuje (o práci s intervalovými hodnotami a jejich limitacích se lze dočíst v části 3.1).

Závěry studie jsou následující. Vnitřní konzistence měřená pomocí Cronbachova alfa se pro škálu svědomitosti a otevřenosti po korekci zvýšila, a to bez ohledu na to, zdali byla využita viněta týkající se stejné či jiné oblasti než té, ke které se vztahuje sebehodnotící otázka. Dále v případě použití vinět týkajících se otevřenosti se velmi zvýšila korelace mezi otevřeností a svědomitostí, diskriminační validita tedy byla ohrožena. Korekce pomocí vinět odpovídajících oblasti sebehodnotících otázek naopak vedla ke snížení korelace mezi svědomitostí a otevřeností, což podporuje diskriminační validitu. Dále původní a korigované skóry pro svědomitost a otevřenost vykazovaly podobné vlastnosti při analýze kriteriální validity. Výsledky však poukazují na to, že korekce škál pomocí vinět odpovídajících oblastí sebehodnotících otázek pomáhá lépe identifikovat vztah k výsledkům konkrétních předmětů. Škála otevřenosti korigovaná pomocí vinět týkajících se otevřenosti více (ve srovnání s korekcí pomocí vinět týkajících se svědomitosti) korespondovala s výsledky v jazykových testech. A podobně škála svědomitosti korigovaná pomocí vinět týkajících se svědomitosti více korespondovala s výsledky v testech z matematiky. Pro zvýšení kriteriální validity se tedy nabízí doporučení užívat viněty popisující stejnou oblast, jaká je v sebehodnotících otázkách.

Hodnocení vlastních znalostí a dovedností

Hodnocení znalosti informačních a komunikačních technologií

Voňková a Hrabák (2015b) zjišťovali s využitím metody ukotvujících vinět (ne)porovnatelnost hodnocení vlastních ICT (information and communication technology) znalostí a dovedností u žáků středních odborných škol.

K hodnocení vlastních znalostí a dovedností užívají následující otázku:

Nejdříve, prosím, ohodnoťte svoji úroveň vědomostí a dovedností spojených s informační a komunikační technologií s desetibodovou škálou 1, ..., 10, kde jednička označovala nejnižší znalosti a dovednosti a desítka označovala nejvyšší znalosti a dovednosti.

Dále zformulovali celkem čtyři ukotvující viněty, které popisují čtyři různé úrovně ICT dovedností. Viněta popisující druhou nejnižší úroveň ICT dovedností je následující (škála je stejná jako u sebehodnotící otázky):

Daniel na počítači umí pracovat s kancelářským balíkem Office. Vytváří pro firmu dokumenty, prezentace nebo tabulky. V Excelu ovládá často používané funkce včetně funkce KDYŽ, COUNTIF, POČET, SUMA, ... Je schopný pracovat s programy, které má v počítači nainstalovány, a umí si další nainstalovat. Je schopen si vyhledat informace na internetu nebo komunikovat se spolupracovníky pomocí e-mailu. Své soubory a složky má upravené ve stromové struktuře složek.

Vzorek žáků tvoří dvě odlišné skupiny – žáci navštěvující střední odborné školy se zaměřením na ICT (ICT žáci, N = 228) a žáci středních odborných škol bez zaměření na ICT, konkrétně žáci s humanitním zaměřením (neICT žáci, N = 147). Autoři ukazují, že metoda ukotvujících vinět pomáhá rozlišit rozdílné užívání škály u těchto skupin studentů a že porovnání korigovaného sebehodnocení u těchto dvou skupin studentů odpovídá porovnání objektivní ICT znalosti a dovednosti (ICT žáci mají vyšší ICT znalosti a dovednosti než neICT žáci). Při porovnání původního (nekorigovaného) sebehodnocení však k tomuto závěru dojít nelze. Závěrem zmiňují, že metoda ukotvujících vinět je tedy velmi nadějným prostředkem pro zvyšování validity výzkumů založených na hodnocení vlastních ICT znalostí a dovedností.

Mezinárodní srovnávací šetření ICILS 2013 (International Computer and Information Literacy Study 2013), ve kterém byla porovnávána úroveň počítačové a informační gramotnosti žáků osmých tříd základních škol ve 21 státech a který přinesl pro Českou republiku velmi pozitivní zjištění, neboť čeští žáci byli v CIL testu (testu počítačové a informační gramotnosti) první, byl zároveň i zdrojem k dalším úvahám o diskrepanci mezi sebehodnocením žáků a jejich výsledkem v testech. Voňková a Hrabák (2015a) porovnávají

výsledek v CIL testu se sebehodnocením základních a pokročilých ICT dovedností pro jednotlivé země. Čeští žáci v tomto ohledu vykazují největší rozdíl mezi hodnocením svých pokročilých ICT znalostí a dovedností a výsledkem v testu (test: 1. místo, hodnocení vlastních pokročilých dovedností: 17. místo), za nimi následují australští žáci (test: 3. místo, sebehodnocení: 20. místo) a dále s odstupem Švýcarsko (test: 10. místo; sebehodnocení: 20. místo). V případě porovnání výsledku v CIL testu s hodnocením základních ICT dovedností předstihla Českou republiku (test: 1. místo; hodnocení vlastních základních dovedností: 9. místo) jen Korea (test: 7., sebehodnocení: 16.). Průměrné sebehodnocení je v České republice v porovnání s ostatními zeměmi tedy nejvíce vzdálené od průměrného výsledku v testu, což lze interpretovat tak, že naši žáci své dovednosti ve srovnání s ostatními zeměmi nejvíce podhodnocují. Mimo České republiky patří k těmto zemím například i Austrálie, Švýcarsko a Korea. Mezi země, kde naopak žáci své ICT dovednosti nejvíce nadhodnocují, patří např. Chile, Chorvatsko a Slovinsko.

Voňková a Hrabák (2015a) dále poukazují nejen na rozdílnost v CIL výsledcích a sebehodnocení na úrovni zemí, ale i v rámci jednotlivých skupin žáků v jednotlivých zemích. Příkladem může být porovnání pohlaví. Dívky dopadly ve všech zemích v CIL testu lépe než chlapci, ale hodnotí své pokročilé ICT dovednosti hůře než chlapci. Navrhují tuto neshodu objasnit pomocí metody ukotvujících vinět. Provádí sekundární analýzu dat sebraných v rámci studie Voňková a Hrabák (2015b). Na základě hodnocení vinět porovnávají standardy pro hodnocení dle dvou kritérií: a) zaměření střední odborné školy – studium na střední škole s ICT zaměřením (ICT) a s humanitním zaměřením (neICT), b) pohlaví. Nejvyšší standardy pro hodnocení mají chlapci na ICT školách, následovaní dívkami na ICT školách a dále dívkami na neICT školách. Nejnižší standardy pro hodnocení mají chlapci na neICT školách. Je nutné podotknout, že se ve studii nepracuje s reprezentativními vzorky uvedených skupin, pro potvrzení či vyvrácení těchto výsledků by tedy bylo ještě vhodné provést analýzu s reprezentativními vzorky.

Hodnocení porozumění psanému anglickému textu

Voňková a Šrámková (2013) ukazují na možnosti aplikace metody ukotvujících vinět při korekci sebehodnocení porozumění psanému textu v anglickém jazyce u studentů střední školy. Diskutují vztah mezi sebehodnocením a korigovaným

sebehodnocením pomocí vinět a dále i vztah mezi (korigovaným) sebehodnocením a výsledkem didaktického testu zjišťujícího schopnost studentů porozumět psanému textu v angličtině. Šetření probíhalo mezi žáky pražských škol SPŠ Zeměměřická a SPŠ Stavební ve věkovém rozmezí od 15 do 20 let. Žákům byl předložen dotazník, který se skládal ze dvou částí. První část byla tvořena sebehodnoticí otázkou a čtyřmi ukotvujícími vinětami, druhá část byla didaktický test zjišťující porozumění psanému textu v angličtině.

Znění sebehodnoticí otázky je:

Označte na stupnici 1 – 10, jak hodnotíte svou schopnost porozumění psaným textům v Aj (schopnost čtení anglických textů). 1 = naprosto nedostatečná; 10 = vynikající.

Ukotvující viněty v první části byly zjednodušené popisy úrovní schopnosti čtení v cizím jazyce podle Společného evropského referenčního rámce (SERR) pro jazyky. Viněty popisovaly úrovně A1, A2, B1 a C1. Příkladem viněty pro úroveň A1 je:

Jana rozumí základním slovům označujícím např. členy rodiny, běžné předměty, nejnámější druhy zvířat atd. Rozumí velmi jednoduchým větám, např. krátkým oznámením na cedulích na ulici, na plakátech nebo v katalogích.

Jak hodnotíte Janinu schopnost porozumění textu?

Žáci hodnotí vinětu na stejné škále jako vlastní schopnosti, tedy na škále desetibodové 1–10, přičemž 1 označuje naprosto nedostatečnou a 10 vynikající schopnost.

Druhá část dotazníku – didaktický test – sestávala ze tří textů seřazených podle náročnosti. První text odpovídal úrovni A2 SERR, druhý text úrovni B1 a třetí text úrovni B2. Podle učebnic, ze kterých se studenti učí, se předpokládalo, že úroveň porozumění textu respondentů se bude pohybovat v rozmezí cca od A1 do B2. Z tohoto důvodu jsou ve vinětách (až na jednu výjimku) popsány nižší úrovně porozumění textu a testová část dotazníku se zaměřuje na testování nižších úrovní porozumění.

Výsledky výzkumu ukazují, že hodnocení vinět vykazovalo mezi studenty velkou variabilitu, což poukazuje na odlišné využití jednotlivých bodů škály.

Korigované sebehodnocení se tedy odlišuje od původního nekorigovaného sebehodnocení. Na příkladu dvou konkrétních studentů autoři ukazují, že korigované sebehodnocení porozumění psanému textu v anglickém jazyce je v souladu s výsledkem v didaktickém testu, zatímco u samotného sebehodnocení tomu tak není.

Hodnocení nekázně

Voňková a kol. (2017) studují využití metody ukotvujících vinět při měření nečestného chování žáků ve školách. Při měření nečestného chování žáků se na toto chování často dotazujeme samotných žáků. Sebehodnocení nečestného chování však může odrážet nejen reálnou úroveň chování, ale i styl užívání hodnoticí škály. Ve studii analyzují autoři hodnocení vlastního podvádění žáků a dopady heterogenity ve stylu užívání škály na porovnání podvádění žáků s různými charakteristikami (pohlaví, známky, vzdělání rodičů). Heterogenita ve stylu užívání škály byla měřena na základě žákovského hodnocení následující ukotvující viněty:

*Honza si dvakrát za poslední měsíc místo odpoledního vyučování, kdy měla třída suplovanou dvouhodinovku tělesné výchovy, vyrazil s kamarády do města. Následně si sám napsal omluvenku, ve které zfalšoval podpis svého otce.
U Honzy hodnotím záškoláctví a podvádění číslicí:*

Škála pro sebehodnocení i hodnocení ukotvující viněty byla pětibodová, přičemž krajní bod 1 označuje, že uvedený přestupek není pro žáka charakteristický (žák je ukázněný), a krajní bod 5 označuje, že uvedený přestupek je pro žáka charakteristický (žák je neukázněný). Ve studii byla analyzována heterogenita ve stylu užívání škály na základě hodnocení vinět popisujících obecnou nekázeň, drzost vůči učiteli a šikanu učitele a také šikanu žáka. Znění všech vinět je uvedeno v příloze 2.

Vzorek tvořili žáci sedmých ročníků všech škol na Praze 6. Celkový počet žáků ve výzkumu byl 670, hodnocení vlastního chování prováděli náhodně vybraní žáci na každé škole (N = 265). Sběr dat byl uskutečněn v roce 2009.

Pomocí metody ukotvujících vinět byl identifikován styl užívání škály při hodnocení vlastního nečestného chování žáků s různými charakteristikami. Vztah mezi charakteristikami žáků a jejich sebehodnocením je výrazně

ovlivněn heterogenitou v užívání škály. Před korekcí o tuto heterogenitu nebyl nalezen (statisticky) významný vztah pohlaví a studijních výsledků k úrovni nečestného chování. Vzdělání rodičů bylo slabým prediktorem. Po korekci se ukazuje, že tyto charakteristiky jsou výraznými prediktory. Jako nejsilnější prediktor se jeví vynikající prospěch, následovaný vzděláním rodičů a pohlavím respondentů. Dále byla diskutována i myšlenka, že pokud se nám podaří identifikovat stejné užívání škály respondentem u různých vinět, je možno zadávat pouze jedinou vinětu a zjednodušit tak sběr výzkumných dat. Empirický výsledek však ukázal, že užitím rozdílných vinět docházíme k různým výsledkům – žáci užívají škálu rozdílně při hodnocení různých druhů kázeňských prohrěšků. Z toho tedy vyplynulo doporučení, že není vhodné aplikovat určitou vinětu na data týkající se rozdílných konceptů nebo domén jednoho konceptu.

Hodnocení životní spokojenosti

V současné době analyzuje autorka této knihy s využitím metody ukotvujících vinět data o spokojenosti se životem u žáků pátých tříd v České republice. Jedná se o novou aplikaci metody ukotvujících vinět – analýza spokojenosti se životem u dětí pomocí metody ukotvujících vinět ještě, pokud je nám známo, nebyla publikována.

K analýze je využit datový soubor získaný v rámci výzkumu CLoSE. Obecný popis vzorku v rámci tohoto výzkumu je uveden například ve Straková a Greger (2013): „Výzkum CLoSE dlouhodobě sleduje žáky, kteří se na jaře roku 2011 zúčastnili ve 4. ročníku mezinárodního výzkumu matematického a přírodovědného vzdělávání (TIMSS) a výzkumu čtenářské gramotnosti (PIRLS). Dotazníkového šetření se zúčastnilo celkem 4438 žáků 5. ročníku ze 163 základních škol v ČR, kteří navštěvovali 215 pátých tříd.“

V rámci dotazníkového šetření byla sebrána i data o životní spokojenosti žáků včetně hodnocení ukotvující viněty. Znění sebehodnotící otázky ke spokojenosti se životem je následující:

Jak jsi celkově spokojený/á se svým životem?

velmi spokojen/a, spokojen/a, ani spokojen/a, ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Ukotvující viněta k hodnocení spokojenosti se životem je následující:

Pavel vychází s rodiči dobře, alespoň jednou měsíčně s nimi vyráží na výlet do přírody. Má dva dobré kamarády, kterým může vše říci. S nimi tráví hodně volného času. Mnoho dalších kamarádů však nemá. Pavel chodí do školy rád, ale tak dvakrát za měsíc je stresovaný kvůli písemkám. Ve škole dostává většinou dvojky. Jednou za měsíc se stane, že dostane i čtyřku. Za to mu rodiče vynadají a trvají na tom, že se musí látku doučit.

*Jak může podle tebe být Pavel celkově spokojený se svým životem?
velmi spokojen, spokojen, ani spokojen, ani nespokojen, nespokojen, velmi nespokojen*

Prozatím máme k dispozici pouze první výsledky analýzy (např. deskriptivní statistiky k hodnocení vinět). Na základě hodnocení ukotvující viněty o životní spokojenosti lze konstatovat, že dívky mají signifikantně vyšší standardy než chlapci. Také žáci, jejichž rodiče mají vyšší vzdělání, mají vyšší standardy pro hodnocení. Na základě těchto výsledků lze očekávat, že po korekci o styl užívání škály se mohou rozdíly v korigované životní spokojenosti mezi těmito skupinami výrazně změnit.

2.2.3. Žákovské hodnocení učitelova řízení třídy a učitelovy podpory žáků

Voňková a kol. (2015) porovnávají žákovské hodnocení učitelova řízení třídy (*teacher classroom management*) v 68 zemích, které se zúčastnily výzkumu PISA 2012. Celkový počet pozorování je v jejich studii přes 310 000, což jsou přibližně dvě třetiny všech zúčastněných žáků ve studii PISA (celkový počet pozorování je přibližně 485 000; existují tři verze žákovských dotazníků, ve dvou verzích jsou zařazeny ukotvující viněty (např. OECD, 2017b)).

Žáci hodnotili řízení třídy jejich učitelem pomocí následujících čtyř otázek:

Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícími tvrzeními?

Učitel umí žáky přimět k tomu, aby poslouchali, co říká.

Učitel ve třídě udržuje kázeň.

Učitel své hodiny začíná včas.

Učitel musí dlouho čekat, než se žáci utiší.

Škála pro odpovědi je čtyřbodová Likertova škála: *rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím.*

Ukotvující viněty byly využity tři, příklad jedné z nich je následující (znění ostatních vinět je v příloze 2):

Přečti si popisy tří učitelů matematiky a pak se rozhodni, do jaké míry souhlasíš s tvrzeními na konci popisů.

Žáci ve třídě paní učitelky Sýkorové jsou v jejích hodinách klidní a ukáznění. Paní učitelka vždy chodí do hodiny včas.

Paní učitelka Sýkorová má nad děním ve třídě kontrolu.

Škála pro odpovědi je stejná jako v případě hodnocení skutečných učitelů. Analýza dat je provedena pomocí parametrického modelu metody ukotvujících vinět (více k tomuto modelu viz část 3.2). Autoři poukazují na signifikantní heterogenitu ve stylu užívání škály při porovnávání zemí a dle této heterogenity klasifikují země například na a) země s obecně nižším standardem při hodnocení učitelova řízení třídy (např. Indonésie, Malajsie, Thajsko, Jordánsko a Rumunsko), či naopak obecně vyšším standardem (např. Šanghaj v Číně, USA včetně amerických států Florida, Massachusetts a Connecticut, které se zúčastnily výzkumu PISA samostatně, dále Británie a Island), b) země, které mají tendenci využívat spíše koncové body škály (např. Litva, Lucembursko, Turecko, Tunisko a Chile), či naopak prostřední body škály (např. Korea, provincie Macao v Číně, čínská Tchaj-pej, Finsko a Rusko), c) země, které volí raději koncový bod škály *rozhodně souhlasím* před *souhlasím* (např. některé země Středního východu, západní Asie, Balkánu a Jižní Ameriky), či naopak volí raději středový bod *souhlasím* před *rozhodně souhlasím* (např. Šanghaj v Číně, Korea, dále Severní Amerika a země severní či západní Evropy jako je Nizozemsko, Norsko, Dánsko).

V článku je dále prezentováno, jak koreluje na úrovni zemí korigovaná

a nekorigovaná hodnocení učitelova řízení třídy s externími proměnnými, jako jsou hrubý domácí produkt, veřejné výdaje na sekundární vzdělávání, procento soukromých škol, existence státem řízené zkoušky při ukončování střední školy a výsledky v matematické a čtenářské gramotnosti zjišťovanými pomocí testů PISA 2012. Zajímavým zjištěním je např. to, že korigované hodnocení je pozitivně korelováno s výsledky v matematické i čtenářské gramotnosti (lepší učitelovo řízení třídy je asociováno s lepšími výsledky v testech), zatímco u nekorigovaného hodnocení jsou korelace signifikantně negativní. Hodnocení kvality učitelova řízení třídy je před korekcí o rozdílné užívání škály signifikantně negativně korelováno s veřejnými výdaji na střední školství (čím vyšší veřejné výdaje, tím horší kvalita řízení třídy), zatímco po korekci docházíme k očekávanému pozitivnímu, sic nesignifikantnímu vztahu. Podobný závěr platí i pro hrubý domácí produkt – před korekcí je vztah k učitelovu řízení třídy negativní, po korekci je vztah signifikantně pozitivní. Tento posun ve výši a signifikanci korelací s externími proměnnými podporuje externí validitu korigovaných hodnocení pomocí metody ukotvujících vinět.

Kyllonen a Bertling (2014a) pracovali stejně jako Voňková a kol. (2015) s daty z výzkumu PISA 2012. Poukazují na problematiku porovnávání zemí na základě odpovědí na dotazníkové položky s hodnoticími škálami. Jako jedno z možných řešení navrhují i metodu ukotvujících vinět. Vypočítali korelaci mezi výsledkem v testu z matematické gramotnosti a podporou žáků ze strany učitele (*teacher support behavior*), a to na úrovni jednotlivých žáků v rámci jedné země (*within country*) a také na úrovni zemí (*between country*). Podporu ze strany učitele uvažují před korekcí, tj. před očištěním o rozdílné užívání škály pomocí vinět, a také po korekci, tj. po korekci o rozdílné užívání škály. Podpora ze strany učitele je měřena pomocí následujících položek zařazených do žakovského dotazníku (OECD, 2017b):

Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícími tvrzeními?

Učitel nám říká, že musíme pilně pracovat.

Když je třeba, učitel se nám více věnuje.

Učitel nám pomáhá s učením.

Učitel nám dává příležitost vyjádřit své názory.

Škála pro odpovědi je čtyřbodová Likertova škála: *rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím.*

Jedna ze tří vinět týkajících se podpory ze strany učitele je následující (další viněty jsou uvedeny v příloze 2):

Přečti si popisy tří učitelů matematiky a pak se rozhodni, do jaké míry souhlasíš s tvrzeními na konci popisů.

Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou.

Paní učitelce Novotné záleží na tom, jak se žáci učí.

Škála pro odpovědi je stejná jako v případě hodnocení skutečných učitelů. Průměrná korelace na úrovni jednotlivců v jednotlivých zemích (*mean within-country correlation*) vzrostla z 0,03 před korekcí na 0,13 po korekci. Korelace na úrovni zemí (*between-country correlation*, $N = 63$) vzrostla z neočekávané negativní hodnoty $-0,45$ před korekcí na $0,29$ po korekci.

Další analýza dat týkající se rozdílného užívání škály v různých zemích, tendence k volbě například extrémních či středových bodů škály, souvislosti rozdílného užívání škály a geografické polohy země apod. autoři ve studii neuvádí. Ve studii uvádí pouze zmíněné hodnoty korelace mezi výsledkem v testu a podporou učitele. Dále při výpočtu používají neparametrický přístup metody ukotvujících vinět. Ten se opírá při práci s tzv. *přesmyčkami* či *nekonzistencemi* o mnoho předpokladů a výsledek pak může být výrazně těmito předpoklady ovlivněn (více viz část 3.1). Poznamenejme, že v této knize je provedena další analýza dat k podpoře ze strany učitele pomocí parametrického modelu (více viz kapitola 6). Jsou analyzovány (statisticky) signifikantní rozdíly mezi zeměmi v užívání škály, jejich tendence k volbě určitých škálových kategorií, souvislost užívání škály s geografickou polohou země a také souvislost mezi korigovanou a nekorigovanou škálou podpory ze strany učitele.

3. Statistické přístupy metody ukotvujících vinět

V této kapitole představíme dva základní statistické přístupy metody ukotvujících vinět: neparametrický přístup a parametrický model. Mimo pedagogický výzkum se užívá především parametrický model, který umí pracovat s pořadím vinět neodpovídajícím jejich přirozenému pořadí, vyjadřovat korigované sebehodnocení na původní škále nekorigovaného sebehodnocení a také dobře porovnávat efekty různých proměnných na korigované sebehodnocení po očištění o efekty jiných proměnných (tj. tak, jak to běžně činí statistické regresní modely). Neparametrický způsob je lákavý tím, že je jednoduchý a začínající výzkumníci se s ním relativně rychle naučí pracovat. Má však některá omezení, která shrneme na konci části 3.1.1 po vysvětlení základního statistického principu neparametrického přístupu.

3.1. Neparametrický (jednoduchý) přístup

3.1.1. Základy neparametrického přístupu pro začínající výzkumníky

Neparametrický přístup metody ukotvujících vinět je založen na principu relativního porovnání (sebe)hodnocení s hodnocením vinět(y). Každému respondentovi, který uvedl své (sebe)hodnocení a hodnocení vinět(y), přidělíme korigované (sebe)hodnocení. Princip získání korigovaného sebehodnocení je jednoduchý.

Představme si situaci, kdy respondent uvedl sebehodnocení a hodnocení jedné viněty. V tomto případě existují pouze tři možné pozice sebehodnocení vůči hodnocení viněty: 1) sebehodnocení je menší než hodnocení viněty, 2) sebehodnocení je stejné jako hodnocení viněty, 3) sebehodnocení je větší než hodnocení viněty. Korigované sebehodnocení může tedy v tomto případě nabývat právě tři hodnot. V případě, že se udává sebehodnocení a hodnocení viněty na pětibodové škále (např. Likertova škála s kategoriemi 1) *rozhodně nesouhlasím*, 2) *nesouhlasím*, 3) *ani souhlas, ani nesouhlas*, 4) *souhlasím* a 5) *rozhodně souhlasím*), mohou například nastat následující situace.

Hodnotí-li respondent 1 sám sebe číslem 4 (*souhlasím*) a vinětu hodnotí číslem 5 (*rozhodně souhlasím*), je korigované sebehodnocení takového respondenta rovno 1 (sebehodnocení je menší než hodnocení viněty). Hodnotí-li jiný respondent 2 sám sebe číslem 3 a vinětu též číslem 3, je jeho korigované sebehodnocení rovno 2 (sebehodnocení je rovno hodnocení viněty). Hodnotí-li další respondent 3 sám sebe číslem 2 a vinětu číslem 1, je jeho korigované sebehodnocení rovno 3 (sebehodnocení je větší než hodnocení viněty).

Všimněme si, že pokud bychom respondenty porovnali podle jejich původního, nekorigovaného sebehodnocení, bylo by jejich pořadí respondent 1, respondent 2 a respondent 3 (seřazeno od největší po nejmenší hodnotu). Pokud bychom respondenty seřadili dle jejich korigovaného sebehodnocení, je jejich pořadí opačné, tj. respondent 3, respondent 2 a respondent 1. Jde o to, že všichni tři respondenti hodnotí *tutéž* ukotvující vinětu. Respondent 1 hodnotí sebe hůře než tuto ukotvující vinětu, respondent 2 hodnotí sebe stejně jako tuto ukotvující vinětu a respondent 3 hodnotí sebe lépe než tuto ukotvující vinětu. Rozdíl mezi nekorigovaným a korigovaným sebehodnocením je dán rozdílným užíváním škálových kategorií. Respondent 1 má tendenci užívat souhlasné kategorie (lze ho označit za optimistického, či také respondenta s nižšími standardy pro hodnocení), respondent 2 užívá středové kategorie a respondent 3 zase nesouhlasné kategorie (lze ho označit za pesimistického, či také respondenta s vyššími standardy pro hodnocení).

Příklad s jednou vinětou je pro vysvětlení jednoduchý, v praxi však spíše pracujeme s více vinětami. Pro více než jednu vinětu musíme řešit, zdali respondent seřadil viněty dle jejich přirozeného pořadí. Výzkumník navrhne dvě viněty, jedna popisuje nízkou úroveň a druhá vysokou úroveň měřeného konceptu. Může se stát, že respondent při svém hodnocení nezachová přirozené pořadí vinět (tj. pořadí navržené výzkumníkem), ale pořadí vinět obrátí, tj. ohodnotí dle výzkumníka horší úroveň popsanou ve vinětě jako lepší a dle výzkumníka lepší úroveň popsanou ve vinětě jako horší. V tom případě je získání korigovaného sebehodnocení o něco komplikovanější. Ale pojďme postupně.

Případ dvou vinět s hodnocením odpovídajícím přirozenému pořadí

V případě, že je přirozené pořadí vinět respondentem zachováno, může nabývat korigované sebehodnocení celkem pěti hodnot: 1) sebehodnocení je menší než

hodnocení první viněty ($s < v_1$), 2) sebehodnocení je rovno hodnocení první viněty ($s = v_1$), 3) sebehodnocení je mezi hodnocením první a druhé viněty ($v_1 < s < v_2$), 4) sebehodnocení je rovno hodnocení druhé viněty ($s = v_2$), 5) sebehodnocení je větší než hodnocení druhé viněty ($v_2 < s$).

Uvedme si konkrétní příklad sebehodnocení a hodnocení dvou vinět. Představme si, že pracujeme se sedmibodovou škálou, například se škálou 1) *rozhodně nesouhlasím*, 2) *nesouhlasím*, 3) *spíše nesouhlasím*, 4) *ani souhlas, ani nesouhlas*, 5) *spíše souhlasím*, 6) *souhlasím*, 7) *rozhodně souhlasím*. Respondent může tedy uvést své hodnocení volbou jedné ze sedmi kategorií. Na obrázku 4 je to znázorněno šedými buňkami $s = 1$ až $s = 7$. Dále si představme, že respondent ohodnotí dvě viněty, vinětu 1 volbou kategorie 2 (šedá buňka $v_1 = 2$ na obrázku 4) a vinětu 2 volbou kategorie 5 (šedá buňka $v_2 = 5$). Přirozené pořadí vinět je zachováno, viněta 1 je hodnocena hůře než viněta 2 ($v_1 < v_2$). V tomto případě je získání korigovaného sebehodnocení snadné, určíme relativní pozici sebehodnocení vůči hodnocení vinět a získáme tím jednu hodnotu korigovaného sebehodnocení (viz obrázek 4). Pokud je sebehodnocení rovno jedné ($s = 1$), je korigované sebehodnocení rovno též jedné ($C = \{1\}$), neboť sebehodnocení je menší než hodnocení první viněty ($s < v_1$). Respondent se tedy hodnotí pod úrovní první viněty. Pokud sebehodnocení je rovno dvěma ($s = 2$), je jeho korigované sebehodnocení rovno dvěma ($C = \{2\}$), neboť sebehodnocení je rovno hodnocení první viněty ($s = v_1$). Respondent se tedy hodnotí na úrovni první viněty. Je-li sebehodnocení rovno třem ($s = 3$), je korigované sebehodnocení rovno třem ($C = \{3\}$), neboť sebehodnocení je mezi hodnocením první a druhé viněty ($v_1 < s < v_2$). Dále je-li sebehodnocení rovno čtyřem ($s = 4$), je nadále mezi hodnocením první a druhé viněty ($v_1 < s < v_2$), takže korigované sebehodnocení je stále rovno třem ($C = \{3\}$) jako v předchozím případě. Je-li sebehodnocení rovno pěti ($s = 5$), dostáváme se na úroveň druhé viněty ($s = v_2$) a korigované sebehodnocení je rovno čtyřem ($C = \{4\}$). Poslední dva případy, sebehodnocení rovno šesti nebo sedmi ($s = 6$ nebo $s = 7$), jsou překódovány na korigované sebehodnocení rovné pěti ($C = \{5\}$), neboť pro oba případy platí, že sebehodnocení je větší než hodnocení druhé viněty ($v_2 < s$).

	1	2	3	4	5	6	7
Sebehodnocení 1	$s = 1$						
Sebehodnocení 2		$s = 2$					
Sebehodnocení 3			$s = 3$				
Sebehodnocení 4				$s = 4$			
Sebehodnocení 5					$s = 5$		
Sebehodnocení 6						$s = 6$	
Sebehodnocení 7							$s = 7$

	1	2	3	4	5	6	7
Hodnocení vinět v_1 a v_2		$v_1 = 2$			$v_2 = 5$		

1) $s < v_1$							
2) $s = v_1$							
3) $v_1 < s < v_2$							
4) $s = v_2$							
5) $v_2 < s$							
Korigované sebehodnocení	$C = \{1\}$	$C = \{2\}$	$C = \{3\}$	$C = \{3\}$	$C = \{4\}$	$C = \{5\}$	$C = \{5\}$

Obrázek 4. Neparametrický přístup pro začínající výzkumníky: Postup získání korigovaného sebehodnocení v případě dvou vinět s hodnocením odpovídajícím přirozenému pořadí

Případ dvou vinět se stejným hodnocením

Respondent však nemusí zachovat přirozené pořadí vinět, hodnocení dvou vinět mu může splynout. Příkladem může být situace na obrázku 5. Respondent zde hodnotí obě viněty číslem 2 ($v_1 = v_2 = 2$). V takovémto případě si zapíšeme všechny splněné relativní pozice sebehodnocení vůči hodnocení vinět. V případě, že sebehodnocení je rovno jedné ($s = 1$), je splněna pouze jedna z pěti podmínek ($s < v_1$), další čtyři podmínky ($s = v_1$, $v_1 < s < v_2$, $s = v_2$ a $v_2 < s$) splněny nejsou. Korigované sebehodnocení je tedy rovno jedné ($C = \{1\}$). V případě, že sebehodnocení je rovno dvěma ($s = 2$), jsou splněny dvě podmínky najednou: druhá podmínka $s = v_1$ a čtvrtá podmínka $s = v_2$. Korigované sebehodnocení nenabývá jednoho čísla (skaláru), jak tomu

bylo v předchozích případech, ale nabývá intervalové hodnoty (v literatuře k metodě ukotvujících vinět se též pro tuto hodnotu užívá označení vektor), kterou zapisujeme $C = \{2,3,4\}$. V ostatních případech, kdy je sebehodnocení rovno třem, čtyřem, pěti, šesti či sedmi je splněna pouze pátá podmínka – sebehodnocení je větší než hodnocení druhé viněty ($v_2 < s$). Korigované sebehodnocení je tedy ve všech těchto ostatních případech rovno pěti ($C = \{5\}$).

	1	2	3	4	5	6	7
Sebehodnocení 1	$s = 1$						
Sebehodnocení 2		$s = 2$					
Sebehodnocení 3			$s = 3$				
Sebehodnocení 4				$s = 4$			
Sebehodnocení 5					$s = 5$		
Sebehodnocení 6						$s = 6$	
Sebehodnocení 7							$s = 7$

	1	2	3	4	5	6	7
Hodnocení vinět v_1 a v_2		$v_1 = 2$ $v_2 = 2$					

1) $s < v_1$							
2) $s = v_1$							
3) $v_1 < s < v_2$							
4) $s = v_2$							
5) $v_2 < s$							
Korigované sebehodnocení	$C = \{1\}$	$C = \{2,3,4\}$	$C = \{5\}$	$C = \{5\}$	$C = \{5\}$	$C = \{5\}$	$C = \{5\}$

Obrázek 5. Neparametrický přístup pro začínající výzkumníky: postup získání korigovaného sebehodnocení v případě dvou vinět se stejným hodnocením

Případ dvou vinět s hodnocením opačným, než je přirozené pořadí

Postup získání korigovaného sebehodnocení v případě dvou vinět s hodnocením opačným, než je přirozené pořadí, zůstává stejný jako ve všech předchozích případech – vždy si určíme, které z pěti podmínek platí. Platí-li jedna podmínka, korigované sebehodnocení nabývá jednoho čísla. Platí-li dvě podmínky, korigované sebehodnocení nabývá intervalové hodnoty. Příklad takové situace je zakreslen na obrázku 6. Viněta 1 je hodnocena číslem 5 ($v_1 = 5$) a viněta 2 číslem 2 ($v_2 = 2$), hodnocení je tedy ve sporu s přirozeným hodnocením – hodnocení viněty 2 je menší než hodnocení viněty 1 ($v_2 < v_1$). Nyní opět budeme případ po případě zjišťovat, které z pěti podmínek ($s < v_1$, $s = v_1$, $v_1 < s < v_2$, $s = v_2$ a $v_2 < s$) jsou splněny. Pokud je sebehodnocení rovno jedné ($s = 1$), je splněna pouze první podmínka ($s < v_1$). Korigované sebehodnocení je tudíž rovno jedné ($C = \{1\}$). V případě, kdy je sebehodnocení rovno dvěma ($s = 2$), jsou splněny dvě podmínky najednou – první podmínka $s < v_1$ a čtvrtá podmínka $s = v_2$. Korigované sebehodnocení nabývá tedy intervalové hodnoty $C = \{1,2,3,4\}$. V případech, kdy sebehodnocení je rovno třem nebo čtyřem ($s = 3$ nebo $s = 4$), jsou splněny dvě krajní podmínky (první a pátá podmínka) – sebehodnocení je menší než hodnocení první viněty ($s < v_1$) a zároveň sebehodnocení je větší než hodnocení druhé viněty ($v_2 < s$). Korigované sebehodnocení tedy nabývá intervalové hodnoty $C = \{1,2,3,4,5\}$. Posune-li se sebehodnocení na pět, jsou splněny dvě podmínky – druhá podmínka $s = v_1$ a pátá podmínka $v_2 < s$. Korigované sebehodnocení je tedy rovno dvěma až pěti ($C = \{2,3,4,5\}$). Poslední dva případy, kdy je sebehodnocení rovno šesti a sedmi, vedou k jedné hodnotě korigovaného sebehodnocení ($C = \{5\}$), neboť je splněna pouze pátá podmínka ($v_2 < s$).

	1	2	3	4	5	6	7
Sebehodnocení 1	$s = 1$						
Sebehodnocení 2		$s = 2$					
Sebehodnocení 3			$s = 3$				
Sebehodnocení 4				$s = 4$			
Sebehodnocení 5					$s = 5$		
Sebehodnocení 6						$s = 6$	
Sebehodnocení 7							$s = 7$

	1	2	3	4	5	6	7
Hodnocení vinět v_1 a v_2		$v_2 = 2$			$v_1 = 5$		

1) $s < v_1$							
2) $s = v_1$							
3) $v_1 < s < v_2$							
4) $s = v_2$							
5) $v_2 < s$							
Korigované sebehodnocení	$C = \{1\}$	$C = \{1,2,3,4\}$	$C = \{1,2,3,4,5\}$	$C = \{1,2,3,4,5\}$	$C = \{2,3,4,5\}$	$C = \{5\}$	$C = \{5\}$

Obrázek 6. Neparametrický přístup pro začínající výzkumníky: postup získání korigovaného sebehodnocení v případě dvou vinět s hodnocením opačným, než je přirozené pořadí

Princip získání korigovaného sebehodnocení by tímto měl být čtenáři alespoň částečně objasněn. Překódování lze provést v Excelu, existuje však i software R, který překódování učiní za nás. V kapitole 4 popisujeme, jak zpracovávat data v tomto softwaru.

V následující části naše úvahy zobecníme a rozšíříme. Nejprve však učiňme tři poznámky. Za prvé, škála původního, nekorigovaného sebehodnocení je jiná než škála korigovaného sebehodnocení. U původního sebehodnocení se řídíme škálou uvedenou v sebehodnotící otázce (např. Likertovou škálou). U korigovaného sebehodnocení je význam jednotlivých kategorií dán relativní pozicí vůči vinětám. Například hodnota korigovaného sebehodnocení tři

vyjadřuje, že sebehodnocení je mezi první a druhou vinětou. Za druhé, v případě sebehodnoticích otázek se škálami s nízkým počtem kategorií a velkým počtem vinět nemusí nikdy korigované sebehodnocení nabýt všech pozic. Například u dichotomické škály 1 a 2 a dvou vinět nenastane případ, kdy by sebehodnocení bylo mezi hodnocením dvou vinět, tj. nenastane $v_1 < s < v_2$. Běžně se používá pětibodová či sedmibodová škála a tři viněty. Za třetí, s intervalovými hodnotami korigovaného sebehodnocení se obtížně pracuje například při počítání průměrů, korelací či jiných deskriptivních statistik. Je tedy otázkou, jakou hodnotu v rámci intervalu zvolit. Kyllonen a Bertling (2014b) a Primi a kol. (2016) volí nejnížší hodnotu intervalu. Kyllonen a Bertling (2014b) argumentují, že při volbě nejnížší hodnoty dostáváme nejvyšší reliabilitu (měřenou Cronbachovým alfa). Primi a kol. (2016) na základě tohoto zjištění volí stejný způsob.

V literatuře je však již dlouho zaveden další způsob práce s intervalovými hodnotami, které popisujeme v části 4.1. Při volbě právě jednoho čísla jako reprezentanta intervalové hodnoty se však můžeme dopouštět systematické chyby. Začínajícím výzkumníkům lze tedy doporučit, aby v případě, že se uchýlí zvolit právě jednu hodnotu jako reprezentanta intervalu, učinili tzv. *sensitivity analysis* k volbě toho či onoho reprezentanta. Například mohou vyzkoušet, jak se závěry mění při volbě nejmenší, středové či největší hodnoty intervalu. Některé intervalové hodnoty mohou být způsobeny i chybou měření (*measurement error*), s čímž umí dobře pracovat parametrický model popsáný v části 3.2.

3.1.2. Korigované sebehodnocení – výčet všech relativních pozic sebehodnocení vůči hodnocení vinět

V této části zobecníme úvahy z minulé části. Doporučujeme prostudovat předchozí část pro snadnější pochopení následujících úvah. Nejprve si uvedeme výčet možností pro korigované sebehodnocení v případě dvou vinět, následně pro případ tří vinět.

V tabulce 1 je pro všechna možná řazení hodnocení dvou vinět v_1 a v_2 a sebehodnocení s zaznamenáno korigované (sebe)hodnocení C pomocí neparametrického přístupu (King & Wand, 2007). Celkově existuje třináct

případů vzájemného postavení hodnocení vinět a sebehodnocení. V každém případě tedy vyhodnotíme, zdali platí jedna z následujících pěti možností: 1) sebehodnocení menší než hodnocení první viněty ($s < v_1$), 2) sebehodnocení rovno hodnocení první viněty ($s = v_1$), 3) sebehodnocení mezi hodnocením vinět ($v_1 < s < v_2$), 4) sebehodnocení rovno hodnocení druhé viněty ($s = v_2$), 5) sebehodnocení větší než hodnocení druhé viněty ($v_2 < s$). Pokud uvedená možnost nastává, je v tabulce zaznamenána 1, v opačném případě 0 (viz sloupce 3 až 7). Výsledné korigované sebehodnocení C je pak buď *skalár*, pokud je v tabulce zaznamenána pouze jedna jednička (tj. nastává pouze jedna z pěti možností), nebo je *vektor* (bývá označován též jako *intervalová hodnota*), pokud jsou v tabulce zaznamenány dvě jedničky (tj. nastávají dvě z pěti možností). V případě vektoru zaznamenáme jako korigované hodnocení čísla všech možností ohraničené čísly možností, kde jsou zaznamenány jedničky. Například je-li $s = v_1$ (možnost 2) a zároveň $s = v_2$ (možnost 4), pak korigované sebehodnocení je rovno 2 až 4, zapisujeme jako $\{2,3,4\}$.

Tabulka 1 *Korigované (sebe)hodnocení C pomocí neparametrického přístupu v případě dvou vinět v_1 a v_2 (s přirozeným řazením $v_1 < v_2$) a sebehodnocením s*

Případ	Odpovědi respondenta	1	2	3	4	5	C
		$s < v_1$	$s = v_1$	$v_1 < s < v_2$	$s = v_2$	$v_2 < s$	
1	$s < v_1 < v_2$	1	0	0	0	0	{1}
2	$s = v_1 < v_2$	0	1	0	0	0	{2}
3	$v_1 < s < v_2$	0	0	1	0	0	{3}
4	$v_1 < s = v_2$	0	0	0	1	0	{4}
5	$v_1 < v_2 < s$	0	0	0	0	1	{5}
6	$s < v_1 = v_2$	1	0	0	0	0	{1}
7	$s = v_1 = v_2$	0	1	0	1	0	{2,3,4}
8	$v_1 = v_2 < s$	0	0	0	0	1	{5}
9	$s < v_2 < v_1$	1	0	0	0	0	{1}
10	$s = v_2 < v_1$	1	0	0	1	0	{1,2,3,4}
11	$v_2 < s < v_1$	1	0	0	0	1	{1,2,3,4,5}
12	$v_2 < s = v_1$	0	1	0	0	1	{2,3,4,5}
13	$v_2 < v_1 < s$	0	0	0	0	1	{5}

Pozn. Převzato z King a Wand (2007, s. 51).

Nyní podrobněji rozeberme jednotlivé případy v tabulce 1:

- Případy 1–5: hodnocení vinět zachovává přirozené pořadí (tedy $v_1 < v_2$) a výsledné korigované sebehodnocení C je skalár. Vzhledem k tomu, že máme dvě viněty, může C nabývat $2 \times 2 + 1 = 5$ hodnot.
- Případy 6–8: hodnocení vinět je totožné (tzv. přesmyčky, *ties*). Korigované sebehodnocení má intervalovou hodnotu (tj. jedná se o výše zmíněný *vektor*), pokud je sebehodnocení rovno hodnocení vinět.
- Případy 9–13: hodnocení vinět je v nesprávném pořadí (tzv. nekonzistence, *inconsistencies*). Korigované sebehodnocení nabývá skaláru jen v případech, kdy je sebehodnocení mimo problematickou oblast nesprávného pořadí vinět (je buď menší, nebo větší než hodnocení obou vinět). Ve zbylých případech je korigované sebehodnocení vektorem.

Poznamenejme, že případy 6, 8, 9 a 13 ústí ve skaláry (jednoznačnou hodnotu korigovaného sebehodnocení), ačkoli se při hodnocení vinět objevují přesmyčky a nekonzistence. Je však celkem přirozené, že respondenti nehodnotí dle přirozeného řazení takové viněty, které jsou hodně vzdálené jejich vlastnímu sebehodnocení – například pro dvoumetrového člověka mohou vcelku přirozeně splývat výšky lidí 140 cm a 150 cm.

Uvedli jsme výčet všech možností relativní pozice sebehodnocení vůči hodnocení dvou vinět, které mohou být respondentem hodnoceny i jinak než v přirozeném pořadí (tj. počítá se s výskytem přesmyček a nekonzistencí). Výčet všech možností relativní pozice sebehodnocení vůči třem vinětám je uveden v tabulce 2.

Tabulka 2 Korigované (sebe)hodnocení C pomocí neparametrického přístupu v případě tří vinět v_1 , v_2 a v_3 (s přirozeným řazením $v_1 < v_2 < v_3$) a sebehodnocením s

Případ	Odpovědi respondenta	C	Případ	Odpovědi respondenta	C
1	$s < v_1 < v_2 < v_3$	{1}	39	$v_3 < s = v_2 < v_1$	{1,2,3,4,5,6,7}
2	$s = v_1 < v_2 < v_3$	{2}	40	$v_3 < v_2 < s < v_1$	{1,2,3,4,5,6,7}
3	$v_1 < s < v_2 < v_3$	{3}	41	$v_3 < v_2 < s = v_1$	{2,3,4,5,6,7}
4	$v_1 < s = v_2 < v_3$	{4}	42	$v_3 < v_2 < v_1 < s$	{7}
5	$v_1 < v_2 < s < v_3$	{5}	43	$s < v_1 = v_2 < v_3$	{1}
6	$v_1 < v_2 < s = v_3$	{6}	44	$s = v_1 = v_2 < v_3$	{2,3,4}
7	$v_1 < v_2 < v_3 < s$	{7}	45	$v_1 = v_2 < s < v_3$	{5}
8	$s < v_1 < v_3 < v_2$	{1}	46	$v_1 = v_2 < s = v_3$	{6}
9	$s = v_1 < v_3 < v_2$	{2}	47	$v_1 = v_2 < v_3 < s$	{7}
10	$v_1 < s < v_3 < v_2$	{3}	48	$s < v_3 < v_1 = v_2$	{1}
11	$v_1 < s = v_3 < v_2$	{3,4,5,6}	49	$s = v_3 < v_1 = v_2$	{1,2,3,4,5,6}
12	$v_1 < v_3 < s < v_2$	{3,4,5,6,7}	50	$v_3 < s < v_1 = v_2$	{1,2,3,4,5,6,7}
13	$v_1 < v_3 < s = v_2$	{4,5,6,7}	51	$v_3 < s = v_1 = v_2$	{2,3,4,5,6,7}
14	$v_1 < v_3 < v_2 < s$	{7}	52	$v_3 < v_1 = v_2 < s$	{7}
15	$s < v_2 < v_1 < v_3$	{1}	53	$s < v_1 = v_3 < v_2$	{1}
16	$s = v_2 < v_1 < v_3$	{1,2,3,4}	54	$s = v_1 = v_3 < v_2$	{2,3,4,5,6}
17	$v_2 < s < v_1 < v_3$	{1,2,3,4,5}	55	$v_1 = v_3 < s < v_2$	{3,4,5,6,7}
18	$v_2 < s = v_1 < v_3$	{2,3,4,5}	56	$v_1 = v_3 < s = v_2$	{4,5,6,7}
19	$v_2 < v_1 < s < v_3$	{5}	57	$v_1 = v_3 < v_2 < s$	{7}
20	$v_2 < v_1 < s = v_3$	{6}	58	$s < v_2 < v_1 = v_3$	{1}
21	$v_2 < v_1 < v_3 < s$	{7}	59	$s = v_2 < v_1 = v_3$	{1,2,3,4}
22	$s < v_2 < v_3 < v_1$	{1}	60	$v_2 < s < v_1 = v_3$	{1,2,3,4,5}
23	$s = v_2 < v_3 < v_1$	{1,2,3,4}	61	$v_2 < s = v_1 = v_3$	{2,3,4,5,6}
24	$v_2 < s < v_3 < v_1$	{1,2,3,4,5}	62	$v_2 < v_1 = v_3 < s$	{7}
25	$v_2 < s = v_3 < v_1$	{1,2,3,4,5,6}	63	$s < v_2 = v_3 < v_1$	{1}
26	$v_2 < v_3 < s < v_1$	{1,2,3,4,5,6,7}	64	$s = v_2 = v_3 < v_1$	{1,2,3,4,5,6}
27	$v_2 < v_3 < s = v_1$	{2,3,4,5,6,7}	65	$v_2 = v_3 < s < v_1$	{1,2,3,4,5,6,7}
28	$v_2 < v_3 < v_1 < s$	{7}	66	$v_2 = v_3 < s = v_1$	{2,3,4,5,6,7}
29	$s < v_3 < v_1 < v_2$	{1}	67	$v_2 = v_3 < v_1 < s$	{7}
30	$s = v_3 < v_1 < v_2$	{1,2,3,4,5,6}	68	$s < v_1 < v_2 = v_3$	{1}
31	$v_3 < s < v_1 < v_2$	{1,2,3,4,5,6,7}	69	$s = v_1 < v_2 = v_3$	{2}
32	$v_3 < s = v_1 < v_2$	{2,3,4,5,6,7}	70	$v_1 < s < v_2 = v_3$	{3}
33	$v_3 < v_1 < s < v_2$	{3,4,5,6,7}	71	$v_1 < s = v_2 = v_3$	{4,5,6}
34	$v_3 < v_1 < s = v_2$	{4,5,6,7}	72	$v_1 < v_2 = v_3 < s$	{7}
35	$v_3 < v_1 < v_2 < s$	{7}	73	$s < v_1 = v_2 = v_3$	{1}
36	$s < v_3 < v_2 < v_1$	{1}	74	$s = v_1 = v_2 = v_3$	{2,3,4,5,6}
37	$s = v_3 < v_2 < v_1$	{1,2,3,4,5,6}	75	$v_1 = v_2 = v_3 < s$	{7}
38	$v_3 < s < v_2 < v_1$	{1,2,3,4,5,6,7}			

Pozn. Převzato z van Soest a Voňková (2014, s. 121).

3.1.3. Intervalové hodnoty korigovaného sebehodnocení

Nyní jsme tedy získali korigované sebehodnocení u jednotlivých respondentů. Cílem však typicky je sumarizace tohoto korigovaného sebehodnocení pro jednotlivé skupiny respondentů, např. pro respondenty ze země A a následně respondenty ze země B. Jako prostředek k této sumarizaci lze využít histogram či tabulku četností. Sestavení histogramu lze provést následovně. Pokud C nabývá pouze skalárních hodnot, je sestavení histogramu přímočaré. Postupujeme tak, jak je to běžné při sestavování histogramu, tj. spočítáme relativní frekvenci p_j hodnot v jednotlivých kategoriích j ($j = 1, 2, \dots, 2J + 1$, kde J je počet vinět) a pro každou kategorii zobrazíme sloupek s velikostí odpovídající příslušné proporci. Pokud C nabývá vektorových (a skalárních) hodnot, máme několik možností, jak se se situací vypořádat. Wand, King a Lau (2011) uvádí čtyři přístupy (ukázka zpracování dat pomocí všech čtyř přístupů je v části 4.1.4):

a) Sestavit histogram bez vektorových hodnot

Vektorové hodnoty vypustíme a zbydou nám pouze skaláry. Opodstatnění tohoto přístupu spočívá v tom, že vektorové hodnoty dostáváme v případě, kdy se nacházíme v problematické oblasti hodnocení vinět (přesmyčky a nekonzistence), což může indikovat to, že respondent například neporozuměl zadání. Avšak pokud vektorové hodnoty vypustíme, ztratíme informaci a můžeme zavést zkreslení způsobené výběrem (výběr individuí, při kterém není dodržena randomizace; což může vést k nereprezentativnímu výběru ze studované populace; může se například stát, že uvedeným vyloučením respondentů s vektorovými hodnotami dojde k vyloučení respondentů s nižším vzděláním).

b) Rozložit prostor rovnoměrně přes všechny prvky vektoru

Pokud má například vektor hodnoty $\{2,3,4,5\}$, pak každé ze čtyř hodnot přiřadíme váhu jedna čtvrtina (2 má váhu jedna čtvrtina, 3 má váhu jedna čtvrtina, 4 má váhu jedna čtvrtina i 5 má váhu jedna čtvrtina). V součtu má tedy respondent váhu 1 stejně jako ostatní respondenti se skalárním korigovaným hodnocením. Opodstatnění tohoto přístupu spočívá v tom, že pokud dostaneme vektorovou hodnotu a nevyužíváme či nemůžeme využít další informace a nechceme respondenta vypustit jako v předchozím případě, jsme nejistí

a každou hodnotu ve vektoru započítáme stejně. Pokud však předpoklad uniformního rozložení není splněn, může být histogram *příliš placatý* a mohou se tím ztrácet jeho základní vlastnosti (šikmost, špičatost, centrální tendence atd.).

Poslední dva přístupy jsou technicky náročnější, jejich plné pochopení vyžaduje hlubší znalost matematiky a statistiky.

c) Započítat každý prvek vektoru respondenta podle proporce *podobných respondentů, kteří zvolili kategorie obsažené ve vektoru*

Podobnost respondentů definujeme pomocí vnějších charakteristik jako je pohlaví, (očekávané) vzdělání či věk. Využíváme tedy další informace o respondentech, kterou máme v datech. Pokud například zvolíme jen jednu proměnnou pohlaví, daný respondent s vektorovou hodnotou {2,3,4} je muž a většina respondentů mužů má v rámci korigovaných hodnot 2, 3 a 4 hodnotu 2, pak největší váha připadá pro respondenta s vektorovou hodnotou právě na hodnotu 2.

K přesnému popisu tohoto způsobu se využívá tzv. *censored ordered probit model* (v češtině lze užívat název *enzorovaný probitový model pro ordinální proměnnou*). Samotný probitový model pro ordinální proměnnou (*ordered probit model*) pro pozorované korigované sebehodnocení $C_i, i = 1, \dots, I$ nabývající hodnotu $c = 1, \dots, 2J + 1$ lze zapsat následujícím způsobem:

$$C_i = c \Leftrightarrow \tau_{c-1} \leq C_i^* < \tau_c,$$

$$C_i^* = \mathbf{X}_i' \boldsymbol{\beta} + \epsilon_i,$$

$$\epsilon_i \sim N(0, \sigma^2),$$

kde C_i^* je latentní spojitá nepozorovaná proměnná, která je modelovaná jako součet lineární kombinace pozorovaných charakteristik respondenta (\mathbf{X}_i) a náhodné chyby ϵ_i . Chyba ϵ_i má normální rozdělení se střední hodnotou nula a rozptylem σ^2 . Pozorované korigované sebehodnocení C_i nabývá hodnotu c , pokud je latentní proměnná mezi parametry τ_{c-1} a τ_c ($-\infty = \tau_0 < \tau_1 < \dots < \tau_{2J} < \tau_{2J+1} = \infty$). Pro identifikaci parametrů se předpokládá, že koeficient β_0 u konstantního členu vektoru \mathbf{X}_i je roven nule a že rozptyl náhodné chyby σ^2 je roven jedné.

Za předpokladu platnosti tohoto modelu je pravděpodobnost pozorované hodnoty $c = 1, \dots, 2J + 1$ následující:

$$\begin{aligned} P(C_i = c) &= P(\tau_{c-1} \leq C_i^* < \tau_c) = P(\tau_{c-1} \leq \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_i < \tau_c) \\ &= P(\tau_{c-1} - \mathbf{X}'_i \boldsymbol{\beta} \leq \epsilon_i < \tau_c - \mathbf{X}'_i \boldsymbol{\beta}) = \Phi(\tau_c - \mathbf{X}'_i \boldsymbol{\beta}) - \Phi(\tau_{c-1} - \mathbf{X}'_i \boldsymbol{\beta}), \end{aligned}$$

kde $\Phi()$ představuje distribuční funkci normálního rozdělení.

V případě, že korigované sebehodnocení C_i není skalárem, ale nabývá intervalové hodnoty $C_i = \{m, \dots, n\}$, můžeme využít zobecnění výše uvedeného zápisu pro probitový model pro ordinální proměnnou

$$C_i = \{m, \dots, n\} \Leftrightarrow \tau_{m-1} \leq C_i^* < \tau_n.$$

Pravděpodobnost pozorované hodnoty $\{m, \dots, n\}$ lze pak odvodit stejným způsobem jako výše a lze ji tedy zapsat jako

$$P(C_i = \{m, \dots, n\}) = \Phi(\tau_n - \mathbf{X}'_i \boldsymbol{\beta}) - \Phi(\tau_{m-1} - \mathbf{X}'_i \boldsymbol{\beta}).$$

Podmíněnou pravděpodobnost prvku j v intervalové hodnotě $\{m, \dots, n\}$ pro respondenta i lze pak zapsat následovně:

$$p_{ij:m:n} = \frac{P(C_i=j)}{P(C_i=\{m,\dots,n\})} \text{ pro } j \in \{m, \dots, n\}.$$

V této podmíněné pravděpodobnosti normalizujeme součet jednotlivých pravděpodobností prvků uvnitř intervalové hodnoty na hodnotu jedna. Mimo intervalovou hodnotu je tedy pravděpodobnost rovna nule. V případě pozorované skalární hodnoty je její pravděpodobnost rovna jedné:

$$p_{ij:j:j} = \frac{P(C_i=j)}{P(C_i=j)} = 1.$$

Podmíněnou pravděpodobnost lze využít pro rozmístění vah jednotlivým prvkům uvnitř intervalové hodnoty pro každého respondenta na základě jeho pozorovaných charakteristik. Pro korigované sebehodnocení respondentů a jejich pozorované hodnoty odhadneme cenzorovaný probitový model pro ordinální proměnnou a na základě podmíněné pravděpodobnosti (místo parametrů $\boldsymbol{\beta}$ použijeme jejich odhady z modelu) rozmístíme váhy jednotlivým prvkům jejich (intervalové) hodnoty. Je tedy možné, že pokud má respondent

korigované sebehodnocení v intervalu $\{2,3,4\}$, mohou být na základě odhadu cenzurovaného modelu pro ordinální proměnnou váhy jednotlivých prvků různé, například $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$ (respondenti s podobnými charakteristikami mají tedy spíše korigované sebehodnocení rovno 2 než 3 či v poslední řadě 4). Takto vypočítané váhy pro jednotlivé prvky $j = 1, \dots, 2J + 1$ sečteme a sestavíme finální histogram.

d) Sestavit histogram dle principu minimální a maximální entropie

Při sestavování histogramu s intervalovými hodnotami lze též zvážit dva hraniční případy – histogram, který je ze všech histogramů pro daná data nejméně uniformní („nejvíce zubatý“), a histogram, který je naopak nejvíce uniformní („nejvíce placatý“). První případ, kdy je histogram nejméně uniformní, vede k nejmenšímu rozlišení mezi respondenty. V extrémním případě jsou všichni respondenti v jedné kategorii, jinými slovy jsou jejich korigovaná sebehodnocení totožná. Druhý případ, kdy je histogram nejvíce uniformní, vede k co největšímu rozlišení mezi respondenty.

Z formálního pohledu lze k těmto případům dojít pomocí minimalizace entropie³ (první případ s nejméně uniformním histogramem) či maximalizace entropie (druhý případ s nejvíce uniformním histogramem). Využívá se především postup založený na minimalizaci entropie. Jedná se v podstatě o nejhorší scénář, mezi respondenty rozliší minimálně. Méně informace o rozdílech mezi respondenty na základě daných dat nelze získat. Intervalové hodnoty jsou při sestavování takového histogramu reprezentovány takovou jednou hodnotou („skalárem“) z daného intervalu, aby co nejvíce z těchto „skalárů“ padlo do co nejmenšího možného počtu kategorií korigovaného sebehodnocení.

V tomto případě je $p_{j:m:m} = 1$ pro nějaké j v každém intervalu $\{m, \dots, n\}$, přičemž volba tohoto j je získána pomocí minimalizace (pro nejméně uniformní rozdělení) či maximalizace (pro nejvíce uniformní rozdělení) funkce entropie

$$H = - \sum_{j=1}^{2J+1} p_j \ln(p_j),$$

³ Entropii lze charakterizovat jako míru neurčitosti systému. Tento pojem se užívá ve fyzice (termodynamice) a matematické informatice.

kde p_j je označení pravděpodobnosti v j -té kategorii histogramu. Pokud nějaká $p_j = 0$, pak definujeme $-0 \ln(0) \equiv 0$ (to odpovídá limitě $\lim_{a \rightarrow 0^+} a \ln a = 0$). Míra entropie H je rovna nula právě tehdy, když existuje j , pro nějž $p_j = 1$. Pro libovolné J (počet vinět) je maximální hodnota H nabyta pro $p_j = \frac{1}{2J+1}$, $j = 1, \dots, 2J+1$, tj. jedná se o rovnoměrné rozdělení. Maximální hodnota H je tedy rovna $\ln(2J+1)$ (lze odvodit $H = -\sum_{j=1}^{2J+1} \frac{1}{2J+1} \ln\left(\frac{1}{2J+1}\right) = -\frac{2J+1}{2J+1} (\ln 1 - \ln(2J+1)) = \ln(2J+1)$).

U všech čtyř přístupů jsou myšlenky pro konstruování histogramů odlišné. Využívají se často dodatečné předpoklady. V prvním přístupu se předpokládá, že vypuštění vektorových hodnot nevede k zavedení systematické chyby. V druhém přístupu se předpokládá, že rovnoměrné rozložení hodnot přes interval nemění základní vlastnosti rozložení korigovaných odpovědí, jako je jeho špičatost. V třetím přístupu se předpokládá, že podobní respondenti definovaní pomocí dalších vnějších pozorovaných charakteristik umožní správné rozložení v rámci jednotlivých kategorií daného vektoru. Navíc se zde přidávají i statistické předpoklady parametrického cenzurovaného modelu pro ordinální proměnnou. Ve čtvrtém přístupu se snažíme sice sestavit histogram pouze pomocí informace, kterou poskytuje sebehodnocení a hodnocení vinět, avšak cílem je vytvořit hraniční případy buď co nejvíce uniformního histogramu, či naopak co nejméně uniformního histogramu. Velké rozdíly mezi histogramy tvořenými uvedenými přístupy signalizují velkou závislost na dodatečných předpokladech a primárně i velkém množství intervalových hodnot. V ideálním případě, kdy se neobjevuje ani systematická chyba ani chyba měření, dostáváme jen skalární hodnoty pro jednotlivé respondenty, a tudíž nepotřebujeme dodatečné předpoklady pro nakládání s intervalovými hodnotami (Voňková & Hrabák, 2015b). V reálných případech se však intervalové hodnoty vyskytují a můžeme doporučit zkonstruovat histogramy pomocí všech uvedených přístupů a vzájemně je porovnat. Příkladem může být porovnávání sebehodnocení ICT znalostí a dovedností u dvou skupin ICT a neICT studentů ve Voňková a Hrabák, 2015b. Mezi všemi uvedenými přístupy byly minimální rozdíly a všechny histogramy tvořené pomocí uvedených přístupů podpořily závěr, že korigované sebehodnocení ICT znalostí a dovedností je u ICT studentů vyšší než u neICT studentů.

3.2. Parametrický model

3.2.1. Specifikace modelu

Jako parametrický model metody ukotvujících vinět se užívá CHOPIT model (*compound hierarchical ordered probit model*) a jeho rozšíření.⁴

Nejprve uvedeme základní specifikaci CHOPIT modelu, který je běžně užíván v aplikacích metody ukotvujících vinět a je též implementován ve statistických softwarech R a STATA. Po představení základní specifikace shrneme i možnosti rozšíření modelu. Plné pochopení parametrického modelu je ve srovnání s neparametrickým řádově náročnější na matematické a statistické znalosti.

Ve většině aplikací metody ukotvujících vinět slouží jako vstupní data jedno sebehodnocení, jedna viněta či tři ukotvující viněty a respondentovy charakteristiky (typicky pohlaví, věk, vzdělání, země apod.). CHOPIT model představíme pro tuto strukturu dat.

CHOPIT model se skládá ze dvou částí. V první části modelujeme sebehodnocení a v druhé části hodnocení vinět (King et al., 2004; van Soest & Voňková, 2014).

Rovnice pro sebehodnocení jsou následující:

$$Y_{si}^* = \mu_i + \epsilon_{si}, \quad i = 1, \dots, I,$$

$$\mu_i = \mathbf{X}_i' \boldsymbol{\beta},$$

⁴ Voňková (2012b) shrnuje na základě Tandon a kol. (2003), ze kterých statistických modelů se vycházelo při budování parametrického modelu metody ukotvujících vinět. Jedná se probitový model pro ordinální proměnnou, který je běžně používán v ekonometrii, a model pro stupňovaný kredit (*partial credit model*), který je běžně používán v psychometrii pro vyhodnocování didaktických testů. Oba modely mají společné rysy, například zachycují odpovědi na ordinální škály. Liší se však v mechanismu generování dat, probitový model pro ordinální proměnnou modeluje pravděpodobnosti volby škálových kategorií s využitím kumulativní distribuční funkce, model pro stupňovaný kredit porovnává dvě sousední kategorie. Ani jeden model však neumí odlišit interpretaci škály respondentem a skutečnou hodnotou pro respondenta, což je již v parametrickém modelu metody ukotvujících vinět řešeno.

$$\epsilon_{si} \sim N(0, \sigma_s^2), \quad \epsilon_{si} \perp \mathbf{X}_i,$$

$$Y_{si} = j \Leftrightarrow \tau_i^{j-1} < Y_{si}^* \leq \tau_i^j, \quad j = 1, \dots, J,$$

$$\tau_i^1 = \mathbf{X}_i' \boldsymbol{\gamma}^1,$$

$$\tau_i^j = \tau_i^{j-1} + \exp(\mathbf{X}_i' \boldsymbol{\gamma}^j), \quad j = 2, \dots, J-1,$$

$$\tau_i^0 = -\infty, \quad \tau_i^J = \infty.$$

μ_i zde představuje proměnnou, která reprezentuje *skutečnou* úroveň měřeného konceptu pro respondenta i . Tuto úroveň vnímá respondent s náhodnou chybou ϵ_{si} . Rozdělení ϵ_{si} je normální s nulovou střední hodnotou a rozptylem σ_s^2 . Součtem proměnné μ_i a chyby ϵ_{si} je reprezentována respondentova vnímaná, nepozorovaná (latentní) úroveň měřeného konceptu Y_{si}^* .

μ_i je modelováno jako lineární kombinace pozorovaných respondentových charakteristik \mathbf{X}_i , přičemž první člen vektoru \mathbf{X}_i je konstanta, která je rovna jedné. Parametry této lineární kombinace $\boldsymbol{\beta}$ jsou jedny ze zásadních parametrů. Ukazují, v jakém vztahu jsou pozorované respondentovy charakteristiky ke *skutečné* úrovni měřeného konceptu. Například v jakém vztahu je věk žáka (\mathbf{X}_i) k jeho *skutečným* kázeňským problémům (μ_i), či v jakém vztahu je pohlaví žáka (\mathbf{X}_i) k jeho *skutečným* nekognitivním schopnostem, jako je vůle a sebekontrola (μ_i). Poznamenejme, že předpokládáme nezávislost náhodné chyby ϵ_{si} s pozorovanými charakteristikami respondenta \mathbf{X}_i .

Y_{si}^* není pozorováno přímo. K dispozici máme Y_{si} , odpověď respondenta i na sebehodnoticí otázku s ordinální škálou o J kategoriích. Škála odpovědí je často pětibodová ($J = 5$), příkladem může být Likertova škála 1) *souhlasím*, 2) *spíše souhlasím*, 3) *ani souhlas, ani nesouhlas*, 4) *spíše nesouhlasím*, 5) *nesouhlasím*. Odpověď respondenta Y_{si} je závislá jednak na jím vnímané úrovni měřeného konceptu Y_{si}^* (součet skutečné úrovně μ_i a chyby ϵ_{si}) a jednak na jeho užívání škály. Odlišnost respondentů $i = 1, \dots, I$ v užívání J -bodové škály je zachycena parametrem τ_i^j , $i = 1, \dots, I$ a $j = 1, \dots, J$. Konkrétní pozorovací mechanismus je takový, že respondent i odpoví na sebehodnoticí otázku volbou kategorie j ($Y_{si} = j$) právě tehdy, když je vnímaná úroveň měřeného konceptu mezi parametry τ_i^{j-1} a τ_i^j ($\tau_i^{j-1} < Y_{si}^* \leq \tau_i^j$). Parametry τ_i^j jsou meze oddělující

jednotlivé škálové kategorie a mohou být pro každého respondenta jiné. Pokud by nebylo umožněno, aby byly mezní parametry pro jednotlivé respondenty odlišné, nemohli bychom zachytit případnou heterogenitu ve stylu užívání škály. Mezní parametry jsou modelovány jako kombinace pozorovaných respondentových charakteristik \mathbf{X}_i . První mez τ_i^1 je lineární kombinací \mathbf{X}_i s koeficienty $\boldsymbol{\gamma}^1$ ($\tau_i^1 = \mathbf{X}_i' \boldsymbol{\gamma}^1$). Další meze τ_i^j , $j = 2, \dots, J - 1$ jsou vždy součtem předchozí meze τ_i^{j-1} a kombinace pozorovaných charakteristik \mathbf{X}_i s koeficienty $\boldsymbol{\gamma}^j$ ($\tau_i^j = \tau_i^{j-1} + \exp(\mathbf{X}_i' \boldsymbol{\gamma}^j)$). Užití exponenciely umožňuje zachovat neklesající posloupnost mezních parametrů ($\tau_i^{j-1} \leq \tau_i^j$). Krajní parametry přirozeně definujeme jako $\tau_i^0 = -\infty$ a $\tau_i^J = \infty$.

V uvedené části modelu pro sebehodnocení není možné identifikovat klíčové parametry $\boldsymbol{\beta}$ popisující vztah mezi respondentovými charakteristikami a *skutečnou* úrovní měřeného konceptu a parametry $\boldsymbol{\gamma}^1$ popisující užívání první škálové kategorie v závislosti na respondentových charakteristikách. Pouze rozdíl těchto parametrů $\boldsymbol{\beta} - \boldsymbol{\gamma}^1$ je identifikován. Pro identifikaci těchto parametrů použijeme hodnocení vinět.

Rovnice pro viněty jsou následující:

$$Y_{vi}^* = \theta_v + \epsilon_{vi}, \quad i = 1, \dots, I, \quad v = 1, \dots, V,$$

$$\text{i.i.d. } \epsilon_{vi} \sim N(0, \sigma_v^2), \quad \epsilon_{vi} \perp \epsilon_{si}, \mathbf{X}_i,$$

$$Y_{vi} = j \Leftrightarrow \tau_i^{j-1} < Y_{vi}^* \leq \tau_i^j, \quad j = 1, \dots, J.$$

θ_v zde představuje parametr, který reprezentuje *skutečnou* úroveň měřeného konceptu pro hypotetickou osobu popsanou ve vinětě v . Tuto úroveň respondent vnímá s náhodnou chybou ϵ_{vi} . ϵ_{vi} jsou nezávislé stejně rozdělené náhodné veličiny (značíme jako i.i.d.), jejich rozdělení je normální s nulovou střední hodnotou a rozptylem σ_v^2 . Předpokládáme nezávislost ϵ_{vi} na ϵ_{si} a také na \mathbf{X}_i . Součtem parametru θ_v a chyby ϵ_{vi} je reprezentována respondentem i vnímaná latentní úroveň měřeného konceptu u hypotetické osoby popsané ve vinětě v . Tento součet je označen jako Y_{vi}^* . Zde je důležité poznamenat, že θ_v se nemění s charakteristikami respondenta (tj. neoznačujeme θ_{vi}). Respondentům je předložena tatáž viněta, hodnotí tedy tutéž úroveň měřeného konceptu. Předpoklad, že všichni respondenti interpretují úroveň hypotetické osoby ve

viněť stejně, se nazývá ekvivalence viněť (více o tomto předpokladu lze najít v části 5.1.2).

Y_{vi}^* není pozorováno přímo. K dispozici máme Y_{vi} , hodnocení hypotetické osoby popsané ve viněť v respondentem i . Hodnocení viněť se provádí stejně jako u sebehodnocení na ordinální škále o J kategoriích. Toto hodnocení je závislé nejen na respondentem i vnímané úrovni měřeného konceptu popsaného ve viněť v (Y_{vi}^*), ale také na užívání škály. Odlišnost užívání škály je modelována stejně jako pro sebehodnocení, tj. je zachycena parametry τ_i^j . Předpoklad, že užívání škály je u sebehodnocení a hodnocení viněť stejné, se nazývá konzistence odpovědí (více o tomto předpokladu lze najít v části 5.1.1). Konkrétní pozorovací mechanismus je stejný jako u sebehodnocení. Respondent i hodnotí hypotetickou osobu ve viněť v volbou kategorie j ($Y_{vi} = j$) právě tehdy, když je vnímaná úroveň měřeného konceptu mezi parametry τ_i^{j-1} a τ_i^j ($\tau_i^{j-1} < Y_{vi}^* \leq \tau_i^j$).

3.2.2. Odhad parametrů modelu metodou maximální věrohodnosti

Odhad parametrů CHOPIT modelu se provádí pomocí metody maximální věrohodnosti (ke statistické metodě maximální věrohodnosti existuje řada publikací jako například Millar, 2011). Vzhledem k předpokladu nezávislosti odpovědí pro jednotlivé respondenty (ϵ_{si} jsou nezávislé stejně rozdělené náhodné veličiny, ϵ_{vi} jsou též nezávislé stejně rozdělené náhodné veličiny s rozptylem σ_v) můžeme věrohodnostní funkci vyjádřit jako součin individuálních přírůstků:

$$L(Y_{si} = y_{si}, Y_{vi} = y_{vi}, i = 1, \dots, I, v = 1, \dots, V, \Omega) = \prod_{i=1}^I P(Y_{si} = y_{si}, Y_{vi} = y_{vi}, v = 1, \dots, V | \Omega),$$

kde y_{si} a y_{vi} jsou pozorované odpovědi jednotlivých respondentů, Y_{si} a Y_{vi} jsou náhodné veličiny uvedené v popisu modelu výše a Ω je množina všech odhadovaných parametrů:

$$\Omega = \{\beta, \gamma^j, \theta_v, \sigma_s, \sigma_v, j = 1, \dots, J, v = 1, \dots, V\}.$$

Vzhledem k předpokladu nezávislosti náhodných chyb v sebehodnocení a hodnocení vinět ϵ_{si} a ϵ_{vi} , $v = 1, \dots, V$ můžeme individuální přírůstek zapsat jako součin pravděpodobností:

$$P(Y_{si} = y_{si}, Y_{vi} = y_{vi}, v = 1, \dots, V | \Omega) \\ = P(Y_{si} = y_{si} | \Omega) \prod_{v=1}^V P(Y_{vi} = y_{vi} | \Omega) .$$

Nyní je tedy nutné vyjádřit pravděpodobnosti volby škálové kategorie j u sebehodnocení a hodnocení vinět. Na základě toho pak sestavíme věrohodnostní funkci. Vyjádření provedeme podrobně, aby z něj bylo mimo jiné patrné, jak postupovat při identifikaci parametrů modelu (v modelu nelze identifikovat všechny parametry, některé musíme zafixovat). Dále by měla být zjevná podobnost mezi vyjádřením pravděpodobností pro sebehodnocení a hodnocení vinět. Pro zjednodušení vynecháme podmínění množinou hledaných parametrů Ω .

Nejprve vyjádříme pravděpodobnosti, že sebehodnocení je rovno škálovým kategoriím $j = 1, \dots, J$:

$$P(Y_{si} = 1) = P(\tau_i^0 < Y_{si}^* \leq \tau_i^1) = P(\tau_i^0 < \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_{si} \leq \tau_i^1) \\ = P(-\infty < \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_{si} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1) = P(-\infty < \epsilon_{si} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1 - \mathbf{X}'_i \boldsymbol{\beta}) \\ = P(-\infty < \epsilon_{si} \leq \mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta})) = P\left(-\infty < \frac{\epsilon_{si}}{\sigma_s} \leq \frac{\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta})}{\sigma_s}\right) \\ = \Phi\left(\frac{\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta})}{\sigma_s}\right)$$

$$P(Y_{si} = 2) = P(\tau_i^1 < Y_{si}^* \leq \tau_i^2) = P(\tau_i^1 < \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_{si} \leq \tau_i^2) \\ = P(\mathbf{X}'_i \boldsymbol{\gamma}^1 < \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_{si} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1 + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)) \\ = P(\mathbf{X}'_i \boldsymbol{\gamma}^1 - \mathbf{X}'_i \boldsymbol{\beta} < \epsilon_{si} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1 - \mathbf{X}'_i \boldsymbol{\beta} + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)) \\ = P(\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) < \epsilon_{si} \leq \mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)) \\ = P\left(\frac{\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta})}{\sigma_s} < \frac{\epsilon_{si}}{\sigma_s} \leq \frac{\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)}{\sigma_s}\right) \\ = \Phi\left(\frac{\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)}{\sigma_s}\right) - \Phi\left(\frac{\mathbf{X}'_i (\boldsymbol{\gamma}^1 - \boldsymbol{\beta})}{\sigma_s}\right)$$

$$P(Y_{si} = j) = \Phi\left(\frac{\mathbf{X}'_i(\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \sum_{k=2}^j \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_s}\right) - \Phi\left(\frac{\mathbf{X}'_i(\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \sum_{k=2}^{j-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_s}\right)$$

$$j = 3, \dots, J - 1$$

$$\begin{aligned} P(Y_{si} = J) &= P(\tau_i^{J-1} < Y_{si}^* \leq \tau_i^J) = P(\tau_i^{J-1} < \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_{si} \leq \tau_i^J) \\ &= P\left(\mathbf{X}'_i \boldsymbol{\gamma}^1 + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k) < \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_{si} \leq \infty\right) \\ &= P\left(\mathbf{X}'_i \boldsymbol{\gamma}^1 - \mathbf{X}'_i \boldsymbol{\beta} + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k) < \epsilon_{si} \leq \infty\right) \\ &= P\left(\mathbf{X}'_i(\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k) < \epsilon_{si} \leq \infty\right) \\ &= P\left(\frac{\mathbf{X}'_i(\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_s} < \frac{\epsilon_{si}}{\sigma_s} \leq \infty\right) \\ &= 1 - \Phi\left(\frac{\mathbf{X}'_i(\boldsymbol{\gamma}^1 - \boldsymbol{\beta}) + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_s}\right) \end{aligned}$$

Poznamenejme, že $\Phi()$ označuje distribuční funkci standardního normálního rozdělení. Pro distribuční funkci obecně platí, že $\Phi(-\infty) = 0$ a $\Phi(\infty) = 1$, což je využito při odvozování první a poslední pravděpodobnosti (tj., že sebehodnocení je rovno první a poslední škálové kategorii).

Podobně jako pro sebehodnocení lze vyjádřit pravděpodobnosti pro hodnocení vinět. Při odvozování lze oproti sebehodnocení v podstatě nahradit vyjádření *skutečné* úrovně měřeného konceptu pro respondenta i ($\mathbf{X}'_i \boldsymbol{\beta}$) *skutečnou* úrovní měřeného konceptu pro hypotetickou osobu popsanou ve viněť v (θ_v) a rozptyl náhodné chyby v sebehodnocení (σ_s) rozptylem náhodné chyby v hodnocení vinět (σ_v):

$$\begin{aligned} P(Y_{vi} = 1) &= P(\tau_i^0 < Y_{vi}^* \leq \tau_i^1) = P(\tau_i^0 < \theta_v + \epsilon_{vi} \leq \tau_i^1) \\ &= P(-\infty < \theta_v + \epsilon_{vi} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1) = P(-\infty < \epsilon_{vi} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v) \\ &= P\left(-\infty < \frac{\epsilon_{vi}}{\sigma_v} \leq \frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v}{\sigma_v}\right) = \Phi\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v}{\sigma_v}\right) \end{aligned}$$

$$\begin{aligned}
P(Y_{vi} = 2) &= P(\tau_i^1 < Y_{vi}^* \leq \tau_i^2) = P(\tau_i^1 < \theta_v + \epsilon_{vi} \leq \tau_i^2) \\
&= P(\mathbf{X}'_i \boldsymbol{\gamma}^1 < \theta_v + \epsilon_{vi} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1 + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)) \\
&= P(\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v < \epsilon_{vi} \leq \mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)) \\
&= P\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v}{\sigma_v} < \frac{\epsilon_{vi}}{\sigma_v} \leq \frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)}{\sigma_v}\right) \\
&= \Phi\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)}{\sigma_v}\right) - \Phi\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v}{\sigma_v}\right)
\end{aligned}$$

$$\begin{aligned}
P(Y_{vi} = j) &= \Phi\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \sum_{k=2}^j \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_v}\right) - \Phi\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \sum_{k=2}^{j-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_v}\right) \\
& \qquad \qquad \qquad j = 3, \dots, J-1
\end{aligned}$$

$$\begin{aligned}
P(Y_{vi} = J) &= P(\tau_i^{J-1} < Y_{vi}^* \leq \tau_i^J) = P(\tau_i^{J-1} < \theta_v + \epsilon_{vi} \leq \tau_i^J) \\
&= P\left(\mathbf{X}'_i \boldsymbol{\gamma}^1 + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k) < \theta_v + \epsilon_{vi} \leq \infty\right) \\
&= P\left(\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k) < \epsilon_{vi} \leq \infty\right) \\
&= P\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_v} < \frac{\epsilon_{vi}}{\sigma_v} \leq \infty\right) \\
&= 1 - \Phi\left(\frac{\mathbf{X}'_i \boldsymbol{\gamma}^1 - \theta_v + \sum_{k=2}^{J-1} \exp(\mathbf{X}'_i \boldsymbol{\gamma}^k)}{\sigma_v}\right)
\end{aligned}$$

Všechny tyto pravděpodobnosti lze krátce vyjádřit pomocí označení pro meze τ_i^j a pro skutečnou úroveň měřeného konceptu μ_i pro respondenta i a θ_v pro hypotetickou osobu ve vině v :

$$P(Y_{si} = j) = \Phi\left(\frac{\tau_i^j - \mu_i}{\sigma_s}\right) - \Phi\left(\frac{\tau_i^{j-1} - \mu_i}{\sigma_s}\right), \quad j = 1, \dots, J,$$

$$P(Y_{vi} = j) = \Phi\left(\frac{\tau_i^j - \theta_v}{\sigma_v}\right) - \Phi\left(\frac{\tau_i^{j-1} - \theta_v}{\sigma_v}\right), \quad v = 1, \dots, V, \quad j = 1, \dots, J.$$

Při odhadu parametrů maximalizujeme logaritmus věrohodnostní funkce (Ω označuje všechny odhadované parametry modelu):

$$\begin{aligned}
& \max_{\Omega} \log L(Y_{si} = y_{si}, Y_{vi} = y_{vi}, i = 1, \dots, I, v = 1, \dots, V, \Omega) \\
&= \max_{\Omega} \log \prod_{i=1}^I (P(Y_{si} = y_{si}, \Omega)) \prod_{v=1}^V P(Y_{vi} = y_{vi}, \Omega) \\
&= \max_{\Omega} \sum_{i=1}^I \log(P(Y_{si} = y_{si}, \Omega)) + \sum_{i=1}^I \sum_{v=1}^V \log(P(Y_{vi} = y_{vi}, \Omega)) .
\end{aligned}$$

Parametry věrohodnostní funkce si lze tímto způsobem odhadnout v libovolném statistickém softwaru. Parametrický model je implementován v softwaru R a STATA, takže pro běžného výzkumníka není nutné odhad modelu programovat. V kapitole 4 bude popsáno, jakým způsobem lze využít již implementovaný model v uvedených softwarech pro odhad parametrů.

3.2.3. Identifikace parametrů modelu

Všechny parametry popsaného CHOPIT modelu nelze zároveň identifikovat. Je nutné zvolit měřítko pro odhadované parametry:

- 1) Volba lokace parametrů: Umístíme jeden z odhadovaných parametrů, tj. zvolíme jeho lokaci. Toto můžeme učinit pomocí fixace prvního členu ve vektoru β , tj. parametru u konstantního členu X_i . Typicky je roven nule. Toto zapisujeme jako $\beta_0 = 0$. Další možností je fixace jednoho z parametrů označujících *skutečnou* úroveň měřeného konceptu ve vinětách, tj. jednoho z parametrů θ_v . Nejčastěji volíme fixování úrovně v první viněti a pokládáme ji rovnou nule, tj. $\theta_1 = 0$.
- 2) Volba rozsahu škály parametrů: Umístíme další parametr tak, aby byl daný rozsah škály pro odhadované parametry. Jednou z možností je zafixovat rozptyl náhodné chyby u sebehodnocení a položit ji rovnou nule, tj. $\sigma_s^2 = 1$. Další možností je fixace dalšího parametru označujícího *skutečnou* úroveň měřeného konceptu ve vinětách. Typicky volíme fixování úrovně v poslední viněti a pokládáme ji rovnou jedné, tj. $\theta_v = 1$.

Jak je uvedeno ve Wand a kol. (2007), běžně používané jsou dvě parametrizace:

- 1) Nastavení $\beta_0 = 0$ a $\sigma_s^2 = 1$: Toto nastavení je v souladu s běžně užívanou parametrizací z probitového modelu pro ordinální proměnnou

a je tudíž užitečné pro porovnání parametrů β v jednodušším probitovém modelu pro ordinální proměnnou a CHOPIT modelu.

- 2) Nastavení $\theta_1 = 0$ a $\theta_V = 1$: Škála pro odhadované parametry je definována pomocí popsaných úrovní měřeného konceptu ve vinětách. Hodnota 0 odpovídá úrovni v první vinětě a hodnota 1 úrovni v poslední vinětě. Toto umožňuje dobře interpretovat hodnotu μ_i , tj. skutečnou úroveň měřeného konceptu pro respondenta i . Při použití tohoto nastavení však musíme mít alespoň dvě viněty a zároveň předpokládat pořadí vinět. Pokud si výzkumník není jistý pořadím vinět, může nejprve odhadnout model pomocí prvního nastavení $\beta_0 = 0$ a $\sigma_s^2 = 1$ a pořadí vinět určit. Pokud je model odhadnut s *nesprávným* pořadím vinět, odhadované parametry mohou mít například velké standardní chyby.

Zajímavou otázkou je, jak funguje identifikace zbylých parametrů modelu. Zprv je nutné si uvědomit, že při využití pouze sebehodnocení (a nikoli hodnocení vinět) můžeme identifikovat pouze rozdíl parametrů $\gamma^1 - \beta$ (při předpokladu, že zafixujeme jednotku škály σ_s), nelze identifikovat jednotlivé parametry γ^1 a β . Toto dává smysl, neboť pouze pomocí sebehodnocení nelze oddělit dvě věci – užívání škály (zachyceno pomocí parametrů první meze γ^1) a skutečnou úroveň měřeného konceptu pro respondenta (zachyceno pomocí parametrů β). Parametry dalších mezí, tj. $\gamma^2, \gamma^3, \dots, \gamma^J$, jsou identifikovány pouze pomocí sebehodnocení. Toto by mělo být patrné z rozpisů pravděpodobností pro sebehodnocení. V případě první pravděpodobnosti $P(Y_{si} = 1)$ určujeme hodnotu distribuční funkce standardního normálního rozdělení v bodě $\frac{X'_i(\gamma^1 - \beta)}{\sigma_s}$. U druhé pravděpodobnosti $P(Y_{si} = 2)$ musíme určit hodnotu distribuční funkce standardního normálního rozdělení v tomto bodě posunutém o $\frac{\exp(X'_i\gamma^2)}{\sigma_s}$, tj. v bodě $\frac{X'_i(\gamma^1 - \beta) + \exp(X'_i\gamma^2)}{\sigma_s}$. Analogicky je tomu pro třetí až J -tou pravděpodobnost. Posuny u těchto pravděpodobností lze tedy též identifikovat, parametry $\gamma^3, \dots, \gamma^J$ určují pouze posun a nevystupují v rozdílu či jiné kombinaci s dalšími parametry. Klíčová je však identifikace jednotlivých parametrů β a γ^1 . Tu nakonec získáme díky hodnocení vinět. Viněty nám pomohou identifikovat parametr γ^1 , což je pak využito pro identifikaci β v rozdílu $\gamma^1 - \beta$. Myšlenku identifikace pomocí vinět lze popsat následujícím způsobem. Uvažujme nastavení $\beta_0 = 0$ a $\sigma_s^2 = 1$. Z tohoto nastavení vyplývá,

že první člen γ_0^1 je identifikován díky sebehodnocení, stejně jako celý vektor γ^2 . Na základě γ^2 pak lze díky druhé pravděpodobnosti pro viněty $P(Y_{vi} = 2)$ identifikovat σ_v . Dále pak na základě tohoto σ_v a první pravděpodobnosti $P(Y_{vi} = 1)$ identifikujeme zbylé členy vektoru γ^1 . (Poznamenejme, že θ_v je identifikováno díky již zmíněným γ_0^1 a σ_v .) Nyní již tedy máme identifikovaný celý vektor γ^1 a díky první pravděpodobnosti pro sebehodnocení $P(Y_{si} = 1)$ můžeme identifikovat i klíčový parametr β . Toto je jeden z možných *hrubých* popisů identifikace parametrů. Pro čtenáře doporučujeme, aby si pečlivě rozepsal věrohodnostní funkci a individuální přírůstky a nad identifikací se samostatně zamyslel.

3.2.4. Rozšíření standardního modelu

1) Přidání individuálního efektu u_i do první meze τ_i^1

$$\tau_i^1 = \mathbf{X}'_i \boldsymbol{\gamma}^1 + u_i$$

Přidání individuálního efektu u_i do první meze τ_i^1 je zavedeno jako rozšíření standardního CHOPIT modelu v článku Kapteyn a kol. (2007). Velká hodnota u_i pro respondenta znamená, že určitou úroveň měřeného konceptu hodnotí nižšími škálovými kategoriemi ve srovnání s ostatními respondenty. Pokud například užíváme Likertovu škálu 1) *souhlasím*, 2) *spíše souhlasím*, 3) *ani souhlas*, 4) *spíše nesouhlasím*, 5) *nesouhlasím*, tak respondent s vysokou hodnotou náhodného efektu u_i hodnotí určitou *skutečnou* úroveň měřeného konceptu volbou kategorie *souhlasím*, zatímco jiní respondenti volí kategorii *spíše souhlasím* či dokonce vyšší kategorie. Nebo takovýto respondent může pro jinou úroveň měřeného konceptu volit středovou kategorii *ani souhlas*, *ani nesouhlas*, zatímco ostatní respondenti již vyjadřují určitou míru nesouhlasu. Obecně o respondentovi, který volí spíše nižší kategorie popisující větší míru souhlasu (tj. takového, jehož pozorované charakteristiky a/nebo individuální efekt přispívají k vysokým hodnotám parametrů mezi τ_i^j , $j = 1, \dots, J - 1$), říkáme, že je buď optimistický nebo že má nižší standardy pro hodnocení. Van Soest a Voňková (2014) zavedli test specifikace CHOPIT modelu opřený o porovnání relativní pozice sebehodnocení mezi vinětami jakožto výstupu neparametrického přístupu s relativní pozicí sebehodnocení mezi vinětami

generovanou na základě odhadnutého parametrického modelu CHOPIT. Klíčovým výsledkem bylo, že specifikace standardního CHOPIT modelu je vždy zamítnuta (analyzováno bylo šest zdravotních domén, jako jsou problémy s chůzí či dýcháním), zatímco rozšíření o individuální efekt přináší mnohem lepší výsledky. Celkově autoři doporučují více flexibilní modely (parametrické či semiparametrické) než je standardní CHOPIT model.

2) Modelování více sebehodnocení

Většina aplikací metody ukotvujících vinět stojí na využití jednoho (sebe)hodnocení, tj. hodnocení jedné (sebe)hodnotící otázky. V pedagogice a psychologii se však často měří daný koncept pomocí více než jedné (sebe)hodnotící otázky. V rámci značení standardního modelu zavedeme i označení pro více sebehodnotících otázek:

$$Y_{si}^* = \mu_i + \epsilon_{si}, \quad i = 1, \dots, I, \quad s = 1, \dots, S.$$

V takovýchto situacích lze použít CHOPIT model s přidáním náhodným efektem η_i pro respondenta i v rovnici pro proměnnou μ_i reprezentující *skutečnou* úroveň měřeného konceptu

$$\mu_i = \mathbf{X}_i' \boldsymbol{\beta} + \eta_i,$$

přičemž předpokládáme, že η_i je nezávislé na pozorovaných charakteristikách respondenta (\mathbf{X}_i) a má normální rozdělení s nulovou střední hodnotou a rozptylem ω^2 :

$$\eta_i \sim N(0, \omega^2).$$

Pro *skutečnou* úroveň měřeného konceptu μ_i zachytí náhodný efekt η_i nepozorovanou heterogenitu mezi respondenty, rozdíly mezi respondenty v pozorovaných charakteristikách \mathbf{X}_i jsou zachyceny parametrem $\boldsymbol{\beta}$. Takovéto rozšíření CHOPIT modelu je například uvedeno v King a kol. (2004) a Wand a kol. (2007). Podobný princip je využit též v běžně známých panelových či hierarchických modelech (např. Cameron & Trivedi, 2005).

Při existenci více sebehodnocení lze uvažovat i o přidání náhodného efektu na úrovni sebehodnocení. V článku Voňková a kol. (2015) je dále umožněno modelovat rozptyl pro každou sebehodnotící otázku zvlášť:

i.i.d. $\epsilon_{si} \sim N(0, \sigma_s^2)$, $s = 1, \dots, S$.

Tímto je možné zachytit rozdílný rozptyl při podávání odpovědí na jednotlivé sebehodnotící otázky. Některé sebehodnotící otázky mohou vykazovat mnohem větší rozptyl než jiné. Jedná se například o otázky s negativní formulací či otázky s reverzní škálou odpovědí (místo škály 1) *souhlasím*, 2) *spíše souhlasím*, 3) *ani souhlas, ani nesouhlas*, 4) *spíše nesouhlasím*, 5) *nesouhlasím* jsou uvedeny kategorie v opačném pořadí, tj. 1) *nesouhlasím*, 2) *spíše nesouhlasím*, 3) *ani souhlas, ani nesouhlas*, 4) *spíše souhlasím*, 5) *souhlasím*).

Peracchi & Rossetti (2012) se sice nezabývali modelováním více sebehodnocení v rámci jedné domény, ale modelovali sebehodnocení v různých doménách. Jednalo se o různé zdravotní domény, jako je bolest, deprese, problémy se spaním, pohybem, soustředěním a dýcháním. Jejich přístup lze však také aplikovat na existenci více sebehodnocení v rámci jedné domény. Ze statistického pohledu je jejich model zobecněním výše uvedených rozšíření. Rozptyl latentní proměnné se může v rámci jejich modelu lišit pro jednotlivé domény. Zároveň kovariance mezi latentními proměnnými pro jednotlivé domény může být nenulová. Konkrétně je jejich model pro vnímanou, nepozorovanou (latentní) proměnnou specifikován rovnicí

$$Y_{ih}^* = \alpha_h + \beta_h' \mathbf{X}_i + \gamma_h \epsilon_i + \sigma_h U_{ih}, \quad h = 1, \dots, H,$$

kde $h = 1, \dots, H$ značí zdravotní doménu, dále $\alpha_h, \beta_h, \gamma_h$ a σ_h jsou neznámé parametry, \mathbf{X}_i je vektor pozorovaných respondentových charakteristik, ϵ_i je nepozorovaný individuální efekt, U_{ih} jsou nepozorované regresní chyby nezávislé napříč doménami. ϵ_i a U_{ih} jsou vzájemně nezávislé a jsou též nezávislé na \mathbf{X}_i , mají nulovou střední hodnotu, jednotkový rozptyl a distribuční funkce G a F_U . Zatímco ϵ_i je společný všem doménám, jeho koeficient γ_h je odlišný pro jednotlivé domény. Vzhledem k tomu, že ϵ_i je konstantní pro jednoho respondenta napříč různými doménami, umožňuje zachytit korelaci latentních proměnných pro různé domény. Kovarianční matice latentních proměnných má následující faktorovou strukturu:

$$\text{Var}(Y_{ih}^* | \mathbf{X}_i) = \sigma_h^2 + \gamma_h^2,$$

$$\text{Cov}(Y_{ih}^*, Y_{il}^* | \mathbf{X}_i) = \gamma_h \gamma_l, \quad h \neq l.$$

Latentní proměnnou Y_{ih}^* nepozorujeme, místo ní máme k dispozici odpovědi na sebehodnotící otázky v jednotlivých doménách

$$Y_{ih} = j \Leftrightarrow \tau_{ih,j-1} < Y_{ih}^* \leq \tau_{ihj}, \quad j = 1, \dots, J + 1,$$

přičemž meze τ_{ihj} jsou modelovány jako

$$\tau_{ihj} = \begin{cases} -\infty, & j = 0 \\ \eta_{h1} + \boldsymbol{\delta}'_{h1} \mathbf{W}_i + \lambda_{h1} \epsilon_i, & j = 1 \\ \tau_{ih,j-1} + \exp(\eta_{hj} + \boldsymbol{\delta}'_{hj} \mathbf{W}_i + \lambda_{hj} \epsilon_i), & j = 2, \dots, J \\ \infty, & j = J + 1, \end{cases}$$

kde η_{hj} , $\boldsymbol{\delta}_{hj}$ a λ_{hj} jsou neznámé parametry, \mathbf{W}_i je vektor pozorovaných respondentových charakteristik, který může zahrnovat i proměnné \mathbf{X}_i . Díky zahrnutí individuálního efektu ϵ_i do mezí τ_{ihj} je umožněno, aby heterogenita ve stylu odpovídání na otázky byla závislá nejen na pozorovaných charakteristikách respondenta, ale také na nepozorované proměnné. ϵ_i může být vnímáno jako jeden faktor společný latentní proměnné a také mezním proměnným τ v různých doménách s koeficienty specifikovanými pro jednotlivé domény $h = 1, \dots, H$ a odpovědní kategorie $j = 1, \dots, J$.

Odpovědi na $K \geq 1$ vinět v jednotlivých dimenzích jsou modelovány následujícím způsobem:

$$Z_{ihk}^* = \xi_{hk} + \omega_{hk} V_{ihk}, \quad k = 1, \dots, K,$$

$$Z_{ihk} = j \Leftrightarrow \tau_{ih,j-1} < Z_{ihk}^* \leq \tau_{ihj}, \quad j = 1, \dots, J + 1,$$

kde latentní chyby V_{ihk} jsou nezávislé a stejně rozdělené náhodné veličiny s nulovou střední hodnotou a distribuční funkcí F_V . Škálové parametry ω_{hk} měří, jak dobře jsou viněty pochopeny. Poznamenejme, že mezní proměnné τ_{ihj} jsou stejné jako pro sebehodnocení, což odpovídá předpokladu konzistence odpovědí (*response consistency*). Latentní proměnná Z_{ihk}^* není závislá na pozorovaných respondentových charakteristikách \mathbf{X}_i ani na nepozorovaném individuálním efektu ϵ_i . To znamená, že až na náhodnou chybu jsou hypotetické situace popsané ve vinětách vnímány stejně všemi respondenty. Tento předpoklad je nazýván ekvivalence vinět (*vignette equivalence*).

3) Modelování časové dimenze

Angelini, Cavapozzi a Paccagnella (2011) rozšířili standardní CHOPIT model o přidání časové dimenze, tj. využili strukturu panelových dat a modelovali užívání škály v průběhu času. Ve srovnání s modely pro více sebehodnocení se v jejich rošíření mohou pozorované charakteristiky respondenta i měnit v čase, tj. místo \mathbf{X}_i užíváme \mathbf{X}_{it} , $t = 1, \dots, T$. Vnímanou latentní proměnnou modelují po vzoru běžných random-effect modelů, tj. náhodnou složku rozkládají na individuální náhodný efekt a idiosynkratickou chybu. Vinětové rovnice rozšiřují o modelování pohlaví u vinětové osoby. Specifikace jejich modelu pro sebehodnocení je následující:

$$Y_{it}^* = \mathbf{X}_{it}'\boldsymbol{\beta}_t + \epsilon_{it}, \quad i = 1, \dots, I, \quad t = 1, \dots, T,$$

$$\epsilon_{it} = \eta_i + \omega_{it}.$$

Vnímaná latentní úroveň zdravotních problémů při výkonu práce Y_{it}^* (v článku modelují reportování problémů v průběhu času) je součtem lineární kombinace pozorovaných respondentových charakteristik \mathbf{X}_{it} s koeficienty $\boldsymbol{\beta}_t$ (tj. koeficienty se mohou měnit v různých časových okamžicích t) a nepozorované náhodné proměnné ϵ_{it} . Nepozorovaná náhodná proměnná ϵ_{it} je součtem individuálního efektu η_i a idiosynkratické chyby ω_{it} . Náhodné složky jsou rozděleny následujícím způsobem:

$$\eta_i \sim N(0, \sigma^2),$$

$$\omega_{it} \sim N(0, 1),$$

η_i nezávislá na ω_{it} ,

ω_{it} nezávislé stejně rozdělené náhodné veličiny $i = 1, \dots, I, t = 1, \dots, T$.

Z uvedeného rozdělení náhodných veličin vyplývá, že rozptyl nepozorované náhodné proměnné ϵ_{it} je $\sigma^2 + 1$ a je korelovaná v různých časových okamžicích, přičemž kovariance pro různé časy je rovna σ^2 :

$$\text{Var}(\epsilon_{it}) = \sigma^2 + 1,$$

$$\text{Cov}(\epsilon_{it}, \epsilon_{is}) = \sigma^2 \quad s \neq t.$$

Vnímaná latentní proměnná Y_{it}^* je pozorována na základě v CHOPIT modelu standardního mechanismu s tím, že se meze mohou měnit v čase:

$$Y_{it} = k \Leftrightarrow \tau_{it}^{k-1} \leq Y_{it}^* < \tau_{it}^k, \quad k = 1, \dots, K,$$

$$\tau_{it}^1 = \boldsymbol{\gamma}_t^1 \mathbf{V}_{it},$$

$$\tau_{it}^k = \tau_{it}^{k-1} + \exp(\boldsymbol{\gamma}_t^k \mathbf{V}_{it}), \quad k = 2, \dots, K - 1.$$

Proměnné \mathbf{V}_{it} mohou obsahovat proměnné \mathbf{X}_{it} a *dummy* proměnnou za čas t .

Zajímavé je modelování hodnocení vinět $j = 1, \dots, J$. Nepozorovaná úroveň zdravotních problémů při výkonu práce popsaná ve vinětě j vnímaná respondentem i v čase t je modelována jako

$$Z_{ijt}^* = \theta_{jt} + \theta_{\text{fem}} \text{Female}_{ijt} + u_{ijt},$$

$$u_{ijt} = \zeta_i + v_{ijt},$$

kde θ_{jt} závisí na vinětě j na čase t , Female_{ijt} je *dummy* proměnná, která je rovna jedné, pokud osoba popsaná ve vinětě j je žena, pohlaví hypotetických osob ve viněťách je randomizováno pro různé respondenty. Angelini a kol. (2011) přidávají poznámku, že pohlaví hypotetické osoby ve vinětě je zavedeno v článku Kapteyn a kol. (2007) a užito i ve van Soest a kol. (2011), přičemž tito autoři argumentují, že respondenti mohou vnímat hypotetickou osobu jinak v závislosti na jejím pohlaví.

Podobně jako u sebehodnocení je nepozorovaná náhodná proměnná u_{ijt} rozdělena na individuální efekt ζ_i a idiosynkratickou chybu v_{ijt} . Longitudiální data umožňují uvolnit předpoklad ekvivalence vinět (*vignette equivalence*), který říká, že situace popsaná ve vinětě je v průměru vnímána všemi respondenty stejným způsobem. Předpoklad ekvivalence vinět je pomocí použitého značení možno zaznamenat jako $u_{ijt} = v_{ijt}$. V ekvivalentní formulaci modelu by bylo možné užít individuální efekt ζ_i v rovnici mezních parametrů τ_{ij}^t s tím, že pak by šel tento individuální efekt interpretovat jako nepozorovaný rozdíl mezi jednotlivými respondenty v hodnocení hypotetických osob ve viněťách. V modelu jsou zahrnuty celkem dva individuální efekty, η_i pro nepozorovanou individuální heterogenitu v sebehodnocení a ζ_i pro

nepozorovaný individuální rozdíl v hodnocení vinět.

Rozdělení náhodných veličin je následující:

$$\zeta_i \sim N(0, \rho^2),$$

$$v_{ijt} \sim N(0, \delta_{jt}^2),$$

ζ_i nezávislé na $v_{ijt}, \eta_i, \omega_{it}$,

v_{ijt} nezávislé stejně rozdělené náhodné veličiny $i = 1, \dots, I, j = 1, \dots, J, t = 1, \dots, T$,

v_{ijt} nezávislé na η_i, ω_{it} .

Z uvedeného rozdělení náhodných veličin vyplývá, že rozptyl u_{ijt} je roven $\rho^2 + \delta_{jt}^2$ a jeho kovariance je rovna ρ^2 v dvou různých časových okamžicích:

$$\text{Var}(u_{ijt}) = \rho^2 + \delta_{jt}^2,$$

$$\text{Cov}(u_{ijt}, u_{ijs}) = \rho^2 \text{ s } s \neq t.$$

Respondent i hodnotí vinětu j v čase t na základě následujícího schématu:

$$Z_{ijt} = k \Leftrightarrow \tau_{it}^{k-1} \leq Z_{ijt}^* < \tau_{it}^k, \quad k = 1, \dots, K.$$

Meze τ_{it}^k jsou modelovány stejně jako pro sebehodnocení, což je založeno na předpokladu konzistence odpovědí (*response consistency*).

4. Zpracování dat ve statistických softwarech

4.1. Neparametrický přístup

V této kapitole se zaměříme na statistické zpracování dat pomocí neparametrického přístupu metody ukotvujících vinět. Cílem je ukázat čtenáři, jak může pomocí statistického softwaru, kde je neparametrický přístup implementován, zpracovat svá vlastní data. Představíme zde vybrané klíčové příkazy pro zpracování dat neparametrickým přístupem. Samozřejmě jako při každém zpracování dat je nutné zjistit vlastnosti datové struktury pomocí vhodných základních deskriptivních statistik (např. tabulka četností, průměr, medián a rozptyl), vhodných statistických testů (např. t-test) či vhodných regresních modelů. Tomuto se věnuje mnoho učebních textů jako je Hendl (2015) a Chráska (2007), čtenáře tedy odkazujeme na tyto texty.

Vzhledem k tomu, že neparametrický přístup je prozatím implementován pouze ve statistickém softwaru R, zaměříme se právě na tento software. R je volně dostupný software, lze si ho stáhnout na webových stránkách <https://www.r-project.org/>. Pokud čtenář nemá vůbec žádné zkušenosti s tímto softwarem, doporučíme mu úvodní knihu Crawley (2005).

V této části textu budeme vycházet z článku Wand a kol. (2011), a to včetně ukázky datového souboru k politické svobodě obyvatel z různých zemí a výsledku zpracování odpovědí na sebehodnoticí otázku a ukotvující viněty. Tento datový soubor je volně dostupný, takže si čtenář může všechny níže uvedené příkazy sám vyzkoušet. Další užitečné informace o neparametrickém přístupu lze najít též v článku King a Wand (2007).

Prvním krokem pro práci s metodou ukotvujících vinět v softwaru R je instalace balíčku s názvem `anchors`. Tuto instalaci provedeme pomocí standardního příkazu:

```
> install.packages("anchors")
```

Balíček je pak vždy nutné před použitím aktivovat příkazem:

```
> library("anchors")
```

Podrobný popis balíčku s jeho funkcemi je uveden ve Wand, King a Lau (2016). Pro demonstrační účely využijeme volně dostupná data k politické svobodě s názvem `freedom`, která jsou součástí balíčku `anchors`. Tato data lze zavést do paměti pomocí následujícího příkazu:

```
> data("freedom")
```

Datový soubor `freedom` obsahuje 11 proměnných a 3500 pozorování. Proměnné zahrnují odpovědi na sebehodnotící otázku (proměnná `self`) a šest ukotvujících vinět (`vign1`, `vign2`, ..., `vign6`).

Sebehodnotící otázka má následující znění:

*self Do jaké míry se cítíte svobodný ve vyjadřování svých myšlenek beze strachu z pronásledování státní moci?
zcela svobodný/á, velmi svobodný/á, středně svobodný/á, trochu svobodný/á
a vůbec ne svobodný/á*

Následujících šest ukotvujících vinět popisuje hypotetické osoby s klesající mírou politické svobody.

vign1 [Kay] nesouhlasí s mnoha politickými rozhodnutími své vlády. Často publikuje své názory v novinách, kde kritizuje rozhodnutí státních představitelů a požaduje změnu. Považuje za nepravděpodobné, že by jí za to hrozily represe ze strany státní moci.

vign2 [Michael] nesouhlasí s mnoha politickými rozhodnutími své vlády. Ačkoli ví, že kritika je přijímána s nevolí, nemyslí si, že by vláda někoho potrestala za vyjadřování kritických názorů. Vyjadřuje beze strachu své názory na většinu otázek bez ohledu na to, kdo ho právě poslouchá.

vign3 [Bob] má jiné politické názory než jeho vláda. Slyšel o lidech, kteří byli uvězněni kvůli vystupování proti vládě. Vládní představitelé někdy ve svých proslovech odsuzují ty, kteří je kritizují. [Bob] občas píše příspěvky na politická témata do novin, ale nikdy je nepodepisuje svým pravým jménem.

vign4 [Connie] se nelíbí postoj vlády k mnoha otázkám. Má kamaráda, který byl uvězněn za přílišnou kritiku vládních představitelů. Proto se [Connie] zdráhá vyjadřovat své názory na veřejnosti.

vign5 [Vito] nesouhlasí s mnoha politickými rozhodnutími své vlády a je velmi opatrný, komu o svých názorech říká. Svě skutečné názory sděluje pouze rodině a blízkým přátelům. Zná několik lidí, kteří byli zadrženi za vyjadřování negativních názorů na veřejnosti.

vign6 [Sonny] žije v obavě, že bude pronásledován za své politické názory. Ti jeho známí, kteří se otevřeně vyjádřili proti vládě, byli zatčeni a zadrženi. [Sonny] nikdy nemluví o ničem, co vláda dělá, dokonce ani doma před svou rodinou.

Po všech uvedených vinětech následovala otázka se stejnou škálou, jako je u sebehodnotící otázky:

*Do jaké míry se dle Vašeho názoru cítí [jméno uvedené ve viněti] svobodný ve vyjadřování svých myšlenek beze strachu z pronásledování státní mocí?
zcela svobodný/á, velmi svobodný/á, středně svobodný/á, trochu svobodný/á
a vůbec ne svobodný/á*

Datový soubor `freedom` dále obsahuje proměnné pohlaví (`sex`; 1 označuje muže, 0 ženy), věk (`age`; věk respondenta uvedený v rocích), vzdělání (`educ`; 1 označuje žádné formální vzdělání, 2 nižší než primární vzdělání, 3 ukončené primární vzdělání, 4 ukončené nižší sekundární vzdělání, 5 ukončená střední škola, 6 ukončená college a 7 ukončené postgraduální vzdělání) a zemi (`country`; 1 označuje Eastasia, 2 Eurasia a 3 Oceania, jedná se o fiktivní země z Orwellova románu 1984). Základní strukturu datového souboru lze zjistit pomocí funkce `str()`:

```
> str(freedom)
'data.frame': 3500 obs. of 11 variables:
 $ sex      : int  0 1 0 1 1 1 1 1 1 0 ...
 $ age      : int  20 55 27 30 25 56 26 54 48 31 ...
 $ educ     : int  4 5 2 1 4 2 6 4 1 4 ...
 $ country: Factor w/ 3 levels "Eastasia","Eurasia",...: 2 3 1 1 3 1 1 1
1 3 ...
 $ self     : int  1 2 1 1 2 2 1 3 3 4 ...
 $ vign1    : int  4 3 2 2 3 3 2 5 5 5 ...
 $ vign2    : int  3 3 1 2 3 2 1 5 4 5 ...
```

```

$ vign3 : int 3 4 4 4 5 4 4 4 5 5 ...
$ vign4 : int 5 4 4 5 5 5 3 5 4 5 ...
$ vign5 : int 3 4 5 5 5 5 3 5 4 5 ...
$ vign6 : int 4 5 5 5 5 4 5 5 4 5 ...

```

4.1.1. Pořadí vinět

Pro výpočet nekorigovaného (sebe)hodnocení pomocí neparametrického přístupu je nutné stanovit pořadí vinět. Vlastní sebehodnocení je pak porovnáváno s hodnocením seřazených vinět.

Výše je uveden seznam vinět týkajících se politické svobody, přičemž viněty jsou v něm seřazeny sestupně, tj. od popisu hypotetických osob s největší mírou politické svobody (`vign1`) k popisu osob s nejnižší mírou politické svobody (`vign6`). Toto pořadí stanovil výzkumník a priori. Je však otázkou, jak seřadí viněty respondenti. Pokud se pořadí stanovené výzkumníkem odlišuje od většiny respondentů, může to poukazovat na problémy s formulací vinět. Viněty mohou být například příliš dlouhé, nesrozumitelné, různými respondenty odlišně pochopené apod. Nebo může prostě výzkumník udělat chybu při stanovení pořadí. Analýza pořadí vinět vyplývající z hodnocení respondentů může být tedy zásadní pro diagnostikování problému s výzkumným nástrojem či sebranými daty.

Pro analýzu pořadí vinět lze použít funkci `anchors.order()`. Strukturu výstupních hodnot této funkce zjistíme pomocí `str(anchors.order())`. Tato funkce aplikovaná pro datový soubor `freedom` má následující výstup:

```

> z1<-anchors.order(~vign1+vign2+vign3+vign4+vign5+vign6,freedom)
> str(z1)
List of 9
 $ labels      : chr [1:3500] "{2,3,5},{1,6},4" "{1,2},{3,4,5},6"
"2,1,{3,4},{5,6}" "{1,2},3,{4,5,6}" ...
 $ ngroup      : int [1:3500] 3 3 4 3 2 4 5 2 2 1 ...
 $ nviol       : num [1:3500] 5 0 1 0 0 3 3 2 7 0 ...
 $ vign        : num [1:3500, 1:6] 4 3 2 2 3 3 2 5 5 5 ...
 .. attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:3500] "109276" "25117" "106877" "69437" ...
 .. ..$ : chr [1:6] "vign1" "vign2" "vign3" "vign4" ...
 $ compare.matrix: num [1:6, 1:6] NA 0.3206 0.08 0.0557 0.0666 ...
 .. attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:6] "1" "2" "3" "4" ...
 .. ..$ : chr [1:6] "<1" "<2" "<3" "<4" ...
 $ freq        : 'table' int [1:615(1d)] 233 129 113 98 97 93 79 63
58 57 ...

```

```

..- attr(*, "dimnames")=List of 1
.. ..$ out: chr [1:615] "{1,2,3,4,5,6}" "{1,2},{3,4,5,6}"
"{1,2},3,{4,5,6}" "{1,2},{3,5},{4,6}" ...
$ ties          : chr "set"
$ main         : chr "Treatment of ties: represent as sets"
$ N            : int 3500
- attr(*, "class")= chr "anchors.order"

```

`anchors.order()`\$labels popisuje pořadí vinět u jednotlivých respondentů. Z výše uvedeného výstupu například vidíme, že první respondent hodnotil viněty 2, 3 a 5 stejně a dal jim ve srovnání s ostatními vinětami nižší hodnocení (viněty jsou uvedeny v jedné množině {2,3,5} a zároveň před dalšími vinětami), nejvyšší hodnocení dal vinětě 4. Viněty 1 a 6 hodnotil shodně, přičemž toto hodnocení je vyšší než u vinět 2, 3 a 5 a nižší než u viněty 4. Původní hodnocení vinět u prvního respondenta bylo: $vign1 = 4$, $vign2 = 3$, $vign3 = 3$, $vign4 = 5$, $vign5 = 3$ a $vign6 = 4$.

Dalším výstupem je `anchors.order()`\$ngroup, který udává pro každého respondenta počet odlišných hodnocení vinět. U prvního respondenta jsou to tři množiny {2,3,5}, {1,6} a {4}. Vzhledem k tomu, že máme šest vinět a škála pro hodnocení vinět je pětibodová, můžeme celkem získat pět odlišných hodnocení, tj. hodnocení alespoň dvou vinět budou totožná. Počet různých hodnocení vinět je obecně omezen počtem bodů hodnotící škály. Pokud je počet vinět větší než počet bodů škály, může být zjevně počet různých hodnocení roven maximálně počtu bodů škály. Pokud je počet vinět menší než počet bodů škály či jsou tyto počty totožné, může být maximální počet odlišných hodnocení roven počtu vinět. Nizký počet odlišných hodnocení typicky poukazuje na to, že respondentovi úroveň vinět splývá a hodnotí je stejně (tedy pokud počet bodů škály je pětibodový, jak tomu typicky bývá; v extrémním případě jednobodové hodnotící škály by si samozřejmě byla všechna hodnocení rovna). To by mělo být varovným signálem pro výzkumníka, který se snažil formulovat viněty s různou úrovní měřeného konceptu. Naopak při dosažení maximálního počtu různých hodnocení vinět sice respondent odlišuje úroveň měřeného konceptu ve vinětách, ale nemůžeme si být jistí, zdali je jeho řazení ve shodě s řazením, které předpokládá výzkumník.

K odhalení počtu porušení pořadí oproti předpokládanému pořadí výzkumníkem slouží další funkce `anchors.order()`\$nviol. Zmiňovaný první respondent porušil předpokládané pořadí výzkumníkem celkem pětkrát. Ačkoli

viněta 1 popisuje nejvyšší úroveň politické svobody (předpoklad výzkumníka), hodnotil ji první respondent hůře (volil vyšší bod škály popisující horší úroveň politické svobody) než vinětu 2, 3 a 5 (srovnej $vign1 = 4$ s $vign2 = 3$, $vign3 = 3$ a $vign5 = 3$). Zde tedy porušil předpokládané pořadí třikrát. Dále pak hodnocení viněty 4 má být nižší než hodnocení vinět 5 i 6, což je opět v případě tohoto respondenta porušeno (srovnej $vign4 = 5$ s $vign5 = 3$ a $vign6 = 4$). Zde je tedy porušeno předpokládané pořadí dvakrát. Další porušení pořadí již nenastala a jejich počet je tedy skutečně roven 5. Poznamenejme, že při stejném hodnocení vinět respondentem (např. $vign1 = vign2$) není touto funkcí `$nviol` porušení pořadí započítáno. Například u druhého respondenta je hodnocení vinět následující $\{1,2\},\{3,4,5\},6$, tj. $vign1 = vign2 < vign3 = vign4 = vign5 < vign6$ a hodnota funkce `$nviol` je rovna 0.

Funkce `anchors.order()` `$vign` shrnuje původní hodnocení vinět respondenty. Jedná se tedy o matici s 3500 řádky (odpovídá 3500 pozorování) a 6 sloupci (odpovídá hodnocení 6 vinět).

Dalším výstupem je `anchors.order()` `$compare.matrix`, který je velmi zásadní pro analýzu pořadí udaného respondenty ve srovnání s předpokládaným pořadím. Udává, kolikrát v datovém souboru pro každou dvojici vinět nastalo porušení pořadí. Například kolik respondentů hodnotilo vinětu 1 hůře než vinětu 2. Pro zobrazení výstupu této funkce pro data `freedom` bychom mohli tedy použít příkaz `anchors.order(~vign1+vign2+vign3+vign4+vign5+vign6,freedom)$compare.matrix`. Další možností je však využití `summary` funkce u `anchors.order`, která ještě zobrazí další doprovodné statistiky a obsahuje i zbytek výstupu `anchors.order()`:

```
> summary(z1,top=10,digits=3)
ANCHORS: SUMMARY OF VIGNETTE ORDERING
Treatment of ties: represent as sets
Number of cases with at least two distinct vignette responses: 3267
and with no violations of natural ordering: 773
and with no more than 1 violation of natural ordering: 1654
and with no more than 2 violation of natural ordering: 2299
```

Proportion of cases a vignette (row) is less than another (column):

	<1	<2	<3	<4	<5	<6
1	NA	0.235	0.651	0.757	0.715	0.775
2	0.321	NA	0.663	0.732	0.707	0.754
3	0.080	0.121	NA	0.457	0.363	0.575
4	0.056	0.080	0.138	NA	0.183	0.374
5	0.067	0.068	0.198	0.339	NA	0.495
6	0.055	0.070	0.081	0.100	0.103	NA


```

Upper tri =      p_{ij} - p_{ji} (negative values suggest misorderings)
Lower tri = 1 - p_{ij} - p_{ji} (big numbers means many ties)
  1      2      3      4      5      6
1      NA -0.085 0.571 0.702 0.649 0.720
2 0.444      NA 0.542 0.652 0.639 0.684
3 0.269 0.405      NA 0.320 0.165 0.494
4 0.215 0.218 0.477      NA -0.156 0.275
5 0.187 0.225 0.171 0.345      NA 0.392
6 0.188 0.440 0.176 0.526 0.402      NA

```

Top 10 orderings (out of 615 unique orderings):

	Frequency	Proportion	Ndistinct	Nviolation
{1,2,3,4,5,6}	233	0.0666	1	0
{1,2},{3,4,5,6}	129	0.0369	2	0
{1,2},3,{4,5,6}	113	0.0323	3	0
{1,2},{3,5},{4,6}	98	0.0280	3	1
{1,2},{3,4,5},6	97	0.0277	3	0
2,1,{3,4,5,6}	93	0.0266	3	1
{1,2},3,{4,5},6	79	0.0226	4	0
{1,2},5,{3,4,6}	63	0.0180	3	2
2,1,3,{4,5},6	58	0.0166	5	1
2,1,{3,5},{4,6}	57	0.0163	4	2

Ve funkci `summary` jsme zvolili omezení zobrazení všech statistik ve výstupu na tři desetinná místa (`digits = 3`) a výčet pouze deseti nejčastějších pořadí uváděných respondenty (`top = 10`). Poznamenejme, že tento výstup je též uveden a analyzován v článku Wand a kol. (2011).

Na konci výstupu je frekvenční tabulka deseti nejčastějších pořadí. Vidíme, že nejčastěji respondentům všechny viněty splynuly a hodnotili je totožně. Jednalo se o 233 respondentů (`Frequency`), což je procentuálně 6,7 % (`Proportion`). Počet odlišných hodnocení je tedy pouze 1 (`Ndistinct`) a počet porušení pořadí je 0 (`Nviolation`). Další nejčastější pořadí bylo totožné hodnocení vinět 1 a 2 s vyšším totožným hodnocením ostatních vinět, tj. $\{1,2\},\{3,4,5,6\}$. Dále následovala pořadí $\{1,2\},3,\{4,5,6\}$ a $\{1,2\},\{3,5\},\{4,6\}$. U posledního jmenovaného pořadí došlo k porušení přirozeného pořadí – viněta 4 je hodnocena hůře než viněta 5.

V úvodu výstupu jsou shrnující statistiky k počtu porušení přirozeného pořadí a rozdílného hodnocení. 3267 respondentů hodnotilo viněty minimálně dvěma různými body škály, 773 respondentů neporušilo přirozené pořadí vinět, dále 1654 respondentů ho porušilo jedenkrát (např. uvedlo viněty v pořadí $\{1,2\},\{3,5\},\{4,6\}$ nebo $2,1,\{3,4,5,6\}$) a 2299 ho porušilo dvakrát (např. $\{1,2\},5,\{3,4,6\}$ nebo $2,1,\{3,5\},\{4,6\}$).

Tato analýza naznačuje, že velmi často docházelo k porušení řazení vinět 4 a 5 a dále i vinět 1 a 2. Střední část výstupu toto dobře shrnuje. V první tabulce jsou uvedeny proporce případů, kdy je hodnocení viněty s indexem uvedeným v řádku nižší než hodnocení viněty uvedené ve sloupci. Například 23,5 % respondentů hodnotilo vinětu 1 lépe než vinětu 2 (toto výzkumník očekává), ale mnohem více respondentů (32,1 %) hodnotilo vinětu 1 hůře. Rozdíl mezi těmito pravděpodobnostmi je záporný, což poukazuje na případné špatně stanovené přirozené pořadí výzkumníkem. Navíc mnoho respondentů ($100 - 23,5 - 32,1 = 44,4$ %) hodnotilo tyto viněty stejně. Poznamenejme však znovu, že vzhledem k pětibodové škále a šesti hodnoceným vinětám muselo dojít ke stejnému hodnocení alespoň dvou vinět. V druhé tabulce jsou shrnuty jak rozdíly v těchto pravděpodobnostech (horní trojúhelníková matice) označující již zmíněné porušení přirozeného pořadí každé dvojice vinět, tak počet totožných hodnocení u každé dvojice vinět. Jak vidíme, nejen v případě vinět 1 a 2 je rozdíl mezi pravděpodobnostmi $p_{ij} - p_{ji}$ záporný ($23,5 - 32,1 = -8,6$ %; ve výstupu je $-8,5$ %, jedná se pouze o zaokrouhlovací chybu). Negativní rozdíl je nalezen i pro viněty 4 a 5, přičemž se jedná ještě o větší rozdíl než u vinět 1 a 2. Podíl respondentů hodnotících vinětu 4 lépe než vinětu 5 byl nižší (18,3 %) než podíl respondentů hodnotících vinětu 5 lépe než vinětu 4 (33,9 %), rozdíl je dvakrát větší než pro viněty 1 a 2 (srovnej $18,3 - 33,9 = -15,6$ % pro viněty 4 a 5 s $23,5 - 32,1 = -8,6$ % pro viněty 1 a 2).

Dalším zajímavým výstupem je, že nejčastější shodné hodnocení dvou vinět nastává pro viněty 4 a 6 (52,6 %). Respondenti v nich tedy nejčastěji vidí totožnou úroveň politické svobody.

Nekonzistence mezi pořadím vinět, které zde udávají respondenti a které stanovuje výzkumník, může mít mnoho důvodů. Jeden z nich může spočívat v problematické formulaci vinět. Viněty 1 a 2 respondentům často splývají (44,4 % je hodnotí stejně) možná proto, že obě popsané osoby vládu kritizují, zveřejňují své názory na veřejnosti a neobávají se potrestání. Dovětky, jak respondent kritizuje vládu (vyžadování změny ve vinětě 1 oproti vinětě 2), mohou respondentovi uniknout. Prohození pořadí vinět 1 a 2 (32,1 % respondentů) může být zase způsobeno tím, jak jsou viněty zakončeny. V poslední větě viněty 1 je uvedeno, že respondent vidí malé důvody k potrestání, tj. nějaké důvody tam jsou, nejedná se čistě o žádné důvody. Oproti tomu je viněta 2 zakončena pozitivně, *Michael* před všemi zveřejňuje

své názory. Samozřejmě se zde jedná jen o interpretaci autora, reálné důvody by bylo možné zjistit například pomocí interview s respondenty. Mimo formulace vinět může problém též spočívat třeba v nedostatku času na sběr dat (respondent nemá dostatek času na přečtení viněty a zamyšlení se nad ní) či neporozumění vinět respondenty (ačkoli je viněta správně zformulovaná, respondenti ji nemusí rozumět např. kvůli nízké úrovni čtenářské gramotnosti – toto by však souviselo s problematickou formulací vinět, kdy výzkumník nepřizpůsobil formulaci vinět cílové populaci).

Dále je též zajímavé, že nezanedbatelné části respondentů všechny viněty splývají (6,66 %, 233 respondentů). Tito respondenti většinou spátrují ve všech vinětách spíše nízkou míru politické svobody – hodnotí hypotetické osoby jako středně svobodné (34,8 %), trochu svobodné (23,2 %) či zcela nesvobodné (26,6 %). Co se týká vinět 4, 5 a 6, tak respondentům buď často splývaly (připomeňme, že viněta 4 byla hodnocena stejně jako viněta 6 celkem 52,6 % respondenty) nebo je hodnotili v opačném pořadí (vinětu 4 hodnotilo hůře 33,9 % respondentů). Důvod opět můžeme spatřovat v problematické formulaci vinět. Například ve vinětě 4 je přítel Connie uvězněn, což je pro ni bolestivá zpráva o jednom z nejbližších lidí. V ostatních vinětách není zas až tak blízká osoba uvězněna. Toto může být pro někoho zásadnější než popsané možnosti vyjadřování svých názorů – ve vinětě 4 vyhýbání se vyjadřování na veřejnosti, ve vinětě 5 sdělování názorů pouze doma a ve vinětě 6 nevyjařování svých názorů za žádných okolností, ani doma před svými nejbližšími.

4.1.2. Výpočet korigovaného sebehodnocení

Po analýze pořadí vinět je nutné stanovit pořadí vinět, se kterým budeme dále pracovat (tedy pokud se nerozhodneme viněty přeformulovat a sbírat data znovu). V článku Wand a kol. (2011), z něhož v tomto příkladu vinět týkajících se politické svobody vycházíme, zvolili pro výpočet korigovaného sebehodnocení čtyři viněty s tím, že první dvě viněty prohodili. Konkrétně zvolili pořadí vign2, vign1, vign3 a vign6. My zde provedeme výpočet korigovaného sebehodnocení pro všechny viněty v původním pořadí daném výzkumníkem a pomocí doprovodných statistik budeme dále analyzovat vhodnost a smysluplnost volby určitého pořadí vinět pro výpočet korigovaného sebehodnocení.

Ve srovnání s funkcí `anchors.order()` užívá funkce pro výpočet korigovaného hodnocení `anchors()` navíc i označení pro sebehodnocení na levé straně formule:

```
> z2<-
anchors(self~vign1+vign2+vign3+vign4+vign5+vign6,freedom,method="C",com
bn=TRUE)
> summary(z2)
```

ANCHORS: SUMMARY OF RELATIVE RANK ANALYSIS:

Overview of C-ranks

Number of cases: 2065 with interval value, 1435 with scalar value

Maximum possible C-rank value: 13

Interval on C-scale: Frequency and proportions Cs to Ce

	N	Prop	MinEnt
1 to 1	323	0.092	1
2 to 2	41	0.012	2
3 to 3	23	0.007	3
4 to 4	64	0.018	4
5 to 5	260	0.074	5
6 to 6	74	0.021	6
7 to 7	58	0.017	7
8 to 8	28	0.008	8
9 to 9	11	0.003	9
10 to 10	31	0.009	10
11 to 11	22	0.006	11
12 to 12	162	0.046	12
13 to 13	338	0.097	13
10 to 12	43	0.012	12
10 to 13	39	0.011	13
1 to 10	26	0.007	8
1 to 11	4	0.001	8
1 to 12	28	0.008	8
1 to 13	12	0.003	8
1 to 4	114	0.033	1
1 to 5	26	0.007	1
1 to 6	14	0.004	1
1 to 7	1	0.000	1
1 to 8	13	0.004	8
1 to 9	2	0.001	8
2 to 10	37	0.011	8
2 to 11	11	0.003	8
2 to 12	125	0.036	8
2 to 13	31	0.009	8
2 to 4	101	0.029	4
2 to 5	50	0.014	5
2 to 6	16	0.005	5
2 to 7	4	0.001	5
2 to 8	21	0.006	8
3 to 10	14	0.004	8
3 to 11	4	0.001	8

3 to 12	6	0.002	8
3 to 13	11	0.003	8
3 to 6	2	0.001	5
3 to 7	10	0.003	5
3 to 8	8	0.002	8
3 to 9	4	0.001	8
4 to 10	33	0.009	8
4 to 11	12	0.003	8
4 to 12	83	0.024	8
4 to 13	70	0.020	8
4 to 6	6	0.002	5
4 to 7	9	0.003	5
4 to 8	10	0.003	8
4 to 9	3	0.001	8
5 to 10	77	0.022	8
5 to 11	26	0.007	8
5 to 12	8	0.002	8
5 to 13	6	0.002	8
5 to 8	27	0.008	8
5 to 9	3	0.001	8
6 to 10	84	0.024	8
6 to 11	36	0.010	8
6 to 12	162	0.046	8
6 to 13	32	0.009	8
6 to 8	15	0.004	8
6 to 9	2	0.001	8
7 to 10	79	0.023	8
7 to 11	37	0.011	8
7 to 12	6	0.002	8
7 to 13	6	0.002	8
8 to 10	106	0.030	8
8 to 11	61	0.017	8
8 to 12	237	0.068	8
8 to 13	51	0.015	8
9 to 13	1	0.000	13

Note: MinEnt is the rank for the interval that minimizes entropy

Summary of C-ranks with ties/intervals broken:

Distribution of ranks omiting interval cases

1	2	3	4	5	6	7	8	9	10	11	12	13
0.225	0.029	0.016	0.045	0.181	0.052	0.04	0.02	0.008	0.022	0.015	0.113	0.236

Distribution of ranks allocating interval cases uniformly

1	2	3	4	5	6	7	8	9	10	11	12	13
0.105	0.045	0.042	0.062	0.108	0.065	0.067	0.087	0.077	0.089	0.058	0.088	0.107

Distribution of ranks allocating interval cases via cpolr
and conditioning on observed ranks

1	2	3	4	5	6	7	8	9	10	11	12	13
0.115	0.030	0.018	0.071	0.133	0.049	0.045	0.141	0.025	0.152	0.017	0.094	0.110

Allocating cases to their MinEnt values produces

1	2	3	4	5	6	7	8	9	10	11	12	13
0.137	0.012	0.007	0.047	0.102	0.021	0.017	0.473	0.003	0.009	0.006	0.059	0.108

Summary of entropy and intervals by subsets of vignettes:

	Vignettes	Estimated entropy	Minimum entropy	Interval Cases	Span avg.	Max. rank
1	123456	2.360	1.736	2065	4.063	13
7	12345	2.168	1.675	1954	3.454	11
6	12456	2.189	1.700	1930	3.349	11
5	23456	2.247	1.700	1763	3.576	11
4	13456	2.250	1.727	1695	3.305	11
3	12356	2.252	1.895	1550	2.930	11
2	12346	2.258	1.915	1544	2.882	11
23	3456	1.921	1.434	1651	3.081	9
21	1235	1.960	1.707	1306	2.392	9
20	1245	1.971	1.623	1790	2.762	9
19	1246	1.994	1.774	1295	2.270	9
18	1234	2.013	1.780	1216	2.302	9
17	1345	2.044	1.675	1538	2.713	9
16	2345	2.051	1.638	1639	2.973	9
15	1236	2.057	1.883	1084	2.163	9
14	1256	2.058	1.856	1176	2.210	9
13	1456	2.060	1.658	1559	2.704	9
12	2456	2.060	1.649	1605	2.870	9
11	1356	2.119	1.875	1113	2.202	9
10	1346	2.126	1.900	1085	2.142	9
9	2356	2.131	1.854	1226	2.448	9
8	2346	2.136	1.874	1205	2.403	9
42	123	1.634	1.493	851	1.785	7
41	456	1.655	1.324	1508	2.502	7
40	126	1.685	1.558	754	1.680	7
39	345	1.712	1.397	1476	2.501	7
38	124	1.719	1.601	814	1.726	7
37	125	1.727	1.610	800	1.715	7
36	346	1.775	1.585	1010	1.933	7
35	356	1.776	1.578	1053	1.993	7
34	135	1.807	1.671	786	1.696	7
33	145	1.829	1.561	1349	2.136	7
32	235	1.832	1.662	960	1.921	7
31	245	1.833	1.565	1443	2.293	7
30	146	1.844	1.724	801	1.644	7
29	246	1.851	1.710	915	1.803	7
28	134	1.856	1.742	687	1.596	7
27	234	1.885	1.736	852	1.830	7
26	156	1.892	1.788	728	1.609	7
25	136	1.904	1.845	532	1.470	7
24	256	1.909	1.790	794	1.731	7
22	236	1.923	1.834	705	1.691	7
57	12	1.094	1.004	550	1.418	5
56	46	1.399	1.291	709	1.467	5
55	45	1.421	1.211	1262	1.953	5
54	13	1.450	1.429	224	1.152	5
53	35	1.457	1.359	694	1.506	5
52	56	1.485	1.403	620	1.421	5
51	23	1.487	1.436	426	1.329	5
50	16	1.492	1.467	196	1.136	5
49	34	1.495	1.414	579	1.405	5
48	26	1.515	1.472	319	1.228	5

47	36	1.528	1.489	425	1.285	5
46	15	1.528	1.508	225	1.154	5
45	14	1.540	1.524	206	1.142	5
44	25	1.559	1.524	371	1.255	5
43	24	1.561	1.529	380	1.272	5
63	1	0.794	0.794	0	1.000	3
62	2	0.879	0.879	0	1.000	3
61	6	0.961	0.961	0	1.000	3
60	3	1.044	1.044	0	1.000	3
59	4	1.061	1.061	0	1.000	3
58	5	1.086	1.086	0	1.000	3

První část výstupu shrnuje přehled korigovaných sebehodnocení c – jejich absolutní a relativní četnosti a také počet intervalových a skalárních hodnot. Intervalových hodnot je 2065 a skalárních hodnot je 1435, což znamená, že u přibližně 60 % respondentů nelze určit právě jednu relativní pozici sebehodnocení vůči hodnocení vinět – celkový počet relativních pozic pro 6 vinět je třináct: 1) $self < vign1$, 2) $self = vign1$, 3) $vign1 < self < vign2$, ..., $vign6$, 13) $vign6 < self$. Důvodem je, že hodnocení vinět je buď nekonzistentní (např. $vign2 < self < vign1 < vign3 < vign4 < vign5 < vign6$) či totožné (např. $vign1 < vign2 < vign3 < self = vign4 = vign5 < vign6$). Poznamenejme, že mohou nastat případy, kdy je hodnocení vinět sice nekonzistentní či totožné, ale korigované sebehodnocení c nenabývá intervalové hodnoty. Například pokud $self < vign1 < vign6 < vign5 < vign4 < vign3 < vign2$, je korigované sebehodnocení c rovno 1, neboť dochází ke splnění pouze první z celkových třinácti relativních pozic. Takovýto respondent ohodnotil svou pozici pod vinětou 1 a ostatní viněty, které hodnotil hůře než první vinětu, zpřeházel. Je možné, že jsou mu svou úrovní natolik vzdálené, že pořadí prostě zpřeházel. Jako když by vzrůstem malý člověk hodnotil výšku velmi vysokých lidí. Mohl by uvést výšky těchto lidí chybně, protože jsou už prostě mimo jeho rozlišovací schopnost. Samozřejmě je též možné, že nekonzistentní hodnocení vinět je důsledkem jejich nepozorného čtení respondenty. Takoví respondenti pak nejsou věrohodní.

Co se týká skalárních hodnot (ve výstupu prvních třináct řádků, kde $c_s = c_e$), nejčastější hodnotou je $c = 13$ (9,7 %) následované hodnotou $c = 1$ (9,2 %). Respondenti tedy označují svou situaci jako horší, než je poslední viněta (sebehodnocení je větší než hodnocení poslední viněty, ostatní viněty jsou hodnoceny buď stejně, nebo níže než poslední šestá viněta), nebo respondenti označují svou pozici jako lepší, než je první viněta (sebehodnocení je menší než

první viněta, ostatní viněty jsou hodnoceny buď stejně, nebo hůře než první viněta). Nejčastější intervalovou hodnotou ($c_s < c_e$) je $c = [8,12]$ (6,8 % respondentů), následované hodnotou $c = [6,12]$ (4,6 % respondentů), ostatní intervalové hodnoty jsou jednotlivě zastoupeny u méně než čtyř procent respondentů.

Poslední sloupec MinEnt v první tabulce se vztahuje k tomu, jak sumarizovat korigované sebehodnocení všech respondentů pomocí tabulky frekvencí (či histogramu). V části 3.1.3 je uvedeno, jak postupovat při sumarizování korigovaného sebehodnocení. V případě výskytu pouze skalárních hodnot je tento postup triviální, jedná se o běžné sestavení tabulky četností či histogramu. V případě intervalových hodnot lze postupovat následovně: a) sestavit histogram bez vektorových hodnot, b) rozložit prostor rovnoměrně přes všechny prvky vektoru, c) započítat každý prvek vektoru respondenta podle proporce *podobných* respondentů, kteří zvolili kategorie obsažené ve vektoru, d) sestavit histogram dle principu minimální a maximální entropie. Shrnutí korigovaného sebehodnocení na základě všech čtyřech postupů lze nalézt ve střední části výstupu (postupně *Distribution of ranks omitting interval cases*, *Distribution of ranks allocating interval cases uniformly*, *Distribution of ranks allocating interval cases via cpolr and conditioning on observed ranks* a *Allocating cases to their MinEnt values produces*).

Vzhledem k existenci velkého množství intervalových hodnot (59 %) je rozložení korigovaného sebehodnocení pro různé postupy netotožné. Kdyby intervalové hodnoty neexistovaly, nezáleželo by na předpokladech uvedených postupů a rozložení by totožné bylo. Zde tedy bude záležet na výzkumníkovi, jaké učiní předpoklady a jaký postup zvolí. Pokud vynechá intervalové hodnoty (postup a), nemusí být respondenti, kteří zbydou, reprezentativním vzorkem dané populace. Nejčastěji jsou zastoupeny krajní kategorie 13 s 23,6 % respondenty (situace respondenta je horší než poslední šestá viněta) a 1 s 22,5 % respondenty (situace respondenta je lepší než první viněta).

V případě rovnoměrného rozložení přes celý interval (postup b) je celkové rozložení hodnot plochého tvaru. Rozdíl mezi nejčetnější kategorií 13 a nejméně početnou kategorií 3 je 6,5 procentního bodu (srovnej 10,7 % pro kategorii 13 s 4,2 % pro kategorii 1). Při volbě cenzurovaného probitového

modelu pro ordinální proměnnou jako parametrického doplňku k neparametrickému přístupu (postup c) vyvstává otázka, do jaké míry jsou splněny předpoklady tohoto parametrického modelu. Poznamenejme, že v případě, kdy nspecifikujeme, jaké vysvětlující proměnné vstupují do parametrického modelu (tj. jak budeme definovat podobnost respondentů a na základě toho přidělovat váhy jednotlivým hodnotám v intervalu), vystupuje jako vysvětlující proměnná pouze konstanta. Pokud zvolíme rozdělení hodnot dle principu minimální entropie, dostaneme nejméně plochý histogram, jedná se o „nejhorší scénář“, kdy většina respondentů je v jedné kategorii, a tudíž nekorigované sebehodnocení mezi respondenty nerozlišuje. V našem modelovém případě se nejčastěji vyskytuje kategorie 8, jejíž zastoupení je 47,3 %. Další hodnoty jsou výrazně méně zastoupené, následuje kategorie 1 s 13,7 % respondentů a kategorie 13 s 10,8 % respondentů. Kategorie 9, 11, 3 a 10 jsou zastoupeny méně než jedním procentem respondentů. (Oproti tomu je v případě rovnoměrného rozložení přes celý interval (postup b) nejméně zastoupená kategorie 3 s 4,2 % respondenty.) Pro každou intervalovou hodnotu je v případě volby postupu s minimální entropií přidělena jedna hodnota. Tu lze ve výstupu najít v první tabulce za frekvencí každé skalární a intervalové hodnoty (vidíme, že většina intervalových hodnot je reprezentována hodnotou 8). Princip minimální entropie nejméně ze všech postupů diferencuje mezi respondenty. Poznamenejme, že naopak dle principu maximální entropie bychom dospěli k rozložení hodnot, které by bylo nejvíce ploché.

4.1.3. Výběr vinět

Zásadní otázkou je, jaké viněty mají největší diskriminační sílu, tj. pomocí jakých vinět nejlépe rozlišíme korigované sebehodnocení mezi jednotlivými respondenty. Viněty reprezentující extrémní situaci nejsou pro běžnou populaci vhodné. Například viněta popisující extrémní politickou nesvobodu nebude vhodná pro evropské země. Lze očekávat, že respondenti by hodnotili svou politickou svobodu lépe než tuto vinětu a všichni by tudíž po korekci byli v jedné kategorii. Takováto viněta je neúčinná. Dodejme, že entropie by pro takovýto případ byla rovna nule. Dále v předvýzkumu máme často zařazeno více vinět a pro hlavní výzkum chceme vybrat ty nejužitečnější viněty, resp. nejužitečnější kombinaci vinět.

V poslední části výstupu `summary(anchors())` uvedeného v 4.1.2 je pro

každou kombinaci vinět (při zachování pořadí určeného výzkumníkem) uvedena odhadnutá entropie, minimální entropie, počet intervalových hodnot, průměrné rozpětí hodnot (v průměru přes kolik kategorií se hodnoty rozpínají) a maximální hodnota korigovaného sebehodnocení (pro 6 vinět to může teoreticky být 13, pro 5 vinět 11, obecně pro J vinět $2J+1$). Vidíme, že v případě volby pěti vinět bychom se měli dle entropie rozhodnout pro kombinaci vinět 12346, následované kombinací 12356, 13456 a 23456, tj. jsou postupně vynechané viněty 5, 4, 2 a 1. Výše jsme již uvedli, že docházelo k častým přesmyčkám mezi vinětou 1 a 2 a také mezi vinětou 4 a 5. Pokud bychom se rozhodli pro výběr tří vinět, jako optimální kombinace se jeví 236. U této kombinace navíc dochází k tomu, že hodnota odhadnuté i minimální entropie je vyšší než pro kombinaci čtyř vinět 3456 (srovnej odhadnutou entropii 1,923 a minimální entropii 1,834 pro kombinaci třech vinět 236 postupně s hodnotami 1,921 a 1,434 pro kombinaci čtyř vinět 3456). Ve výzkumných šetřeních se často užívají tři viněty. V tomto případě by tedy výzkumník měl zvážit kombinaci vinět 236. V případě volby dvou vinět se jedná o kombinaci 24 či 25. V případě volby jedné viněty, což je též častý případ v praxi, by se nabízela dle hodnoty entropie viněta 5 nebo 4. Jedna viněta znamená menší zátěž pro respondenty (čtení příběhů může být náročné). Typicky se však v této situaci s jednou vinětou využívá parametrický model s dodatečnými statistickými předpoklady.

4.1.4. Grafické znázornění korigovaného sebehodnocení

Korigované sebehodnocení pomocí neparametrického přístupu je pro základní orientaci vhodné graficky zobrazit. V následujícím příkladu zobrazíme korigované sebehodnocení pro tři viněty 2, 3 a 6, které jsou nejvhodnější kombinací tří vinět na základě hodnot odhadnuté či minimální entropie. Zobrazení provedeme pro dvě skupiny respondentů – respondenty z Eastasie ($N = 2705$) a respondenty z Oceánie ($N = 609$). Obě skupiny zobrazíme pomocí všech čtyř výše uvedených postupů (vynechání intervalových hodnot, rovnoměrné rozložení hodnot v intervalu, využití parametrického cenzorovaného probitového modelu pro ordinální proměnnou a alokace na základě minimální entropie). V parametrickém doplňku cenzurovaného probitového modelu pro ordinální proměnnou zvolíme jako vysvětlující proměnné pohlaví, věk a vzdělání. Následující kód vede k obrázku 7:

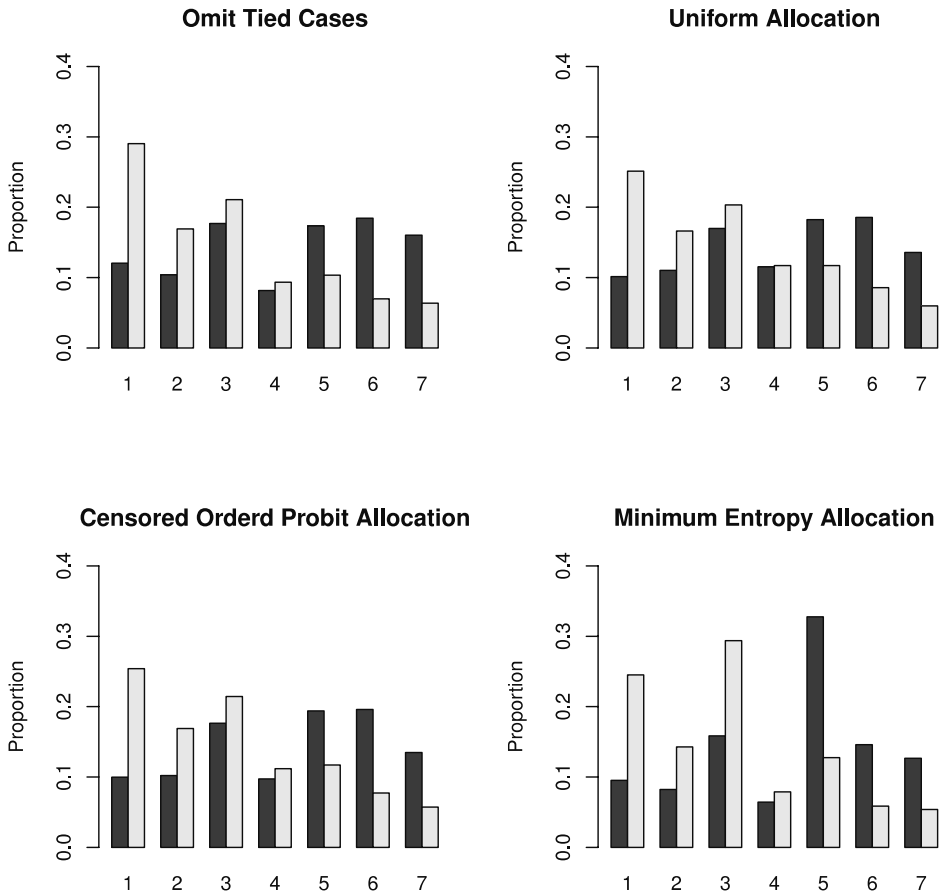
```

> fo<-
list(self=self~1,vign=cbind(vign2,vign3,vign6)~1,cpolr=~sex+age+educ)
> z3e<-anchors(fo,freedom,method="C",subset=country=="Eastasia")
> z3o<-anchors(fo,freedom,method="C",subset=country=="Oceania")

> pdf("freedomEO.pdf")
> par(mfrow=c(2,2))
> ylim<-c(0,0.4)
> barplot(z3e,z3o,ties="omit",ylim=ylim,xlab="",main="Omit Tied Cases")
> barplot(z3e,z3o,ties="uniform",ylim=ylim,xlab="",main="Uniform
Allocation")
> barplot(z3e,z3o,ties="cpolr",ylim=ylim,xlab="",main="Censored Orderd
Probit Allocation")
> barplot(z3e,z3o,ties="minentropy",ylim=ylim,xlab="",main="Minimum
Entropy Allocation")
> dev.off()

```

Na první pohled je zřejmé, že rozložení vycházející z prvních tří postupů (vynechání intervalových hodnot, rovnoměrné rozložení hodnot v intervalu a využití parametrického doplňku) vedou k podobným histogramům. Pouze rozložení s minimální entropií je odlišné. Je nejméně ploché, přičemž nejvíce ukazuje rozdíl mezi skupinami respondentů z Eastasie a Oceánie. Rozložení korigovaného sebehodnocení je pro respondenty z Eastasie sešikmené vpravo (respondenti se často hodnotí okolo úrovně poslední viněty, ta je v tomto příkladu viněta 6) a pro respondenty z Oceánie naopak sešikmené vlevo (respondenti se často hodnotí okolo úrovně první viněty, ta je v tomto příkladu označena jako viněta 2). Toto tedy poukazuje na horší úroveň politické svobody v Eastasii. Porovnání histogramů jsme provedli pouze na základě deskriptivních statistik. V literatuře další porovnání v kontextu metody ukotvujících vinět diskutována nejsou (připomeňme, že neparametrický přístup je využíván zřídka, většina publikovaných aplikací využívá parametrický model).



Obrázek 7. Ukázka zpracování dat v R: histogram korigovaného sebehodnocení pomocí neparametrického způsobu

Pozn. Pro respondenty z Eastasie je rozdělení označeno černou barvou, pro respondenty z Oceánie šedou. Větší hodnoty poukazují na horší úroveň politické svobody. Pro korekci jsou využity tři viněty, korigovaná škála je tedy sedmibodová.

Porovnání dvou distribucí však lze obecně učinit i různými statistickými testy jako je Kolmogorův-Smirnovův test s nulovou hypotézou, že sebehodnocení dvou skupin respondentů pocházejí ze stejného rozdělení pravděpodobností. Pomocí tohoto testu by tedy šlo porovnat rozdělení původního nekorigovaného sebehodnocení pro obě skupiny a pak i korigovaného sebehodnocení. Pro

nekorigované sebehodnocení respondentů z Eastasie a Oceánie můžeme vypočítat základní deskriptivní statistiky (minimum, první kvartil, medián, třetí kvartil, maximum a průměr) a zmiňovaný Kolmogorův-Smirnovův test následně:

```
> attach(freedom)
> se<-self[country=="Eastasia"] #Eastasia - nekorigovane sebehodnoceni
> so<-self[country=="Oceania"] #Oceania - nekorigovane sebehodnoceni
> summary(se)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  3.000   4.000   3.631  5.000   5.000

> summary(so)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000   3.000   2.954  4.000   5.000

> ks.test(se,so,alternative="less")
```

Two-sample Kolmogorov-Smirnov test

```
data: self[country == "Eastasia"] and self[country == "Oceania"]
D^- = 0.24548, p-value < 2.2e-16
alternative hypothesis: the CDF of x lies below that of y
```

Warning message:

```
In ks.test(self[country == "Eastasia"], self[country == "Oceania"], :
  p-value will be approximate in the presence of ties
```

Průměrné sebehodnocení respondentů z Eastasie je větší než u respondentů z Oceánie: 3,63 pro Eastasii a 2,95 pro Oceánii. Podobně první kvartil, medián i třetí kvartil jsou též větší pro respondenty z Eastasie (první kvartil = 3, medián = 4 a třetí kvartil = 5 pro respondenty z Eastasie versus první kvartil = 2, medián = 3 a třetí kvartil = 4 pro respondenty z Oceánie). Kolmogorův-Smirnovův test s alternativní hypotézou, že kumulativní distribuční funkce sebehodnocení respondentů z Eastasie leží pod kumulativní distribuční funkcí sebehodnocení respondentů z Oceánie, má p-hodnotu menší než $2,2 \times 10^{-6}$. Všechny tyto statistiky ukazují, že respondenti z Eastasie hodnotí svou politickou svobodu signifikantně hůře než respondenti z Oceánie (škála původního sebehodnocení je pětibodová od *zcela svobodný/á* po *vůbec ne svobodný/á*, a tudíž větší hodnota na škále znamená nižší politickou svobodu).

Pro korigované sebehodnocení provedeme porovnání distribucí s umístěním intervalových hodnot pomocí principu minimální entropie (nejhorší scénář, mezi respondenty nejméně odlišuje):

```

Ne<-sum(country=="Eastasia")
No<-sum(country=="Oceania")
kse<-rep(seq(1,7),times=summary(z3e)$minentropy*Ne) #Eastasia -
korigovane sebehodnoceni pomoci minimalni entropie
kso<-rep(seq(1,7),times=summary(z3o)$minentropy*No) #Oceanie -
korigovane sebehodnoceni pomoci minimalni entropie

> summary(kse)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   3.000   5.000   4.392   6.000   7.000

> summary(kso)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   2.000   3.000   3.084   4.000   7.000

> ks.test(kse,kso,alternative="less")

      Two-sample Kolmogorov-Smirnov test

data:  kse and kso
D^- = 0.36213, p-value < 2.2e-16
alternative hypothesis: the CDF of x lies below that of y

Warning message:
In ks.test(kse, kso, alternative = "less") :
  p-value will be approximate in the presence of ties

```

Pro korigované sebehodnocení dostáváme podobný výsledek jako pro původní sebehodnocení. Popisné statistiky i výsledek Kolmogorova-Smirnovova testu ukazují, že politická svoboda je horší v případě respondentů Eastasie. V literatuře jsou samozřejmě známé i případy, kdy korigované a nekorigované sebehodnocení jsou od sebe zcela odlišné (např. King a kol., 2004; Voňková a Hrabák, 2015b).

Porovnání distribucí před korekcí a po korekci se běžně v literatuře nečiní, je na ukázkou provedeno zde, aby byla čtenáři nastíněna cesta, jak korigovaná sebehodnocení u dvou skupin porovnávat. Neparametrický přístup je spíše zaměřen na získání popisných statistik pro řazení vinět, četnost nekonzistentního hodnocení vinět, výběru vinět apod. Pro zjištění statisticky významných rozdílů mezi skupinami (po očištění o vlivy dalších proměnných) se běžně užívá parametrický model CHOPIT.

4.2. Parametrický model

V softwaru R je též implementovaný parametrický CHOPIT model. V návaznosti na zpracování volně dostupných dat `freedom` neparametrickým

přístupem v R ukážeme odhad parametrů CHOPIT modelu také na těchto datech. Základní funkce, která je k odhadu parametrů modelu k dispozici, se nazývá `chopit` (tj. stejně jako `model`, ale s použitím malých písmen). Tato funkce je implementována v balíčku `anchors` stejně jako funkce k neparametrickému přístupu.

Specifikace vzorce zahrnuje části pro: a) sebehodnocení (ve vzorci označeno jako `self`) – potřebujeme uvést název proměnné pro sebehodnocení v datech (často si tuto proměnnou v datovém souboru výzkumníci nazývají `self`, stejně je tomu i v našem datovém souboru) a názvy pozorovaných charakteristik respondenta (jako například pohlaví `sex` a státní příslušnost `country`) pro vysvětlení *skutečné* úrovně měřeného konceptu u respondenta, b) viněty (ve vzorci označeno jako `vign`) – specifikujeme názvy proměnných reprezentujících odpovědi na jednotlivé viněty, c) meze v sebehodnocení a ve vinětách (ve vzorci označeno jako `tau`) – uvádíme názvy pozorovaných charakteristik respondenta, které zachycují (mohou zachytit) heterogenitu ve stylu užívání škály. Dle specifikace CHOPIT modelu v části 3.2 se tyto proměnné překrývají s proměnnými pro vysvětlení *skutečné* úrovně měřeného konceptu v sebehodnocení. Obecně je však možné parametry mezi odhadnout i pro nad/podmnožinu proměnných pro modelování *skutečné* úrovně. Příkazy pro odhad modelu se sebehodnocením `self`, vinětami `vign1`, ..., `vign6` a vysvětlujícími proměnnými pro skutečnou úroveň měřeného konceptu a mezi `sex`, `educ` a `country` jsou následující:

```
> fo<-  
list(self=self~sex+educ+factor(country), vign=cbind(vign1,vign2,vign3,vi  
gn4,vign5,vign6)~1,tau=~sex+educ+factor(country),use.linear=FALSE)  
> cout<-chopit(fo,data=freedom)
```

Funkce `summary()` je užitečná mimo jiné pro tisk odhadů parametrů a *log-likelihood*. Jejím výstupem je:

```
> summary(cout)  
ANCHORS: SUMMARY OF RELATIVE CHOPIT ANALYSIS:  
  
Model formula:  
$self  
self ~ sex + educ + factor(country)
```

```
$vign
cbind(vign1, vign2, vign3, vign4, vign5, vign6) ~ 1
```

```
$tau
~sex + educ + factor(country)
```

```
$use.linear
[1] FALSE
$cpolr
~1
<environment: 0x99656b8>
```

Coefficients:

	coeff	se
gamma.cut1.(Intercept)	-1.7123	0.0562
gamma.cut1.sex	0.0539	0.0228
gamma.cut1.educ	0.0137	0.0067
gamma.cut1.factor(country)Eurasia	0.0461	0.0504
gamma.cut1.factor(country)Oceania	-0.1371	0.0307
gamma.cut2.(Intercept)	0.7232	0.0283
gamma.cut2.sex	-0.0401	0.0205
gamma.cut2.educ	-0.0152	0.0060
gamma.cut2.factor(country)Eurasia	-0.0279	0.0450
gamma.cut2.factor(country)Oceania	0.0293	0.0273
gamma.cut3.(Intercept)	0.7034	0.0238
gamma.cut3.sex	-0.0222	0.0167
gamma.cut3.educ	0.0113	0.0051
gamma.cut3.factor(country)Eurasia	0.0275	0.0374
gamma.cut3.factor(country)Oceania	-0.0988	0.0217
gamma.cut4.(Intercept)	0.6199	0.0217
gamma.cut4.sex	0.0444	0.0159
gamma.cut4.educ	0.0166	0.0048
gamma.cut4.factor(country)Eurasia	0.0593	0.0364
gamma.cut4.factor(country)Oceania	0.0202	0.0210
sigma.random.effect	1.0000	NaN
sigma.self	1.0000	NaN
sigma.vign1	0.7978	0.0183
sigma.vign2	0.9998	0.0239
sigma.vign3	0.7554	0.0173
sigma.vign4	0.8348	0.0208
sigma.vign5	0.7271	0.0172
sigma.vign6	1.3321	0.0419
theta.vign1	-1.0128	0.0515
theta.vign2	-1.1309	0.0533
theta.vign3	-0.1710	0.0498
theta.vign4	0.2447	0.0514
theta.vign5	0.0239	0.0500
theta.vign6	1.0337	0.0656
beta.(Intercept)	0.0000	NaN
beta.sex	0.1396	0.0387
beta.educ	-0.0556	0.0116
beta.factor(country)Eurasia	0.4672	0.0896
beta.factor(country)Oceania	-0.7134	0.0513

```
-Log-likelihood of CHOPIT: 32597.57
Partition of CHOPIT -Log-likelihood by question:
```


	-LL	N
Self (self)	5178.345	3467
vign1	5067.674	3467
vign2	5238.978	3467
vign3	4794.419	3467
vign4	4362.268	3467
vign5	4509.207	3467
vign6	3446.680	3467

Number of cases that contribute at least partially to likelihoods:

- a) in self-responses: 3467
- b) in vign-responses: 3467

Odhadnuté parametry β ukazují po očištění o pohlaví a vzdělání na signifikantně odlišnou úroveň korigovaného sebehodnocení v Eastasii (referenční skupina), Eurasii a Oceánii. Největší hodnota β parametru je v Eurasii ($\text{coeff} = 0,4672$, $\text{se} = 0,0896$), dále následuje Eastasie (referenční, s hodnotou 0) a nakonec Oceánie ($\text{coeff} = -0,7134$, $\text{se} = 0,0513$). Vyšší hodnoty korigovaného sebehodnocení ukazují na horší úroveň politické svobody.

Koeficienty γ ukazují na odlišnosti v užívání škály mezi měřenými skupinami respondentů. Parametr $\gamma.\text{cut}1$ reprezentuje hranici mezi prvními dvěma body škály zcela svobodný/á a velmi svobodný/á. Vyšší hodnoty parametru znamenají, že respondenti označují určitou úroveň politické svobody spíše kategorií 1 (zcela svobodný/á) než kategorií 2 (velmi svobodný/á), což lze interpretovat tak, že jsou (na této části škály) optimističtější či že mají nižší standardy pro hodnocení. Nejvyšší hodnotu parametru $\gamma.\text{cut}1$ mají respondenti z Eurasie ($\text{coeff} = 0,0461$, $\text{se} = 0,0504$), dále následuje referenční Eastasie a nejnižší hodnotu parametru má Oceánie ($\text{coeff} = -0,1371$, $\text{se} = 0,0307$). Respondenti z Eurasie mají tedy při volbě mezi prvními dvěma kategoriemi nejnižší standardy pro hodnocení (jsou neoptimističtější), následují je respondenti z Eastasie, nejvyšší standardy pro hodnocení mají respondenti z Oceánie (jsou nejpesimističtější). Ostatní parametry $\gamma.\text{cut}$ reprezentují posun mezi od předcházející meze ($\tau_j = \tau_{j-1} + \exp(\gamma.\text{cut}j)$). Na základě hodnot parametrů $\gamma.\text{cut}j$ $j > 1$ lze říci, že buď jsou odlišnosti v posunu mezi při srovnání skupin respondentů nesignifikantně odlišné, nebo že má Eurasie nejvyšší posun a Oceánie nejnižší. Lze tedy konstatovat, že při hodnocení libovolné úrovně politické svobody volí

respondenti z Eurasie spíše nižší body škály (jsou optimističtí) ve srovnání s ostatními respondenty. Oproti tomu respondenti z Oceánie volí spíše vyšší body škály při hodnocení libovolné úrovně (jsou pesimističtí).

Další parametry se vztahují k vinětám, θ reprezentuje průměrné hodnocení viněty očištěné o odlišné užívání škály, σ je rozptylem θ . Signifikantně největší rozptyl hodnocení je zaznamenán u viněty 6 (1,3321, $se = 0,0419$), naopak nejmenší rozptyl je nalezen u viněty 5 (0,7271, $se = 0,0172$). Co se týká parametrů θ , nejnižší hodnotu má parametr $\theta.vign2$ (-1,1309, $se = 0,0533$), následovaný hodnotami $\theta.vign1$ (-1,0128, $se = 0,0515$), $\theta.vign3$ (-0,1710, $se = 0,0498$), $\theta.vign5$ (0,0239, $se = 0,0500$), $\theta.vign4$ (0,2447, $se = 0,0514$) a nakonec $\theta.vign6$ (1,0337, $se = 0,0656$). Nejnižší hodnota poukazuje na nejvyšší úroveň popsané politické svobody hypotetického jedince. Nejvyšší úroveň politické svobody je tedy dle respondentů po očištění o odlišné užívání škály popsána ve vinětě 2. Naopak nejnižší úroveň je popsána ve vinětě 6. Stejně jako při analýze pořadí vinět pomocí neparametrického přístupu zaznamenáváme přehození pořadí vinět 1 a 2 a také vinět 4 a 5.

Ve výstupu je též uvedena hodnota Log-likelihood, kterou lze používat například při srovnávání různých podmodelů CHOPIT modelu. Dále je uvedena i parcializace Log-likelihood dle jednotlivých otázek.

Parametrický model lze též odhadnout v softwaru STATA. Ukázka kódu pro jednoduchý případ s jedním sebehodnocením, jednou vinětou a jednou vysvětlující proměnnou pohlaví je uveden níže. Čtenář by tím měl získat základní orientaci pro zápis kódu CHOPIT modelu ve STATA (další informace k odhadu CHOPIT modelu ve STATA lze najít například v Rabe-Hesketh & Skrondal (2002)):

```
clear
insheet using "/home/data0.csv"
oprobit self_rec gender_male
esttab using "/home/table_probit.doc", replace
rename v1 id
rename self_rec xsay2
rename vign_rec xsay1
reshape long xsay, i(id) j(item)
drop if xsay==.
```

```

tab item, gen(i)
sort id item
*list id xsay i1-i2
rename i2 self
for var gender_male: gen s_X=self*X
gen vign=1-self
eq het: vign self
constraint def 1 [lns1]self=0
eq thresh: gender_male
gllamm xsay s_gender_male i1, i(id) link(soprobit) s(het) constr(1)
ethresh(thresh) init
outreg2 using "/home/table_chopit.doc", replace dec(3) ctitle("chopit")
label

```

Po odhadu modelu CHOPIT lze dále pracovat s odhadnutými parametry. Můžeme sestavit rozdělení korigovaného sebehodnocení na původní pětibodové škále, zobrazit porovnání odhadnutých mezí, porovnat odhadnuté parametry CHOPIT modelu s odhadnutými parametry probitového modelu pro ordinální proměnnou apod. Takovýto rozbor činíme v kapitole 6, kde je představena nová aplikace metody ukotvujících vinět pro porovnání učitelovy podpory žáků v zemích účastnících se PISA 2012.

5. Formulace vinět a jejich umístění v dotazníku

Přestože metoda ukotvujících vinět má již mimo pedagogiku četná využití, doporučení pro formulaci vinět až na výjimky neexistují. Několik rad lze například najít na webových stránkách Gary Kinga s často kladenými dotazy od výzkumníků zabývajících se metodou <http://gking.harvard.edu/vign/faq> (King, 2017). Vybraná metodologická doporučení jsou publikována v článku Voňková a kol. (2016). Níže uvedeme přehled těchto doporučení týkajících se základních předpokladů metody ukotvujících vinět (konzistence odpovědí a ekvivalence vinět), diskriminační síly vinět, počtu vinět a respondentů, pořadí (sebe)hodnocení a vinět a zakončíme několika poznámkami k obsahu vinět.

5.1. Předpoklady metody ukotvujících vinět

Metoda ukotvujících vinět má dva základní předpoklady – konzistenci odpovědí (*response consistency*) a ekvivalenci vinět (*vignette equivalence*). Předpoklad konzistence odpovědí je definován tak, že každý respondent používá škálu v (sebe)hodnoticích otázkách stejným způsobem jako při hodnocení vinět. Předpoklad ekvivalence vinět je definován tak, že všichni respondenti interpretují viněty stejným způsobem (King et al., 2004). Nedodržením předpokladů můžeme dostat nespolehlivé výsledky. V této části oba předpoklady definujeme, popíšeme základní myšlenky statistických testů těchto předpokladů a diskutujeme doporučení pro jejich dodržení již při samotné formulaci vinět.

5.1.1. Předpoklad konzistence odpovědí (*response consistency*)

Konzistence odpovědí je jedním z klíčových předpokladů metody ukotvujících vinět. Tento předpoklad říká, že respondent používá škálu v (sebe)hodnoticích otázkách stejným způsobem jako při hodnocení vinět. Například předpokládáme, že žák hodnotící vlastní chování a chování hypotetických žáků ve vinětě je k sobě i k žákům ve viněťách stejně přísný.

Studie, které se primárně zaměřují na testování předpokladu konzistence odpovědí, jsou Kapteyn a kol. (2011) a Au a Lorgelly (2014). Další studie Bago d'Uva a kol. (2011) a Peracchi a Rossetti (2013) představují testy obou základních předpokladů metody ukotvujících vinět, tj. konzistence odpovědí i ekvivalence vinět. Existují i studie, které se přímo nezaměřují na testování těchto dvou předpokladů, ale testují, zdali je korigované sebehodnocení v porovnání s nekorigovaným sebehodnocením více korelováno s nějakou objektivní mírou. Pokud by tomu tak nebylo, znamená to, že nějaký z předpokladů metody (např. zmíněné základní předpoklady, specifikace modelu apod.) je porušen. Mezi takové studie patří například Voňková a Hullegie (2011). Vedle uvedených klíčových předpokladů je důležité, aby specifikace modelu, který ve studii používáme, byla správná, odpovídala datům. Na testování specifikace parametrického modelu CHOPIT se zaměřuje například studie van Soest a Voňková (2014).

(Statistické) testování předpokladu

Zde si představíme myšlenky testů týkajících se předpokladu konzistence odpovědí. Výzkumníkům v pedagogice by mohly být blízké testy popsané v Kapteyn a kol. (2011) a Au a Lorgelly (2014). Myšlenky obou studií jsou přímočaré. Připojíme i stručný popis testů v Bago d'Uva a kol. (2011) a Peracchi a Rossetti (2013), které patří spíše do pokročilejších oblastí statistiky.

Replikační viněta (*replica vignette*)

Kapteyn a kol. (2011) popisují základní myšlenku testu následujícím způsobem: a) v první fázi dotazníkového šetření sbíráme data k sebehodnocení respondenta (např. hodnocení vlastních problémů s pohybem) a také informaci o respondentově objektivní situaci (např. respondent je schopen vyjít dvě patra, aniž by se zadýchal; tuto informaci můžeme získat pomocí toho, že respondentovi v dotazníku položíme otázku, kolik pater je schopen ujit), b) na základě informace o respondentově objektivní situaci sestrojíme vinětu, která reflektuje respondentovu objektivní situaci, tato speciální viněta byla v článku nazvána replikační viněta (*replica vignette*) (např. [Jan] je schopen ujit dvě patra schodů, aniž by se zadýchal... Jak byste hodnotili u [Jana] problémy s pohybem?), c) v druhé fázi dotazníkového šetření zjistíme respondentovo hodnocení replikující viněty. Pokud platí předpoklad konzistence vinět (škálu

užíváme stejným způsobem v sebehodnocení i vinětech), tak by neměl existovat systematický rozdíl mezi sebehodnocením respondenta a jeho hodnocením replikační viněty.

Autoři provedli dva výzkumy k testování předpokladu. Data pro první výzkum byla sebrána na vzorku respondentů American Life Panel v roce 2008, a to pro šest zdravotních domén (spánek, pohyb, dýchání, koncentrace, afekt a bolest). Druhý výzkum se uskutečnil v roce 2010 pro tři zdravotní domény (pohyb, dýchání a koncentrace). Předpoklad byl jednak testován jednoduchými neparametrickými testy, Wilcoxonovým pořadovým testem a znaménkovým testem. Nulová hypotéza pro Wilcoxonův pořadový test zní, že rozdělení rozdílu mezi sebehodnocením a hodnocením replikující viněty je symetrické kolem nuly. Nulová hypotéza pro znaménkový test zní, že medián rozdílu mezi sebehodnocením a hodnocením replikující viněty je roven nule. Další test vycházel ze specifikace parametrického modelu ukotvujících vinět. V podstatě bylo testováno, zdali parametry mezi v první části modelu k sebehodnocení jsou rovny parametrům v druhé části modelu k vinětě. K porovnání parametrů byl využit test poměru maximální věrohodnosti (*likelihood ratio test*).

Na základě deskriptivních statistik došli autoři k závěru, že předpoklad konzistence odpovědí funguje dobře pro domény spánek, pohyb a afekt (data k bolesti nebyla analyzována, u domén koncentrace a dýchání se předpoklad jevil jako zjevně nesplněný). Formální statistický test opírající se o specifikaci parametrického modelu nezamítl předpoklad konzistence odpovědí pouze pro spánek. Slabší test předpokladu konzistence odpovědí (parametry *první* meze v první části modelu k sebehodnocení jsou rovny parametrům *první* meze v druhé části modelu k vinětě) nebyl zamítnut pro žádnou testovanou doménu.

Kvalitativní analýza předpokladu

Další zmíněnou studií prezentující test konzistence odpovědí je Au a Lorgelly (2014). Jako první provádí kvalitativní analýzu předpokladu konzistence odpovědí. Kladou si otázku, jakým způsobem respondenti hodnotí viněty a co by mohlo vést k odlišnému způsobu hodnocení vinět a hodnocení sama sebe. Základ dotazníku tvořil instrument EQ-5D (Euroqol Group, 1990), ve kterém hodnotí respondenti vlastní zdraví v jednotlivých dimenzích. Pro pět z těchto dimenzí (problémy s pohybem, sebek péče, běžné aktivity, bolest / zdravotní

obtíže a úzkost/deprese) byly využity existující viněty, které byly například zařazeny do výzkumů ELSA (The English Longitudinal Study of Ageing), SHARE (Survey of Health, Ageing and Retirement). Pro každou z pěti dimenzí zdraví byly vybrány tři viněty popisující odlišnou úroveň zdravotních problémů (mírné, střední a závažné problémy). Pro dotazníkové šetření byly vytvořeny dvě verze dotazníků A a B. Ve verzi A bylo použito 15 vinět (5 dimenzí x 3 viněty) a ve verzi B byly konstruovány 3 multidimenzionální viněty, které zahrnují všech 5 dimenzí. Příklad viněty z verze A pro dimenzi problémů s pohybem je následující:

Vicky má velmi oteklé nohy. Na chůzi delší než 50 metrů vynakládá úsilí, cítí, že její nohy jsou těžké. Prosím zaškrtni jednu z následujících možností, která nejlépe popisuje Vickyiny problémy s pohybem: 1) nemá problémy s chůzí, 2) má mírné problémy s chůzí, 3) má středně závažné problémy s chůzí, 4) má zásadní problémy s chůzí, 5) je neschopná chůze.

Příklad viněty z verze B je následující:

William má velmi oteklé nohy. Na chůzi delší než 50 metrů vynakládá úsilí, cítí, že jeho nohy jsou těžké. Dokáže si umýt obličej a učesat vlasy, ale nedokáže se celý umýt bez pomoci. Potřebuje asistenci při oblékání přes hlavu, ale umí si obléci oděv patřící na dolní část těla. Žije sám a v posledním měsíci měl žaludeční problémy. V posledních třech týdnech neuklízěl dům, přestal vařit a potřebuje někoho, aby mu nakoupil. William má bolest v zádech, která vyvolává nepohodlí při změně polohy těla. Je neschopen stát nebo sedět více než půl hodiny. Léky trochu zmírňují jeho bolest, ale bolest nikdy nezmizí. Cítí napětí a v každé chvíli je na pokraji sil. Má deprese skoro každý den a cítí beznaděj. Má nízkou sebeúctu, je neschopen si užívat života a cítí, že se stal pro ostatní překážkou.

Za multidimenzionálními vinětami následovala otázka na zhodnocení celkového zdravotního stavu hypotetické osoby na pětibodové škále a dále i otázky na jednotlivé dimenze zdraví. Sebehodnoticí otázky následovaly v obou verzích vždy po vinětách, nejprve byla kladena otázka k celkovému zdravotnímu stavu respondenta a následně i jeho zdravotnímu stavu v jednotlivých dimenzích. První fáze výzkumu zahrnovala hloubkové polostrukturované rozhovory s respondenty, kteří vyplnili dotazník.

Cílem bylo pochopit, jak respondenti přemýšlejí při hodnocení vinět. Pro pochopení realizovatelnosti hodnocení vinět ze strany respondentů byli participanti tázáni, jak náročné pro ně bylo hodnocení vinět. Následně měli na škále 1–10 zhodnotit a) náročnost pochopení vinět a b) množství koncentrace, které úkol vyžadoval (*feasibility*). Následovaly další dotazy, které se již specificky zaměřovaly na hodnocení platnosti předpokladu konzistence odpovědí. Respondenti byli požádáni, aby popsali své myšlenky při hodnocení hypotetických osob ve vinětách. Pokud v jejich odpovědi nezaznělo, zda předpokládali, že hypotetická osoba je stejného věku, následoval na tento předpoklad dotaz (*same age*). Negativní odpověď znamenala dodatečné dotazování na to, zdali si respondenti vzpomínají na instrukci uvedenou před vinětami, kde byli žádáni, aby předpokládali, že hypotetická osoba má stejný věk a stejná další určení (pohlaví, rasa aj.). Respondenti byli též dotazováni, zdali si představovali sebe při hodnocení hypotetických individuí a porovnávali své zdraví se zdravím hypotetických individuí (*imagined self*) a zdali se cítili být ovlivněni příběhy z vinět při následném hodnocení vlastního zdraví (*influenced by vignettes*). Navíc participanti byli dotázáni, zda si myslí, že hodnotili své vlastní zdraví na stejné škále jako zdraví hypotetické osoby. Tato finální otázka dle autorů měla poskytnout informaci o tom, s jakou pravděpodobností platí předpoklad *response consistency*, zatímco předchozí dotazy směřovaly k získání informace ke stylům odpovídání na otázky v dotazníku a na důvody, proč by předpoklad *response consistency* měl či neměl být platný. Po dosažení saturace dat v hloubkových rozhovorech (kvalitativní část) byly v online průzkumu pilotovány dotazníkové otázky o stylu odpovídání na ukotvující viněty. Jejich formulace vycházela ze závěrů hloubkových rozhovorů a jejich finální podoba měla být využita v hlavním dotazníkovém šetření. Klíčová otázka na ověření platnosti *response consistency* v dotazníkovém šetření byla:

Do jaké míry souhlasíš s následujícím tvrzením? Svě zdraví jsem hodnotil/a na stejné škále (stejným způsobem) jako zdraví hypotetických osob.

1) rozhodně souhlasím, 2) spíše souhlasím, 3) ani souhlas, ani nesouhlas, 4) spíše nesouhlasím, 5) rozhodně nesouhlasím

Po první fázi s hloubkovými rozhovory následovala druhá fáze s dotazníkovým šetřením bez navazujících rozhovorů, dotazníkové šetření zahrnovalo otázky, které byly zformulovány na základě zjištění z první fáze.

K participaci na výzkumu byli vyzváni všichni dospělí ve věku 18 a více let, kteří plynule mluví anglicky. Výzva k participaci byla zveřejněna na stránkách Monash University v Austrálii. Studie se mohli zúčastnit studenti, akademická obec, administrativní a techničtí pracovníci i veřejnost. Celkově se studie zúčastnilo 105 participantů, 34 z nich se účastnilo hloubkových rozhovorů (saturace dat byla dosažena po 30 rozhovorech, tj. po 30 rozhovorech nepřinášel další sběr dat nové podstatné informace) a dalších 71 participantů se účastnilo online dotazníkového šetření bez rozhovorů. Většina respondentů byly ženy (70 %), věkový rozsah byl 18 až 72 let. Ve srovnání s australskou populací byli respondenti mladí, vzdělaní a etnicky různorodí (35 % nebylo narozeno v Austrálii), což je pravděpodobně výsledek popsané procedury získávání participantů do výzkumu.

Co se týká *feasibility*, respondenti udávali, že výzkum pro ně byl jednoduchý, rychlý a jasný. Tři respondenti ve věku 18–24 let však uvedli, že měli problém si představit hypotetickou osobu jejich věku s chronickou bolestí nebo problémy s chůzí. Jednomu respondentovi se obtížně hodnotilo zdraví hypotetické osoby v jednotlivých dimenzích bez znalosti dalších aspektů jejího zdraví. Toto vede k zamyšlení nad kompromisem mezi jednoduchostí a krátkou formulací viněty na jedné straně a nedostatkem znalosti kontextu na druhé straně. Na desetibodové škále 1 *velmi náročné na pochopení* až 10 *velmi snadné na pochopení* hodnotili respondenti obtížnost pochopení vinět z obou verzí dotazníku srovnatelně – z verze A (viněty k jednotlivým dimenzím zdraví) v průměru 8,3 a z verze B (viněty k obecnému stavu zdraví) v průměru 8,6. Verze B však byla pro respondenty náročnější z hlediska soustředění než verze A – verze B byla v průměru hodnocena 7,4, zatímco verze A 5,6 na desetibodové škále, kde 1 označuje *málo koncentrace* a 10 označuje *hodně koncentrace*. Respondenti se tedy museli na komplexní viněty popisující obecný stav zdraví mnohem více soustředit.

Co se týká samotného předpokladu *response consistency*, na otázku, zdali vlastní zdraví hodnotili na stejné škále jako zdraví hypotetické osoby, prokázali respondenti s verzí B mnohem větší konzistenci (49 % s uvedeným tvrzením rozhodně souhlasilo) než respondenti verze A (29 % rozhodně souhlasilo). Předpoklad konzistentních odpovědí tedy fungoval lépe v případě multidimenzionálních vinět z verze B. Nesouhlasně (kategorie *rozhodně nesouhlasím* a *nesouhlasím*) odpovědělo 15 % respondentů s verzí B a 19 %

respondentů s verzí A. Důvodem bylo například to, že se respondenti snaží být ve svém případě optimističtí a nestěžovat si.

Pouze 37 % respondentů předpokládalo, že hypotetická osoba ve vinětě je stejného věku (rozdíly mezi verzemi nebyly signifikantní), přestože jim bylo v úvodu před hodnocením viněť řečeno, že tento předpoklad mají učinit (*same age*). Respondenti si sice úvod k viněťám přečetli, pak ho ale při hodnocení viněť zapoměli zohlednit. Z toho tedy vyplývá doporučení připomínat respondentům tuto instrukci častěji. Téměř polovina respondentů (49 %) si při hodnocení viněty představovala sebe v situaci popsané ve vinětě (*imagined self*). Důvodem bylo dle vyjádření respondentů například to, že je snadnější hodnotit něco, co je či není zlé, když by se to přihodilo jim samotným. Na druhou stranu také téměř polovina (47 %) respondentů si sebe při hodnocení hypotetické osoby ve vinětě nepředstavovala. Respondenti uváděli, že měli na mysli náhodně vybranou osobu popsanou ve vinětě. Polovina respondentů uvedla, že při hodnocení vlastního zdraví byla ovlivněna přečtením viněť (*influenced by vignettes*), přičemž tento podíl byl signifikantně větší pro verzi B než verzi A.

V logistické regresi byla vysvětlována proměnná dodržení předpokladu konzistence odpovědí (1 = rozhodně souhlasím s tím, že jsem hodnotil/a zdraví hypotetické osoby stejným způsobem jako své zdraví, 0 = ostatní případy). Vysvětlující proměnné byly charakteristiky respondentů: věk, pohlaví, zaměstnanecký status, rodinný status, vzdělání, místo narození, objektivní míra zdraví měřená VAS skórem a EQ-5D indexem (Euroqol Group, 2012) a proměnné *same age*, *imagined self* a *influenced by vignettes*. Z hlediska metodologického výrazně podporuje předpoklad konzistence odpovědí komplexnost viněť užitých ve verzi B a představa sebe samého při hodnocení hypotetických osob ve viněťách. Tento výsledek též poukazuje na to, že představa sebe samého při hodnocení hypotetických osob se zdá být důležitější pro dodržení předpokladu konzistence odpovědí než představa stejného věku u hypotetických osob jako u sebe samého. Signifikantními prediktory pro předpoklad konzistence odpovědí byla dotazníková verze B.

Test s využitím objektivní proměnné

Au a Lorgelly (2014) tímto představili postup, jak by mohl být předpoklad konzistence odpovědí testován s využitím kvalitativní metodologie. Bago d'Uva

a kol. (2011) zavádí statistický test obou předpokladů, konzistence vinět a ekvivalence vinět. Vychází ze specifikace parametrického modelu CHOPIT (viz část 3.2). Pro provedení testu předpokladu konzistence odpovědí musí být splněny dvě podmínky: a) β v části modelu pro sebehodnocení je identifikováno pomocí externí objektivní míry (podobný předpoklad je i ve van Soest et al., 2011), b) předpoklad ekvivalence vinět není porušen. Samotný test se pak provádí pomocí porovnání parametrů mezi v části modelu pro sebehodnocení a parametrů mezi v části modelu pro hodnocení vinět. V případě platnosti předpokladu konzistence vinět je rozdíl mezi těmito mezemi nulový. Co se týká testu ekvivalence vinět, musí být splněny následující dvě podmínky: a) k dispozici jsou hodnocení minimálně dvou vinět, b) užívání škály při hodnocení různých vinět je stejné (oproti tomu Voňková & Hullegie (2011) ukazují, že užívání škály při hodnocení různých vinět v rámci jedné domény je pro některé zdravotní domény odlišné). Test ekvivalence vinět se pak provede tak, že se přidají regresory do rovnice latentní proměnné Y_v^* a testuje se, zdali mají nulový efekt.

Společný test obou předpokladů

Statistické testování obou předpokladů představuje i studie Peracchi a Rossetti (2013). Zavádí společný test obou předpokladů vycházející z faktu, že parametrický model je přeidentifikován při předpokladu platnosti obou předpokladů. Vyčísľují počet restrikcí modelu, které vedou k přeidentifikování modelu. V případě jedné viněty se jedná o jednu takovou restrikcí, v případě dvou vinět o tři takovéto restrikcí. Základním empirickým výsledkem studie je, že pokud se test provede pro jednotlivé skupiny respondentů, zamítnutí platnosti takovýchto restrikcí je méně pravděpodobné. Nevýhoda testu je, že může vést k zamítnutí kvůli specifikaci modelu nebo proto, že jsou v modelu vynechány relevantní proměnné.

Doporučení pro dodržení předpokladu při formulaci vinět

Co se týká doporučení pro dodržení tohoto předpokladu při formulaci vinět v pedagogickém výzkumu, můžeme odkázat na studii Voňková a kol. (2017), která se zaměřuje na měření kázně žáků pomocí jejich sebehodnocení a hodnocení ukotvujících vinět. V českém jazyce je jejich doporučení shrnuto v již zmiňované studii Voňková a kol. (2016). Předpoklad byl podpořen nejen

při formulaci vinět, ale také při samotném sběru dat. Žáci účastníci se výzkumu byli požádáni, aby hodnotili sami sebe stejně přísně jako žáky popsané ve vinětách, což je v podstatě samotný předpoklad. Dále dotazník použitý v tomto výzkumu byl anonymní a studentům bylo při sběru dat zdůrazněno, že data sebraná v tomto výzkumu budou využita pouze pro výzkumné účely a nebudou tudíž ukázána ani jejich učitelům ani řediteli. Třídní učitel byl při sběru dat přítomen (vyplňoval učitelský dotazník, kde hodnotil chování jednotlivých žáků i ukotvující viněty), ale seděl v první lavici a byl navíc požádán, aby během sběru dat nechodil třídou a nenahlížel do odpovědí žáků. Toto mělo redukovat následující dva efekty. Zprvé, aby studenti nehodnotili sami sebe o mnoho lépe, než jak tomu ve skutečnosti je, a nechtěli učiteli ukázat, jak je jejich chování vzorné. Zadruhé, aby studenti nehodnotili hypotetické žáky ve vinětách hůře, než jak si myslí, že se skutečně chovají, tj. aby neudávali sociálně žádoucí odpověď (Fisher, 1993; Wei & Huang, 2005), že popsané chování ve vinětách je pro ně nepřijatelné a nemělo by být akceptováno. Jinými slovy, mělo být zabráněno tomu, aby žáci vlastní chování v tomto výzkumu nadhodnocovali a naopak chování žáků ve vinětách podhodnocovali. Obecně může docházet i k opačnému efektu, tj. že respondenti sami sebe podhodnocují a naopak viněty nadhodnocují. King a kol. (2004) uvádí, že k porušení *response consistency* může dojít například v situaci, kdy se respondenti nepovažují za až tak dobré či důležité jako osoby popsané ve vinětách. Měření různých konceptů může přinášet odlišné tendence respondentů a je tudíž nutné zvážit specifika daného konceptu.

Při testování předpokladu konzistence odpovědí můžeme v praxi postupovat iterativním způsobem: provedeme dotazníkové šetření se sebehodnotící otázkou a vinětami, sebraná data podrobíme testu s nulovou hypotézou, že předpoklad *response consistency* je dodržen. Je-li hypotéza zamítnuta, viněty či sebehodnotící otázku upravíme či přeformulujeme. Provedeme další dotazníkové šetření s nově upravenými či přeformulovanými vinětami a opět testujeme, zdali není předpoklad *response consistency* zamítnut. Opakujeme do té doby, než nulová hypotéza není zamítnuta. Tento postup je sice časově náročnější, nicméně nás může dovést k žádoucímu výsledku.

5.1.2. Předpoklad ekvivalence vinět (*vignette equivalence*)

Ekvivalence vinět je druhý z klíčových předpokladů metody ukotvujících vinět. Tento předpoklad je definován tak, že všichni respondenti interpretují viněty stejným způsobem (King et al., 2004). Stejně jako pro předpoklad konzistence odpovědí platí, že pokud není předpoklad dodržen, dochází k chybné korekci sebehodnocení.

Statistické testování předpokladu

Klíčové studie Bago d'Uva a kol. (2011) a Peracchi a Rossetti (2013) pro testování tohoto předpokladu jsou představeny v předchozí části o předpokladu konzistence odpovědí. V praxi lze opět postupovat tak, že formulujeme viněty a testujeme, zdali je předpoklad splněn. Pokud tomu tak není, vinětu přeformulujeme a opět provedeme testování.

Doporučení pro dodržení předpokladu při formulaci vinět

Pro doporučení dodržení předpokladu ekvivalence vinět můžeme stejně jako pro předpoklad konzistence odpovědí odkázat na studii Voňková a kol. (2017), v českém jazyce jsou tato doporučení shrnuta ve Voňková a kol. (2016). Studenti účastníci se výzkumu měření nekázně s využitím metody ukotvujících vinět byli požádáni, aby si pozorně přečetli chování studenta popsáno ve vinětě a hodnotili toto chování pouze a jen na základě toho, co přečetli, a nedoplňovali do viněty žádnou další informaci. Délka viněty byla přibližně na tři řádky (výzkumu se účastnili žáci 7. třídy ZŠ). Ve vinětě týkající se podvádění a záškoláctví byl popsán kontext (ke kázeňskému přestupku došlo v situaci, kdy třída měla suplovanou dvouhodinovou tělesnou výchovu) i frekvence uvedeného jevu (k záškoláctví a následnému napsání omluvenky došlo dvakrát za poslední měsíc).

Pokud formulujeme vinětu k multidimenzionálnímu konceptu a v popisu viněty vynecháme jednu či více dimenzí, pak při aplikaci metody musíme předpokládat, že užívání škály je ve vynechaných dimenzích srovnatelné s užíváním škály v dimenzích popsaných ve vinětě. K tomuto je ovšem zapotřebí důkladného odůvodnění.

Grol-Prokopczyk a kol. (2015) se zamýšlí nad tím, jaké důvody mohou zapříčinit porušení předpokladu ekvivalence vinět. V případě užití vinět

v mezinárodním srovnání je nutné pečlivě přistoupit k překladu vinět do různých jazyků. Gramaticky korektní překlad zdrojového textu ještě negarantuje ekvivalentní pochopení konceptu v různých zemích či skupinách. Autoři se odvolávají na schéma ve studii Pan a Fond (2014), kde zdroje kvalitního překladu vycházejí nejen z 1) lingvistických pravidel, ale také 2) kulturních norem a 3) běžné sociální praxe. Pro dodržení těchto požadavků je někdy nutné odchýlit se od pořadí slov či dokonce obsahu zdrojového textu. Užívané pojmy mohou navíc mít v odlišných kulturách odlišný pojmový obsah. Popis bolesti způsobené nadměrným užíváním počítače může mít jiný význam v technologicky vyspělé ekonomice ve srovnání s ekonomikou, kde jsou počítače užívané zřídka. Popis kvality zraku s odkazem na čtení například novin může vyvolat jiné interpretace v zemích s dramaticky odlišnou úrovní gramotnosti. Grol-Prokopczyk a kol. (2015) dále uvádí, že ti, kteří navrhují nové viněty, jsou v potenciálně výborné pozici, neboť se již mohou vyvarovat známých nešvarů při formulaci vinět. Viněty by měly být univerzálně pochopitelné. Měly by být validizovány jednotlivě.⁵ Dalším faktem je, že výzkum ukotvujících vinět prováděli doposud především kvantitativně založení výzkumníci a nabízí se tudíž možnost více spolupracovat s experty v oblasti překladu a se znalostí lokálních kultur. Překonání problémů s překladem založeným na rozhovorech s respondenty může pomoci překonat tento problém. To nicméně bude souviset i se zásadními výdaji na další výzkum. Výzkumníci, kteří provádějí sekundární analýzy, by měli provádět testy ekvivalence vinět a konzistence odpovědí, které jsou dosud publikované v literatuře.

Voňková a kol. (2016) shrnují diskuzi ze studie Bonsang a van Soest (2012) o tom, kdy může dojít k porušení tohoto předpokladu. Uvádí, že se jedná o situace, kdy formulace vinět je nejednoznačná. Respondenti s odlišnými charakteristikami si pak mohou chybějící informaci doplnit dle vlastního uvážení či zkušeností, což pak může vést k systematickým rozdílům ve způsobu interpretace viněty. K porušení tohoto předpokladu může dle Bonsang a van Soest (2012) dojít i v případě, kdy měřený koncept je multidimenzionální

⁵ Pro mnoho pojmů existuje více vinět, které byly v analýzách pro identifikaci škály užívány najednou, výjimku tvoří například Voňková a Hullegie (2011), kde byly viněty pro tři zdravotní problémy s pohybem, dýchavičností a soustředěním na věci validizovány jednotlivě a očištěná sebehodnocení pomocí jednotlivých vinět byla srovnávána s objektivní mírou daného zdravotního problému.

a formulace viněty nepokryje komplexnost konceptu, což pak může vést k nežádoucímu doplnění informace jednotlivými respondenty dle jejich vlastního uvážení. Ve vinětě bychom sice mohli multidimenzionální koncept popsat podrobněji, nicméně je nutné mít na paměti, že délka viněty musí být pro respondenty únosná. V pedagogickém výzkumu je zachování rozumné délky vinět zásadní a musí být přizpůsobeno věkové kategorii dětských respondentů.

5.2. Diskriminační síla, počet a pořadí vinět

5.2.1. Diskriminační síla vinět

Voňková a kol. (2016) k diskriminační síle vinět uvádí: „Zeptáme-li se náhodného výběru žáků na základní škole, zdali umí integrovat a derivovat, nepřinese to mnoho informací o jejich znalosti matematiky. Podobně je tomu u formulace vinět. Úroveň měřeného konceptu ve vinětě by měla odpovídat úrovni v populaci. Představme si, že budeme měřit nějaký koncept na desetibodové škále, a že většina sebehodnocení se bude pohybovat v rozmezí jedna až pět a zároveň většina hodnocení viněty se bude pohybovat v rozmezí šest až deset. V takovémto případě nejsme schopni pomocí hodnocení vinět očistit užívání škály v sebehodnocení, většina sebehodnocení je pod úrovní hodnocení vinět.“

Je sice samo o sobě zajímavé sledovat variabilitu hodnocení viněty i na opačném konci škály (různí respondenti hodnotí tentýž příběh různým způsobem), nicméně pro účely korekce sebehodnocení pomocí metody ukotvujících vinět je tato informace irelevantní, ani jeden z přístupů, parametrický model a neparametrický přístup, ji nevyužije. Museli bychom využít předpokladu, že respondenti užívají kategorie na jednom konci škály stejným způsobem jako na druhém konci škály (např. že libovolnou úroveň konceptu hodnotí určitý respondent vždy vyššími škálovými kategoriemi a jiný respondent vždy nižšími škálovými kategoriemi). Tento předpoklad by však byl minimálně diskutabilní a vyžadoval by například další empirické ověření a opodstatnění založené na teorii vztahující se k měřenému konceptu.

Pro zjištění diskriminační síly vinět je vhodné provést předvýzkum v cílové populaci. Na základě získaných dat sestavíme dva histogramy – histogram pro sebehodnocení a histogram s hodnocením vinět. Tyto histogramy porovnáme. Pokud na první pohled vidíme, že rozdělení jsou zcela odlišná (například sebehodnocení je v rozmezí horního bodu škály až prostředního bodu a hodnocení vinět je v rozmezí prostředního bodu škály až dolního bodu), lze vyvodit, že viněty mají mizivou diskriminační sílu (Voňková a kol., 2016). Mimo prvního vhledu na základě porovnání histogramů existuje i formální postup výpočtu, který popisujeme v části 4.1.3.

5.2.2. Počet vinět a počet respondentů

Volba počtu vinět souvisí především s heterogenitou populace, na kterou se ve výzkumu zaměřujeme. Čím je populace heterogennější, tím více vinět bychom měli formulovat (Voňková et al., 2016; King, 2017). Zároveň však nezáleží jen na počtu vinět, ale také na jejich diskriminační síle. Při volbě počtu vinět lze přemýšlet podobně jako při volbě počtu respondentů ve výzkumných šetřeních. Nezáleží jen na počtu respondentů (analogicky nezáleží jen na počtu vinět), ale také na tom, jaké respondenty vybíráme, resp. jakým způsobem činíme výběr, aby byl náš vzorek reprezentativní (analogicky jakou diskriminační sílu mají viněty). A i kdyby viněty měly dobrou diskriminační sílu, neplatí čím více, tím lépe. Příliš dlouhý dotazník a následně zahlcený respondent je obecně problém. King a Wand (2007) doporučují zařadit do předvýzkumu větší počet vinět, následně ověřit jejich diskriminační sílu a vybrat ty, které nám přináší nejvíc informace. Typicky se sestavují k jednomu měřenému konceptu tři viněty nebo jedna viněta.

Pokud používáme neparametrický přístup, musíme mít k dispozici hodnocení vinět od všech respondentů v našem výběru. U každého respondenta musíme totiž zjistit relativní pozici jeho sebehodnocení vůči jeho hodnocení vinět. Pokud používáme parametrický CHOPIT model, stačí mít k dispozici hodnocení vinět pouze od náhodného výběru respondentů v našem výběru (King, 2017). Tedy ne všichni respondenti musí hodnotit i viněty. Pokud jsme v této situaci, je důležité, abychom pomocí hodnocení vinět v našem podvýběru byli schopni identifikovat parametry v první mezi (tj. hranici mezi první a druhou kategorií, v CHOPIT modelu označené jako τ^1). Je dokonce i možné

identifikovat heterogenitu užívání škály pomocí hodnocení vinět sebraných v jednom vzorku a to následně využít pro korekci odlišného užívání škály v jiném vzorku. Toto lze učinit za předpokladu, že heterogenita užívání škály je v obou vzorcích totožná. Příkladem by mohlo být užití dat z výzkumu PISA 2012 pro korekci heterogenity užívání škály dat získaných z výzkumů v jednotlivých zemích účastnících se PISA. V PISA 2012 byla sbírána data k hodnocení vinět týkajících se učitelova řízení třídy a učitelovy podpory žáků. V případě, že nějaká země sebere v rámci svého vlastního výzkumu data od žáků k hodnocení těchto konceptů a nesebere data k hodnocení vinět, může využít data k hodnocení vinět získaných v PISA výzkumu. Zde je v podstatě popsána myšlenka, jak velmi flexibilně a úsporně užívat data hodnocení vinět.

Další úvahy můžeme rozvíjet pro počet hodnocených vinět. Pokud bychom využili myšlenku adaptivního testování, mohli bychom se respondentů ptát jen na podvýběr našich vinět (King, 2017). Tedy každý respondent by mohl dostat k hodnocení pouze omezený počet vinět. K využití tohoto přístupu bychom museli mít k dispozici větší množství vinět a předpokládat jejich pořadí (to by šlo získat například v předvýzkumu). Respondenta bychom nechali zhodnotit například středovou vinětu. Pokud by jeho sebehodnocení bylo vyšší, zvolili bychom v dalším kroku k zhodnocení vinětu, která by odpovídala lépe úrovni tohoto sebehodnocení. Relativně brzy bychom pak měli získat hodnocení viněty, která odpovídá sebehodnocení. Myšlenka adaptivního testování při sběru dat k hodnocení vinět by však vyžadovala vyvinutí nového statistického aparátu pro analýzu takovýchto dat. Při užití parametrického modelu nemusíme mít hodnocení všech vinět u respondenta. Je možné, aby respondent hodnotil pouze náhodný výběr vinět. V případě, že se rozhodneme nechat různé respondenty hodnotit pouze různé podvýběry vinět, musíme si být jisti, že zajistíme identifikaci všech parametrů v modelu. Není například vhodné nechat systematicky hodnotit první dvě viněty ženy a další dvě viněty muže. Pokud by sebehodnocení mužů bylo bližší hodnocení prvních dvou vinět, nemuselo by být možné očistit jejich sebehodnocení a technicky bychom se mohli dostat do problému identifikace parametrů hranic mezi jednotlivými kategoriemi (parametry γ).

5.2.3. Pořadí (sebe)hodnocení a vinět

Při sestavování dotazníku je otázkou, v jakém pořadí by měla být sebehodnoticí otázka vůči ukotvujícím vinětám. Jinými slovy, zdali by měla být sebehodnoticí otázka před či za ukotvujícími vinětami. Zde lze rozvinout několik úvah. Sebehodnoticí otázka před ukotvujícími vinětami má několik výhod. Respondent není ovlivněn při svém sebehodnocení vinětami, odpovídá tak, jak by běžně odpovídal. Doporučení klást sebehodnoticí otázku před vinětami uvádí na svých webových stránkách Gary King (King, 2017).

V článku Hopkins a King (2010) se však spíše doporučuje opačné pořadí. Klademe-li viněty před sebehodnoticí otázku, umožní to respondentům ukotvit své odpovědi na sebehodnoticí otázku a pomůže jim to ujasnit si daný koncept. Toto však spatřujeme jako limitující pro samotnou formulaci sebehodnoticích otázek. Pokud by bylo vhodné dodat respondentům k sebehodnoticí otázce další informaci pomocí vinět, je otázkou, zda je formulace samotné sebehodnoticí otázky dostatečná a nebylo by například vhodné uvést ještě nějakou doplňující informaci v sebehodnoticí otázce. Při kladení sebehodnoticích otázek za viněty by se pak do budoucna muselo počítat i s tím, že vždy bude nutné nechávat respondenta hodnotit viněty společně se sebehodnoticí otázkou a sebehodnoticí otázka by pak nemohla sama o sobě být kladena. Přenositelnost identifikace heterogenity užívání škály z jednoho vzorku se sebehodnocením i vinětami do druhého, kde je jen sebehodnocení, by pak nebyla možná.

Buckley (2008) empiricky ukazuje, že na pořadí sebehodnocení vůči hodnocení vinět záleží a doporučuje randomizaci tohoto pořadí, tj. náhodně vybraní respondenti uvádí odpovědi nejprve na sebehodnoticí otázku, zbytek respondentů hodnotí nejprve viněty. Tímto lze mimo jiné zjistit, zdali v daném zkoumaném případě na pořadí záleží, a jaký způsob a jaké pořadí vede k validním výsledkům (např. lze si klást otázku, zda je korelace s externími proměnnými, u nichž se předpokládá těsný vztah k měřenému konceptu, vyšší, pokud je sebehodnocení před vinětami).

Dále je velmi vhodné, aby viněty nebyly v dotazníku za sebou kladeny v jejich přirozeném pořadí. Pokud respondent špatně ukotví svou první odpověď a pak si uvědomuje, že jsou viněty uvedeny v určitém pořadí a chce dostát tomuto pořadí, začne tím chybně odpovídat i na zbylé viněty. Je velmi pravděpodobné, že pokud v dotazníku uvedeme viněty v jejich přirozeném pořadí, dostaneme

méně smyček (viněty hodnotí respondent stejně) či nekonzistentních hodnocení vinět (viněty hodnotí respondent v opačném pořadí, než je jejich přirozené pořadí), což může dát výzkumníkovi falešnou představu o tom, že viněty jsou v tomto smyslu kvalitní a že pořadí jím zamýšlené s ním sdílí i respondenti. Dobře zformulované viněty kladené v dotazníku v náhodném pořadí by u respondentů, kteří své odpovědi neodbývají a dobře se zamýšlí nad otázkami, měly vést k minimu přesmyček a nekonzistentních hodnocení. Samozřejmě zde hraje roli i náhodná chyba, která může být příčinou takovýchto hodnocení. Nicméně tu lze zachytit pomocí parametrického CHOPIT modelu.

5.3. K obsahu vinět

Při využití vinět pro mezinárodní porovnávání je nutné mimo kvalitního překladu zajistit i kulturně univerzální obsah. To souvisí s předpokladem ekvivalence vinět. V případě, že bude nějaké skupině respondentů obsah vzdálen, neporozumí mu, bude interpretovat viněty odlišným způsobem (jakoby každý četl jiný příběh), dojde k porušení ekvivalence vinět.

Viněty musí být relevantní pro cílovou skupinu. Je vhodné volit takový obsah, který je cílové skupině blízký. Ve vinětách bychom neměli, pokud je to možné, užívat hodnotící adjektiva, neboť ta by pak mohla ovlivňovat respondenty při jejich hodnocení. Obecně se na formulování vinět vztahují stejná pravidla jako na formulování běžných dotazníkových položek. Jedná se například o vyhnutí se sugestivnímu obsahu či důrazu na jasnost a srozumitelnost.

Voňková a kol. (2016) v diskuzi k formulaci vinět uvádějí, že při formulaci vinět je nutné mít na paměti jejich primární cíl – identifikace heterogenity užívání škály v (sebe)hodnoticích otázkách. Ta je následně využita pro korekci sebehodnocení, ale to už je technická záležitost (aplikujeme neparametrický či parametrický přístup a sebehodnocení korigujeme). Z obecného pohledu lze konstatovat, že viněta nemusí pokrývat celý koncept, nemusí být v tomto smyslu univerzální. Viněta může pokrývat jen určitou část konceptu, např. jednu dimenzi v případě multidimenzionálních konceptů. Pokud se díky krátké vinětě podaří identifikovat heterogenitu užívání škály při dodržení všech v předchozích částech diskutovaných předpokladů (konzistence odpovědí, ekvivalence vinět a diskriminační síla vinět), jedná se o velmi žádoucí stav. Dodržení všech uvedených předpokladů není však vždy snadné, jedná se

v podstatě o limitaci metody při její aplikaci na měření určitých jevů. Vše vyžaduje obezřetnost při formulování vinět a sběru dat. Statistické testování předpokladů uvedené v literatuře pomáhá zjistit, zdali jsme předpoklady splnili a dosáhli tak žádoucího stavu. Z pohledu univerzality je vhodné se zamýšlet nad tím, zda respondent užívá škálu stejným způsobem při hodnocení různých konceptů.

6. Nová aplikace parametrického modelu: porovnání učitelovy podpory žáků v zemích z výzkumu PISA 2012

6.1. Cíle výzkumu

V této části je uvedena nová aplikace parametrického modelu metody ukotvujících vinět v pedagogickém výzkumu založená na sekundární analýze dat z výzkumu PISA 2012. Rozšiřuje tím základní deskriptivní analýzu uvedenou v Kyllonen a Bertling (2014a). Zaměřuje se na analýzu heterogenity ve stylu odpovídání u žáků v různých zemích a její dopady na žákovské hodnocení učitelovy podpory (*teacher support behavior*). Hlavní cíle této analýzy jsou následující:

1. analýza heterogenity ve stylu odpovídání žáků na dotazníkové položky týkající se učitelovy podpory žáků na úrovni zemí,
2. korekce žákovského hodnocení učitelovy podpory o heterogenitu ve stylu hodnocení na úrovni zemí,
3. porovnání korigovaného a nekorigovaného žákovského hodnocení učitelovy podpory na úrovni zemí,
4. analýza vztahu (ne)korigovaného žákovského hodnocení učitelovy podpory s výsledkem žáků v testu gramotnosti na úrovni zemí.

Prezentace výsledků bude následující. Nejprve popíšeme data a uvedeme specifikaci parametrického modelu CHOPIT pro tuto analýzu dat. Dále budeme analyzovat heterogenitu ve stylu odpovídání na položky na základě žákovského hodnocení ukotvujících vinět týkajících se učitelovy podpory. Rozlišíme země, v nichž mají žáci vyšší standardy pro hodnocení podpory ze strany jejich učitele, tj. užívají nižší („pesimističtější“) škálové kategorie pro hodnocení určitého stavu učitelovy podpory ve srovnání s ostatními (například pojmenují určitý stav učitelovy podpory jako nedostatečný, zatímco ostatní ho pojmenují jako výborný). Analogicky rozlišíme i země s nižšími standardy pro hodnocení. Dále

uvedeme země s tendencí volby extrémních kategorií a naopak středových kategorií na základě hodnocení vinět. Po analýze stylu odpovídání na položky se zaměříme na dopady tohoto stylu na žákovské hodnocení podpory ze strany jejich učitele. Porovnáme korigované a nekorigované hodnocení žáků na úrovni zemí a zjistíme, v jakých zemích došlo k zásadní či naopak minimální změně po provedení korekce ve stylu odpovídání na položky. Následně budeme analyzovat korelaci mezi (ne)korigovaným hodnocením a externí proměnnou, výsledkem v testu matematické gramotnosti, což lze považovat za indikátor externí validity. V poslední části porovnáme získané výsledky pro žákovské hodnocení učitelovy podpory s výsledky žákovského hodnocení učitelova řízení třídy (*teacher classroom management*) ve studii Voňková a kol. (2015).

6.2. Data z výzkumu PISA 2012

V této analýze jsou využita data z výzkumu PISA 2012. Výzkum PISA patří obecně k významným zdrojům informací o vzdělávacích systémech v různých zemích. Porovnává nejen výsledky žáků v testech čtenářské, matematické a přírodovědné gramotnosti, ale poskytuje i údaje ze žákovských, rodičovských a školních dotazníků. O mezinárodních srovnávacích výzkumech vyšla v České republice celá řada publikací (např. Kaščák & Pupala, 2011; Münich & Federičová, 2015; Potužníková, Lokajíčková, & Janík, 2014; Straková, 2016; Štech, 2015). Straková (2016) ve své přehledové monografii shrnuje metodologii, přínosy, rizika a příležitosti mezinárodních srovnávacích výzkumů v obecné rovině.

6.2.1. Důležitost zařazení ukotvujících vinět do žákovského dotazníku PISA 2012

Z našeho pohledu je klíčové, že v roce 2012 byly do žákovského dotazníku výzkumu PISA (OECD, 2017b) zařazeny též ukotvující viněty. Lze to vnímat jako jeden ze zásadních metodologických počínů skupiny výzkumníků připravujících dotazníkové šetření PISA, kteří zaregistrovali potenciál metody ukotvujících vinět při porovnávání zemí či jiných skupin respondentů mimo pedagogický výzkum.

Kyllonen a Bertling (2014a) například uvádějí, že v mezinárodních výzkumech jako PISA existuje celá řada jevů měřená pomocí dotazníkových položek s posuzovacími škálami, které jsou s výsledky žáků v testech sice pozitivně korelovány na úrovni jednotlivých zemí, ale negativně korelovány mezi jednotlivými zeměmi. V PISA 2003 byla například průměrná korelace v jednotlivých zemích (*mean within-country correlation*) mezi výsledkem v testu matematické gramotnosti a matematickým sebepojetím (*mathematics self-concept*; založeno na pěti dotazníkových položkách jako je *Učím se matematiku rychle se škálou rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím*) rovna 0,40. Zároveň však byla korelace mezi průměrným výsledkem v testu matematické gramotnosti v jednotlivých zemích ($N = 37$) a průměrným matematickým sebepojetím v jednotlivých zemích (*between-country correlation*) rovna $-0,20$. To znamená, že země s nižším matematickým sebepojetím vykazovaly lepší výsledky v testech matematické gramotnosti a naopak země s vyšším matematickým sebepojetím vykazovaly horší výsledky v testech matematické gramotnosti. Podobné korelace byly získány i pro další koncepty, jako je zájem o matematiku a postoj ke škole. Matematicky je samozřejmě možné dostat zápornou korelaci na agregovaných datech, i přes průměrnou pozitivní korelaci v rámci jednotlivých skupin. V praxi je však takovýto výsledek minimálně zarážející a vyžaduje dalšího zkoumání. Například Freedman (1999) poukazuje na dobře známé diskrepance týkající se pozitivních korelací mezi protestantismem a počtem sebevražd v Evropě, a též imigrací a gramotností ve Spojených státech. Je možné, že minimálně určitou část této nekonzistence lze připsat rozdílům v odlišném užívání škály v různých zemích a kulturách. Metoda ukotvujících vinět by právě takovéto nekonzistence či dokonce paradoxy mohla tudíž alespoň částečně objasnit.

Do žakovského dotazníku PISA 2012 byly zařazeny dvě sady ukotvujících vinět. První sada se třemi vinětami se týkala učitelova řízení třídy (*teacher classroom management*). Druhá sada taktéž s třemi vinětami se právě týkala učitelovy podpory žáků (*teacher support behavior*). Analýza první sady vinět a její dopady na žakovské hodnocení řízení třídy jejich učitelem je uvedena ve studii Voňková a kol. (2015). V naší analýze žakovského hodnocení učitelovy podpory budeme vycházet ze specifikace modelu a postupu analýzy dat uvedené v této studii.

Data z žákovského dotazníku byla sbírána ve všech 68 zúčastněných zemích/regionech (jejich výčet včetně zkratky je uveden v tabulce 3). Sběr dat v rámci výzkumu PISA je učiněn se záměrem, aby výběr žáků byl reprezentativním vzorkem patnáctiletých žáků v jednotlivých zemích (více o sběru dat OECD, 2014). Co se týká žákovského dotazníku, byly administrovány tři verze, přičemž ukotvující viněty byly zahrnuty ve dvou z těchto tří verzí. Při administraci jednotlivých verzí dotazníku byl uplatněn rotační systém. V rámci jednotlivých zúčastněných škol ve výzkumu PISA byly tyto tři verze mezi žáky rotovány. Ačkoli tedy jen část vybraných žáků hodnotila viněty, je tento podvýběr podobný ve svých charakteristikách celkovému výběru (více např. Voňková a kol., 2015). V naší analýze pracujeme s daty od celkem 306 699 žáků ze 68 zemí/regionů.

Tabulka 3 *Seznam analyzovaných zemí/regionů z výzkumu PISA 2012 včetně jejich zkratk*

Zkratka	Název země	Zkratka	Název země
ALB	Albánie	LIE	Lichtenštejnsko
ARE	Spojené arabské emiráty	LTU	Litva
ARG	Argentina	LUX	Lucembursko
AUS	Austrálie	LVA	Lotyšsko
AUT	Rakousko	MAC	Macao – Čína
BEL	Belgie	MEX	Mexiko
BGR	Bulharsko	MNE	Montenegro
BRA	Brazílie	MYS	Malajsie
CAN	Kanada	NLD	Nizozemsko
CHE	Švýcarsko	NOR	Norsko
CHL	Chile	NZL	Nový Zéland
COL	Kolumbie	PER	Peru
CRI	Kostarika	POL	Polsko
CZE	Česká republika	PRT	Portugalsko
DEU	Německo	QAT	Katar
DNK	Dánsko	QCN	Šanghaj – Čína
ESP	Španělsko	QRS	Perm (Ruská federace)
EST	Estonsko	QUA	Florida (Spojené státy americké)
FIN	Finsko	QUB	Connecticut (Spojené státy americké)
FRA	Francie	QUC	Massachusetts (Spojené státy americké)
GBR	Velká Británie	ROU	Rumunsko
GRC	Řecko	RUS	Ruská federace
HKG	Hongkong	SGP	Singapur
HRV	Chorvatsko	SRB	Srbsko

Zkratka	Název země	Zkratka	Název země
HUN	Maďarsko	SVK	Slovensko
IDN	Indonésie	SVN	Slovinsko
IRL	Irsko	SWE	Švédsko
ISL	Island	TAP	Čínská Tchaj-pej
ISR	Izrael	THA	Thajsko
ITA	Itálie	TUN	Tunisko
JOR	Jordánsko	TUR	Turecko
JPN	Japonsko	URY	Uruguay
KAZ	Kazachstán	USA	Spojené státy americké
KOR	Korea	VNM	Vietnam

6.2.2. Deskriptivní statistiky hodnoticích otázek a ukotvujících vinět

Pro hodnocení učitelovy podpory žáků jsou v žakovském dotazníku PISA 2012 využity následující čtyři položky:

Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícími tvrzeními?

H1: *Učitel nám říká, že musíme pilně pracovat.
rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím*

H2: *Když je třeba, učitel se nám více věnuje.
rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím*

H3: *Učitel nám pomáhá s učením.
rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím*

H4: *Učitel nám dává příležitost vyjádřit své názory.
rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím*

Dále žakovský dotazník zahrnuje i tři ukotvující viněty:

Přečti si popisy tří učitelů matematiky a pak se rozhodni, do jaké míry souhlasíš s tvrzeními na konci popisu.

V1: *Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou.*

Paní učitelce Novotné záleží na tom, jak se žáci učí.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

V2: *Pan učitel Novák dává domácí úkoly z matematiky jednou za týden.*

Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou.

Panu učiteli Novákovi záleží na tom, jak se žáci učí.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

V3: *Paní učitelka Ježková dává domácí úkoly z matematiky jednou za týden.*

Opravené je žákům před zkoušením nebo písemkou nevrací.

Paní učitelce Ježkové záleží na tom, jak se žáci učí.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Poznamenejme, že pro zlepšení interpretace výsledků prezentujeme výsledky na obrácené škále, tj. na škále 1) *rozhodně nesouhlasím*, 2) *nesouhlasím*, 3) *souhlasím*, 4) *rozhodně souhlasím*. Co se tedy týká interpretace hodnocení podpory žáků ze strany jejich učitele, platí, že čím vyšší hodnota na škále (a potažmo čím vyšší hodnota odhadnutých koeficientů β v CHOPIT modelu), tím vyšší úroveň učitelovy podpory žáků. Co se týká hodnocení hypotetických učitelů v ukotvujících vinětách platí, že čím vyšší hodnocení ukotvujících vinět, tím jsou žáci optimističtější, jinými slovy mají nižší standardy pro hodnocení.

Pro první orientaci v datech prezentujeme základní deskriptivní statistiky pro čtyři položky týkající se hodnocení učitelovy podpory žáků. V tabulce 4 je hodnocení učitelovy podpory pro všechny dotázané žáky ze zemí z výzkumu PISA 2012. Počet pozorování je 306 699, k výpočtu jsou použity finální žákovské váhy uvedené v datovém souboru k žákovským dotazníkům. Pro orientaci je uvedeno i průměrné hodnocení, přestože jsou odpovědi udávány na ordinální škále. Hodnocení všech čtyř položek je podobné – u všech položek je nejčastěji zastoupená odpověď *souhlasím* (volí ji přibližně polovina žáků), druhou nejpočetnější kategorií je *rozhodně souhlasím* (volí ji necelých 40 % žáků). Toto v průměru poukazuje na celkově dobré hodnocení učitelovy podpory žáků. Ostatní žáci volí kategorie s nesouhlasem, přičemž *rozhodně nesouhlasím* volí 2 % (otázka H1) až 5 % (otázka H4) žáků.

Tabulka 4 *Hodnocení učitelovy podpory pro všechny dotázané žáky ze zemí z výzkumu PISA 2012: relativní četnosti (v %) a průměrné hodnocení*

	1 Rozhodně nesouhlasím	2 Nesouhlasím	3 Souhlasím	4 Rozhodně souhlasím	Průměr
H1: <i>Učitel nám říká, že musíme pilně pracovat.</i>	2,19	8,96	51,36	37,49	3,24
H2: <i>Když je třeba, učitel se nám více věnuje.</i>	3,29	10,23	49,22	37,25	3,20
H3: <i>Učitel nám pomáhá s učením.</i>	2,46	8,27	49,20	40,07	3,27
H4: <i>Učitel nám dává příležitost vyjádřit své názory.</i>	5,04	12,83	47,25	34,88	3,12

Pozn. Počet pozorování je 306 699. Četnosti jsou uvedeny v procentech. K výpočtu četností jsou využity celkové žákovské váhy uvedené v datovém souboru k žákovskému dotazníku PISA 2012 (OECD, 2017a). I přes užití ordinální škály je pro orientaci uvedeno průměrné hodnocení.

V tabulce 5 jsou uvedeny deskriptivní statistiky pro jednotlivé země k první otázce (H1: *Učitel nám říká, že musíme pilně pracovat*). Pro každou zemi jsou konkrétně uvedeny relativní četnosti (v procentech) volby jednotlivých kategorií (1) *rozhodně nesouhlasím*, 2) *nesouhlasím*, 3) *souhlasím*, 4) *rozhodně souhlasím*). K výpočtu jsou využity žákovské váhy (*final student weight*) uvedené v datovém souboru k žákovskému dotazníku (OECD, 2017a). I přes užití ordinální škály odpovědí je v tabulce uvedeno průměrné hodnocení. Výsledný průměr lze tudíž považovat za orientační. Dále je zde pro orientaci uvedeno pořadí zemí dle průměru, četnosti volby *rozhodně souhlasím*, četnosti volby (*rozhodně*) *souhlasím* (tj. dvou kategorií *rozhodně souhlasím* a *souhlasím*), četnosti volby *rozhodně nesouhlasím* a četnosti volby (*rozhodně*) *nesouhlasím*.

Nejvyšší průměrné hodnocení učitelovy podpory uvedli žáci v Albánii

(průměrné hodnocení je 3,64) následované Tuniskem, Jordánskem, Ruskem a Spojenými arabskými emiráty. Naopak nejnižší průměrné hodnocení je ve Švýcarsku (průměrné hodnocení je 2,85) následované Německem, Finskem, Koreou a Rakouskem. Pokud bychom země seřadili podle frekvence kategorie *rozhodně souhlasím*, je pět zemí s nejvyšším průměrným hodnocením zároveň i pětící zemí s nejvyšší frekvencí této kategorie. V Albánii je kategorie *rozhodně souhlasím* zastoupena v 67,27 procentech hodnocení žáků. Země s nejnižší frekvencí této kategorie nekopíruje přesně pořadí pětice zemí s nejnižším průměrem, nicméně existuje zde podobnost. Mezi tuto pětici zemí patří Finsko, Nizozemsko, Korea, Dánsko a Švýcarsko. Ve Finsku je relativní četnost *rozhodně souhlasím* 17,19, což je ve velkém kontrastu se zeměmi s nejvyšším zastoupením této kategorie, kde dosahuje kolem 60 procent. Pořadí zemí dle frekvence souhlasných kategorií *rozhodně souhlasím* a *souhlasím* je též podobné předchozím výčtům. Pro orientaci uvedme, že pořadí zemí je dle frekvence (*rozhodně*) *souhlasím* následující: Albánie (97,92 %), Kazachstán, Thajsko, Vietnam a Ruská federace, ..., Lucembursko, Finsko, Rakousko, Švýcarsko a Německo (68,84 %). Na základě deskriptivních statistik první otázky lze říci, že relativně mnoho vyspělých evropských zemí má na základě hodnocení žáků velmi malou podporu ze strany jejich učitele. Naopak mezi země s nejvyšší podporou patří překvapivě Tunisko či některé asijské země, jako je Jordánsko.

Tabulka 5 Deskriptivní statistiky pro jednotlivé země u první položky k učitelově podpoře žáků (H1): Učitel nám říká, že musíme pilně pracovat.

	Kategorie škály				Počet pozorování	Pořadí					
	Rozhodně nesouhlasím	Nesouhlasím	Souhlasím	Rozhodně souhlasím		Průměr	Průměr	Rozhodně souhlasím	(Rozhodně) souhlasím	Rozhodně nesouhlasím	(Rozhodně) nesouhlasím
ALB	0,76	1,32	30,65	67,27	2562	3,64	1	1	1	60	68
ARE	2,00	4,04	38,79	55,17	7192	3,47	5	5	9	40	60
ARG	2,76	9,59	53,38	34,27	3553	3,19	37	35	38	23	31
AUS	2,45	10,68	53,22	33,65	9117	3,18	39	37	42	30	27
AUT	5,61	22,58	45,07	26,75	2979	2,93	64	55	66	3	3
BEL	3,14	14,12	58,68	24,06	5262	3,04	55	61	54	16	15
BGR	2,10	6,50	46,59	44,81	3285	3,34	15	16	21	37	48
BRA	2,73	8,19	42,88	46,19	11678	3,33	18	14	29	24	40
CAN	2,52	8,56	53,19	35,73	13634	3,22	32	31	31	28	38
CHE	5,20	24,22	51,12	19,46	7203	2,85	68	64	67	4	2
CHL	3,14	11,36	52,20	33,30	4324	3,16	44	40	47	15	22

	Kategorie škály				Počet pozorování	Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím		Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
COL	2,06	8,72	49,45	39,78	5358	3,27	25	21	28	39	41
CRI	3,42	7,92	49,65	39,02	2828	3,24	28	22	33	12	36
CZE	3,03	12,52	57,36	27,09	3389	3,09	52	54	50	19	19
DEU	5,08	26,08	44,16	24,68	2650	2,88	67	58	68	5	1
DNK	2,25	16,77	62,31	18,67	4528	2,97	63	65	57	33	12
ESP	2,60	10,21	51,30	35,89	16399	3,20	36	30	41	27	28
EST	1,54	12,01	53,70	32,75	3133	3,18	43	42	45	54	24
FIN	2,30	22,97	57,55	17,19	5591	2,90	66	68	65	31	4
FRA	4,57	14,5	56,82	24,10	2915	3,00	57	60	58	6	11
GBR	1,52	6,62	51,74	40,13	8096	3,30	21	20	19	55	50
GRC	2,73	13,33	51,52	32,42	3331	3,14	46	45	51	25	18
HKG	1,92	8,57	64,54	24,97	3004	3,13	48	56	24	44	45
HRV	3,01	9,31	54,14	33,53	3271	3,18	38	39	37	21	32
HUN	3,19	13,94	58,65	24,21	3126	3,04	54	59	53	14	16
IDN	0,70	5,12	53,42	40,76	3631	3,34	14	19	8	63	61
IRL	1,96	7,23	47,56	43,25	3279	3,32	19	17	23	42	46
ISL	2,81	11,89	56,92	28,37	2205	3,11	51	49	48	22	21
ISR	1,90	5,41	41,66	51,03	3103	3,42	9	8	16	47	53
ITA	1,61	4,60	55,10	38,69	20087	3,31	20	23	12	52	57
JOR	2,46	3,67	31,77	62,10	4258	3,54	3	3	10	29	59
JPN	4,44	15,65	54,96	24,95	4153	3,00	58	57	60	7	9
KAZ	0,56	3,16	45,88	50,40	3778	3,46	6	9	2	66	67
KOR	5,72	16,00	60,76	17,51	3304	2,90	65	66	62	2	7
LIE	2,67	20,52	49,61	27,20	180	3,01	56	53	63	26	6
LTU	1,91	8,63	44,40	45,06	3003	3,33	17	15	25	46	44
LUX	6,25	17,65	48,13	27,97	3273	2,98	62	50	64	1	5
LVA	0,67	6,11	55,36	37,86	2777	3,30	22	26	14	64	55
MAC	1,65	6,65	63,78	27,91	3518	3,18	41	51	20	51	49
MEX	2,26	10,44	54,26	33,04	21912	3,18	40	41	40	32	29
MNE	4,01	13,87	48,58	33,54	2906	3,12	49	38	56	8	13
MYS	1,02	6,12	44,79	48,06	3359	3,40	11	11	15	59	54
NLD	1,86	14,99	65,76	17,39	2764	2,99	60	67	52	48	17
NOR	3,57	15,89	58,21	22,34	2832	2,99	59	63	59	10	10
NZL	1,79	10,21	53,52	34,48	2698	3,21	34	34	35	50	34
PER	1,84	9,29	55,16	33,71	3585	3,21	33	36	32	49	37
POL	3,57	13,76	54,00	28,66	2987	3,08	53	48	55	9	14
PRT	1,27	6,40	53,99	38,34	3658	3,29	23	25	18	58	51
QAT	3,26	7,35	37,36	52,03	6435	3,38	13	7	26	13	43
QCN	0,71	4,14	48,54	46,61	3449	3,41	10	12	6	61	63
QRS	0,51	4,60	45,69	49,20	1138	3,44	8	10	7	67	62
QUA	3,10	8,58	50,46	37,86	1229	3,23	30	27	34	18	35
QUB	2,11	8,85	53,57	35,47	1065	3,22	31	32	30	36	39
QUC	2,22	10,23	55,07	32,48	1103	3,18	42	44	39	34	30
ROU	3,12	9,01	45,98	41,90	3305	3,27	26	18	36	17	33
RUS	0,59	3,91	39,78	55,72	3411	3,51	4	4	5	65	64
SGP	1,56	4,60	47,36	46,48	3638	3,39	12	13	11	53	58

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
SRB	2,07	11,29	50,56	36,08	2939	3,21	35	29	44	38	25
SVK	1,47	13,57	57,27	27,69	2965	3,11	50	52	49	56	20
SVN	2,14	11,18	56,28	30,41	3699	3,15	45	46	43	35	26
SWE	3,56	17,40	56,11	22,93	2913	2,98	61	62	61	11	8
TAP	1,44	7,29	56,53	34,75	4005	3,25	27	33	22	57	47
THA	0,41	3,93	57,29	38,37	4366	3,34	16	24	3	68	66
TUN	3,03	3,71	29,93	63,34	2613	3,54	2	2	13	20	56
TUR	1,95	5,56	39,26	53,23	3144	3,44	7	6	17	43	52
URY	1,99	11,63	57,29	29,09	3159	3,13	47	47	46	41	23
USA	1,92	8,81	53,09	36,18	3186	3,24	29	28	27	45	42
VNM	0,71	3,70	62,97	32,62	3279	3,28	24	43	4	62	65
Celkem	2,19	8,96	51,36	37,49	306699	3,24					

Pozn. V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové žákovské váhy uvedené v datovém souboru k žákovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet žáků, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec, označeno jako *(rozhodně) souhlasím*), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec, označeno jako *(rozhodně) nesouhlasím*).

Co se týká ostatních otázek H2, H3 a H4, uvádíme jejich deskriptivní statistiky pro úsporu místa a přehlednost textu v tabulkách 14, 15 a 16 v příloze 3. Jak již bylo zmíněno, rozdělení četností jednotlivých kategorií je v průměru za všechny žáky v datovém souboru pro všechny hodnotící otázky podobná. U jednotlivých zemí lze však najít v hodnocení těchto otázek i velké rozdíly – například Izrael by byl dle pořadí průměrného hodnocení první otázky na 9. místě mezi ostatními zeměmi, zatímco dle druhé otázky by byl až na 49. místě. Velký rozdíl je například i u amerického státu Massachusetts (dle první otázky na 42. místě, dle druhé otázky na 6. místě). Většina zemí se však v pořadí dle průměrného hodnocení první a druhé otázky příliš neliší. Průměrný rozdíl pořadí je přibližně 10,5. Podobně je tomu při vzájemném srovnání ostatních otázek. Největší průměrný rozdíl v pořadí hodnocení dvou otázek je nalezen pro dvojici H2 a H4 (rozdíl je roven 12,7). Naopak nejpodobnější si jsou hodnocení

otázek H2 a H3. Ve statistických modelech (probitový model a CHOPIT model) jsou všechny čtyři otázky používány k odhadu jednoho ukazatele učitelovy podpory žáků. S tímto záměrem byly otázky k učitelově podpoře žáků navrhovány. Uvedený rozbor pořadí těchto otázek nasvědčuje tomu, že tento konstrukt je vhodné vytvořit. Avšak je nutné poznamenat, že deskriptivní statistiky slouží pouze pro první orientaci. Navíc není bez další informace vůbec zřejmé, do jaké míry jsou odpovědi žáků založeny na skutečné úrovni učitelovy podpory a do jaké míry jsou ovlivněny odlišným užíváním škály.

Rozdílné užívání škály lze zjistit na základě rozdílného hodnocení ukotvujících vinět. V tabulce 6 je uvedeno rozdělení hodnocení všech tří ukotvujících vinět pro všechny dotázané žáky. Tato velmi stručná tabulka poukazuje na to, jak rozdílná hodnocení jedné situace se mohou vyskytnout v odpovědích žáků. Někteří žáci například hodnotí první vinětu (*Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrací ještě před zkoušením nebo písemkou. Paní učitelce Novotné záleží na tom, jak se žáci učí.*) pomocí volby nejnižší škálové kategorie *rozhodně nesouhlasím* (3,77 %), jiní žáci odpovídají *rozhodně souhlasím* (36,56 %). Přirozené pořadí vinět je však v průměrném hodnocení žáků zachováno, první viněta má nejvyšší průměrné hodnocení 3,17, druhá pak 2,81 a třetí má nejnižší průměrné hodnocení 1,95. Rozdělení hodnocení první viněty (*rozhodně nesouhlasím* 3,77 %, *nesouhlasím* 11,53 %, *souhlasím* 48,14 % a *rozhodně souhlasím* 36,56 %) je v porovnání s dalšími vinětami nejvíce podobné rozdělení hodnocení *skutečných* učitelů žáků (připomeňme rozdělení otázky H1 *rozhodně nesouhlasím* 2,19 %, *nesouhlasím* 8,96 %, *souhlasím* 51,36 % a *rozhodně souhlasím* 37,49 %). Rozdělení hodnocení první viněty a hodnotících otázek jsou obdobně podobné i na úrovni jednotlivých zemí, což poukazuje na velkou diskriminační sílu první viněty. Druhá a především třetí viněta mají potenciál dobře rozlišit užívání nižších škálových kategorií, které se v hodnotících otázkách vyskytují méně často.

Tabulka 6 *Hodnocení tří ukotvujících vinět pro všechny dotázané žáky ze zemí z výzkumu PISA 2012: relativní četnosti (v %) a průměrné hodnocení*

	1 Roz- hodně nesou- hlasím	2 Nesou- hlasím	3 Sou- hlasím	4 Roz- hodně sou- hlasím	Průměr
V1: <i>Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou. Paní učitelce Novotné záleží na tom, jak se žáci učí.</i>	3,77	11,53	48,14	36,56	3,17
V2: <i>Pan učitel Novák dává domácí úkoly z matematiky jednou za týden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou. Panu učiteli Novákovi záleží na tom, jak se žáci učí.</i>	5,65	26,73	49,06	18,55	2,81
V3: <i>Paní učitelka Ježková dává domácí úkoly z matematiky jednou za týden. Opravené je žákům před zkoušením nebo písemkou nevrací. Paní učitelce Ježkové záleží na tom, jak se žáci učí.</i>	41,16	32,43	16,18	10,22	1,95

Pozn. Počet pozorování je 306 699. Četnosti jsou uvedeny v procentech. K výpočtu četností jsou využity celkové žákovské váhy uvedené v datovém souboru k žákovskému dotazníku PISA 2012 (OECD, 2017a). I přes užití ordinální škály je pro orientaci uvedeno průměrné hodnocení.

K rozboru heterogenity ve stylu odpovídání na úrovni zemí prezentujeme v následující tabulce 7 hodnocení první viněty (průměrné hodnocení dalších dvou vinět v jednotlivých zemích je prezentováno v tabulkách 17 a 18 v příloze 3). Stejně jako v případě otázek H1 až H4 uvádíme u jednotlivých zemí nejen frekvence kategorií škály (*rozhodně souhlasím, souhlasím, nesouhlasím a rozhodně nesouhlasím*), ale pro orientaci též průměrné hodnocení, pořadí dle

průměrného hodnocení, pořadí dle frekvence kategorií *rozhodně souhlasím*, *(rozhodně) souhlasím*, *rozhodně nesouhlasím* a *(rozhodně) nesouhlasím*. Dle průměrného hodnocení první viněty patří mezi nejoptimističtější země/regiony (země/regiony s nejvyšším hodnocením) Ruská federace, Perm, Vietnam, Thajsko, Brazílie, Uruguay a Portugalsko. Naopak země s nejnižším hodnocením jsou Tunisko, Albánie, Norsko, Šanghaj, Švédsko, Rumunsko, Francie a Korea. Rozdělení hodnocení některých zemí je spíše rovnoměrné (například v Tunisku jsou všechny čtyři škálové kategorie zastoupeny srovnatelně, tj. kolem 25 %), což může být způsobeno i náhodným stylem odpovídání. Ve většině zemí je však rozdělení pravostranně sešikmené, převažuje zastoupení vyšších škálových kategorií, tj. žáci hodnotí paní učitelku Novotnou jako učitelku, které záleží na tom, jak se žáci učí. Pořadí zemí dle hodnocení jednotlivých vinět se však liší. Nejvíce se dle očekávání od sebe liší pořadí hodnocení první viněty popisující nejvyšší úroveň učitelovy podpory žáků a třetí viněty popisující nejnižší úroveň. Průměrný rozdíl v pořadí zemí je pro první a třetí vinětu 28 míst, což je při celkovém počtu 68 analyzovaných zemí zásadní rozdíl. Toto poukazuje například na to, že některé země mají vysoké standardy při hodnocení první viněty, ale nízké standardy při hodnocení třetí viněty (Albánie, Tunisko a Rumunsko). Jiné země zase vykazují nízké standardy při hodnocení první viněty, ale vysoké standardy při hodnocení třetí viněty (Rusko, Vietnam a Singapur). Užívání škálových kategorií je tedy nutné zkoumat odděleně, nelze plošně říci, že jednotlivé země mají stejné standardy při hodnocení všech tří vinět. Při rozboru odhadů statistických modelů budeme zkoumat užívání jednotlivých škálových kategorií nejprve odděleně.

Tabulka 7 *Deskriptivní statistiky pro jednotlivé země u první ukotvující viněty (VI):* Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrací ještě před zkoušením nebo písemkou.

	Kategorie škály					Počet pozorování	Pořadí					
	Rozhodně nesouhlasím	Nesouhlasím	Souhlasím	Rozhodně souhlasím	Průměr		Průměr	Rozhodně souhlasím	(Rozhodně) souhlasím	Rozhodně nesouhlasím	(Rozhodně) nesouhlasím	
ALB	18,93	29,94	28,84	22,29	2562	2,55	67	59	67	2	2	
ARE	4,16	8,95	39,06	47,83	7192	3,31	9	6	23	31	46	
ARG	3,39	7,43	46,30	42,89	3553	3,29	13	13	14	40	55	
AUS	3,99	15,00	49,86	31,16	9117	3,08	41	38	45	34	24	
AUT	4,36	13,04	45,80	36,80	2979	3,15	31	20	40	29	29	

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
BEL	4,26	14,57	54,19	26,98	5262	3,04	47	49	44	30	25
BGR	4,02	13,10	51,03	31,85	3285	3,11	40	37	38	33	31
BRA	1,66	6,56	42,71	49,07	11678	3,39	4	3	8	62	61
CAN	4,08	12,84	48,23	34,85	13634	3,14	35	30	36	32	33
CHE	4,53	14,27	50,94	30,26	7203	3,07	44	43	43	26	26
CHL	4,45	12,61	44,62	38,32	4324	3,17	29	16	37	27	32
COL	6,89	18,80	43,71	30,60	5358	2,98	51	40	54	14	15
CRI	3,67	9,20	38,91	48,21	2828	3,32	8	5	20	38	49
CZE	2,88	11,66	54,83	30,62	3389	3,13	37	39	29	45	40
DEU	2,17	10,79	52,43	34,61	2650	3,19	23	31	21	52	48
DNK	2,20	10,45	58,12	29,22	4528	3,14	32	44	19	51	50
ESP	3,88	9,74	49,30	37,08	16399	3,20	22	18	26	35	43
EST	1,72	10,74	57,03	30,50	3133	3,16	30	41	17	60	52
FIN	2,48	12,41	60,34	24,77	5591	3,07	43	55	31	48	38
FRA	13,70	20,15	39,55	26,60	2915	2,79	62	52	63	4	6
GBR	6,23	20,11	46,48	27,18	8096	2,95	52	48	55	16	14
GRC	7,92	21,26	47,78	23,05	3331	2,86	58	58	59	11	10
HKG	5,27	21,11	54,54	19,08	3004	2,87	57	66	56	18	13
HRV	1,60	5,02	56,37	37,01	3271	3,29	11	19	5	63	64
HUN	2,14	7,67	53,66	36,54	3126	3,25	18	24	12	55	57
IDN	1,70	8,21	53,31	36,78	3631	3,25	15	21	13	61	56
IRL	3,30	15,17	53,37	28,16	3279	3,06	45	46	41	41	28
ISL	4,37	12,79	47,36	35,48	2205	3,14	34	28	39	28	30
ISR	3,30	9,20	42,94	44,56	3103	3,29	12	11	18	42	51
ITA	2,16	7,04	54,60	36,20	20087	3,25	16	25	10	53	59
JOR	10,57	11,73	32,73	44,97	4258	3,12	38	9	48	7	21
JPN	2,89	12,34	49,45	35,32	4153	3,17	28	29	32	44	37
KAZ	4,71	13,91	44,73	36,64	3778	3,13	36	22	42	24	27
KOR	7,17	21,60	54,77	16,45	3304	2,80	61	68	58	13	11
LIE	3,53	12,74	58,01	25,72	180	3,06	46	54	34	39	35
LTU	5,99	13,97	46,41	33,63	3003	3,08	42	34	46	17	23
LUX	4,90	11,63	47,80	35,67	3273	3,14	33	26	35	23	34
LVA	4,98	15,18	53,26	26,59	2777	3,01	48	53	47	21	22
MAC	3,75	19,29	57,04	19,93	3518	2,93	53	64	51	37	18
MEX	3,79	10,16	50,49	35,56	21912	3,18	26	27	27	36	42
MNE	6,49	16,29	49,14	28,08	2906	2,99	50	47	50	15	19
MYS	2,01	7,20	49,90	40,89	3359	3,30	10	14	11	59	58
NLD	1,23	5,40	60,75	32,61	2764	3,25	17	36	6	64	63
NOR	14,63	24,72	39,24	21,41	2832	2,67	66	62	66	3	3
NZL	5,15	17,31	50,63	26,91	2698	2,99	49	51	49	20	20
PER	8,03	22,60	46,12	23,24	3585	2,85	60	57	60	10	9

	Kategorie škály					Počet pozorování	Pořadí				
	Rozhodně nesouhlasím	Nesouhlasím	Souhlasím	Rozhodně souhlasím	Průměr		Průměr	Rozhodně souhlasím	(Rozhodně) souhlasím	Rozhodně nesouhlasím	(Rozhodně) nesouhlasím
POL	5,22	19,25	53,59	21,95	2987	2,92	55	61	53	19	16
PRT	2,13	6,54	47,63	43,70	3658	3,33	7	12	9	57	60
QAT	4,94	10,64	39,57	44,85	6435	3,24	19	10	33	22	36
QCN	9,27	28,13	43,47	19,13	3449	2,72	65	65	65	9	4
QRS	0,43	5,30	39,92	54,36	1138	3,48	2	2	2	68	67
QUA	2,41	11,18	52,41	34,00	1229	3,18	25	33	25	49	44
QUB	2,50	11,47	48,72	37,31	1065	3,21	21	17	28	47	41
QUC	2,26	11,17	52,31	34,27	1103	3,19	24	32	24	50	45
ROU	10,72	26,12	38,65	24,51	3305	2,77	63	56	64	6	5
RUS	0,67	3,98	34,69	60,66	3411	3,55	1	1	1	67	68
SGP	2,74	8,09	48,99	40,18	3638	3,27	14	15	15	46	54
SRB	7,86	20,79	44,40	26,94	2939	2,90	56	50	57	12	12
SVK	3,06	11,67	56,62	28,65	2965	3,11	39	45	30	43	39
SVN	2,10	8,90	52,42	36,58	3699	3,23	20	23	16	58	53
SWE	9,49	24,11	47,51	18,89	2913	2,76	64	67	62	8	7
TAP	4,70	19,75	53,54	22,02	4005	2,93	54	60	52	25	17
THA	1,17	5,19	47,41	46,23	4366	3,39	5	7	4	65	65
TUN	25,61	28,48	25,19	20,72	2613	2,41	68	63	68	1	1
TUR	12,16	21,41	35,96	30,47	3144	2,85	59	42	61	5	8
URY	2,13	5,20	47,02	45,65	3159	3,36	6	8	7	56	62
USA	2,16	10,94	54,07	32,83	3186	3,18	27	35	22	54	47
VNM	0,82	5,03	45,52	48,63	3279	3,42	3	4	3	66	66
Celkem	3,77	11,53	48,14	36,56	306699	3,17					

Pozn. V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové žákovské váhy uvedené v datovém souboru k žákovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet žáků, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec).

6.3. Model CHOPIT a jeho rozšíření pro analýzu dat

V naší aplikaci pracujeme s více než jednou hodnotící otázkou, takže nelze využít standardní specifikaci CHOPIT modelu pro jednu (sebe)hodnotící otázku. Musíme využít rozšíření CHOPIT modelu pro více (sebe)hodnocení. Pro

podrobný popis modelu a jeho rozšíření odkazujeme čtenáře na část 3.2. Zde využíváme zavedené značení z této kapitoly. Náš model vychází ze specifikace modelu CHOPIT pro čtyři hodnotící otázky uvedeného ve Voňková a kol. (2015).

Pro hodnotící otázky k učitelově podpoře žáků vycházíme při odhadu z těchto rovnic:

$$Y_{si}^* = \mu_i + \epsilon_{si}, \quad i = 1, \dots, I, \quad s = 1, 2, 3, 4,$$

$$\mu_i = \mathbf{X}_i' \boldsymbol{\beta},$$

$$\text{i.i.d. } \epsilon_{si} \sim N(0, \sigma^2), \quad s = 1, 2, 3, 4,$$

$$Y_{si} = j \Leftrightarrow \tau_i^{j-1} < Y_{si}^* \leq \tau_i^j, \quad j = 1, 2, 3, 4,$$

$$\tau_i^1 = \mathbf{X}_i' \boldsymbol{\gamma}^1,$$

$$\tau_i^j = \tau_i^{j-1} + \exp(\mathbf{X}_i' \boldsymbol{\gamma}^j), \quad j = 2, \dots, J - 1,$$

$$\tau_i^0 = -\infty, \quad \tau_i^J = \infty.$$

Za \mathbf{X}_i dosazujeme *dummy* proměnné za jednotlivé země/regiony. Referenční zemí je USA. Pro identifikaci parametrů předpokládáme, že konstanta β_0 je rovna nule a σ^2 je roven jedné. Poznamenejme, že předpokládáme stejný rozptyl σ^2 pro všechny čtyři hodnotící otázky. Ve studii Voňková a kol. (2015) byl též odhadnut model s různými rozptyly pro jednotlivé hodnotící otázky týkající se učitelova řízení třídy, odhady parametrů tohoto modelu se ve srovnání s odhady parametrů modelu s předpokladem společného rozptylu zásadně nelišily. Počet škálových kategorií je roven čtyřem ($j = 1, 2, 3, 4$): 1) *rozhodně nesouhlasím*, 2) *nesouhlasím*, 3) *souhlasím*, 4) *rozhodně souhlasím*.

Pro vinětovou část modelu vycházíme při odhadu z následujících rovnic:

$$Y_{vi}^* = \theta_v + \epsilon_{vi}, \quad i = 1, \dots, I, \quad v = 1, 2, 3,$$

$$\text{i.i.d. } \epsilon_{vi} \sim N(0, \sigma_v^2), \quad \epsilon_{vi} \perp \epsilon_{si}, \mathbf{X}_i,$$

$$Y_{vi} = j \Leftrightarrow \tau_i^{j-1} < Y_{vi}^* \leq \tau_i^j, \quad j = 1, 2, 3, 4.$$

Při odhadu parametrů maximalizujeme logaritmus věrohodnostní funkce, která je také uvedena v části 3.2.2. Zde však zahrnujeme k odhadu i finální váhy jednotlivých respondentů ve výzkumu PISA 2012 (v kódovacím manuálu k žákovskému dotazníku PISA je tato proměnná značená jako W_FSTUWT , více viz OECD (2014) a OECD (2017a)). O konstrukci vah nejen ve výzkumu PISA, ale také v dalších mezinárodních výzkumech, se lze například dočíst ve Straková (2016). Při odhadu parametrů tedy maximalizujeme logaritmus věrohodnostní funkce

$$\max_{\Omega} \sum_{i=1}^I w_i \log \left(\prod_{s=1}^4 P(Y_{si} = y_{si}, \Omega) \prod_{v=1}^3 P(Y_{vi} = y_{vi}, \Omega) \right),$$

kde w_i jsou zmíněné finální váhy respondentů. Zmíněná specifikace s vahami nebyla ještě prozatím nikde odhadnuta.

6.4. Heterogenita ve stylu odpovídání a její dopad na hodnocení učitelovy podpory žáků

6.4.1. Heterogenita ve stylu užívání jednotlivých kategorií škály

V prvním kroku představení výsledků se budeme věnovat odhadnutým parametrům mezi CHOPIT modelu. Meze odrážejí heterogenitu ve stylu užívání škály u žáků v jednotlivých zemích při hodnocení učitelovy podpory žáků. V tabulce 8 jsou prezentovány odhady parametrů γ^j , $j = 1, 2, 3$ a v další tabulce 9 jsou na základě těchto parametrů vypočítány odhady všech tří mezí τ^j , $j = 1, 2, 3$ doplněné o pořadí v jednotlivých zemích. Připomeňme, že první mez označuje hranici mezi kategoriemi *rozhodně nesouhlasím* a *nesouhlasím*, dále druhá mez označuje hranici mezi *nesouhlasím* a *souhlasím* a třetí mez označuje hranici mezi *souhlasím* a *rozhodně souhlasím*. Obecně vyšší hodnoty mezí poukazují na to, že je určitá úroveň učitelovy podpory žáků označena za horší úroveň. Jinými slovy, vyšší hodnoty poukazují na vyšší standardy při hodnocení učitelovy podpory žáků. Takovéto hodnocení lze také označit jako pesimistické (raději pojmenuje žák daný stav jako horší – zvolí nižší kategorie na škále).

Tabulka 8 *Odhady koeficientů mezi γ^j CHOPIT modelu*

	γ^1	s.e.	signif.	γ^2	s.e.	signif.	γ^3	s.e.	signif.
konstanta	-2,055	0,012	***	-0,021	0,013		0,313	0,007	***
ALB	0,491	0,021	***	-0,148	0,020	***	-0,163	0,012	***
ARE	-0,066	0,019	***	-0,212	0,016	***	-0,104	0,008	***
ARG	-0,193	0,021	***	-0,192	0,018	***	-0,018	0,009	
AUS	-0,159	0,018	***	0,034	0,015	*	0,038	0,008	***
AUT	0,144	0,021	***	-0,012	0,017		-0,243	0,011	***
BEL	-0,032	0,019		-0,075	0,016	***	0,038	0,008	***
BGR	-0,237	0,021	***	0,042	0,017	*	-0,006	0,009	
BRA	-0,260	0,018	***	-0,079	0,015	***	-0,025	0,008	**
CAN	0,088	0,017	***	-0,095	0,014	***	-0,036	0,007	***
CHE	0,049	0,018	**	0,019	0,015		-0,081	0,008	***
CHL	-0,220	0,020	***	-0,039	0,017	*	-0,065	0,009	***
COL	-0,164	0,019	***	0,015	0,016		-0,037	0,008	***
CRI	-0,092	0,023	***	-0,292	0,021	***	-0,091	0,010	***
CZE	-0,371	0,021	***	0,116	0,016	***	0,088	0,009	***
DEU	0,162	0,022	***	0,021	0,017		-0,183	0,011	***
DNK	-0,401	0,020	***	0,189	0,016	***	0,174	0,008	***
ESP	0,208	0,017	***	-0,097	0,014	***	-0,074	0,007	***
EST	0,086	0,021	***	0,118	0,018	***	0,033	0,010	***
FIN	-0,033	0,019		0,190	0,015	***	0,066	0,008	***
FRA	0,319	0,021	***	-0,214	0,018	***	-0,090	0,010	***
GBR	-0,054	0,019	**	0,007	0,015		-0,002	0,008	
GRC	0,195	0,020	***	-0,039	0,017	*	-0,049	0,009	***
HKG	0,166	0,021	***	0,047	0,017	**	0,204	0,009	***
HRV	-0,069	0,022	**	0,055	0,017	**	0,009	0,009	
HUN	-0,093	0,021	***	0,016	0,017		0,013	0,009	
IDN	-0,831	0,023	***	0,241	0,019	***	0,249	0,009	***
IRL	0,172	0,021	***	-0,037	0,018	*	-0,037	0,010	***
ISL	-0,156	0,023	***	-0,040	0,020	*	0,035	0,010	***
ISR	-0,170	0,021	***	-0,090	0,018	***	-0,127	0,010	***
ITA	-0,016	0,017		-0,036	0,014	**	0,009	0,007	
JOR	0,149	0,019	***	-0,390	0,018	***	-0,224	0,009	***
JPN	-0,052	0,020	*	-0,092	0,016	***	0,040	0,008	***
KAZ	-0,043	0,020	*	0,119	0,018	***	0,027	0,010	**
KOR	0,015	0,020		-0,008	0,016		0,236	0,008	***
LIE	0,457	0,056	***	-0,157	0,049	**	-0,089	0,033	**
LTU	0,256	0,020	***	-0,129	0,018	***	-0,169	0,010	***
LUX	0,169	0,020	***	-0,122	0,016	***	-0,173	0,010	***
LVA	0,046	0,021	*	0,120	0,019	***	0,062	0,010	***
MAC	0,220	0,020	***	0,082	0,017	***	0,193	0,009	***
MEX	0,000	0,017		-0,029	0,014	*	-0,028	0,007	***
MNE	-0,026	0,021		-0,054	0,017	**	-0,048	0,009	***

	γ^1	s.e.	signif.	γ^2	s.e.	signif.	γ^3	s.e.	signif.
MYS	-0,233	0,022	***	0,056	0,019	**	0,059	0,009	***
NLD	-0,286	0,023	***	0,067	0,018	***	0,138	0,009	***
NOR	0,081	0,020	***	-0,045	0,017	**	0,046	0,009	***
NZL	-0,190	0,023	***	0,098	0,018	***	0,056	0,009	***
PER	-0,067	0,020	***	0,036	0,017	*	0,042	0,009	***
POL	0,095	0,021	***	0,022	0,017		0,013	0,009	
PRT	-0,236	0,021	***	-0,002	0,018		0,092	0,009	***
QAT	-0,057	0,018	**	-0,219	0,016	***	-0,144	0,008	***
QCN	0,496	0,020	***	0,078	0,018	***	0,057	0,010	***
QRS	-0,035	0,032		0,060	0,027	*	-0,025	0,014	
QUA	0,025	0,030		-0,091	0,024	***	-0,043	0,013	***
QUB	-0,005	0,032		0,011	0,027		-0,016	0,014	
QUC	0,011	0,033		0,002	0,029		-0,004	0,014	
ROU	0,095	0,020	***	-0,077	0,017	***	-0,118	0,009	***
RUS	0,089	0,022	***	-0,042	0,019	*	-0,126	0,010	***
SGP	0,089	0,022	***	-0,154	0,020	***	0,080	0,009	***
SRB	-0,027	0,020		0,010	0,017		-0,053	0,009	***
SVK	-0,324	0,022	***	0,126	0,017	***	0,094	0,009	***
SVN	0,093	0,020	***	0,018	0,016		-0,014	0,009	
SWE	-0,171	0,021	***	0,075	0,017	***	0,030	0,009	**
TAP	-0,189	0,020	***	0,056	0,016	***	0,176	0,008	***
THA	-0,301	0,021	***	0,019	0,019		0,146	0,009	***
TUN	0,616	0,020	***	-0,354	0,019	***	-0,313	0,011	***
TUR	0,177	0,020	***	-0,207	0,018	***	-0,129	0,010	***
URY	-0,392	0,022	***	0,000	0,018		0,023	0,010	*
USA	0			0			0		
VNM	-0,405	0,024	***	0,085	0,021	***	0,155	0,010	***

Pozn. Ve sloupcích γ^j , $j = 1,2,3$ jsou odhady parametrů mezi pro každou zemi. Ve sloupcích označených s.e. je standardní chyba těchto parametrů. Ve sloupcích signif. je označena statistická signifikance těchto parametrů na základě následujícího pravidla: $p < 0,001$ značíme ***, $p < 0,01$ značíme ** a $p < 0,05$ značíme *. Referenční zemí je USA, proto u parametrů této země nejsou uvedeny standardní chyby. Meze τ^j lze pak vypočítat na základě rovnic specifikujících CHOPIT model ($\tau_i^1 = \mathbf{X}'_i \boldsymbol{\gamma}^1$, $\tau_i^2 = \tau_i^1 + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^2)$, $\tau_i^3 = \tau_i^2 + \exp(\mathbf{X}'_i \boldsymbol{\gamma}^3)$). Obecně vyšší hodnoty mezi poukazují na to, že je určitá úroveň učitelovy podpory žáků označena za horší úroveň. Jinými slovy, vyšší hodnoty poukazují na vyšší standardy při hodnocení učitelovy podpory žáků. Takovéto hodnocení lze také označit jako pesimistické (raději pojmenuje žák daný stav jako horší – zvolí nižší kategorie na škále).

Tabulka 9 *Odhady mezi τ^j v CHOPIT modelu a pořadí zemí dle těchto odhadnutých mezi*

	τ^1	Pořadí	τ^2	Pořadí	τ^3	Pořadí
ALB	-1,564	3	-0,720	2	0,441	9

	τ^1	Pořadí	τ^2	Pořadí	τ^3	Pořadí
ARE	-2,121	43	-1,328	59	-0,096	63
ARG	-2,248	55	-1,440	66	-0,097	64
AUS	-2,214	49	-1,201	45	0,219	41
AUT	-1,912	16	-0,944	14	0,128	55
BEL	-2,087	36	-1,178	44	0,242	34
BGR	-2,292	59	-1,271	53	0,088	59
BRA	-2,315	60	-1,411	64	-0,077	62
CAN	-1,968	22	-1,077	30	0,242	33
CHE	-2,006	25	-1,008	20	0,252	30
CHL	-2,275	56	-1,334	61	-0,053	60
COL	-2,219	50	-1,225	48	0,092	58
CRI	-2,147	46	-1,416	65	-0,168	68
CZE	-2,426	64	-1,327	58	0,165	50
DEU	-1,893	14	-0,893	8	0,245	32
DNK	-2,457	66	-1,274	54	0,353	18
ESP	-1,847	8	-0,958	16	0,311	20
EST	-1,969	23	-0,867	7	0,546	7
FIN	-2,088	37	-0,904	9	0,557	5
FRA	-1,736	5	-0,946	15	0,303	21
GBR	-2,109	41	-1,123	36	0,241	35
GRC	-1,860	9	-0,918	11	0,384	14
HKG	-1,889	13	-0,863	6	0,813	3
HRV	-2,124	45	-1,090	32	0,290	24
HUN	-2,148	47	-1,153	40	0,232	38
IDN	-2,886	68	-1,640	68	0,113	57
IRL	-1,883	11	-0,939	13	0,378	16
ISL	-2,211	48	-1,270	52	0,145	53
ISR	-2,225	51	-1,330	60	-0,126	65
ITA	-2,071	33	-1,127	37	0,253	29
JOR	-1,906	15	-1,244	49	-0,152	67
JPN	-2,107	40	-1,214	47	0,209	43
KAZ	-2,098	39	-0,994	19	0,409	12
KOR	-2,040	28	-1,068	27	0,662	4
LIE	-1,598	4	-0,762	4	0,489	8
LTU	-1,799	6	-0,938	12	0,216	42
LUX	-1,886	12	-1,019	21	0,130	54
LVA	-2,009	26	-0,905	10	0,550	6
MAC	-1,835	7	-0,772	5	0,885	2
MEX	-2,055	30	-1,103	34	0,226	39
MNE	-2,081	34	-1,154	41	0,150	52
MYS	-2,288	57	-1,253	50	0,197	46
NLD	-2,341	61	-1,294	55	0,275	28
NOR	-1,974	24	-1,038	23	0,393	13
NZL	-2,245	54	-1,165	42	0,281	26

	τ^1	Pořadí	τ^2	Pořadí	τ^3	Pořadí
PER	-2,122	44	-1,107	35	0,319	19
POL	-1,960	18	-0,959	17	0,426	10
PRT	-2,291	58	-1,314	56	0,185	47
QAT	-2,112	42	-1,326	57	-0,143	66
QCN	-1,559	2	-0,501	1	0,947	1
QRS	-2,090	38	-1,050	24	0,284	25
QUA	-2,030	27	-1,136	39	0,173	49
QUB	-2,060	32	-1,070	28	0,275	27
QUC	-2,044	29	-1,063	26	0,298	22
ROU	-1,960	17	-1,053	25	0,162	51
RUS	-1,966	20	-1,027	22	0,178	48
SGP	-1,966	21	-1,127	38	0,354	17
SRB	-2,083	35	-1,093	33	0,202	44
SVK	-2,380	63	-1,269	51	0,233	37
SVN	-1,962	19	-0,966	18	0,382	15
SWE	-2,226	52	-1,171	43	0,238	36
TAP	-2,244	53	-1,209	46	0,421	11
THA	-2,356	62	-1,358	62	0,224	40
TUN	-1,439	1	-0,752	3	0,247	31
TUR	-1,878	10	-1,082	31	0,120	56
URY	-2,447	65	-1,468	67	-0,070	61
USA	-2,055	31	-1,076	29	0,291	23
VNM	-2,460	67	-1,394	63	0,202	45

Pozn. Meze τ^j lze vypočítat na základě rovnic specifikujících CHOPIT model ($\tau_i^1 = \mathbf{X}_i' \boldsymbol{\gamma}^1, \tau_i^2 = \tau_i^1 + \exp(\mathbf{X}_i' \boldsymbol{\gamma}^2), \tau_i^3 = \tau_i^2 + \exp(\mathbf{X}_i' \boldsymbol{\gamma}^3)$). Odhady parameterů $\boldsymbol{\gamma}^j$ jsou uvedeny v tabulce 8. Obecně vyšší hodnoty mezi poukazují na to, že je určitá úroveň učitelovy podpory žáků označena za horší úroveň. Jinými slovy, vyšší hodnoty poukazují na vyšší standardy při hodnocení učitelovy podpory žáků. Takovéto hodnocení lze také označit jako pesimistické (raději pojmenuje žák daný stav jako horší – zvolí nižší kategorie na škále).

Odhadnuté koeficienty u jednotlivých zemí jsou statisticky signifikantně odlišné pro všechny tři meze, heterogenita ve stylu užívání škály tedy zásadně ovlivňuje porovnávání učitelovy podpory žáků v jednotlivých zemích. Žáci v jednotlivých zemích se tedy signifikantně odlišují v tom, jak pojmenovávají *určitou* úroveň učitelovy podpory žáků. Žáci z některých zemí vnímají určitou úroveň učitelovy podpory jako výbornou, žáci z jiných zemí jen jako dobrou.

Hodnoty jednotlivých zemí lze z věcného pohledu nejlépe interpretovat na základě jejich blízkosti k odhadnutým koeficientům θ_j , $j = 1,2,3$, tj. na základě blízkosti k odhadnutým efektům jednotlivých vinět. Odhadnutý koeficient θ_1

označuje úroveň první viněty, tj. viněty označující nejvyšší úroveň učitelovy podpory žáků. Jeho hodnota je $-0,154$ se standardní chybou $0,015$. Úroveň druhé viněty je odhadnuta na $-0,697$ se standardní chybou $0,015$ a pro třetí vinětu je odhadnutý parametr roven $-1,922$ se standardní chybou $0,015$. Hodnoty první meze (hranice mezi *rozhodně nesouhlasím* a *nesouhlasím*) se pohybují v rozpětí $-2,886$ (Indonésie) a $-1,439$ (Tunisko). Průměrná hodnota je $-2,076$. Tato mez je nejvíce srovnatelná s hodnotou třetí viněty popisující nejhorší úroveň učitelovy podpory žáků. Druhá mez (hranice mezi *nesouhlasím* a *souhlasím*) se pohybuje v rozpětí $-1,640$ (Indonésie) a $-0,501$ (Šanghaj). Průměrná hodnota je $-1,113$, což je úroveň blížíící se spíše druhé vinětě než třetí vinětě. Třetí mez (hranice mezi *souhlasím* a *rozhodně souhlasím*) se pohybuje v rozmezí $-0,168$ (Kostarika) a $0,947$ (Šanghaj), průměrná hodnota je $0,254$, což je nad úrovní první viněty. Odhadnuté koeficienty lze dále interpretovat i tak, že pokud je například odhadnutá hodnota třetí meze τ^3 (hranice mezi *souhlasím* a *rozhodně souhlasím*) větší než odhadnutý parametr θ_1 (viz příklad Šanghaje $0,947 > -0,154$), tak žáci z takovéto země označují situaci popsanou v první vinětě spíše pomocí středové kategorie *souhlasím* než *rozhodně souhlasím*. Oproti tomu je například v Kostarice odhadnutá třetí mez τ^3 ($-0,168$) nižší než odhadnutý koeficient θ_1 ($-0,154$), což znamená, že žáci v této zemi označují situaci popsanou ve vinětě spíše pomocí koncové kategorie *rozhodně souhlasím*.

Pro přehlednost zařazujeme nejen tabulku 9 s odhadnutými mezemi, ale také geografické rozdělení zemí dle těchto mezí zobrazené na obrázcích 8 (první mez), 9 (druhá mez) a 10 (třetí mez).

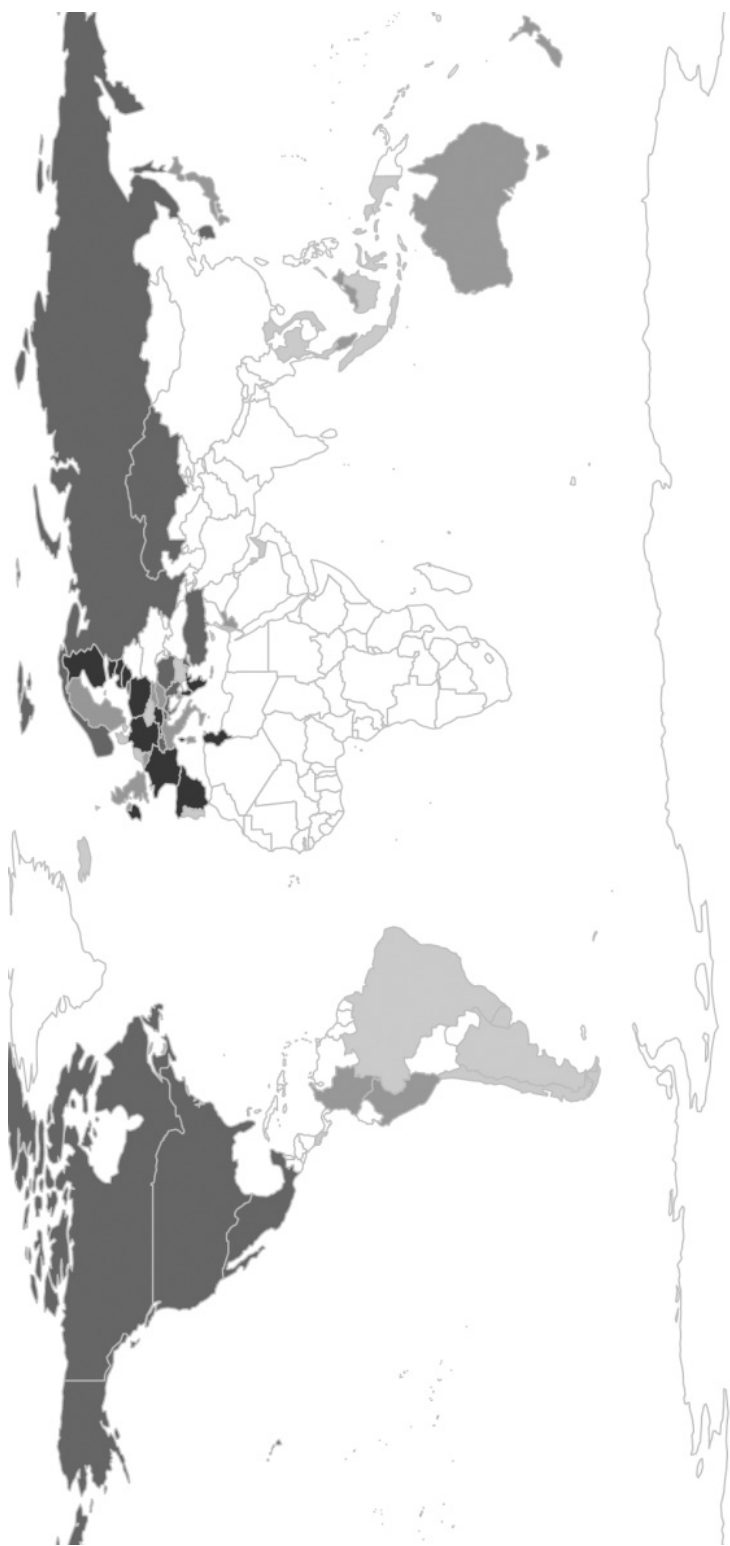
Z hlediska výše hodnot jednotlivých mezí lze země rozdělit do následujících šesti skupin:

1. **Vysoká hodnota první meze (hranice mezi *rozhodně nesouhlasím* a *nesouhlasím*):** Žáci v těchto zemích volí ve srovnání s žáky z ostatních zemí spíše koncový bod škály *rozhodně nesouhlasím* než *nesouhlasím* při hodnocení učitelovy podpory žáků. Jejich hodnocení lze na základě hodnoty první meze označit za pesimistické, určitou úroveň učitelovy podpory označují za nejhorší (volí škálovou kategorii *rozhodně nesouhlasím*), zatímco žáci z jiných zemí by tuto úroveň označili pomocí vyšších škálových bodů (zvolili by například

nesouhlasím). Mezi tyto země patří například Tunisko, Šanghaj, Albánie, Lichtenštejnsko, Francie, Litva, Macao, Španělsko, Řecko a Turecko. Na obrázku 8 jsou označeny černou barvou.

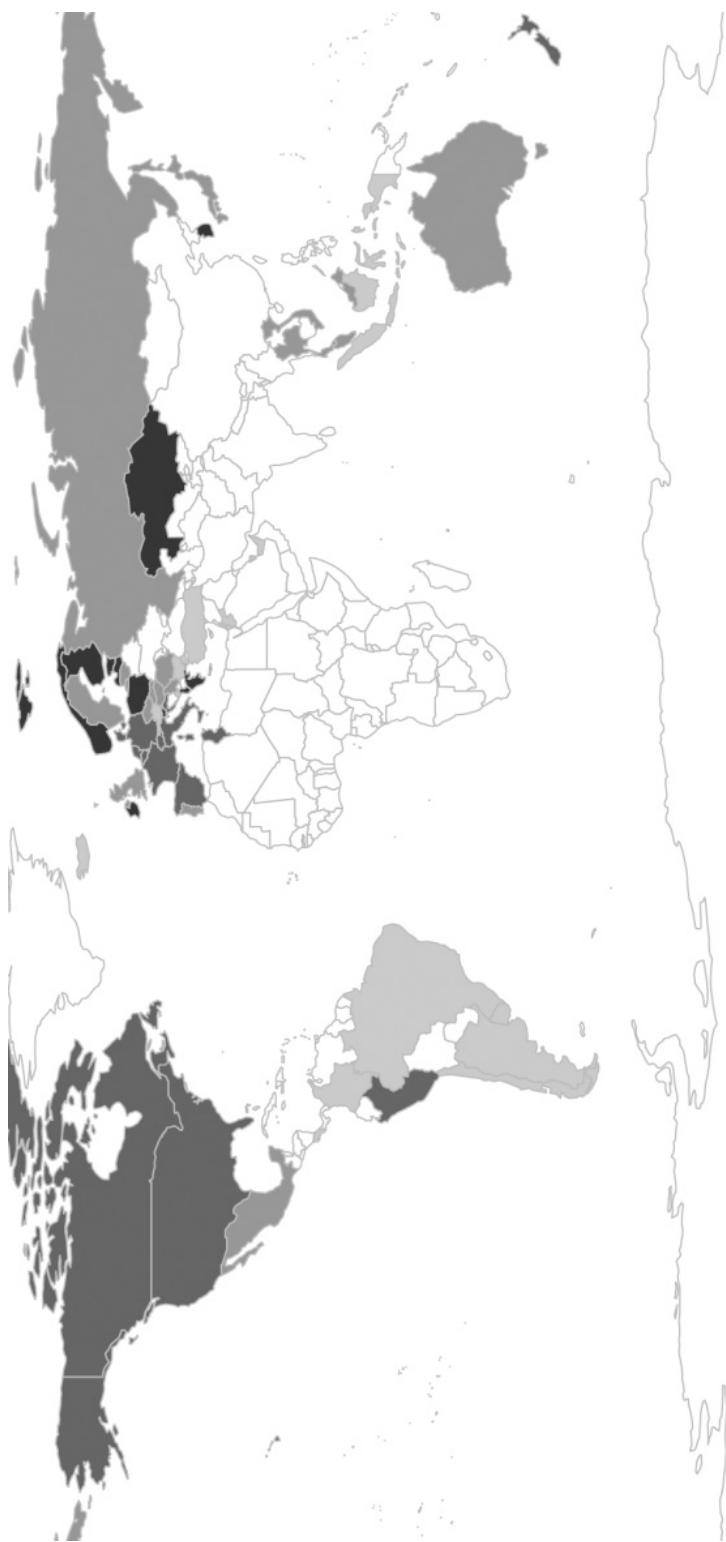
2. **Nízká hodnota první meze (hranice mezi rozhodně nesouhlasím a souhlasím):** Žáci v těchto zemích volí ve srovnání s žáky z ostatních zemí spíše středový bod škály *nesouhlasím* než *rozhodně nesouhlasím* při hodnocení učitelovy podpory žáků. Na základě hodnoty první meze lze říci, že hodnocení žáků z těchto zemí je optimistické. Další interpretací může být i to, že žáci z těchto zemí užívají méně často extrémní bod škály *rozhodně nesouhlasím* a hodnotí situace spíše pomocí středové kategorie *souhlasím*. Mezi tyto země patří například Indonésie, Vietnam, Dánsko, Uruguay, Česká republika, Slovensko, Thajsko, Nizozemsko, Brazílie a Bulharsko. Na obrázku 8 jsou označeny nejslabším odstínem šedé barvy.
3. **Vysoká hodnota druhé meze (hranice mezi nesouhlasím a souhlasím):** Obecně je druhá mez zajímavá z pohledu přechodu mezi nesouhlasem a souhlasem. Kdyby škála byla dvoubodová, tj. obsahovala volby pouze mezi souhlasem a nesouhlasem, hrála by tato hranice primární roli. Vysoká hodnota druhé meze označuje, že žáci v těchto zemích volí ve srovnání s žáky z ostatních zemí spíše *nesouhlasím* než *souhlasím* při hodnocení učitelovy podpory žáků. Jejich hodnocení lze na základě hodnoty druhé meze označit za pesimistické. Další interpretací je, že mají, co se týká druhé meze, vyšší standardy pro hodnocení ve srovnání s ostatními zeměmi. Mezi tyto země patří například Šanghaj, Albánie, Tunisko, Lichtenštejnsko, Macao, Hongkong, Estonsko, Německo, Finsko a Lotyšsko. Na obrázku 9 jsou označeny černou barvou.
4. **Nízká hodnota druhé meze (hranice mezi nesouhlasím a souhlasím):** Žáci v těchto zemích volí ve srovnání s žáky z ostatních zemí spíše *souhlasím* než *nesouhlasím* při hodnocení učitelovy podpory žáků. Jejich hodnocení lze na základě hodnoty druhé meze označit za optimistické, přiklánějí se spíše ke kladné části škály (ke kategoriím souhlasu). Mezi tyto země patří například Indonésie, Uruguay, Argentina, Kostarika, Brazílie, Vietnam, Thajsko, Chile, Izrael a Spojené arabské emiráty. V pořadí jedenáctá je Česká republika. Na obrázku 9 jsou označeny nejslabším odstínem šedé barvy.

5. **Vysoká hodnota třetí meze (hranice mezi *souhlasím* a *rozhodně souhlasím*):** Žáci v těchto zemích volí ve srovnání s žáky z ostatních zemí spíše středový bod škály *souhlasím* než *rozhodně souhlasím*, přiklánějí se tedy spíše ke středové kategorii souhlasu. Co se tedy týká hodnoty třetí meze, mají žáci z těchto zemí vyšší standardy pro hodnocení. Mezi tyto země patří například Šanghaj, Macao, Hongkong, Korea, Finsko, Lotyšsko, Estonsko, Lichtenštejnsko, Albánie a Polsko. Na obrázku 10 jsou označeny černou barvou.
6. **Nízká hodnota třetí meze (hranice mezi *souhlasím* a *rozhodně souhlasím*):** Žáci v těchto zemích volí ve srovnání s žáky z ostatních zemí spíše koncový bod škály *rozhodně souhlasím* než *souhlasím*, přiklánějí se tedy spíše k extrémní škálové kategorii souhlasu. Co se tedy týká hodnoty třetí meze, mají žáci z těchto zemí nižší standardy pro hodnocení. Jejich hodnocení lze tedy, co se týká třetí meze, označit za optimistické. Mezi tyto země patří například Kostarika, Jordánsko, Katar, Izrael, Argentina, Spojené arabské emiráty, Brazílie, Uruguay, Chile a Bulharsko. Na obrázku 10 jsou označeny nejslabším odstínem šedé barvy.



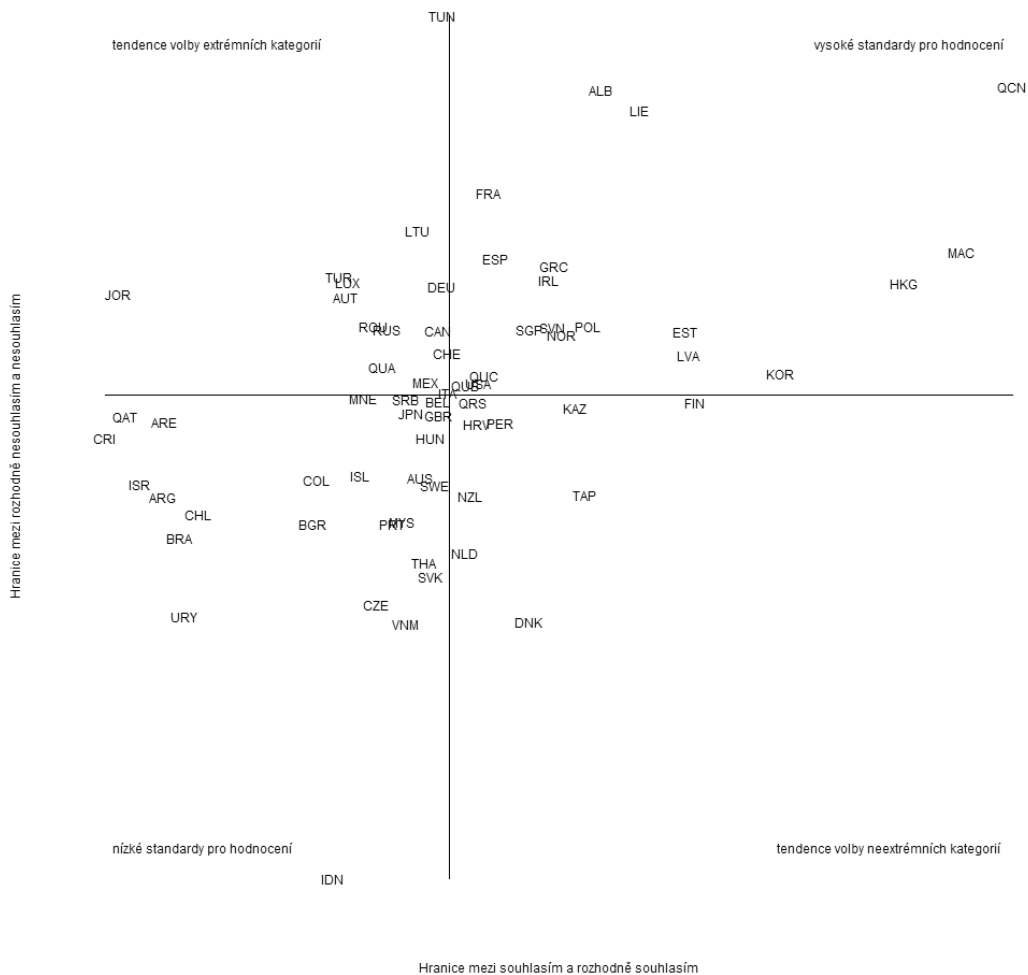
Obrázek 9. Geografické rozdělení zemí dle odhadnuté druhé meze τ^2 CHOPIT modelu

Pozn. Obecně vyšší hodnoty mezi poukazují na vyšší standardy při hodnocení učitelovy podpory žáků. Druhá mez označuje hranici mezi kategoriemi *nesouhlasím* a *souhlasím*. Vyšší hodnoty druhé meze poukazují na to, že žáci spíše nesouhlasí (druhá kategorie) než souhlasí (třetí kategorie) s tím, že danému učiteli záleží, jak se žáci učí. Země/regiony jsou rozděleny na čtyři skupiny dle hodnoty druhé meze – černá barva označuje čtvrtinu zemí/regionů s největšími hodnotami druhé meze (nejvyššími standardy), nejslabší odstín šedé označuje čtvrtinu zemí s nejmenšími hodnotami druhé meze (nejnižšími standardy). Bíle označené země se výzkumu PISA nezúčastnily.



Obrázek 10. Geografické rozdělení zemí dle odhadnuté třetí meze τ^3 CHOPIT modelu

Pozn. Obecně vyšší hodnoty mezí poukazují na vyšší standardy při hodnocení učitelovy podpory žáků. Třetí mez označuje hranici mezi kategoriemi *souhlasím* a *rozhodně souhlasím*. Vyšší hodnoty třetí meze poukazují na to, že žáci spíše souhlasí (třetí kategorie) než rozhodně souhlasí (čtvrtá kategorie) s tím, že danému učiteli záleží, jak se žáci učí. Země/regiony jsou rozděleny na čtyři skupiny dle hodnoty třetí meze – černá barva označuje čtvrtinu zemí/regionů s nejvyššími hodnotami třetí meze (nejvyššími standardy), nejslabší odstín šedé označuje čtvrtinu zemí s nejmenšími hodnotami třetí meze (nejnižšími standardy). Bíle označené země se výzkumu PISA nezúčastnily.



Obrázek 11. *Vztah mezi první mezí (hranice mezi rozhodně nesouhlasím a nesouhlasím) a třetí mezí (hranice mezi souhlasím a rozhodně souhlasím)*

Pozn. Na ose x je zanesena odhadnutá třetí mez CHOPIT modelu – hranice mezi *souhlasím* a *rozhodně souhlasím*. Na ose y je zanesena odhadnutá první mez CHOPIT modelu – hranice mezi *rozhodně nesouhlasím* a *nesouhlasím*. Osy se protínají v průměrných hodnotách příslušných mezí. Kvadranty v grafu jsou interpretovány dle úrovně standardů pro hodnocení (více viz popis obrázku v textu).

6.4.2. Vztah mezi stylem užívání různých kategorií škály

Vedle analýzy jednotlivých hodnot mezi je též zajímavé zkoumat jejich vztah. Korelace mezi první a druhou mezi je rovna 0,880, mezi druhou a třetí mezi je to 0,739. Nejnižší korelace je mezi první a třetí mezi, její hodnota je rovna 0,425. Tato korelace není vysoká a poukazuje na to, že mohou existovat země s vysokou hodnotou první meze a nízkou hodnotou třetí meze, či naopak s nízkou hodnotou první meze a vysokou hodnotou první meze. Obrázek 11 ukazuje vztah mezi první mezi (hranice mezi *rozhodně nesouhlasím* a *nesouhlasím*) a poslední třetí mezi (hranice mezi *souhlasím* a *rozhodně souhlasím*). Z věcného pohledu je užitečné rozdělit země dle výše uvedených mezi na čtyři skupiny (poznamenejme, že pro finální korigované hodnocení učitelovy podpory rozhoduje i výše druhé meze, porovnání korigovaného a nekorigovaného hodnocení je uvedeno v části 6.4.3 při diskuzi odhadnutých parametrů CHOPIT modelu a probitového modelu pro ordinální proměnnou):

1. **Vysoké hodnoty obou mezi** (první kvadrant na obrázku 11): Zde se nachází země s **vysokými standardy pro hodnocení**, jinými slovy pesimistické země. Určitý stav učitelovy podpory žáků tedy označují ve srovnání s ostatními zeměmi jako horší. Označují ho tedy pomocí nižších škálových kategorií, tj. volí spíše označení *rozhodně nesouhlasím* než *nesouhlasím*, či *souhlasím* než *rozhodně souhlasím*. Mezi takovéto země patří Šanghaj, Macao, Hongkong, Albánie a Lichtenštejnsko. Dále lze jmenovat i Koreu, Estonsko, Lotyšsko, Polsko, Řecko, Irsko, Španělsko, Francii, Slovensko, Norsko a Singapur.
2. **Vysoké hodnoty první meze a nízké hodnoty třetí meze** (druhý kvadrant na obrázku 11): Zde se nachází země s **tendencí volby extrémních kategorií**. Vysoké hodnoty první meze poukazují na tendenci označovat, zda *danému* učiteli záleží na tom, jak se žák učí, spíše volbou kategorie *rozhodně nesouhlasím* než kategorie *nesouhlasím*. Nízké hodnoty třetí meze poukazují na tendenci označovat, zda *danému* učiteli záleží na tom, jak se žák učí, spíše volbou kategorie *rozhodně souhlasím* než kategorie *souhlasím*. Mezi takovéto země patří například Jordánsko, které se především vyznačuje tendencí volby *rozhodně souhlasím* spíše než *souhlasím*. Dále mezi tyto země patří též Turecko, Lucembursko, Rakousko, Rumunsko a Ruská federace. Extrémním

případem je Tunisko, které má nejvyšší hodnotu hranice mezi *rozhodně nesouhlasím* a *nesouhlasím*, tj. má tendenci oproti ostatním zemím označovat velký rozsah úrovní učitelovy podpory pomocí nejnižší kategorie škály *rozhodně nesouhlasím*. Při volbě mezi kategoriemi *souhlasím* a *rozhodně souhlasím* se nachází v průměru mezi analyzovanými zeměmi.

3. **Nízké hodnoty obou mezi** (třetí kvadrant na obrázku 11): Zde se nachází země s **nízkými standardy pro hodnocení**, jinými slovy optimistické země, co se týká hodnot první a třetí meze. Určitý stav učitelovy podpory žáků tedy označují ve srovnání s ostatními zeměmi jako lepší, tj. označují ho pomocí vyšších škálových kategorií (tj. volí spíše *nesouhlasím* než *rozhodně nesouhlasím* a také spíše *rozhodně souhlasím* než *souhlasím*). Mezi takového země patří například Indonésie, Uruguay, Brazílie, Chile, Argentina, Bulharsko, Kolumbie, Izrael a Island. Dále lze jmenovat i Vietnam, Českou republiku a Slovensko. Tyto země spolu s Indonésií (extrémní případ) a Uruguay patří mezi země s nejnižší hranicí mezi *rozhodně nesouhlasím* a *nesouhlasím*, tj. mají tendenci nepoužívat koncový bod škály *rozhodně nesouhlasím*, a ve srovnání s ostatními brzy přechází k užívání kategorie blíže středu *nesouhlasím*. Co se týká užívání kategorií *souhlasím* a *rozhodně souhlasím*, nejsou tyto tři země extrémní, nacházejí se spíše blíže průměru. Dále mezi země s nejnižší hranicí mezi *souhlasím* a *rozhodně souhlasím* patří Katar, Kostarika a Spojené arabské emiráty. Tyto tři země se z hlediska hodnoty hranice mezi *rozhodně nesouhlasím* a *nesouhlasím* nacházejí spíše blíže průměru.
4. **Nízké hodnoty první meze a vysoké hodnoty třetí meze** (čtvrtý kvadrant na obrázku 11): Zde jsou země s **tendencí volby středových (tj. neextrémních) kategorií**. Nízké hodnoty první meze poukazují na tendenci označovat, zda *danému* učiteli záleží na tom, jak se žák učí (učitel podporuje žáky), spíše volbou kategorie *nesouhlasím* než kategorie *rozhodně nesouhlasím*. Vysoké hodnoty třetí meze poukazují na tendenci označovat, zda *danému* učiteli záleží na tom, jak se žák učí, spíše volbou kategorie *souhlasím* než kategorie *rozhodně souhlasím*. Mezi tyto země patří Tchaj-pej a Dánsko. Nizozemsko patří společně s Dánskem v této skupině mezi země s nejnižší hranicí mezi *rozhodně*

nesouhlasím a *nesouhlasím* (tendence volby *nesouhlasím*), zatímco Finsko patří v této skupině mezi země s nejvyšší hranicí mezi *souhlasím* a *rozhodně souhlasím* (tendence volby *souhlasím*).

6.4.3. Korigované versus nekorigované hodnocení učitelovy podpory na úrovni zemí

V předchozí části jsme analyzovali heterogenitu užívání škály napříč zeměmi. Některé země mají vysoké standardy pro hodnocení učitelovy podpory žáků, jiné mají naopak tyto standardy nízké. Některé země využívají krajní body škály pro hodnocení většího rámce úrovní učitelovy podpory žáků více než jiné země (například Jordánsko, Turecko a Lucembursko). Naopak existují země, které označují větší spektrum učitelovy podpory žáků pomocí středových bodů škály (například Tchaj-pej a Dánsko). Identifikovaná heterogenita užívání škály je mezi zeměmi statisticky i věcně významná. Existují země, které pojmenují určitý stav učitelovy podpory žáků jako dobrý (souhlasí, že *danému* učiteli záleží na tom, jak se žáci učí), a jiné země, které ho označí za špatný (nesouhlasí, že *danému* učiteli záleží na tom, jak se žáci učí). Tato heterogenita má tedy vliv na porovnání úrovně učitelovy podpory žáků pomocí hodnotících otázek. Některé země mohou na základě tohoto hodnocení vykazovat zdánlivě vysoké úrovně učitelovy podpory žáků, jiné zase mohou být chybně zařazeny mezi země s nejhorší úrovní učitelovy podpory.

V této části porovnáme korigované (očistěné o rozdílné užívání škály) a nekorigované (původní, neočistěné o rozdílné užívání škály) hodnocení učitelovy podpory žáků. Toto provedeme pomocí odhadu β parametrů CHOPIT modelu (parametry zachycující učitelovu podporu žáků, jež jsou korigovány o rozdílné užívání škály) a probitového modelu pro ordinální proměnnou (parametry zachycující učitelovu podporu žáků, jež naopak nejsou korigovány o rozdílné užívání škály). Tyto parametry jsou společně se statistickou signifikancí uvedeny v tabulce 10. Dále je v této tabulce uvedeno i pořadí zemí dle velikosti těchto odhadnutých parametrů.

Tabulka 10 Porovnání odhadů koeficientů β u CHOPIT modelu (korigované efekty zemí) a probitového modelu pro ordinální proměnnou (nekorigované efekty zemí)

	Pořadí		Rozdíl	CHOPIT			Oprobit		
	CHOPIT	Oprobit		β	s.e.	signif.	β	s.e.	signif.
ALB	2	1	-1	0,729	0,022	***	0,543	0,010	***
ARE	36	9	-27	-0,154	0,018	***	0,186	0,007	***
ARG	58	33	-25	-0,394	0,020	***	-0,027	0,008	**
AUS	41	39	-2	-0,196	0,017	***	-0,101	0,007	***
AUT	67	68	1	-0,663	0,021	***	-0,674	0,009	***
BEL	62	60	-2	-0,439	0,019	***	-0,375	0,008	***
BGR	35	20	-15	-0,141	0,021	***	0,064	0,008	***
BRA	49	21	-28	-0,290	0,017	***	0,064	0,007	***
CAN	23	26	3	-0,026	0,017		0,001	0,007	
CHE	55	58	3	-0,339	0,018	***	-0,351	0,007	***
CHL	54	34	-20	-0,338	0,019	***	-0,029	0,008	***
COL	31	14	-17	-0,079	0,018	***	0,106	0,008	***
CRI	52	17	-35	-0,314	0,022	***	0,091	0,009	***
CZE	65	59	-6	-0,556	0,021	***	-0,359	0,009	***
DEU	64	66	2	-0,502	0,022	***	-0,582	0,009	***
DNK	53	51	-2	-0,336	0,019	***	-0,273	0,008	***
ESP	33	44	11	-0,103	0,016	***	-0,175	0,006	***
EST	15	48	33	0,042	0,020	*	-0,191	0,009	***
FIN	28	50	22	-0,059	0,018	**	-0,270	0,008	***
FRA	48	61	13	-0,282	0,021	***	-0,379	0,009	***
GBR	26	29	3	-0,056	0,018	**	-0,007	0,007	
GRC	22	42	20	-0,020	0,020		-0,144	0,008	***
HKG	6	41	35	0,269	0,020	***	-0,136	0,009	***
HRV	47	49	2	-0,273	0,020	***	-0,264	0,008	***
HUN	60	56	-4	-0,405	0,020	***	-0,338	0,009	***
IDN	29	5	-24	-0,077	0,021	***	0,201	0,009	***
IRL	30	47	17	-0,078	0,020	***	-0,190	0,008	***
ISL	40	35	-5	-0,193	0,023	***	-0,049	0,009	***
ISR	59	36	-23	-0,400	0,021	***	-0,052	0,008	***
ITA	34	38	4	-0,139	0,016	***	-0,099	0,006	***
JOR	21	3	-18	-0,019	0,019		0,337	0,008	***
JPN	61	57	-4	-0,436	0,020	***	-0,339	0,008	***
KAZ	3	2	-1	0,489	0,020	***	0,384	0,009	***
KOR	46	64	18	-0,270	0,020	***	-0,455	0,009	***
LIE	25	63	38	-0,041	0,062		-0,401	0,028	***
LTU	10	11	1	0,136	0,020	***	0,136	0,009	***

	Pořadí		CHOPIT				Oprobit		
	CHOPIT	Oprobit	Rozdíl	β	s.e.	signif.	β	s.e.	signif.
LUX	66	65	-1	-0,586	0,021	***	-0,563	0,008	***
LVA	8	32	24	0,205	0,020	***	-0,017	0,009	
MAC	4	43	39	0,316	0,019	***	-0,164	0,009	***
MEX	16	16	0	0,041	0,016	**	0,093	0,006	***
MNE	50	46	-4	-0,294	0,021	***	-0,187	0,009	***
MYS	18	10	-8	0,034	0,021		0,155	0,009	***
NLD	68	67	-1	-0,730	0,022	***	-0,590	0,010	***
NOR	43	53	10	-0,233	0,022	***	-0,315	0,009	***
NZL	39	40	1	-0,173	0,022	***	-0,126	0,009	***
PER	17	24	7	0,037	0,020		0,028	0,009	**
POL	45	62	17	-0,267	0,020	***	-0,393	0,009	***
PRT	37	30	-7	-0,160	0,020	***	-0,010	0,008	
QAT	51	23	-28	-0,310	0,018	***	0,044	0,007	***
QCN	1	8	7	0,823	0,020	***	0,193	0,009	***
QRS	11	13	2	0,126	0,030	***	0,125	0,013	***
QUA	38	37	-1	-0,160	0,028	***	-0,070	0,011	***
QUB	13	22	9	0,049	0,029		0,059	0,012	***
QUC	12	19	7	0,077	0,029	**	0,070	0,012	***
ROU	32	31	-1	-0,082	0,020	***	-0,013	0,009	
RUS	9	4	-5	0,155	0,022	***	0,222	0,009	***
SGP	7	6	-1	0,234	0,021	***	0,198	0,008	***
SRB	44	45	1	-0,239	0,021	***	-0,182	0,009	***
SVK	63	55	-8	-0,462	0,021	***	-0,331	0,009	***
SVN	42	54	12	-0,225	0,019	***	-0,328	0,008	***
SWE	56	52	-4	-0,359	0,021	***	-0,285	0,009	***
TAP	14	28	14	0,043	0,019	*	0,000	0,008	
THA	20	12	-8	-0,001	0,019		0,126	0,008	***
TUN	5	7	2	0,284	0,021	***	0,194	0,009	***
TUR	24	18	-6	-0,029	0,021		0,076	0,008	***
URY	57	25	-32	-0,365	0,021	***	0,008	0,009	
USA	19	27	8	0,000			0,000		
VNM	27	15	-12	-0,058	0,022	**	0,099	0,010	***

Pozn. Ve sloupcích signif. je označena statistická signifikance těchto parametrů na základě následujícího pravidla: $p < 0,001$ značíme ***, $p < 0,01$ značíme ** a $p < 0,05$ značíme *. Referenční zemí je USA, proto u parametrů této země nejsou uvedeny standardní chyby.

Odhadnuté parametry β v CHOPIT modelu jsou statisticky signifikantní, některé země mají signifikantně vyšší úroveň učitelovy podpory žáků než USA

(USA je referenční země při odhadu parametrů). Jedná se například o Šanghaj. Některé země mají naopak signifikantně nižší úroveň učitelovy podpory žáků. Mezi takové země patří mimo jiné i Česká republika. Z věcného pohledu jsou některé země na úrovni druhé viněty popisující střední úroveň učitelovy podpory žáků (například Nizozemsko, Rakousko, Lucembursko a Česká republika), jiné země jsou výrazně nad úrovní první viněty popisující nejvyšší úroveň učitelovy podpory žáků (například již zmíněná Šanghaj).

Zásadní otázkou však je, nakolik se pro jednotlivé země od sebe liší korigované a nekorigované hodnocení, tj. jakou roli sehrávají viněty při korekci. Pořadí některých zemí se po očištění zásadně mění. Některé země se výrazně posunují v pořadí nahoru, což je výsledkem vysokých standardů hodnocení. Mezi takovéto země například patří Macao. Po korekci je na 4. místě, před korekcí je na 43. místě. Po korekci si tedy polepší o 39 míst ve srovnání s ostatními zeměmi. Mezi další země, které si výrazně v pořadí polepší je Lichtenštejnsko (po korekci na 25. místě, před korekcí na 63. místě, celkové zlepšení je tedy o 38 míst). Dále je to Hongkong (6, 41, zlepšení tedy o 35 míst), Estonsko (15, 48, zlepšení tedy o 33 míst), Lotyšsko (8, 32, zlepšení tedy o 24 míst), Finsko (28, 50, zlepšení tedy o 22 míst), Řecko (22, 42, zlepšení tedy o 20 míst), Korea (46, 64, zlepšení tedy o 18 míst), Polsko (45, 62, zlepšení tedy o 17 míst), Irsko (30, 47, zlepšení tedy o 17 míst).

Zjevně existují i země, jejichž pořadí se po očištění zhorší, což je výsledkem jejich nízkých standardů pro hodnocení. Mezi takové země patří Kostarika (po korekci je na 52. místě, před korekcí na 17. místě, pohorší si tedy o 35 míst), Uruguay (57, 25, pohorší si tedy o 32 míst), Brazílie (49, 21, pohorší si tedy o 28 míst), Katar (51, 23, pohorší si tedy o 28 míst), Spojené arabské emiráty (36, 9, pohorší si tedy o 27 míst), Argentina (58, 33, pohorší si tedy o 25 míst), Indonésie (29, 5, pohorší si tedy o 24 míst), Izrael (59, 36, pohorší si tedy o 23 míst), Chile (54, 34, pohorší si tedy o 20 míst) a Jordánsko (21, 3, pohorší si tedy o 18 míst).

Je vhodné analyzovat nejen změnu pořadí, ale také celkové rozdělení (percentuálně zastoupení jednotlivých kategorií škály) hodnocení učitelovy podpory žáků. Obrázek 12 ukazuje rozdělení hodnocení učitelovy podpory žáků v jednotlivých zemích po korekci o rozdílné užívání škály, které je generováno pomocí CHOPIT modelu. Pravděpodobnosti jednotlivých kategorií jsou

vypočítány pomocí následujícího vzorce $P(Y_{Si} = j) = \Phi\left(\frac{\tau_F^j - \mu_i}{\sigma_s}\right) - \Phi\left(\frac{\tau_F^{j-1} - \mu_i}{\sigma_s}\right)$, $j = 1, 2, 3, 4$, do něhož dosazujeme za τ_F^j odhadnutou mez fixní (referenční) země. Jako referenční země je zvoleno USA. Meze jsou tudíž pro všechny země stejné, což nám umožňuje korektní porovnání. Za μ_i dosazujeme odhadnutý koeficient β dané země a σ_s je pro identifikaci rovno jedné. Pro více podrobností ke specifikaci modelu odkazujeme na část 3.2 s popisem parametrického modelu. Percentuální zastoupení první kategorie (*rozhodně nesouhlasím*) se pohybuje od 0,2 (Šanghaj) po 9,3 (Nizozemsko). Nejvíce frekventovanými kategoriemi jsou však kategorie *souhlasím* (její zastoupení se pohybuje od 26,8 % po 48,1 %) a *rozhodně souhlasím* (zastoupení této kategorie se pohybuje od 15,4 % po 70,3 %).

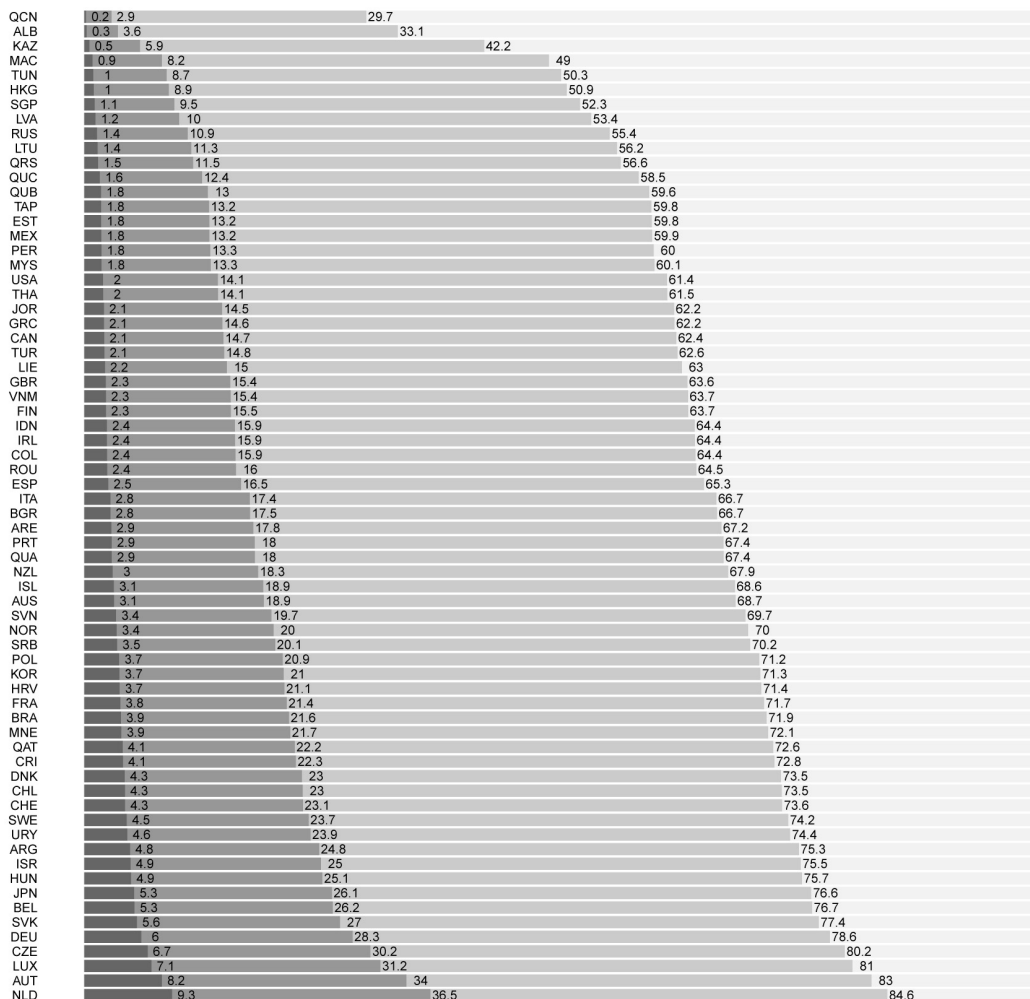
Obrázek 13 ukazuje rozdělení hodnocení učitelovy podpory žáků v jednotlivých zemích před korekcí o rozdílné užívání škály. Toto rozdělení je též generováno pomocí CHOPIT modelu. Pravděpodobnosti jednotlivých kategorií jsou vypočítány pomocí vzorce $P(Y_{Si} = j) = \Phi\left(\frac{\tau_i^j - \mu_i}{\sigma_s}\right) - \Phi\left(\frac{\tau_i^{j-1} - \mu_i}{\sigma_s}\right)$, $j = 1, 2, 3, 4$, do něhož dosazujeme za τ_i^j odhadnuté meze dané země i (ne tedy referenční země), za μ_i odhadnutý koeficient β dané země a σ_s je pro identifikaci rovno nule. Pro více podrobností opět odkazujeme na část 3.2 s popisem parametrického modelu.

Rozdělení před korekcí a po korekcí se pro mnoho zemí liší. Například pro Macao, které má největší posun v úrovni učitelovy podpory před korekcí a po korekcí (vypočítány na základě β parametrů probitového modelu pro ordinální proměnnou a CHOPIT modelu), se též zásadně mění celkové rozdělení. Relativní kumulativní četnosti v procentech jsou po korekcí následující: 0,9 (*rozhodně nesouhlasím*), 8,2 (*nesouhlasím*), 49 (*souhlasím*) a 100 (*rozhodně souhlasím*). Před korekcí jsou tyto kumulativní četnosti následující: 1,6 (*rozhodně nesouhlasím*), 13,8 (*nesouhlasím*), 71,5 (*souhlasím*) a 100 (*rozhodně souhlasím*). Před korekcí jsou tedy více zastoupeny nižší škálové kategorie, označující horší úroveň učitelovy podpory žáků. Opačným příkladem, kdy jsou před korekcí více zastoupeny vyšší škálové kategorie označující lepší úroveň učitelovy podpory a kdy tedy po korekcí úroveň učitelovy podpory klesne, je

Kostarika. Relativní kumulativní četnosti po korekci jsou následující: 4,1 (*rozhodně nesouhlasím*), 22,3 (*nesouhlasím*), 72,8 (*souhlasím*) a 100 (*rozhodně souhlasím*). Před korekcí jsou tyto kumulativní relativní četnosti následující: 3,3 (*rozhodně nesouhlasím*), 13,5 (*nesouhlasím*), 55,8 (*souhlasím*) a 100 (*rozhodně souhlasím*).

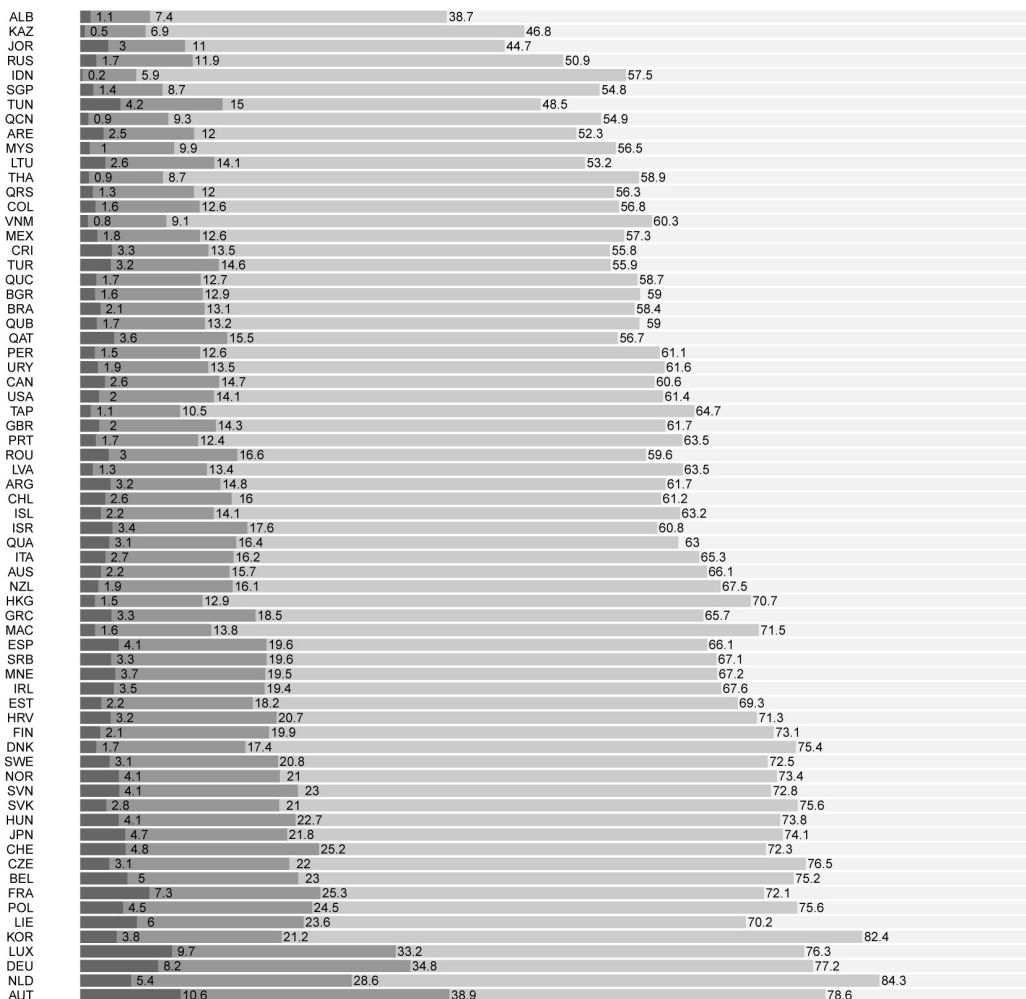
Pro Českou republiku se po korekci situace zhorší v tom smyslu, že naroste procentuální zastoupení nižších škálových kategorií označujících horší úroveň učitelovy podpory žáků. Relativní kumulativní četnosti jsou po korekci následující: 6,7 (*rozhodně nesouhlasím*), 30,2 (*nesouhlasím*), 80,2 (*souhlasím*) a 100 (*rozhodně souhlasím*), zatímco před korekcí je to 3,1 (*rozhodně nesouhlasím*), 22 (*nesouhlasím*), 76,5 (*souhlasím*) a 100 (*rozhodně souhlasím*).

Jak je výše uvedeno, pro korektní porovnání nezatížené rozdílným používáním škály volíme jednu fixní škálu, na níž vyjadřujeme rozdělení všech zemí. V našem případě jsme zvolili škálu USA. Kdyby tedy všechny země používaly škálu stejným způsobem jako USA, je jejich rozdělení učitelovy podpory takové, jaké je uvedeno na obrázku 12. Stejně tak by šla jako referenční země zvolit jakákoli jiná země. Interpretace korigovaného rozdělení je pak nutné učinit s ohledem na to, jakou referenční zemi volíme. Kdybychom například zvolili jako referenční zemi Českou republiku a vyjádřili rozdělení učitelovy podpory na této škále, bylo by rozdělení pro Českou republiku totožné před korekcí a po korekci. V našem případě je rozdělení před korekcí a po korekci totožné pro USA.



Obrázek 12. Rozdělení hodnocení učitelovy podpory žáků v jednotlivých zemích po korekci o rozdílné užívání škály

Pozn. Rozdělení hodnocení učitelovy podpory žáků jsou generována pomocí CHOPIT modelu. Pravděpodobnosti jednotlivých kategorií jsou vypočítány pomocí následujícího vzorce $P(Y_{si} = j) = \Phi\left(\frac{\tau_F^j - \mu_i}{\sigma_s}\right) - \Phi\left(\frac{\tau_F^{j-1} - \mu_i}{\sigma_s}\right)$, $j = 1, 2, 3, 4$, do něhož dosazujeme za τ_F^j odhadnutou mez fixní (referenční země). Jako referenční země je zvoleno USA. Dále za μ_i pak dosazujeme odhadnutý koeficient β dané země, σ_s je pro identifikaci rovno nule. Černou barvou je označena první kategorie (rozhodně nesouhlasím), nejsvětlejší odstín šedé označuje poslední, čtvrtou kategorii (rozhodně souhlasím).



Obrázek 13. Rozdělení hodnocení učitelovy podpory žáků v jednotlivých zemích před korekcí o rozdílné užívání škály

Pozn. Rozdělení hodnocení učitelovy podpory žáků jsou generována pomocí CHOPIT modelu. Pravděpodobnosti jednotlivých kategorií jsou vypočítány pomocí následujícího vzorce $P(Y_{si} = j) = \Phi\left(\frac{\tau_i^j - \mu_i}{\sigma_s}\right) - \Phi\left(\frac{\tau_i^{j-1} - \mu_i}{\sigma_s}\right)$, $j = 1, 2, 3, 4$, do něhož dosazujeme za τ_i^j odhadnuté meze dané země i . Dále pak za μ_i odhadnutý koeficient β dané země, σ_s je pro identifikaci rovno nule. Černou barvou je označena první kategorie (*rozhodně nesouhlasím*), nejsvětější odstín šedé označuje poslední, čtvrtou kategorii (*rozhodně souhlasím*).

6.5. Porovnání výsledků analýzy s dalšími studii

Jak je patrné z odhadů parametrů mezi modelem CHOPIT, je heterogenita ve stylu užívání škály mezi zeměmi významná. Hranice mezi jednotlivými kategoriemi jsme následně zakreslili na světové mapy (obrázky 8, 9 a 10). Z těchto obrázků je mimo jiné též patrné, že rozdíly v užívání škály do určité míry souvisí i s geografickou polohou. Černá barva na obrázcích znázorňuje vysoké standardy při užívání škály (volba nižších škálových kategorií). Dalším indikátorem celkově nízkých či vysokých standardů je součet pořadí jednotlivých mezí uvedených v tabulce 9. Dle tohoto kritéria jsou vysoké standardy zaznamenány především v některých evropských zemích jako Albánie, Lichtenštejnsko, Řecko, Estonsko, Irsko, Francie, Lotyšsko, Španělsko, Polsko a Finsko. Dále mezi tyto země patří i asijské země/regiony jako Šanghaj, Macao a Hongkong. Naopak nízké standardy jsou dle tohoto kritéria v zemích jižní Ameriky, jako je například Uruguay, Brazílie, Argentina a Chile. Dále sem patří i země jihovýchodní Asie, jako je Indonésie, ale i některé evropské země, jako je Bulharsko. Zajímavé je srovnání s analýzou Voňková et al. (2015), kde byly srovnávány země z výzkumu PISA 2012 též podle heterogenity ve stylu užívání škály s tím, že analýza byla číneňna s jinou sadou ukotvujících vinět, viněty se týkaly učitelova řízení třídy. Při analýze heterogenity ve stylu užívání škály lze v těchto dvou analýzách pozorovat shodné prvky. Země jižní Ameriky a jihovýchodní Asie byly při studiu užívání škály pomocí hodnocení učitelova řízení třídy též shledány jako země s nízkými standardy. V případě evropských zemí při analýze učitelova řízení třídy byly jako země s nejvyššími standardy zaznamenány země severní a střední Evropy, zatímco v naší analýze nejsou země s nejvyššími standardy soustředěny pouze v této části Evropy. Bulharsko je zemí, která vykazuje nízké standardy jak v analýze učitelova řízení třídy, tak v naší analýze učitelovy podpory žáků.

V naší analýze jsme zkoumali i země s tendencí volby extrémních a středových kategorií. I zde shledáváme podobnosti mezi analýzami učitelova řízení třídy a učitelovy podpory žáků. V obou analýzách patří mezi země s tendencí volby extrémních kategorií Lucembursko, Turecko a Tunisko. Co se týká tendence užívání středových kategorií v obou analýzách, je jasným příkladem Tchaj-pej.

Co se týká tendence užívat škálovou kategorii *rozhodně souhlasím* před kategorií *souhlasím*, patří mezi jasné příklady v obou analýzách Jordánsko, Katar a Spojené arabské emiráty. A naopak co se týká tendence užívat spíše *souhlasím* než *rozhodně souhlasím*, je pro obě analýzy jasným příkladem Šanghaj. Dalším výrazným společným závěrem obou analýz je, že žáci z Indonésie dávají signifikantně větší přednost užívání škálové kategorie *nesouhlasím* před kategorií *rozhodně nesouhlasím*.

Užívání škály je tedy do určité míry podobné v obou zkoumaných oblastech. Pro korektní očištění o odlišné užívání škály lze však doporučit užívat viněty, které popisují stejný koncept jako hodnotící otázky (více Primi et al., 2016; Voňková et al., 2017).

Korigované a nekorigované hodnocení učitelovy podpory žáků je vhodné porovnat s nějakým vnějším kritériem. V naší analýze jsme si zvolili výsledek v PISA 2012 testu matematické gramotnosti. Lze očekávat, že země s lepším výsledkem v testu budou vykazovat i vyšší úroveň učitelovy podpory žáků (žáci hodnotili podporu učitele matematiky). Korelace výsledku v testu matematické gramotnosti a nekorigovaného hodnocení učitelovy podpory (odhadnuté β parametry probitového modelu pro ordinální proměnnou) je záporná, konkrétně je rovna $-0,469$. Toto lze označit za minimálně neočekávané zjištění. Korelace s korigovaným hodnocením učitelovy podpory výrazně vzrůstá a nabývá kladné hodnoty $0,044$. Ani tuto korelaci nelze označit za vysokou, avšak je alespoň kladná. Co se týká korelací jednotlivých hranic mezi škálovými kategoriemi a výsledkem v testu, jsou všechny korelace kladné (korelace mezi první a výsledkem v testu je $0,245$, korelace mezi druhou a výsledkem v testu je $0,450$ a korelace mezi třetí a výsledkem v testu je $0,661$). Toto poukazuje na to, že země s lepšími výsledky v matematice mají i vyšší standardy pro hodnocení.

Ve studii Voňková a kol. (2015) jsou uvedeny korelace mezi (ne)korigovaným hodnocením učitelova řízení třídy (v PISA 2012 byly zařazeny viněty mimo učitelovy podpory žáků též k učitelovu řízení třídy) a výsledkem v testu matematické gramotnosti. Tyto korelace jsou do určité míry podobné výše uvedeným korelacím s tím, že korelace mezi korigovaným hodnocením a výsledkem v testu je výrazně větší. Konkrétní hodnoty korelací jsou

následující: korelace mezi nekorigovaným hodnocením učitelova řízení třídy a výsledkem v testu je rovna $-0,325$ a korelace mezi korigovaným hodnocením učitelova řízení třídy a výsledkem v testu je $0,452$.

Je nutné poznamenat, že přesnou míru učitelových kvalit – úrovně učitelovy podpory žáků a úrovně učitelova řízení třídy – je náročné získat. Zde využíváme hodnocení žáků. Hodnocení žáků se stává důležitým kritériem pro hodnocení učitelů. Například v sedmi státech v USA (Aljaška, Georgia, Havaj, Iowa, Massachusetts, New York a Utah) je žákovské hodnocení součástí systému hodnocení učitelů. Dalších pět států (Colorado, Connecticut, Mississippi, Missouri a Nové Mexiko) dovoluje žákovské hodnocení při hodnocení učitelů využívat. V těchto konkrétních případech však nejsou hodnocení žáků korigována o případné rozdílné užívání škály. V naší analýze jsme ukázali, že přinejmenším na úrovni zemí je tato korekce důležitá, korelace s výsledky žáků se výrazně zvýší po očištění o rozdílné užívání škály.

Bylo však i ukázáno, že rozdílné užívání škály hraje roli nejen při porovnávání zemí, ale i při porovnání různých škol. Na konferenci ECER 2016 prezentovala H. Voňková první závěry výzkumu vztahu vybraných charakteristik škol a učitelova řízení třídy v evropských zemích (Voňková, 2016). Zdroj dat je PISA 2012, otázky a viněty jsou stejné jako v analýze vztahu učitelova řízení třídy na úrovni zemí (v příloze 2 uvádíme tyto otázky a viněty). Oproti analýze učitelova řízení třídy na úrovni zemí (Voňková et al., 2015) jsou v této analýze na úrovni školy zahrnuty pouze evropské země ($N = 33$). Mezi vybranými charakteristikami škol byly zahrnuty následující: a) soukromé versus veřejné školy, b) počet obyvatel města či vesnice, kde se škola nachází, c) počet škol v blízkosti dané školy, do kterých mohou žáci případně též docházet (konkurence škol), d) počet žáků ve třídě v dané škole, e) využití výsledků žáků na dané škole v procedurách akontability (*accountability procedures*, o tzv. akontabilitě např. Veselý, 2011), f) změna platu učitelů na dané škole v závislosti na jejich hodnocení, g) podíl učitelů na dané škole se vzděláním ISCED 5A. Heterogenita ve stylu odpovídání žáků byla napříč školami s těmito charakteristikami významná. Před očištěním o rozdílné užívání škály nižší podíl učitelů se vzděláním ISCED 5A souvisel s lepší úrovní řízení třídy učitelem. V zásadě byl podíl kvalifikovaných učitelů nejsilnějším faktorem, ale vztah (kvalifikovanější učitelé zhoršují situaci) byl v rozporu s očekáváním. Po

očištění o rozdílné užívání škály hrál podíl těchto učitelů mnohem menší roli. Tento výsledek můžeme interpretovat tak, že žáci na školách s nižším počtem kvalifikovaných učitelů mají nižší standardy pro hodnocení úrovně řízení třídy učitelem (hodnotí situaci jako výbornou ve srovnání s jinými, kteří ji hodnotí pouze jako dobrou). Jako nejdůležitější proměnné (po očištění o rozdílné užívání škály) byly shledány *kompetice škol* a to, zdali je škola soukromá. Na soukromých školách, které o své žáky musely soutěžit s více školami, bylo řízení třídy nejlepší. Co se týká heterogenity ve stylu odpovídání, ve všech analyzovaných evropských zemích sehrálo očištění o rozdílné užívání škály při srovnání různých typů škol roli, výjimkou je pouze Lotyšsko, kde je vztah charakteristik školy a řízení třídy učitelem srovnatelný před korekcí a po korekci. Největší roli odlišného užívání škály bylo shledáno v následujících zemích: Slovinsko, Německo, Dánsko, Francie a Island.

V dalším výzkumu učitelovy podpory žáků by tedy bylo zajímavé učinit podobnou analýzu jako pro učitelovo řízení třídy na úrovni škol. Další analýzy by mohly směřovat i na úroveň žáků – analýza vztahu hodnocení ukotvujících vinět s charakteristikami žáků by mohla ukázat na základní zdroje heterogenity ve stylu užívání škály na úrovni jednotlivců.

7. Alternativní postupy korekce stylu odpovídání na položky

Primárním cílem metody ukotvujících vinět je korekce (sebe)hodnocení různých skupin respondentů o jejich odlišné užívání škály. Ke korekci odlišného užívání škály by mohly pomoci i přístupy popsané v této kapitole. Popíšeme zde čtyři přístupy (OECD, 2014; Buckley, 2009).

První z nich je identifikace **tendence k volbě určitých škálových kategorií** a následné očištění odpovědí o tuto tendenci. Tato technika stojí na předpokladu, že pokud má respondent tendenci k nějakému stylu odpovídání na položky (např. volí často škálové kategorie souhlasu), děje se tak bez ohledu na obsah otázek. Dále tato technika vychází i z toho, že systematická volba určitých kategorií (například kategorií souhlasu) odráží především tendenci respondenta takto zacházet se škálovými body a že je nutné o tuto systematickou volbu respondentovy odpovědi očistit. Nicméně příčina volby některých kategorií škály může jen odrážet reálnou situaci a není tudíž nutné o tuto tendenci následně data očišťovat.

Druhým přístupem je **technika přehánění** (*overclaiming technique*). Vychází z toho, že někteří lidé se snaží ukazovat své znalosti v mnohem lepším světle a u pojmů, které ve skutečnosti neexistují, konstatují, že je znají dobře. Respondentům je předložen seznam pojmů, u kterých se mají vyjádřit, do jaké míry jsou s nimi obeznámeni. V seznamu existujících pojmů se objevují i pojmy neexistující. Respondentova familiárnost s pojmy v dané oblasti je následně očištěna o familiárnost s neexistujícími pojmy.

Třetím přístupem je **hodnocení nucenou volbou** (*forced choice assessment*). Respondent je požádán o volbu jednoho z předložených tvrzení. Například může být respondent požádán zvolit z určité nabídky jeden způsob práce v matematice, který ho nejlépe vystihuje. Dále se může jednat i o volbu nejméně a nejvíce preferované možnosti, a nebo o seřazení všech možností (pořadové otázky). I kdyby měl respondent tendenci k volbě některých škálových kategorií a v případě hodnotících škál u každé alternativy by zvolil

například *souhlasím*, tak v případě *forced choice assessment* je doslova nucen zvolit jednu možnost či uvést (dílčí) pořadí možností.

Čtvrtým přístupem, jak redukovat mezikulturní rozdíly ve stylu odpovídání, je **test posouzení situace** (*situational judgement test*). Respondentům je předložen popis nějaké situace a v následných otázkách se vyjadřuje, jak by se v dané situaci zachoval. Oproti tradičním otázkám v osobnostních dotaznících na chování jedince se ukazuje být tento typ otázek méně citlivý na mezikulturní rozdíly. Popis a příklady těchto alternativ k metodě ukotvujících vinět jsou uvedeny v následujícím textu.

Je nutno dodat, že některé pojmy lze měřit i jiným způsobem než přímým dotazováním se respondenta, čímž se redukuje subjektivní posuzování situací a měření tudíž není zatíženo referenčním zkreslením (*reference bias*), sociálně žádoucími odpověďmi, tendencí k volbě určitých škálových kategorií, uváděním klamných odpovědí apod. Příkladem je měření nekognitivních dovedností dětí jako sebekontrola či vůle. Sebekontrolu lze měřit pomocí dotazníků, které mohou být distribuovány učitelům žáků či samotným dětem. Alternativou je však i provedení určitého úkolu a následné měření chování testovaného (*performance tasks*). Asi nejznámějším z nich je tzv. *marshmallow test* (Mischel, 2014) zaměřený na testování sebekontroly u předškolních dětí. Základem je vytvoření situace, v níž je dětem oddálena odměna. Duckworth a Yeager (2015) popisují tento test a jeho implikace následujícím způsobem. Na začátku testu jsou dětem nabídnuty různé bonbóny, čokoládové sladkosti či jiné pochutiny, mezi nimiž si dítě vybere své nejoblíbenější. Následně experimentátor ponechá pouze ty nejoblíbenější pochutiny a vytvoří dvě hromádky, jednu menší (např. jeden bonbón) a druhou větší (např. dva bonbóny). Dětem jsou pak nabídnuty dvě možnosti – vzít si ihned malou hromádku, nebo počkat (experimentátor může například odejít na chodbu, činit tam něco, co s celou věcí nesouvisí, a pak se vrátit) a vzít si větší hromádku. Otázkou primárně není, jakou volbu dítě učiní, ale jak dlouho je schopno čekat na větší odměnu. Doba čekání koreluje s rodičovským hodnocením sebekontroly, lepšími výsledky ve standardizovaných testech, nižší mírou lehkovážnosti a zdravou tělesnou vahou.

Performance tasks však také mají své nevýhody. Provedení zadaného úkolu v určité chvíli může být ovlivněno vnitřní motivací, emocemi a myšlenkami

děti. Například není zjevné, co je příčinou toho, že si dítě přestane hrát poté, co je o to dospělým požádáno. Je to důsledek jeho sebekontroly či to spíše souvisí s respektováním autority dospělého (Aronson & Carlsmith, 1963)? Ačkoli je úkol sám o sobě „objektivním“, interpretace chování v této situaci je subjektivní, čehož si musí být výzkumník vědom. Dále chování v takovéto uměle navozené situaci nemusí být vždy v souladu s chováním v každodenním životě. Další nevýhodou je citlivost takovýchto testů na faktory, které jsou irelevantní vůči cíli výzkumu (*task-impurity problem*). Duckworth a Yeager (2015) udává příklad Stroopova testu (*Stroop test*), který ukazuje, jak snadno je člověk rozptýlen svými automatickými reakcemi a návyky při provádění nějakého úkolu. Podstata spočívá v tom, že testovanému jsou předložena jména barev vytisklá v různé barvě. Někdy jméno barvy koresponduje s vytištěnou barvou (slovo *červená* je vytisknuto červeně), jindy tomu tak není (slovo *červená* je vytisknuto zeleně). V případě, že jméno barvy neodpovídá barvě tisku, trvá vyjmenování jmen barev déle a je chybovější. Navíc v takovémto případě je Stroopův test testem tzv. *executive functions* (zahrnují například kontrolu pozornosti, pracovní paměť, řešení problému apod). Ale v tomto úkolu nejsou požadovány pouze tyto funkce. Rychlý a přesný výkon vyžaduje též práci s barvami, verbální artikulaci, motivaci dávat pozor apod. Dále tyto úkoly mohou být zcela umělé ve smyslu, že člověk by se jim v reálné situaci zcela vyhnul. V případě *marshmallow test* není dětem povoleno vstát ze židle, hrát si s hračkami či číst knihy, přikrýt pochutiny talířem nebo ubrouskem atd. V reálné situaci by mohla být využita libovolná z těchto taktik k oddálení získání odměny.

V kontextu měření pomocí metody ukotvujících vinět a alternativních způsobů měření rozvíjí G. King diskuzi k měření některých konceptů v lékařství (King, 2017). Tato diskuze je inspirující i pro pedagogiku. Jeho diskuze se zaměřuje na to, zda je zapotřebí ukotvujících vinět v případě, kdy je dostupné přímé měření, např. nějaký lékařský test. Obecně neplatí, že pokud je dostupný lékařský test, dáme tomuto testu přednost. V některých případech může vést sebehodnocení korigované pomocí ukotvujících vinět k menší chybě měření než výsledek lékařského nebo nějakého fyzického testu. V některých případech může být lékařský test sice velmi přesný, ale náklady na sběr dat mohou být velmi vysoké. Alternativní způsoby mohou být relativně přesné a mnohem levnější. V některých případech by pro měření byl sice test dostupný, ale jeho užití by

neodpovídalo dodržení etických pravidel. Dotazníková šetření s využitím metody ukotvujících vinět mohou být řešením situace. Navíc dodejme, že objektivní testy jsou typicky velmi úzce zaměřené (množství cukru v krvi), zatímco slovní otázka je obecná (celkové zdraví) a lépe odráží, co chceme zjistit.

7.1. Tendence k volbě určitých škálových kategorií

Jeden z alternativních způsobů korekce odlišného užívání škály v pedagogickém výzkumu je tendence k volbě určitých škálových kategorií.

V tabulce 11 uvádíme shrnutí tendencí k volbě určitých škálových kategorií dle Baumgartner a Steenkamp (2001) včetně jejich definice, anglické terminologie, teoretického vysvětlení a způsobu měření.

Tabulka 11 *Tendence k volbě určitých škálových kategorií*

Styl odpovídání	Definice a anglická terminologie	Teoretické vysvětlení	Způsob měření
Tendence volit souhlasné kategorie	Tendence uvádět souhlasnou odpověď bez ohledu na obsah otázky. V anglické terminologii se užívá <i>acquiescence response style</i> (ARS), <i>yea-saying</i> nebo <i>positivity</i> .	<ul style="list-style-type: none"> Charakteristika určitých extrovertů, kteří mají tendenci impulzivně přijímat všechna tvrzení. Nekritické schvalování tvrzení respondenty s nízkými kognitivními schopnostmi nebo nízkým socio-ekonomickým statusem. Častěji se vyskytuje u otázek, které jsou nejednoznačné, vágní, nebo se týkají záležitostí, s nimiž si jsou respondenti nejistí. Častěji se vyskytují též v případech, kdy si respondent nemůže dostatečně promyslet odpověď kvůli vyrušování, nedostatku času apod. 	<p>Dva obecné přístupy:</p> <ul style="list-style-type: none"> Míra souhlasu s mnoha položkami, které jsou obsahově heterogenní (toto vyžaduje dostupnost různých baterií otázek, které jsou obsahově co nejvíce odlišné). Míra souhlasu jak s pozitivně tak s negativně formulovanými otázkami v rámci jedné baterie otázek. Speciální případ jsou páry položek, které jsou obsahově identické, ale jedna otázka je formulována pozitivně a druhá negativně.
Tendence volit nesouhlasné kategorie	Tendence uvádět nesouhlasnou odpověď bez ohledu na obsah	<ul style="list-style-type: none"> Charakteristika určitých introvertů, kteří se snaží vyhnout nějaké situaci. 	<ul style="list-style-type: none"> Stejně jako u tendence k souhlasu, až na to, že měříme míru nesouhlasu místo míry souhlasu.

	otázky. V anglické terminologii se užívá <i>disacquiescence response style</i> (DARS), <i>disagreement tendency</i> , <i>na-saying</i> nebo <i>negativity</i> .		
Tendence volit spíše souhlasné kategorie	Tendence uvádět spíše souhlasné odpovědi než nesouhlasné odpovědi. V anglické terminologii se užívá <i>net acquiescence response style</i> (NARS) a <i>directional bias</i> .	[Viz vysvětlení pro tendenci k souhlasu a tendenci k nesouhlasu.]	<ul style="list-style-type: none"> • Obecně míra souhlasu mínus míra nesouhlasu. Často měřená pomocí obsahově heterogenních položek.
Tendence volit extrémní kategorie	Tendence volit koncové body škály (extrémní kategorie) bez ohledu na obsah otázek. V anglické terminologii <i>extreme response style</i> (ERS).	<ul style="list-style-type: none"> • Odraz přisnosti, netolerance k nejednoznačnosti a dogmatismu. • Charakteristika respondentů s méně diferencovanými kognitivními strukturami. • Větší pro významné podněty (tj. podněty, které jsou pro respondenta důležité). 	<ul style="list-style-type: none"> • Počet či proporce heterogenních otázek, na něž respondent odpoví volbou extrémních kategorií na předložené škále. Otázky by měly být nekorelované a měly by mít stejnou proporcii extrémních odpovědí. Průměrná odpověď na otázku by se navíc měla pohybovat kolem středu škály.
Tendence volit určitý rozsah odpovědí	Tendence volit úzký či naopak široký rozsah odpovědí kolem středové kategorie. V anglické terminologii <i>response range</i> (RR).	[Lze odvodit z tendence volit extrémní kategorie.]	<ul style="list-style-type: none"> • Standardní odchylka odpovědí na obsahově heterogenní otázky.
Tendence volit středové kategorie	Tendence volit středové kategorie bez ohledu na obsah otázek. V anglické terminologii	<ul style="list-style-type: none"> • Kvůli vyhýbání se odpovědi (touze skrýt pravý názor), neschopnosti se rozhodnout nebo indiferenci (nedostatků zájmu o téma). 	<ul style="list-style-type: none"> • Počet nebo proporce heterogenních otázek, u nichž respondent volí při odpovědi středovou kategorii.

	<i>midpoint responding</i> (MPR).		
Náhodný styl odpovídání	Tendence odpovídat na otázky nedbale, náhodně či nesmyslně. V anglické terminologii <i>noncontingent responding</i> (NCR).	<ul style="list-style-type: none"> Kvůli nedostatku motivace číst instrukce a interpretovat otázky řádným způsobem. 	<ul style="list-style-type: none"> Součet absolutních rozdílů mezi odpověďmi na dvojici otázek, které jsou silně korelovány, vykazují v průměru podobnou odpověď.

Pozn. Upraveno dle Baumgartner a Steenkamp (2001).

V literatuře jsou nejčastěji zmiňované čtyři styly odpovídání na položky: tendence volit souhlasné kategorie (*acquiescence response style*, ARS), tendence volit nesouhlasné kategorie (*disacquiescence response style*, DARS), tendence volit extrémní kategorie (*extreme response style*, ERS) a náhodný styl odpovídání (*noncontingent responding*, NCR).

Buckley (2009) popisuje způsob měření těchto stylů. Vychází z Bachman a O'Malley (1984) a Baumgartner a Steenkamp (2001). Pomocí dat z šetření PISA 2006 ukazuje, jak ignorování odlišného užívání škály napříč různými zeměmi vede k zavádějícím statistikám. Nejprve bylo pomocí ad hoc analýzy faktorové struktury a korelací sady otázek z žákovského dotazníku PISA 2006 týkajících se potěšení studovat přírodní vědy (*science enjoyment*, 5 otázek), hodnoty přírodních věd (*science value*, 10 otázek), environmentální zodpovědnosti (*environmental responsibility*, 7 otázek), užitečnosti pro kariéru v přírodních vědách (*usefulness for science career*, 4 otázky), přírodních věd v budoucnu (*science in future*, dvě sady otázek po 4 a 5) a učení se přírodním vědám (*science learning*, 6 otázek) vybráno pět heterogenních otázek (průměrná korelace mezi odpověďmi na tyto otázky je 0,158). Všechny otázky byly měřené na čtyřbodové Likertově škále: 1) *rozhodně souhlasím*, 2) *souhlasím*, 3) *nesouhlasím*, 4) *rozhodně nesouhlasím*. Následně jsou míry tendence volit souhlasné kategorie (ARS) a tendence volit nesouhlasné kategorie (DARS) vypočítány tímto způsobem:

$$\widehat{ARS}_i = \frac{1}{n_q} \sum_{j=1}^{n_q} 1[x_{ij} = 1],$$

$$\widehat{DARS}_i = \frac{1}{n_q} \sum_{j=1}^{n_q} 1[x_{ij} = 4],$$

kde i označuje studenta, n_q je počet otázek (v našem případě $n_q = 5$), x_j jsou odpovědi na otázky j , $1[\cdot]$ je charakteristická funkce (nabývá hodnoty 1 nebo 0 dle toho, zda je splněno nebo nesplněno kritérium v závorce). V případě míry \widehat{ARS} se jedná o podíl odpovědí *rozhodně souhlasím* v rámci pěti vybraných heterogenních otázek. V případě \widehat{DARS} se analogicky jedná o podíl odpovědí *rozhodně nesouhlasím*.

Míra tendence volit extrémní kategorie (ERS) je pak pro studenta i součtem uvedených měř ARS a DARS:

$$\widehat{ERS}_i = \widehat{ARS}_i + \widehat{DARS}_i.$$

Pro výpočet náhodného stylu odpovídání (NCR) je nutno postupovat odlišně. Watkins a Cheung (1995) navrhují konstruovat tuto míru pomocí takových dvojic otázek, na něž jsou odpovědi silně korelovány, mají podobný průměr a jsou skórovány stejným způsobem. Buckley (2009) zvolil pět párů takovýchto otázek, jejichž korelace se pohybovala v rozmezí 0,7165 – 0,7925, všechny byly vybrány ze sad otázek na využití přírodních věd. Po volbě těchto otázek je výpočet NCR již jednoduchý:

$$\widehat{NCR}_i = \sum_{j=1}^{n_q} |x_{ij} - y_{ij}|,$$

přičemž n_q je zde počet párů odpovědí na velmi podobné otázky x_j a y_j u studenta i (v našem případě $n_q = 5$). Myšlenka je zde taková, že pokud se odpovědi na velmi podobné otázky od sebe velmi liší, může to poukazovat na chybový či náhodný styl odpovídání na položky.

Dle hodnot ARS, DARS, ERS a NCR lze země z výzkumu PISA 2006 rozdělit do následujících skupin (Buckley, 2009):

- **Vysoké a nízké hodnoty ARS:** Mezi zeměmi z výzkumu PISA 2006

vykazují velkou tendenci volit souhlasné kategorie (vysoké hodnoty ARS) Jordánsko, Tunisko a Katar. Naopak nízké hodnoty ARS mají např. Nizozemsko, Izrael, Dánsko, Norsko, Švédsko a Česká republika. Studenti v těchto zemích tedy nemají tendenci odpovídat souhlasně na otázky týkající se různých oblastí.

- **Vysoké a nízké hodnoty DARS:** Co se týká tendence volit nesouhlasné kategorie (DARS), odlehlým pozorováním je Japonsko, má signifikantně největší míru tendence volit nesouhlasné kategorie. S odstupem za ním následují Austrálie a Korea. Nejnižší hodnoty DARS a tedy nejmenší tendence volit nesouhlasné kategorie má Thajsko, Portugalsko a Rusko.
- **Vysoké a nízké hodnoty ERS:** S vysokými hodnotami ARS a DARS souvisí i tendence volit extrémní kategorie (ERS). Studenti v Kataru, Jordánsku, Tunisku a Tchaj-peji mají největší tendenci užívat obou koncových bodů škály bez ohledu na obsah otázek. Naopak nejmenší tendenci užívat koncové body škály mají Dánsko, Izrael, Nizozemsko, Lotyšsko a Česká republika.
- **Vysoké a nízké hodnoty NCR, srovnání s ERS:** Zajímavé je srovnání tendence volit extrémní kategorie (ERS) s náhodným stylem odpovídání (NCR) – obě míry dosahují vysokých hodnot v Kataru, Tunisku a Jordánsku, zatímco Tchaj-pej dosahuje vysokých hodnot ERS, NCR je zde jedna z nejnižších ze všech studovaných zemí. To poukazuje na náhodné a nedbalé vyplňování dotazníku v Kataru, Tunisku a Jordánsku, zatímco studenti v Tchaj-peji jsou ve svých odpovědích konzistentní, na podobné otázky odpovídají podobným způsobem (tj. volí podobné škálové kategorie), přičemž se často přiklánějí k volbě extrémních kategorií.

Buckley (2009) s odkazem na Technickou zprávu PISA 2006 (OECD, 2009) poukazuje na paradoxní hodnoty negativních korelací na úrovni zemí mezi výsledkem v testu přírodovědné gramotnosti s otázkami týkajícími se postojů k přírodním vědám, jako je zájem o přírodní vědy, potěšení ze studia přírodních věd, názoru na obecnou hodnotu přírodních věd a názoru na osobní přínos přírodních věd. Podobné výsledky byly zároveň zaznamenány i pro čtenářskou

gramotnost v PISA 2000 a matematickou gramotnost v PISA 2003. Jedna z možných interpretací je, že se jedná o zkreslení způsobené agregací (*aggregation bias*) – vztah mezi proměnnými studovaný pomocí agregovaných proměnných (v tomto případě na úrovni zemí) neodpovídá vztahu mezi proměnnými na úrovni jednotlivců v těchto skupinách. V rámci jednotlivých zemí může platit, že čím vyšší výsledek v testu, tím větší zájem o daný předmět, tj. korelace mezi výsledkem v testu a zájmem je kladná. Oproti tomu po zprůměrování výsledku v testu v jednotlivých zemích (tj. výpočtu průměrného výsledku v zemi 1, 2, 3 atd.) a odpovědi ohledně zájmu o daný předmět a následném výpočtu korelace mezi těmito průměrnými hodnotami může vyjít korelace záporná.

V našem kontextu však je zásadní otázkou, zdali k vychýleným odhadům a matoucím závěrům ohledně vztahu mezi výsledky studentů v testech z určitých předmětů a jejich postoji k učení těchto předmětů vedou rozdíly v užívání škály. Buckley (2009) užívá jednoduchý klasický model a předpokládá lineární vztah mezi odpověďmi na postojové otázky a výše definovanými styly odpovídání na položku. Označme y_i pozorované průměrné skóre (průměrnou hodnotu všech odpovědí v sadě otázek k jedné oblasti, např. k zájmu o přírodní vědy) studenta i a dále y_i^* nepozorované pravdivé skóre. Model pak zapíšeme jako

$$y_i = y_i^* + \beta_1ARS_i + \beta_2DARS_i + \beta_3NCR_i.$$

Odhad y_i^* lze získat odhadem lineárního regresního modelu

$$y_i = \beta_0 + \beta_1\widehat{ARS}_i + \beta_2\widehat{DARS}_i + \beta_3\widehat{NCR}_i + \epsilon_i$$

a následným výpočtem

$$\widehat{y}_i^* = \widehat{\beta}_0 + \widehat{\epsilon}_i.$$

Po odhadu tohoto modelu můžeme porovnat vztah mezi výsledkem v testu přírodovědné gramotnosti a postojem k přírodním vědám, který je jednak nekorigovaný (průměrná odpověď na postojové položky) a jednak korigovaný (využije se odhad nepozorovaného pravdivého skóre pomocí výše uvedeného modelu). Buckley (2009) potvrdil paradoxní lineární vztah mezi výsledkem v testu a nekorigovanými postoji – čím vyšší průměrné hodnoty skóre v zemích,

tím horší postoj. Na druhou stranu vztah mezi korigovanými postoji a výsledkem v testu se ukázal být komplikovanější – nelze potvrdit lineární vztah (k analýze vztahu mezi těmito proměnnými byla použita nejen lineární regrese, ale též regrese s kvadratickým členem a lokální regrese), vztah mezi proměnnými je kvadratický. Například pro škálu hodnota přírodních věd je vztah s výsledkem v testu kvadratický. Na jedné části škály tedy platí očekávaný vztah – s klesajícím výsledkem v testu se postoje k přírodním vědám zhoršují. V další části škály však tento vztah neplatí.

Podobným způsobem lze analyzovat vztah nejen na úrovni zemí, ale také na úrovni studentů v rámci jednotlivých zemí. V předchozích analýzách se ukazovalo, že na úrovni studentů v jednotlivých zemích neplatí paradoxní vztah, tj. platilo, čím lepší výsledek v testu, tím lepší jsou postoje studentů. Toto Buckley (2009) sice potvrzuje, ale jen pro nekorigované skóre postojů. Pro korigované skóre postojů pomocí výše uvedeného modelu se ukazuje vztah s výsledkem v testu komplikovanější – v uvažovaných zemích (Japonsko, Tunisko a USA) je vztah podobně jako pro agregované proměnné kvadratický s maximem uprostřed škály.

Nelze tedy dále tvrdit, že na úrovni studentů v jednotlivých zemích je vztah lineární a korelace nabývá očekávaných záporných hodnot (vyšší hodnota na škále postojů značí horší postoj, vyšší skór v testu značí lepší znalosti) a že na úrovni zemí je korelace kladná. Vztah mezi proměnnými je po korekci o užívání škály jak na úrovni zemí, tak v rámci jednotlivých zemí komplikovanější, většinou má kvadratický tvar s maximem pro mediánovou hodnotu postojů.

7.2. Technika přehánění (*overclaiming technique*)

Technika přehánění (*overclaiming technique*) je technika, která může být užita pro odhad obeznámenosti respondenta s neexistujícími koncepty (OECD, 2014). Technika přehánění vychází ze signálně-detekční teorie (*signal detection theory*), která se v psychologii využívá pro studium rozhodovacích procesů za podmínek nejistoty. Příkladem může být určování vzdáleností v přítomnosti mlhy (Ivancevic et al., 2009). Určení vzdálenosti objektu je založeno na

vizuálním zpracování, které je však *rušeno* přítomností mlhy. Mlha redukuje jasné kontury objektu, a proto může být objekt vnímán v mnohem větší vzdálenosti, než ve skutečnosti je. Pokud aplikujeme tuto teorii na situaci, kdy podnět buď je či naopak není přítomen a pozorující klasifikuje každý podnět, že buď byl či nebyl přítomen, pak lze odpovědi pozorujícího třídít způsobem uvedeným v tabulce 12.

Tabulka 12 *Signálně-detekční teorie – podnět a odpověď přítomny či nepřítomny*

	Odpověď: podnět není přítomen	Odpověď: podnět je přítomen
Podnět přítomen	Minutí (<i>miss</i>)	Zásah (<i>hit</i>)
Podnět není přítomen	Správné zamítnutí (<i>correct rejection</i>)	Planý poplach (<i>false alarm</i>)

Tato teorie je používána i v experimentech zaměřených na paměť (např. Stanislaw a Todorov, 1999). Testovanému je například předložen seznam pojmů k nastudování. Následně je vytvořen nový seznam kombinující *staré* pojmy s novými pojmy, které předtím na seznamu nebyly. Testovaný se pak u každého pojmu z nového seznamu vyjadřuje, zdali byl na studijním seznamu či nikoli. Odpovědi testovaného lze pak klasifikovat dle výše uvedené logiky: a) pojem byl na seznamu, testovaný však konstatuje, že tomu tak nebylo (*miss*), b) pojem byl na seznamu a testovaný toto skutečně konstatuje (*hit*), c) pojem nebyl na seznamu, testovaný konstatuje, že tam pojem skutečně nebyl (*correct rejection*), d) pojem nebyl na seznamu, avšak testovaný konstatuje, že tam byl (*false alarm*).

Tendence nadhodnocování znalostí se měří v dotazníkových šetřeních pomocí otázek na znalost neexistujících konceptů. Například žákovský dotazník použitý v PISA 2012 (OECD, 2017b) zahrnoval matematické pojmy, které ve skutečnosti neexistují. Žáci pak byli dotázáni, do jaké míry uvedené pojmy znají. Konkrétně se jednalo o otázku s následujícím zněním: *Přečti si následující seznam matematických pojmů a uveď, do jaké míry je znáš.* Následoval výčet šestnácti pojmů, z nichž tři ve skutečnosti neexistují. Mezi existující pojmy zahrnuté v tomto seznamu patří například *exponential function*,

divisor, quadratic function, linear equation, rational number, radicals, polygon, congruent figure, cosine, arithmetic mean a probability. Mezi pojmy, které v matematice neexistují, byly zahrnuty *subjunctive scaling, declarative fraction a proper number*. Škála, na níž žáci v této otázce odpovídali, byla pětibodová od 1) *never heard of it* po 5) *know it well, understand the concept*.

Overclaiming index, počítaný jako průměrná hodnota odpovědí na obeznámenost se třemi neexistujícími matematickými pojmy (*foil familiarity*), byl variabilní napříč zeměmi. He a van de Vijver (2016) udávají hodnoty tohoto indexu pro země z výzkumu PISA 2012. Nejnižších hodnot nabývá pro Koreu (1,41), Švédsko (1,61), Island (1,72), Finsko (1,74), Tchaj-pej (1,74), Vietnam (1,84) a Českou republiku (1,98). Naopak nejvyšších hodnot, tj. největší obeznámenosti s neexistujícími pojmy, dosáhly Albánie (3,37), Jordánsko (3,22), Kazachstán (3,02), Thajsko (2,97), Katar (2,94) a Indonésie (2,90). Dodejme, že Heine a kol. (1999) obecně poukazují na to, že lidé z východoasijských zemí jsou více sebekritičtí než lidé ze západní Evropy či Severní Ameriky. Bylo by zajímavé toto porovnat s výsledkem analýzy obeznámenosti s neexistujícími koncepty.

V datovém souboru PISA 2012 lze najít dvě proměnné obeznámenosti s matematickými koncepty. Jedná se nejprve o nekorigovanou obeznámenost s matematickými koncepty (FAMCON, *familiarity with mathematical concepts*), která je průměrem odpovědí na všech šestnáct otázek ohledně znalosti matematických konceptů, tj. existujících i neexistujících konceptů. Další proměnnou je očištěná obeznámenost s matematickými koncepty (FAMCONC, *adjusted topic familiarity*), která je rozdílem mezi FAMCON a *foil familiarity* (OECD, 2014). V průměru byla obeznámenost s matematickými koncepty rovna 3,6, obeznámenost s neexistujícími koncepty byla signifikantně nižší, a to 2,3 (Kyllonen & Bertling, 2014a).

Někteří autoři se snaží pomocí techniky přehánění vysvětlit paradoxní negativní vztah mezi motivací žáků učit se matematiku a výsledkem v testu matematické gramotnosti na úrovni zemí. Například He a van de Vijver (2016) počítali vztah mezi vnitřní/vnější motivací a výsledkem v testu před očištěním a po očištění o obeznámenost s neexistujícími koncepty. Mimo techniky přehánění použili též očištění o rozdílný styl odpovídání (konkrétně ERS (*extreme response style*) a MPR (*midpoint responding*)) a metodu ukotvujících vinět (výsledky k metodě

ukotvujících vinět jsou v části 2.2.2). Výsledky užití uvedených technik jsou shrnuty v tabulce 13.

Tabulka 13 *Korelace mezi motivací učit se matematiku a výsledkem v testu matematické gramotnosti po očištění o rozdílné užívání škály*

Korelace s výsledkem v testu	Individuální úroveň		Úroveň zemí	
	Vnitřní motivace	Vnější motivace	Vnitřní motivace	Vnější motivace
Před očištěním	0,17	0,13	-0,53	-0,52
Po očištění o ERS a MPR	0,15	0,11	-0,45	-0,44
Po očištění pomocí techniky přehánění	0,20	0,15	-0,32	-0,45
Po očištění pomocí ukotvujících vinět	0,26	0,20	-0,11	-0,04

Pozn. Zdroj He a van de Vijver (2016).

Na úrovni studentů v jednotlivých zemích je mediánová hodnota korelací rovna 0,17 pro vnitřní motivaci a 0,13 pro vnější motivaci. Po očištění pomocí techniky přehánění se tato hodnota ještě mírně zvýší: z 0,17 na 0,20 (vnitřní motivace) a z 0,13 na 0,15 (vnější motivace). Na úrovni zemí dojde k ještě vyššímu zvýšení: z -0,53 na -0,32 (vnitřní motivace) a z -0,52 na -0,45 (vnější motivace). Jak je ukázáno v tabulce 13, největší potenciál ze zkoumaných metod ve smyslu posunu k očekávanému kladnému vztahu mezi motivací a výsledkem v testu má metoda ukotvujících vinět.

Technika přehánění byla též užitá na samotné porovnání výsledku v testu matematické gramotnosti a obeznámenosti s matematickými koncepty. Kyllonen a Bertling (2014a) ukazují, že průměrná korelace mezi obeznámeností s matematickými pojmy a výsledkem v testu matematické gramotnosti je na úrovni studentů v jednotlivých zemích z výzkumu PISA 2012 rovna 0,45 před očištěním (korelace mezi FAMCON proměnnou a výsledkem v testu matematické gramotnosti) a 0,44 po očištění (korelace mezi FAMCONC

proměnnou a výsledkem v testu). Tyto korelace jsou tedy v zásadě totožné. Na úrovni zemí je však rozdíl zcela zásadní. Před očištěním je tato korelace rovna 0,17, zatímco po očištění je to 0,58.

7.3. Hodnocení nucenou volbou (*forced choice assessment*)

Některé problémy otázek s hodnoticími škálami mohou být řešeny pomocí hodnocení nucenou volbou (*forced choice assessment*). Respondent volí jedno z předložených tvrzení, které nejlépe vystihuje jeho situaci. Jednotlivá tvrzení tedy nejsou hodnocena na jednotlivých hodnoticích škálách, ale respondent vybírá jedno nejvýstižnější tvrzení. Příklad takovéto otázky v žákovském dotazníku PISA 2012 na strategie učení se matematice je následující (v OECD (2017a) je označena jako ST53Q01):

Z každé trojice tvrzení vyber jedno, které nejlépe vystihuje tvůj způsob práce v matematice. Z trojice zaškrtni pouze jeden čtvereček.

- Když se učím na prověrku z matematiky, snažím se určit si, které části učiva jsou nejdůležitější.*
- Když se učím na prověrku z matematiky, nové pojmy v matematice se snažím pochopit tak, že si je dávám do souvislosti s tím, co už znám.*
- Když se učím na prověrku z matematiky, snažím se naučit se co nejvíc věci z paměti.*

V této otázce si student vybírá jednu ze tří strategií, které používá při přípravě na prověrku z matematiky: určení nejdůležitějších částí učiva, vztahování k osvojeným pojmům a pamětní učení. Příklady dalších otázek na strategie učení s nucenou volbou jednoho nejvýstižnějšího popisu lze najít v příloze 4. Obecně lze strategie učení v těchto otázkách rozdělit do třech oblastí – strategie učení se z paměti (*memorization*), strategie kontroly (*control strategies*) a strategie elaborace (*elaboration*).

Nejjednodušší formou těchto otázek je volba jedné ze dvou možností. Dalším typem jsou pořadové otázky, kde respondent udává pořadí u všech nabízených alternativ. Další možností je uvedení pouze nejvíce a nejméně preferované alternativy, tj. jedná se pouze o uvedení parciálního pořadí.

Při volbě jednoho nejuvýstižnějšího tvrzení se zjevně redukuje problém s rozdílným užíváním hodnoticích škály, jako je například tendence volit kategorie souhlasu (ARS) a kategorie nesouhlasu (DARS), tendence volit extrémní kategorie (ERS), tendence volit rámec odpovědi (RR) a tendence volit středovou kategorii (MPR). Respondent, který by vykazoval nějakou z uvedených tendencí při hodnocení jednotlivých tvrzení na hodnoticích škále (např. by rozhodně souhlasil se všemi tvrzeními, tj. vykazoval ARS), je v případě volby jednoho nejuvýstižnějšího tvrzení nucen volit právě jednu možnost.

Takovýto typ otázek dále vyžaduje větší pozornost respondenta, neboť se musí zamýšlet nad všemi možnostmi najednou, držet je v paměti a volit tu nejvhodnější. To může být výhodou, neboť to vede k hlubšímu promyšlení volby odpovědi. Zároveň to však může být i nevýhodou, neboť respondent nemusí být ochoten o otázce přemýšlet déle, což ho nakonec může od odpovědi odradit (je nutné volit únosný počet únosně dlouhých tvrzení).

Bylo též ukázáno, že takovéto otázky jsou více odolné vůči udávání klamných odpovědí a také přispívají ke zvýšení validity měření daného konceptu ve srovnání s otázkami s hodnoticími škálami. Jackson, Wroblewski a Ashton (2000) například ukázali, že v osobnostních dotaznících je uvádění klamných odpovědí častější v situacích, kdy se lidé uchází o zaměstnání, ve srovnání se situacích, kdy už jsou v zaměstnání. Redukce tohoto nežádoucího efektu je redukována, pokud jsou otázky prezentovány ve formě *forced choice assessment*. Korelace dotazníkových odpovědí týkajících se kontraproduktivního chování na pracovišti s následným pozorovaným nežádoucím chováním na pracovišti (např. pozdní příchody, užívání alkoholu a vyhýbání se práci) byla při použití *forced choice assessment* vysoká.

Další věcí je, že formulace zadání takovéto otázky (vyber jedno z následujících tvrzení) definuje požadavek pro respondenta mnohem jasněji než v případě zadání otázek s hodnoticími škálami (do jaké míry souhlasíš s následujícím

tvrzením) (OECD, 2014).

Forced choice assessment má i přes své výhody též několik nevýhod. V první řadě, jak již vyplývá z pojmenování tohoto typu otázek, se jedná o riziko, které je spojeno s nucením respondenta vybrat si z předložené nabídky alternativ. Jednak předložená nabídka nemusí být vyčerpávající a respondent nenalezne pro sebe nejvhodnější odpověď. Dále pak může být objektivně indiferentní mezi nabízenými možnostmi a nabízená škála mu nedovolí dát více alternativ na stejnou úroveň (v případě Likertovy škály by vykázal stejnou preferenci na možnosti, mezi nimiž je indiferentní). Dále nezjistíme, do jaké míry je zvolená možnost v *forced choice assessment* výstižná pro respondenta, tj. nemáme k dispozici absolutní hodnocení. Zvolená možnost nemusí respondenta dobře vystihovat, je pouze nejvhodnější možností z předložené nabídky. Také může hrát roli pořadí nabízených možností, speciálně pokud jsou dlouhé, respondent nemusí dočíst všechny do konce.

Hodnocení nucenou volbou pomohlo podobně jako ukotvující viněty vyřešit též paradoxní vztah mezi výsledkem v testu matematické gramotnosti z výzkumu PISA 2012 a proměnnými popisujícími strategie učení se matematice. Kyllonen a Bertling (2014a) udávají, že korelace na úrovni zemí mezi výsledkem v matematice a strategií kontroly byla $-0,47$ při využití otázek s Likertovou škálou a $0,60$ při využití hodnocení nucenou volbou.

7.4. Test posouzení situace (*situational judgement test*)

Test posouzení situace (*situational judgement test*) hodnotí schopnost jedince volit vhodná řešení situací v běžném životě. V dotazníkových otázkách tohoto typu je prezentována situace a následují otázky na řešení této situace. Příkladem může být otázka z dotazníku PISA 2012 týkající se problémové situace (v OECD (2017a) otázky kódovány jako ST101Q01, ST101Q02, ST101Q03 a ST101Q05):

Představ si, že chceš jít s bratrem do zoo. Nevíš ale, jak se tam dostanete.

Co uděláš? U každého návrhu vyber možnost, která tě nejlépe vystihuje.

(V každém řádku zaškrtni pouze jeden čtvereček.)

	<i>Určitě bych to udělal/a</i>	<i>Pravděpodobně bych to udělal/a</i>	<i>Pravděpodobně bych to neudělal/a</i>	<i>Určitě bych to neudělal/a</i>
<i>Podívám se, jestli je cesta popsaná v brožurce o zoo.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vyhledám nejlepší trasu na mapě.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Nechám na bratrovi, aby zjistil, jak se tam dostaneme.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vím, kde přibližně zoo je, a navrhnu tedy, abychom vyrazili.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

V tomto dotazníku byly ještě další dvě problémové situace týkající se problému s odesláním textové zprávy na mobilu a užitím automatu na prodej jízdenek. Celé znění těchto otázek je uvedeno v příloze 5. Řešení problémových situací v těchto otázkách odráží tři strategie – systematické řešení (*systematic strategies*), nesystematické řešení (*unsystematic strategies*) a řešení spočívající v pomoci dalších osob (*seeking help*).

Test posouzení situace může být vhodný pro porovnávání různých etnik. Zároveň je méně závislý na kognitivních schopnostech jedince než tradiční otázky (OECD, 2014). Často se používají při přijímání zaměstnanců (Lievens & Soete, 2015). Uchazeči o zaměstnání jsou prezentovány hypotetické a náročné situace, které mohou na pracovišti vzniknout. Následně jsou mu předkládána různá řešení situace, z nichž pak vybírá to nejvhodnější. Tyto testy se typicky

týkají řešení problémových situací, rozhodovacího procesu a dovedností při jednání s lidmi. Formát odpovědí u otázek zaměřených na chování v dané situaci typicky bývá: a) výběr nejlepší možnosti nebo nejlepší a zároveň nejhorší možnosti a případně seřazení všech možností, b) uvedení u každé možnosti, zda je přijatelná či nikoli, c) hodnocení na Likertově škále.

Příkladem otázky, která může být položena při vstupu do zaměstnání, je následující (přeloženo z WikiJob, 2017):

Na konci hektického dne v práci shodou okolností pošleš e-mail s důvěrnou informací o jednom z klientů nesprávné osobě.

Kterou z následujících možností by bylo nejlepší učinit?

A – odejít z kanceláře a řešit problém zítra

B – rozhodnout se přehlédnout chybu, poslat e-mail správné osobě a nechat takto celou záležitost být

C – ihned poslat další e-mail „nesprávné“ osobě, nebo pokud je to možné, tak jí zavolat a vysvětlit chybu. Pak poslat e-mail správné osobě.

D – najít svého manažera, vyvětlit mu, co se stalo a nechat ho zabývat se tímto problémem

Chan a Schmitt (1997) také diskutovali rozdíl mezi bělochy a černochoy v testech posouzení situace, přičemž situace byly prezentovány pomocí videa či slovním popisem. Rozdíly mezi participanty různých kultur se výrazně zmenšily v případě, kdy byly situace prezentovány pomocí videa. Dle Patterson a kol. (2012) vykazují testy posouzení situace dobrou reliabilitu a predikční validitu při zjišťování různých osobnostních vlastností, jako je empatie.

8. Závěr

Využití metody ukotvujících vinět není v pedagogickém výzkumu prozatím běžné. Je však pravděpodobné, že její rozšíření výrazně vzroste, a to z několika klíčových důvodů. Zaprvé je v jiných sociálně vědních oblastech demonstrován její potenciál pro zvyšování validity měření jevů. Zadruhé, mnoho studií v rámci pedagogického výzkumu poukazuje na problematiku rozdílného užívání škály jednotlivých respondentů, které ohrožuje porovnávání různých skupin (Clarke, 2001; Chen et al., 1995; Watkins & Cheung, 1995). Metoda ukotvujících vinět je slibnou odpovědí na tuto otázku. Zatřetí jsou ukotvující viněty zařazeny do žákovských dotazníků mezinárodních srovnávacích šetření PISA 2012 i PISA 2015, což je z praktického pohledu důležité, neboť výzkumníci z celého světa mohou s daty pracovat a řešit neporovnatelnost údajů o žácích nejen na mezinárodní úrovni, ale také v rámci jednotlivých států.

K rozšíření užití této metody by měla přispět i tato publikace. V prvních dvou kapitolách jsou popsány teoretické základy této metody a její aplikace v pedagogice a dalších sociálních vědách. Ve třetí a čtvrté kapitole jsou popsány statistické přístupy metody, jako je neparametrický přístup a parametrický model CHOPIT s následnou ukázkou zpracování dat ve statistických softwarech. Pátá kapitola se věnuje formulaci vinět a jejich umístění v dotazníku, diskutují se v ní též podrobně dva klíčové předpoklady metody – konzistence odpovědí a ekvivalence vinět. Šestá kapitola prezentuje novou aplikaci parametrického modelu pro porovnání učitelovy podpory žáků v zemích účastnících se výzkumu PISA 2012. V poslední sedmé kapitole jsou uvedené alternativní postupy k metodě ukotvujících vinět při korekci stylu odpovídání na položky, jako je tendence k volbě určitých škálových kategorií, technika přehánění (*overclaiming technique*), hodnocení nucenou volbou (*forced choice assessment*) a test posouzení situace (*situational judgement test*).

Pro začínající uživatele této metody, kterých bude vzhledem k její nízké četnosti užití většina, je ve třetí kapitole zařazen popis jednoduchého statistického vyhodnocení dat. Toto vyhodnocení je však na deskriptivní úrovni (chybí například signifikantní rozdíly mezi skupinami, ověření předpokladů metody a vhodnost využití modelu), pro podrobnější analýzu je vhodné přejít například k užívání parametrického modelu. Podrobný popis zpracování dat ve

statistických softwarech (metoda je implementována v softwarech R a STATA), a to včetně celého kódu s jeho výstupy a komentářem těchto výstupů, je uveden ve čtvrté kapitole. Začínajícím výzkumníkům lze též doporučit druhou kapitolu s podrobným popisem klíčových aplikací mimo pedagogický výzkum a dále i všech aplikací v pedagogickém výzkumu. To by mělo přispět nejen k lepšímu pochopení celé problematiky a principu metody ukotvujících vinět, ale také k nastínění možností dalších dosud neexistujících aplikací.

V rámci statistických modelů metody a jejich aplikací v sociálních vědách existuje mnoho výzev. Jednou z nich je formulace vinět, které je věnována pátá kapitola. Většina vinět v existujícím výzkumu byla sestrojena pouze pro dospělou populaci, pro dětskou populaci jich existuje jen několik. Konstrukci vinět i pro dospělou populaci je navíc věnováno velmi málo pozornosti. Pokud se však tato fáze výzkumu podcení, může to vést k porušení základních předpokladů metody, jako je konzistence odpovědí a ekvivalence vinět, což ve výsledku může vést k zavádějícím závěrům.

Další výzvou je dodržení a testování zmíněných klíčových předpokladů, konzistence odpovědí a ekvivalence vinět. Z pohledu aplikování metody na různé oblasti zkoumání lze konstatovat, že dodržení těchto předpokladů není snadné, v podstatě to patří k limitaci metody. Je tedy nutné se předpoklady zabývat ve všech fázích výzkumu (konstrukce dotazníku, sběr dat, analýza dat a interpretace výsledků). Z pohledu statistiky můžeme konstatovat, že v literatuře existují statistické testy těchto předpokladů. Pro výzkumníky v sociálních vědách však mohou být příliš obtížné, což by mohlo vést k tomu, že předpoklady testovány nebudou. Bylo by vhodné, kdyby existovaly statistické balíčky, které by běžnému uživateli umožnily předpoklady snadno testovat. Navíc i samotné testy jsou založeny na předpokladech, které nejsou v různých aplikacích přímo splnitelné, a je nutné se zaměřit i na testování těchto podpůrných předpokladů nebo alespoň slovní zdůvodnění, proč by tyto podpůrné předpoklady měly být splněny. Z pohledu kvalitativního výzkumu by bylo vhodné zaměřit se na testování klíčových předpokladů například pomocí metod individuálních či skupinových rozhovorů, jak je to ukázáno v Au a Lorgelly (2014). Uvedená studie se však zaměřuje jen na testování předpokladu konzistence odpovědí a nikoli na předpoklad ekvivalence vinět, což může být též předmětem dalšího výzkumu.

Při formulaci vinět musí výzkumník velmi dobře rozumět zkoumanému jevu, znát nejen jeho definici, ale detailně chápat jeho podstatu. Na pozadí tohoto porozumění vybírá vhodný příklad, který by měřený jev dobře reprezentoval (může se stát, že při tomto přemýšlení precizuje vymezení jevu). Tomuto příkladu musí velmi dobře rozumět všichni respondenti (z různých zemí, skupin apod.) a měli by ho interpretovat stejně (aby byl dodržen předpoklad ekvivalence vinět). Proces porovnávání jedinců či jejich skupin na základě dat (získaných pomocí dotazníkového šetření s ukotvujícími vinětami) by neměl být konečnou fází výzkumu, primárně by se mělo jednat o to, že na základě validních dat lépe porozumíme (pedagogické) realitě a na základě toho budeme schopni navrhnout konkrétní doporučení či změny. Porozumění tedy hraje důležitou roli pro výzkumníka, respondenty a práci s daty. Tyto úvahy by mohly být zdrojem pro další výzkum, a to nejen na poli metodologickém, ale i filozofickém (inspirací může být např. Chvál (2003), který v kontextu hermeneutiky rozebírá vnitřní hodnocení školy).

Další nepochybnou výzvou jsou nové aplikace, které by mohly dále přispět k řešení mnohdy nečekaných paradoxních výsledků, jako je negativní korelace mezi motivací učit se matematiku a výsledkem v testu z matematiky napříč zeměmi (He & van Vijver, 2016). Metodu ukotvujících vinět by bylo vhodné využít pro zvýšení validity měření mnoha psychologických a pedagogických jevů, jako je například školní klima, sebekázeň a vůle. Lze vyslovit přání, aby metoda ukotvujících vinět byla využita pro korekci všech (sebe)posuzovacích škál. I když nedostáváme na první pohled paradoxní výsledky, můžeme dostávat podhodnocené či naopak nadhodnocené efekty proměnných založených na sebeposuzovacích škálách. Následná interpretace výsledků založených na nekorigovaných škálách a užití takovýchto výsledků při rozhodování vede k chybám.

Metoda ukotvujících vinět může být využita jak v longitudiálních výzkumech (můžeme jí zachytit, jak se mění užívání škály jednotlivých respondentů či jejich skupin v čase), tak i v experimentálním výzkumu (šlo by například změřit případnou změnu odlišného užívání škály respondentů v pretestu a posttestu). V longitudiálním výzkumu již metoda ukotvujících vinět použita byla, v experimentálním výzkumu však nikoli a pro výzkumníky to může představovat novou výzvu.

Pokud jsme si správně vědomi, dosud neexistuje žádná systematická monografie zabývající se pouze metodou ukotvujících vinět. A to ani ve světové literatuře a ani mimo pedagogický výzkum. Tato monografie je tudíž první svého druhu. Lze jen doufat, že přispěje k rozšíření znalosti problematiky odlišného užívání škály a jejího potenciálního řešení pomocí metody ukotvujících vinět.

Literatura

Angelini, V., Cavapozzi, D., & Paccagnella, O. (2011). Dynamics of reporting work disability in Europe. *Journal of the Royal Statistical Society: Series A*, 174(3), 621–638.

Angelini, V., Cavapozzi, D., Corazzini, L., & Paccagnella, O. (2012a). Age, health and life satisfaction among older Europeans. *Social Indicators Research*, 105(2), 293–308.

Angelini, V., Cavapozzi, D., & Paccagnella, O. (2012b). Cross-country differentials in work disability reporting among older Europeans. *Social Indicators Research*, 105(2), 211–226.

Angelini, V., Cavapozzi, D., Corazzini, L., & Paccagnella, O. (2012c). Do Danes and Italians rate life satisfaction in the same way? Using vignettes to correct for individual-specific scale biases. *Oxford Bulletin of Economics and Statistics*, 76(5), 643–666.

Aronson, E., & Carlsmith, J. M. (1963). Effect of the severity of threat on the devaluation of forbidden behavior. *Journal of Abnormal and Social Psychology*, 66(6), 584–588.

Au, N., & Lorgelly, P. K. (2014). Anchoring vignettes for health comparisons: An analysis of response consistency. *Quality of Life Research*, 23(6), 1721–1731.

Bago d’Uva T., Lindeboom, M., O’Donnell, O., & van Doorslaer, E. (2011). Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. *Journal of Human Resources*, 46(4), 872–903.

Bago d’Uva T., O’Donnell, O., & van Doorslaer, E. (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology*, 37(6), 1375–1383.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-White differences in response styles. *Public Opinion Quarterly*, 48(2), 491–509.

Baumgartner, H., & Steenkamp, J. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156.

Beranová, V. (2015). Využití metody ukotvujících vinět: subjektivní hodnocení úrovně anglického jazyka žáků základních škol a osmiletých gymnázií (Diplomová práce). Dostupné z https://is.cuni.cz/studium/dipl_uc/index.php?id=a4c96a98a12cf67ae9eeb39ec2510d4d&tid=1&do=xdownload&fid=120194289&did=147735&vdetailu=1

Bonsang, E., & Van Soest, A. (2012). Satisfaction with social contacts of older Europeans. *Social Indicators Research*, 105(2), 273–292.

Buckley, J. (2008). Survey context effects in anchoring vignettes. Dostupné z <http://www.polmeth.wustl.edu/files/polmeth/surveyartifacts.pdf>

Buckley, J. (2009). Cross-National Response Styles in International Educational Assessments: Evidence from PISA 2006. Dostupné z https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf

Buckley, J., & Schneider, M. (2007). *Charter schools: Hope or hype?* United Kingdom: Princeton University Press.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics Methods and Applications*. New York, USA: Cambridge University Press.

Camili, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: SAGE Publication, Inc.

Clarke, L. (2001). Extreme response style in cross-cultural research. *International Marketing Review*, 18(3), 301–324.

Crawley, M. J. (2005). *Statistics: An Introduction using R*. Chichester, UK: John Wiley & Sons, Ltd.

D'Ailly, H. (2003). Children's autonomy and perceived control in learning: A model of motivation and achievement in Taiwan. *Journal of Educational Psychology*, 95(1), 84–96.

Deregowski, J.B., & Serpell, R. (1971). Performance on a sorting task: A cross-cultural experiment. *International Journal of Psychology*, 6(4), 273–281.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251.

Educational Research Anchors Group. (2017, leden). The anchoring vignette method in educational research. Dostupné z <http://eragroup.pdf.cuni.cz/index.html>

Erikson, D. (1982). *The British Columbia story: Antecedents and consequences of aid to private schools*. Los Angeles: Institute for the Study of Private Schools.

Euroqol Group. (1990). EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208.

Euroqol Group. (2012). Interim scoring for the EQ-5D-5L: EQ-5D-5L crosswalk index value Calculator. Dostupné z <http://www.euroqol.org/news/news/article/interim-scoring-for-the-eq-5d-5l-eq-5d-5l-crosswalk-index-value-calculator.html>

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2), 303–315.

Freedman, D. A. (1999). Ecological inference and the ecological fallacy. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* 6 (s. 4027–4030). Oxford: Elsevier.

Grol-Prokopczyk, H., Freese, J., & Hauser, R. M. (2011). Using anchoring

vignettes to assess group differences. *Journal of Health and Social Behavior*, 52(2), 246–261.

Grol-Prokopczyk, H., Verdes-Tennant, E., McEniry, M., & Ispány, M. (2015). Promises and pitfalls of anchoring vignettes in health survey research. *Demography*, 52(5), 1703–1728.

He, J., & van de Vijver, F. J. R. (2016). The motivation-achievement paradox in international educational achievement tests: Towards a better understanding. In R. B. King & A. B. I. Bernardo (Eds.), *The psychology of Asian Learners. A Festschrift in Honor of David Watkins* (s. 253–268). Singapore: Springer Science.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106(4), 766–794.

Hendl, J. (2015). *Přehled statistických metod*. Praha: Portál.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, New Jersey: Lawrence Erlbaum.

Hopkins, D. J., & King, G. (2010). Improving anchoring vignettes designing surveys to correct interpersonal incomparability. *The Public Opinion Quarterly*, 74(2), 201–222.

Hrabák, J. (2014). *Metoda ukotvujících vinět: hodnocení znalosti informačních a komunikačních technologií* (Diplomová práce). Dostupné z https://is.cuni.cz/studium/dipl_uc/index.php?id=a4c96a98a12cf67ae9eeb39ec2510d4d&tid=3&do=xdownload&fid=120175176&did=147470&vdetailu=1

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3), 296–309.

Hunter, J. E., & Schmidt, F. L. (2000). Racial and gender bias in ability and

achievement tests: Resolving the apparent paradox. *Psychology, Public Policy, and Law*, 6(1), 151–158.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82(1), 143–159.

Chen, Ch., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170–175.

Chráška, M. (2007). *Metody pedagogického výzkumu*. Praha: Grada.

Chvál, H. (2003). Metodologické základy vnitřního hodnocení školy (Disertační práce). Praha: Pedagogická fakulta Univerzity Karlovy.

Ivancevic T. T., Jovanovic, B., Djukic, S., Djukic, M., Markovic, S. (2009). *Complex Sports Biodynamics: With Practical Applications in Tennis*. Springer-Verlag Berlin Heidelberg.

Jackson, D. C., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13(4), 371–388.

John, O. P., Naumann, L., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: Discovery, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (s. 114–158). New York: Guilford.

Kagitcibasi, C. (1996). *Family and human development across cultures: A View from the other side*. Hillsdale, NJ: Lawrence Erlbaum.

Kapteyn, A., Smith, J. P., & van Soest, A. (2007). Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review*, 97(1), 461–473.

Kapteyn, A., Smith, J. P., & van Soest, A. (2010). Life satisfaction. In E. Diener, J. E. Helliwell, & D. Kahneman (Eds.), *International differences in well-being* (s. 70–104). Oxford: Oxford University Press.

Kapteyn, A., Smith, J. P., van Soest, A., & Voňková, H. (2011). Anchoring vignettes and response consistency (RAND Working Paper No. WR-840). Dostupné z http://www.rand.org/content/dam/rand/pubs/working_papers/2011/RAND_WR840.pdf

Kaščák, O., & Pupala, B. (2011). PISA v kritickej perspektive. *Orbis Scholae*, 5(1), 53–68.

King, G. (2017, leden). Anchoring vignettes website. Dostupné z <http://gking.harvard.edu/vig>

King, G., Murray, C., Salomon, J., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 567–583.

King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis*, 15(1), 46–66.

Kok, R., Avendano, M., Bago d’Uva, T., & Mackenbach, J. (2012). Can reporting heterogeneity explain differences in depressive symptoms across Europe? *Social Indicators Research*, 105(2), 191–210.

Králová, K. (2015). Subjektivní hodnocení úrovně anglického jazyka u vysokoškolských studentů pedagogických a ekonomických oborů: využití metody ukotvujících vinět (Diplomová práce). Dostupné z https://is.cuni.cz/studium/dipl_uc/index.php?id=a4c96a98a12cf67ae9eeb39ec2510d4d&tid=2&do=xdownload&fid=120194104&did=132415&vdetailu=1

Kristensen, N., & Johansson, E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics*, 15(1), 96–117.

Kyllonen, P. C., & Bertling, J. (2014a). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *A handbook of international large-scale assessment data analysis: Background, technical issues, and methods of data analysis* (s. 277–285). London, UK: Chapman Hall/CRC Press.

Kyllonen, P. C., & Bertling, J. P. (2014b). *Draft report: anchoring vignettes reduce bias in noncognitive rating scale responses*. Princeton, NJ: ETS/OECD.

Lievens, F., & de Soete, B. (2015). Situational judgement test. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (s. 13–19). Oxford: Elsevier.

Marsh, H. W. (1987). The Big-Fish-Little-Pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280–295.

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1), 213–231.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*. Chichester, UK: John Wiley & Sons, Ltd.

Mischel, W. (2014). *The Marshmallow test: Mastering self-control*. New York, NY: Little, Brown.

Murray, C. J. L., Özalin, E., Tandon, A., Salomon, J. A., Sadana, R., & Chatterji, S. (2003). Empirical evaluation of the anchoring vignette approach in health surveys. In C. J. L. Murray & D. Evans (Eds.), *Health systems performance assessments: Debates, methods and empiricism* (s. 369–399). Geneva: World Health Organization.

Münich, D., & Federičová, M. (2015). Srovnání žákovské oblíbenosti školy a matematiky pohledem mezinárodních šetření. *Pedagogická orientace*, 25(4), 557–582.

OECD. (2009). PISA 2006 Technical report: Validation of the embedded attitudinal scales. OECD Publishing: Paris. Dostupné z <http://dx.doi.org/10.1787/9789264048096-18-en>

OECD. (2013). PISA 2012 results: What makes a school successful? Resources, policies and practices, Anchoring vignettes in PISA 2012 student questionnaire. OECD Publishing: Paris. Dostupné z <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-IV.pdf>

OECD. (2014). PISA 2012 Technical report: Context questionnaire development. OECD Publishing: Paris. Dostupné z https://www.oecd.org/pisa/pisaproducts/PISA%202012%20Technical%20Report_Chapter%203.pdf

OECD. (2017a). Database – PISA 2012: Codebook for student questionnaire. Dostupné z https://www.oecd.org/pisa/pisaproducts/PISA12_stu_codebook.pdf

OECD. (2017b). Database – PISA 2012: Student questionnaire form B. Dostupné z https://www.oecd.org/pisa/pisaproducts/PISA12_StQ_FORM_B_ENG.pdf

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning – 2nd edition*. Thousand Oaks, CA: SAGE Publication, Inc.

Pan, Y., & Fond, M. (2014). Evaluating multilingual questionnaires: A sociolinguistic perspective. *Survey Research Methods: Journal of the European Research Association*, 8(3), 181–194.

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O’Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education*, 46(9), 850–868.

Peracchi, F., & Rossetti C. (2012). Heterogeneity in health responses and anchoring vignettes. *Empirical Economics*, 42(2), 513–538.

Peracchi, F., & Rossetti, C. (2013). The heterogeneous thresholds ordered response model: Identification and inference. *Journal of the Royal Statistical Society: Series A*, 176(3), 703–722.

Potužníková, E., Lokajíčková, V., & Janík, T. (2014). Mezinárodní výzkumy školního vzdělání v České republice: zjištění a výzvy. *Pedagogická orientace*, 24(2), 185–221.

Primi, R., Zanon, C., Santos, D., De Druyt, F. & John, O. P. (2016). Anchoring vignettes: Can they make adolescent self-reports of social-emotional skills more reliable, discriminant, and criterion-valid? *European Journal of Psychological Assessment*, 32(1), 39–51.

Rabe-Hesketh, S., & Skrondal, A. (2002). Estimating Chopit models in gllamm: Political efficacy example from King et al. (2002). Dostupné z <http://www.gllamm.org/chopit.pdf>.

Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *British Medical Journal*, 328(7434), 258–260.

Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology*, 70(3), 365–380.

SHARE-ERIC (2016). Questionnaire Wave 2. Dostupné z <http://www.share-project.org/data-access-documentation/questionnaires/questionnaire-wave-2.html>

Sheppard, R., Han, K., Colarelli, S. M., & Dai, G. (2006). Differential item functioning by sex and race in the Hogan personality inventory. *Assessment*, 13(4), 442-453.

Sirven, N., Santos-Eggimann, B., & Spagnoli, J. (2012). Comparability of health care responsiveness in Europe. *Social Indicators Research*, 105(2), 255–271.

Stanislaw H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, 31(1), 137–149.

Straková, J. (2016). *Mezinárodní výzkumy výsledků vzdělávání. Metodologie, přínosy, rizika a příležitosti*. Praha: Univerzita Karlova v Praze – Pedagogická fakulta.

Straková, J. & Greger, D. (2013). Faktory ovlivňující přechod žáků 5. ročníků na osmileté gymnázium. *Orbis Scholae*, 7(3), 73–85.

Super, C.M. (1983). Cultural variation in the meaning and uses of children's intelligence. In: J. B. Deregowski, S. Dziurawiec, & R. C. Annis (Eds.), *Explorations in cross-cultural psychology* (s. 199–212). Lisse, The Netherlands: Swets & Zeitlinger.

Štech, S. (2015). Proč se kritizuje PISA? *Pedagogická orientace*, 25(4), 605–612.

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society: Series A*, 174(3), 575–595.

Van Soest, A. & Voňková, H. (2014). Testing the specification of parametric models using anchoring vignettes. *Journal of the Royal Statistical Society: Series A*, 177(1), 115–133.

Van de Vijver, F. J. R. & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée*, 54, 119–135.

Veselý, A. (2011). Konceptuální rámec pro analýzu vztahu vzdělávací politiky a vzdělávacích výsledků. *Orbis Scholae*, 5(1), 23–52.

Vlčková, P. (2016). Měření subjektivní duševní pohody učitelů na základní škole s využitím metody ukotvujících vinět (Diplomová práce). Dostupné z https://is.cuni.cz/studium/dipl_uc/index.php?id=a4c96a98a12cf67ae9eeb39ec2510d4d&tid=4&do=xdownload&fid=120238746&did=161238&vdetailu=1

Voňková, H. (2012a). Metoda ukotvujících vinět a možnosti využití v pedagogice. *Orbis Scholae*, 6(1), 27–40.

Voňková, H. (2012b). Metoda ukotvujících vinět a předcházející metodologické a statistické přístupy. In V. Ježková (Ed.), *Kvalita ve vzdělávání: Sborník příspěvků z XX. výroční konference České asociace pedagogického výzkumu* (s. 100–108). Praha: Univerzita Karlova.

Voňková, H. (2013). Subjektivní hodnocení problémů s pohybem: Užití parametrického modelu metody ukotvujících vinět. *Orbis Scholae*, 7(1), 49–66.

Voňková, H. (2016). The influence of the selected school characteristics on teacher's classroom management in European countries: student's assessments and anchoring vignettes. ECER conference 2016. Dostupné z <http://www.eera-ecer.de/ecer-programmes/conference/21/contribution/36954/>

Voňková, H., Bendl, S., & Papajoanu, O. (2017). How students report dishonest behavior in school: Self-assessment and anchoring vignettes. *Journal of Experimental Education*, 85(1), 36–53.

Voňková, H., & Hrabák, J. (2015a). Řešení nesrovnatelnosti hodnocení ICT znalostí a dovedností žáků skrze ukotvující viněty. *Pedagogika*, 65(3), 274–291.

Voňková, H., & Hrabák, J. (2015b). The (in) comparability of ICT knowledge and skill self-assessments among upper secondary school students: The use of the anchoring vignette method. *Computers & Education*, 85, 191–202.

Voňková, H., & Hullegie, P. (2011). Is the anchoring vignette method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society: Series A*, 174(3), 597–620.

Voňková, H., Papajoanu, O., & Bendl, S. (2016). Aplikace metody ukotvujících vinět v pedagogickém výzkumu: přehled literatury a metodologická doporučení. *Pedagogická orientace*, 26(3), 537–559.

Voňková, H., & Šrámková, E. (2013). Zjišťování porozumění textu v angličtině: Jaké možnosti dává využití metody ukotvujících vinět? In L. Círus (Ed.), *Efektivita vzdělávání v proměnách společnosti: Sborník příspěvků XXI. celostátní konference České asociace pedagogického výzkumu* (s. 92–97). Ústí nad Labem: Univerzita J. E. Purkyně.

Voňková, H., Zamarro, G., DeBerg, V., & Hitt, C. (2015). Comparisons of student perceptions of teacher's performance in the classroom: Using parametric anchoring vignette methods for improving comparability (EDRE Working Paper No. 2015–01). Dostupné z <http://www.uaedreform.org/downloads/2015/05/comparisons-of-student-perceptions-of-teachersperformance-in-the-classroom-using-parametric-anchoring-vignette-methods-for-improving-comparability.pdf>

Wand, J., King, G., & Lau, O. (2011). Anchors: Software for anchoring vignette data. *Journal of Statistical Software*, 42(3), 1–25.

Wand, J., King, G., & Lau, O. (2016). Package 'anchors'. Dostupné z <https://cran.r-project.org/web/packages/anchors/anchors.pdf>

Watkins, D. (2000). Learning and teaching: A cross-cultural perspective. *School Leadership & Management*, 20(2), 161–173.

Watkins, D., & Cheung, S. (1995). Culture, gender, and response bias: Analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, 26(5), 490–504.

Wei, H. S., & Huang, C. K. (2005). Reviewing school bullying research: Empirical findings and methodical considerations. *NTTU Educational Research Journal*, 16(1), 69–112.

West, M., Kraft, M. A., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148–170.

WikiJob. (2017). Situational Judgement Test. Dostupné z <https://www.wikijob.co.uk/content/aptitude-tests/test-types/situationaljudgement-test>

1. Příloha: Příklady ukotvujících vinět ve výzkumech zdraví a vnímané politické účinnosti (*political efficacy*)

V této příloze jsou ukázky vinět užitých ve výzkumech zdravotního stavu obyvatel ve výzkumu Survey of Health, Ageing and Retirement in Europe (SHARE-ERIC, 2016), následují viněty pro zdravotní problémy při výkonu práce (Kapteyn et al., 2007) a vnímanou politickou účinnost (*political efficacy*) (King et al., 2004).

Tělesné potíže a bolesti

Sebehodnoticí otázka

Celkově vzato, jak velké tělesné potíže nebo bolesti jste měl/a v průběhu posledních 30 dní?

žádné, malé, středně silné, silné, extrémní

Ukotvující viněta

Úvod:

Popíšeme Vám teď několik osob s různě velkými zdravotními problémy. Zajímá nás, jak byste zhodnotil/a zdraví těchto osob. Představte si, že tyto osoby jsou ve Vašem věku a mají stejnou životní historii jako Vy.

Pan Pavel má jednou za měsíc bolesti hlavy, které přejdou, když si vezme tabletu. Když má bolesti hlavy, je schopen vykonávat svoje každodenní úkoly.

Jak velké tělesné potíže nebo bolesti má podle Vašeho názoru pan Pavel?

žádné, malé, středně silné, silné, extrémní

Problémy se spánkem

Sebehodnoticí otázka

Jak silné poruchy spánku jste měl/a v průběhu posledních 30 dní?

žádné, malé, středně silné, silné, extrémní

Ukotvující viněta

Paní Lída večer dobře usne, ale dvakrát týdně se v noci probudí a po zbytek noci už nemůže usnout.

Jak velké poruchy spánku má podle Vašeho názoru paní Lída?

žádné, malé, středně silné, silné, extrémní

Problémy s pohybem

Sebehodnoticí otázka

Celkově vzato, jak velké potíže nebo bolesti jste měl/a v průběhu posledních 30 dní při pohybu?

žádné, malé, středně silné, silné, extrémní

Ukotvující viněta

Pan Rudolf ujede bez problémů vzdálenosti do 200 metrů, ale po jednom kilometru nebo vystoupení schodů do jednoho patra se cítí unaven. S každodenními činnostmi, jako donést si domů nakoupené potraviny, problémy nemá.

Jak velké problémy má podle Vašeho názoru pan Rudolf s pohybem?

žádné, malé, středně silné, silné, extrémní

Problémy se soustředěním a vzpomínáním si na věci

Sebehodnoticí otázka

Celkově vzato, jak velké potíže jste měl/a v průběhu posledních 30 dní se soustředěním nebo se vzpomínáním si na věci?

žádné, malé, středně silné, silné, extrémní

Ukotvující viněta

Paní Anna se dovede soustředit, když se dívá na televizi, když čte časopis, hraje karty nebo šachy. Jednou za týden zapomene, kde má klíče nebo brýle, ale za pět minut je zase najde.

Jak velké problémy má podle Vás paní Anna s koncentrací nebo se vzpomínáním na věci?

žádné, malé, středně silné, silné, extrémní

Problémy s dýchavičností

Sebehodnoticí otázka

Jak velké potíže jste měl/a v průběhu posledních 30 dní s dýchavičností?

žádné, malé, středně silné, silné, extrémní

Ukotvující viněta

Pan Josef nemá žádné problémy, když jde pomalu. Když jde 20 metrů do kopce nebo když vyjde jedno patro schodů, snadno ztrácí dech.

Jak velké potíže má podle Vás pan Josef s dýchavičností?

žádné, malé, středně silné, silné, extrémní

Smutek a sklíčenost

Sebehodnoticí otázka

Celkově vzato, jak velké problémy jste v průběhu posledních 30 dní měl/a s tím, že jste se cítil/a smutný/á, sklíčený/á nebo deprimovaný/á?

žádné, malé, středně silné, silné, extrémní

Ukotvující viněta

Paní Hanu baví práce a ve svém soukromém životě je celkem vzato šťastna. Jednou za tři neděle se den nebo dva cítí zničená a ztrácí zájem o věci, které normálně dělá ráda. Je však schopna vykonávat každodenní běžné činnosti.

Jak velké problémy má podle Vás paní Hana s tím, že se cítí smutná, sklíčená nebo deprimovaná?

žádné, malé, středně silné, silné, extrémní

Zdravotní potíže omezující výkon práce

Sebehodnoticí otázka

*Máte nějaké zdravotní potíže nebo problémy, jež omezují druh nebo rozsah práce (činnosti), které můžete vykonávat?
žádné, malé, středně silné, silné, extrémní*

Ukotvující viněty

Pana Václava jeho práce vcelku baví. Jednou za tři neděle se cítí jeden nebo dva dny sklíčený a ztrácí zájem o věci, které normálně dělá rád – je však schopen své každodenní pracovní činnosti dál vykonávat.

*Jak velké jsou zdravotní potíže pana Václava, jež omezují druh nebo rozsah jeho práce, kterou může vykonávat?
žádné, malé, středně silné, silné, extrémní*

Pan Kamil trpí bolestmi v zádech, zejména v práci mu záda často ztuhnou. Po menší dávce léků je to však lepší. Kromě těchto obecných potíží žádné jiné bolesti nemá.

*Jak velké jsou zdravotní potíže pana Kamila, jež omezují druh nebo rozsah jeho práce, kterou může vykonávat?
žádné, malé, středně silné, silné, extrémní*

Paní Zdena měla v minulosti problémy se srdcem a bylo jí řečeno, že by si měla sledovat hladinu cholesterolu. Když má v práci stres, má někdy bolesti v hrudníku a příležitostně i v pažích.

*Jak velké jsou zdravotní potíže paní Zdeny, jež omezují druh nebo rozsah její práce, kterou může vykonávat?
žádné, malé, středně silné, silné, extrémní*

Spokojenost s příjmy

Sebehodnoticí otázka

*Jak jste spokojen/a s celkovými příjmy Vaší domácnosti?
velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a,*

velmi nespokojen/a

Ukotvující viněty

Úvod:

Popíšeme Vám teď některé aspekty ze života jiných lidí. Zajímá nás, jak tyto aspekty ze života jiných lidí budete hodnotit Vy z Vašeho pohledu. Pokud nebude uvedeno jinak, představte si, že tyto osoby jsou ve Vašem věku a mají stejnou životní historii jako Vy. Když si nebudete jisti, uveďte prosím takové hodnocení, jaké budete považovat za nejlepší.

Paní Milada je vdaná a má dvě děti, celkový příjem její domácnosti po zdanění činí 48.000 korun měsíčně.

Jak si myslíte, že je paní Milada spokojena s celkovým příjmem své domácnosti? velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Pan Jan je ženatý a má dvě děti, celkový příjem domácnosti po zdanění činí 24.000 korun měsíčně.

Jak si myslíte, že je pan Jan spokojen s celkovým příjmem své domácnosti? velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Spokojenost se sociálními kontakty

Sebehodnoticí otázka

Jak jste spokojen/a s Vašimi sociálními kontakty (s rodinou, přáteli atd.)? velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Ukotvující viněty

Pan Milan je svobodný, ale vychází dobře se svými příbuznými a má hodně přátel. Často spolu chodí na sportovní utkání nebo jdou na večeri do restaurace.

Jak si myslíte, že je pan Milan spokojen se svými sociálními kontakty (rodina, přátelé atd.)?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Paní Marie je vdaná už mnoho let. V poslední době tráví se svým mužem málo času a často se hádají. Zdá se, že dávají přednost tomu, trávit čas s jinými lidmi než spolu. Oba mají hodně přátel.

Jak si myslíte, že je paní Marie spokojena se svými sociálními kontakty (rodina, přátelé atd.)?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Spokojenost s každodenními činnostmi

Sebehodnoticí otázka

Jak jste spokojen/a s Vašimi každodenními činnostmi (např. s Vaší prací, pokud pracujete)?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Ukotvující viněty

Paní Stanislava pracuje čtyři dny v týdnu a její zaměstnání jí nepřipadá stresující. Nemůže si vybírat, co dělá – o tom rozhoduje její šéf. Má pocit, že její pracovní místo je velmi jisté.

Jak si myslíte, že je paní Stanislava spokojena se svým zaměstnáním?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Pan Miroslav pracuje na plný úvazek pět dní v týdnu; svou práci si může organizovat podle svého, ale často je pod velkým tlakem, aby splnil své úkoly včas. Pracuje pro velkou společnost a má pocit, že jeho pracovní místo je jisté.

Jak si myslíte, že je pan Miroslav spokojen se svým zaměstnáním?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Spokojenost se životem

Sebehodnoticí otázka

Jak jste spokojen/a s Vaším životem obecně?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Ukotvující viněty

Paní Anně je 72 let a je vdova. Její celkový čistý příjem činí asi 15.000 korun měsíčně. Bydlí ve vlastním domě a má hodně přátel. Dvakrát týdně hraje bridž a pravidelně jezdí s několika přáteli na dovolenou. V poslední době trpí artritidou, takže má při práci v domě a na zahradě bolesti.

Jak si myslíte, že je paní Anna spokojena se svým životem?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Panu Karlovi je 63 let. Před dvěma roky mu zemřela žena a on na ni stále ještě velmi často myslí. Má čtyři děti a deset vnuků, kteří ho pravidelně navštěvují. Karel se svým platem vyjde, ale nemá peníze na zvláštní výdaje jako např. na drahé dárky pro vnuky. Kvůli problémům se srdcem musel před nedávnem přestat pracovat. Snadno se unaví. Jinak nemá žádná vážná zdravotní omezení.

Jak si myslíte, že je pan Karel spokojen se svým životem?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Vliv na rozhodování samosprávy v obci

Sebehodnoticí otázka

Jak velký vliv máte na to, aby se obec zabývala tématy, která Vás zajímají?

obrovský vliv, velký vliv, jistý vliv, malý vliv, vůbec žádný vliv

Ukotvující viněty

Paní Alena trpí tím, že jedna továrna v místě znečišťuje vzduch. Znečištění není nebezpečné, ale někdy ošklivě páchne. Alena a její sousedi podporují v nadcházejících volbách kandidáta opozice, který slíbil, že se touto záležitostí bude zabývat. V obci je mnoho lidí, kteří jsou také toho názoru, že tento kandidát opozice nad současným poslancem pravděpodobně zvítězí.

*Jak velký vliv má paní Alena na to, aby se obec zabývala věcmi, které ji zajímají?
obrovský vliv, velký vliv, jistý vliv, malý vliv, vůbec žádný vliv*

Pan Tomáš trpí tím, že jedna továrna v místě znečišťuje vzduch. Znečištění není nebezpečné, ale někdy ošklivě páchne. Starosta se problémem zabýval, ale rozhodl, že v současné době je nejdůležitějším politickým tématem rozvoj průmyslu a nikoliv čistý vzduch.

Jak velký vliv má pan Tomáš na to, aby se obec zabývala věcmi, které ho zajímají?

obrovský vliv, velký vliv, jistý vliv, malý vliv, vůbec žádný vliv

Kvalita zdravotní péče

Sebehodnoticí otázka

V některých zemích se na termín u lékaře musí dlouho čekat a pro určité druhy ošetření existují seznamy čekatelů. Jak byste celkově hodnotil/a Vaši situaci, jak dlouhá je doba, po kterou musíte čekat na určité lékařské ošetření?

velmi krátká, krátká, únosná, dlouhá, velmi dlouhá

Ukotvující viněta

Pan Martin má potíže se žaludkem, má velké bolesti a musí na operaci. Je ihned převezen do nemocnice, kde je pod stálou kontrolou a dostává silné léky proti bolesti. Ale protože jeho onemocnění není životu nebezpečné, musí čekat na operaci čtyři dny.

Jak byste celkově hodnotil/a dobu, po kterou musí pan Martin čekat na lékařský zákrok?

velmi krátká, krátká, únosná, dlouhá, velmi dlouhá

Sebehodnoticí otázka

Jak byste celkově hodnotil/a stav lékařských ordinací a nemocnic, které jste navštívil/a?

velmi dobrý, dobrý, průměrný, špatný, velmi špatný

Ukotvující viněta

Pan Ludvík byl týden v místní nemocnici, protože dostal vysokou horečku. Měl samostatný pokoj, který byl čistý, ale malý a toaleta byla na chodbě ve vzdálenosti několika metrů. Bylo léto a horko a na pokoji nebyla ani klimatizace ani ventilátor, Ludvík si však mohl přivést vlastní stolní ventilátor z domova.

*Jak byste celkově hodnotil/a stav nemocnice, do které byl pan Ludvík převezen?
velmi dobrý, dobrý, průměrný, špatný, velmi špatný*

Sebehodnoticí otázka

Jak byste celkově hodnotil/a to, jak srozumitelně s Vámi hovoří lékaři a zdravotní sestry a jak Vás dovedou seznámit s rozhodnutími o Vaší léčbě?

velmi dobře, dobře, průměrně, špatně, velmi špatně

Ukotvující viněta

Paní Eliška si před několika měsíci zlomila ruku a musela jít několikrát na rentgen. Nejprve ji lékař a sestry informovali o její zlomenině a vysvětlili jí, o jakou léčbu v jejím případě půjde. Při poslední návštěvě ji změřili a zvažili, změřili jí tlak a odebrali krev, ale neřekli jí, proč to dělají. Paní Eliška se na to po vyšetření zeptala lékaře a on jí všechno vysvětlil.

*Jak byste celkově hodnotil/a, jak srozumitelně lékaři a sestry s paní Eliškou hovořili a jak ji zasvětili do rozhodnutí o její léčbě?
velmi dobře, dobře, průměrně, špatně, velmi špatně*

Zdravotní problémy při výkonu práce (Kapteyn et al., 2007)

Sebehodnoticí otázka

Máte nějaké zdravotní poškození nebo nějaký zdravotní problém, které vás omezují ve vykonávání určitého typu či množství placené práce?

Dva typy škál odpovědí:

- *ano, ne*
- *vůbec ne; ano, jsem mírně limitován; ano, jsem středně limitován; ano, jsem vážně limitován; ano, jsem extrémně limitován, nemohu pracovat*

Následovaly ukotvující viněty pro bolest, emoční problémy, onemocnění srdce. Příklad ukotvujících vinět pro bolest:

[Katie] příležitostně cítí při práci bolest v zádech, ale v posledních několika měsících tento pocit neměla. Pokud bolest pocítí, typicky trvá jen několik málo dní.

*Je [Katie] omezena ve výkonu své práce?
vůbec, mírně, středně, vážně, extrémně*

[Mark] má bolest v zádech a nohách, která je přítomná téměř pořád. Bolest se zhorší, když pracuje. Ačkoli mu léky pomáhají, cítí se nepříjemně při pohybu a držení a zvedání věcí v práci.

*Je [Mark] omezen ve výkonu své práce?
vůbec, mírně, středně, vážně, extrémně*

Vnímaná politická účinnost (*political efficacy*) (King et al., 2004)

Sebehodnoticí otázka

*Do jaké míry máte možnost ovlivňovat vládu v záležitostech, které se Vás týkají?
vůbec, málo, středně, velmi, neomezeně*

Ukotvující viněty

[Alison] nemá dostatek pitné vody. Společně se svými sousedy podporuje v následujících volbách opozičního kandidáta, který slibuje, že se touto záležitostí

bude zabývat. Zdá se, že mnoho lidí v její oblasti má stejně jako ona pocit, že opoziční kandidát porazí současného zastupitele.

Do jaké míry má [Alison] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Imelda] nemá dostatek pitné vody. Společně se svými sousedy se snaží na tento problém upozornit sběrem podpisů pod petici. Plánují ji předložit všem politickým stranám před nadcházejícími volbami.

Do jaké míry má [Imelda] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Jane] nemá dostatek pitné vody, protože úřady provádějí změny v územním plánu za účelem jeho průmyslového rozvoje. V kampani k nadcházejícím volbám opoziční strana slíbila, že se touto záležitostí bude zabývat, ale Jane má pocit, že je zbytečné volit opozici, protože současní zastupitelé jistě vyhrají.

Do jaké míry má [Jane] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Toshiro] nemá dostatek pitné vody. Je zde skupina místních vedoucích představitelů, kteří s problémem mohou něco dělat, ale ti říkají, že v současné době je nejdůležitější průmyslový rozvoj a nikoli pitná voda.

Do jaké míry má [Toshiro] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

[Moses] nemá dostatek pitné vody. Ráda by to změnila, ale nemůže volit, a má pocit, že nikdo z vlády tuto záležitost neřeší. V tichosti trpí a doufá, že se něco v budoucnu stane.

Do jaké míry má [Moses] možnost ovlivňovat vládu v záležitostech, které se jí týkají?

vůbec, málo, středně, velmi a neomezeně

2. Příloha: Příklady ukotvujících vinět v pedagogickém výzkumu

Komunita školy – spolupráce rodiče s ředitelem, učitelem a ostatními rodiči (Buckley, 2008)

Úvod k dotazníku:

Vzhledem k tomu, že různí lidé často reagují odlišně na události v jejich škole, rád bych Vám položil několik otázek ke škole obecně. Budete požádáni přemýšlet o několika situacích ve škole. Neexistují žádné správné či chybné odpovědi, rád bych znal Vaši první reakci.

Spolupráce rodičů a ředitele

Sebehodnoticí otázka

*Jak byste zhodnotili spolupráci mezi ředitelem a rodiči na škole Vašeho dítěte?
A, ..., F*

(poznámka: škálu rodiče dobře znají, běžně se využívá na amerických školách)

Ukotvující viněty

Představte si následující situaci: Ve třídě je dítě, které pravidelně vyrušuje při vyučování a je učitelem pravidelně posíláno k řediteli.

Ředitel této školy zavolá rodičům tohoto žáka, pošle jim i dopis o problémovém chování jejich dítěte a snaží se aktivně najít příčiny jeho problémového chování. Jak byste zhodnotili spolupráci mezi ředitelem a rodiči na této škole?

A, ..., F

Ředitel této školy pošle rodičům dopis o problémovém chování jejich dítěte, ale situaci již dále nijak neřeší.

Jak byste zhodnotili spolupráci mezi ředitelem a rodiči na této škole?

A, ..., F

Ředitel této školy nechá žáka sedět několik hodin v kanceláři vedení školy, ale rodičům nic nesdělí.

Jak byste zhodnotili spolupráci mezi ředitelem a rodiči na této škole?

A, ..., F

Spolupráce rodiče a učitele jeho dítěte

Sebehodnoticí otázka

Jak byste zhodnotili spolupráci mezi učitelem a rodiči na škole Vašeho dítěte?

A, ..., F

Ukotvující viněty

Zde je další situace. Dítěti má být slavnostně udělena cena.

Na této škole pošle učitel rodičům žáka dopis, dále rodičům zavolá a snaží se učinit vše pro to, aby se rodiče slavnostního předávání zúčastnili.

Jak byste zhodnotili spolupráci mezi učitelem a rodiči na této škole?

A, ..., F

Na této škole pošle učitel rodičům žáka dopis, ve kterém je upozorňuje na toto slavnostní předání.

Jak byste zhodnotili spolupráci mezi učitelem a rodiči na této škole?

A, ..., F

Na této škole řekne učitel žákovi, aby pozval své rodiče, ale neučiní nic dalšího.

Jak byste zhodnotili spolupráci mezi učitelem a rodiči na této škole?

A, ..., F

Vzájemná spolupráce rodičů

Sebehodnoticí otázka

Jak byste zhodnotili spolupráci rodičů na škole Vašeho dítěte?

A, ..., F

Ukotvující viněty

Co se stane, pokud se přihodí následující věc. Je polovina roku a ředitel, který byl velmi populární, náhle odejde.

Na této škole se setká hodně rodičů, aby zjistili, proč ředitel odešel, a ujistili se, že se budou moci vyjádřit k výběru dalšího ředitele.

Jak byste zhodnotili spolupráci rodičů na této škole?

A, ..., F

Na této škole se setká několik rodičů, aby zjistili, proč ředitel odešel, a ujistili se, že se budou moci vyjádřit k výběru dalšího ředitele.

Jak byste zhodnotili spolupráci rodičů na této škole?

A, ..., F

Na této škole se setká jen velmi málo rodičů, aby zjistili, proč ředitel odešel, a ujistili se, že se budou moci vyjádřit k výběru dalšího ředitele.

Jak byste zhodnotili spolupráci rodičů na této škole?

A, ..., F

Učitelovo řízení třídy (OECD, 2017b)

Sebehodnoticí otázky

Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícími tvrzeními?

Učitel umí žáky přimět k tomu, aby poslouchali, co říká.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Učitel ve třídě udržuje kázeň.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Učitel své hodiny začíná včas.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Učitel musí dlouho čekat, než se žáci utiší.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Ukotvující viněty

Přečti si popisy tří učitelů matematiky a pak se rozhodni, do jaké míry souhlasíš s tvrzeními na konci popisů.

Žáci ve třídě paní učitelky Sýkorové jsou v jejích hodinách klidní a ukáznění.

Paní učitelka vždy chodí do hodiny včas.

Paní učitelka Sýkorová má nad děním ve třídě kontrolu.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Žáci v hodinách paní učitelky Pokorné často vyrušují. Paní učitelka vždy chodí do hodiny o pět minut dříve.

Paní učitelka Pokorná má nad děním ve třídě kontrolu.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Žáci ve třídě pana učitele Krátkého v jeho hodinách často vyrušují. Pan učitel kvůli tomu často chodí do hodiny s pětiminutovým zpožděním.

Pan učitel Krátký má nad děním ve třídě kontrolu.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Učitelova podpora žáků (OECD, 2017b)

Sebehodnoticí otázky

Když si představíš svého učitele matematiky, do jaké míry souhlasíš s následujícími tvrzeními?

Učitel nám říká, že musíme pilně pracovat.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Když je třeba, učitel se nám více věnuje.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Učitel nám pomáhá s učením.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Učitel nám dává příležitost vyjádřit své názory.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Ukotvující viněty

Přečti si popisy tří učitelů matematiky a pak se rozhodni, do jaké míry souhlasíš s tvrzeními na konci popisů.

Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou.

Paní učitelce Novotné záleží na tom, jak se žáci učí.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Pan učitel Novák dává domácí úkoly z matematiky jednou za týden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou.

Panu učiteli Novákovi záleží na tom, jak se žáci učí.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Paní učitelka Ježková dává domácí úkoly z matematiky jednou za týden. Opravené je žákům před zkoušením nebo písemkou nevrací.

Paní učitelce Ježkové záleží na tom, jak se žáci učí.

rozhodně souhlasím, souhlasím, nesouhlasím, rozhodně nesouhlasím

Vybrané kázeňské přestupky (Voňková et al., 2017)

Hodnocení jednotlivých přestupků provedl na stupnici 1, 2, 3, 4, 5, kde 1 označuje, že daný přestupek není pro žáka charakteristický (žák je ukázněný) a 5 označuje, že je daný přestupek pro žáka charakteristický (žák je neukázněný).

V dotazníku následoval výčet přestupků, které žáci hodnotili u náhodně vybrané poloviny spolužáků. Tato polovina mohla zahrnout i žáka samotného. Heterogenita ve stylu odpovídání byla zkoumána na základě žákovského hodnocení vinět týkajících se podvádění, obecné nekázně, drzosti vůči učiteli a šikany učitele a také šikany žáka.

Ukotvující viněty

Honza si dvakrát za poslední měsíc místo odpoledního vyučování, kdy měla třída suplovanou dvouhodinovku tělesné výchovy, vyrazil s kamarády do města. Následně si sám napsal omluvenku, ve které zfalšoval podpis svého otce.

U Honzy hodnotím záškoláctví a podvádění číslicí:

Pavel je při hodině roztěkaný, nedává pozor, nikdy neví, co se při hodině děje, ruší, a tím zdržuje výklad učitele. Často používá sprostá slova. Občas je vůči učitelům drzý. Když se nudí, čmárá po lavici. Párkrát už byl za školou, omluvenku si napsal sám.

U Pavla hodnotím jeho celkovou neukázněnost číslicí:

Ve většině hodin dějepisu ruší Filip učitele při výkladu látky neustálým šeptáním se svým sousedem. Na napomenutí učitele nereaguje. Na konci výkladu vždy učitel žákům řekne, aby si do sešitu opsali z tabule základní údaje. Filip učitele ignoruje a dává najevo, že ho to nezajímá. Občas se navíc například snaží pobavit třídu tím, že na začátku vyučování nahlašuje učiteli chybějící žáky, kteří ve skutečnosti nechybějí.

U Filipa hodnotím drzost vůči učiteli a psychickou šikanu učitele číslicí:

Eva je v čele pětičlenné party spolužaček, která se téměř každý den vysmívá Lence kvůli jejímu oblečení. Eva navádí členky své party, aby se s Lenkou vůbec nebavily. Občas se stává, že Eva úmyslně do Lenky vrazí tak, aby upadla. Nejenže se jí Eva neomluví, ale ještě se jí směje.

U Evy hodnotím psychickou a fyzickou šikanu spolužačky Lenky číslicí:

ICT znalosti a dovednosti (Voňková & Hrabák, 2015)

Sebehodnoticí otázka

Nejdříve, prosím, ohodnoťte svoji úroveň vědomostí a dovedností spojených s informační a komunikační technologií.

Vyberte jedno číslo na škále od 1 do 10.

1 – minimum (nejnižší vědomosti a dovednosti), 10 – maximum (nejvyšší)

1, ..., 10 (1 – nejnižší, 10 – nejvyšší)

Ukotvující viněty

Nyní, prosím, ohodnoťte úroveň vědomostí a dovedností osob uvedených v krátkých příbězích.

Lukáš dovede svůj počítač spustit a ovládat klávesnicí a myší. Umí si spustit oblíbené hry. Je schopný si na internetu přečíst e-maily nebo vyhledat internetové stránky, které ho zajímají. Umí si z internetu stáhnout soubor na plochu, otevřít ho a také ho dovede vytisknout.

1, ..., 10 (1 – nejnižší, 10 – nejvyšší)

Daniel na počítači umí pracovat s kancelářským balíkem Office. Vytváří pro firmu dokumenty, prezentace nebo tabulky. V Excelu ovládá často používané funkce včetně funkce KDYŽ, COUNTIF, POČET, SUMA, ... Je schopný pracovat s programy, které má v počítači nainstalovány a umí si další nainstalovat. Je schopen si vyhledat informace na internetu nebo komunikovat se spolupracovníky pomocí e-mailu. Své soubory a složky má upravené ve stromové struktuře složek.

1, ..., 10 (1 – nejnižší, 10 – nejvyšší)

Anna ráda na počítači upravuje fotografie, které fotí digitálním fotoaparátem. Ty pak nahrává na svou vlastní stránku, kterou vytvořila. Nabízí tak dalším lidem její výtvořky k prohlédnutí. Pro svoji stránku vytváří texty, prezentace a tabulky v různých kancelářských aplikacích nebo natáčí krátká videa, která upravuje v programu pro střih videa. Všechny materiály pak na své internetové stránce nabízí fanouškům focení. Také s nimi komunikuje. Kvůli své stránce má na harddisku jasně oddělené soubory pro sdílení a soubory osobní.

1, ..., 10 (1 – nejnižší, 10 – nejvyšší)

Pavčina umí pracovat s kompletním kancelářským balíkem Office. Umí také programovat v několika programovacích jazycích (C++, Java, PHP). Vytváří programy a nabízí je na své interaktivní webové stránce, která umožňuje přihlašování uživatelů. Její programy využívají i firmy. Některé programy nabízí zdarma, jiné nabízí v internetovém obchodě, který sama vytvořila, za úplaty. Dále nabízí e-learningové kurzy programování pomocí elektronických konferencí nebo výukových videí, které sama pro zájemce připravuje.

1, ..., 10 (1 – nejnižší, 10 – nejvyšší)

Porozumění psaným textům v anglickém jazyce (Voňková & Šrámková, 2013)

Sebehodnoticí otázka

Označte na stupnici 1 – 10, jak hodnotíte svou schopnost porozumění psaným textům v Aj (schopnost čtení anglických textů).

1, ..., 10

1 = naprosto nedostatečná; 10 = vynikající

Ukotvující viněty

Jana rozumí základním slovům označujícím např. členy rodiny, běžné předměty, neznámější druhy zvířat atd. Rozumí velmi jednoduchým větám, např. krátkým oznámením na cedulích na ulici, na plakátech nebo v katalogích.

Jak hodnotíte Janinu schopnost porozumění textu?

1, ..., 10

Petr je schopen číst velmi krátké, jednoduché texty. Umí vyhledat konkrétní informaci např. v jídelním lístku, reklamním letáku nebo v jízdním řádu. Rozumí krátkým a jednoduchým příběhům a krátkým a jednoduchým dopisům od kamaráda z ciziny.

Jak hodnotíte Petrovu schopnost porozumění textu?

1, ..., 10

Simona rozumí textům o známých tématech, které jsou psány běžně používaným, každodenním jazykem. Rozumí popisům událostí, pocitů a přání v dopisech, které dostává od kamarádky z ciziny.

Jak hodnotíte Simoninu schopnost porozumění textu?

1, ..., 10

Marek rozumí dlouhým a složitým textům z odborné literatury i z prózy, dokáže ocenit rozdíly ve stylech, jakými jsou texty napsány. Rozumí specializovaným článkům a dlouhým technickým instrukcím.

Jak hodnotíte Petrovu schopnost porozumění textu?

1, ..., 10

Životní spokojenost (dotazník z výzkumu CLoSE <http://czechlongitudinal.blogspot.cz/>, formulace od Voňkové)

Sebehodnoticí otázka

Jak jsi celkově spokojený/á se svým životem?

velmi spokojen/a, spokojen/a, ani spokojen/a ani nespokojen/a, nespokojen/a, velmi nespokojen/a

Ukotvující viněta

Pavel vychází s rodiči dobře, alespoň jednou měsíčně s nimi vyráží na výlet do přírody. Má dva dobré kamarády, kterým může vše říci. S nimi tráví hodně volného času. Mnoho dalších kamarádů však nemá. Pavel chodí do školy rád, ale tak dvakrát za měsíc je stresovaný kvůli písemkám. Ve škole dostává většinou dvojky. Jednou za měsíc se stane, že dostane i čtyřku. Za to mu rodiče vynadají a trvají na tom, že se musí látku doučit.

Jak může podle tebe být Pavel celkově spokojený se svým životem?

velmi spokojen, spokojen, ani spokojen ani nespokojen, nespokojen, velmi nespokojen

Svědomitost (Primi et al., 2016)

Použitý dotazník je brazilský dotazník SENNA 1.0 obsahující otázky týkající se hodnocení vlastních dovedností s úzkým vztahem k taxonomii Big Five (John, Nauman, & Soto, 2008).

Ukotvující viněty

Alice má ve svých věcech nepořádek, nesnáší uklízení domu a obvykle nevypracuje svůj domácí úkol.

Jak je Alice podle tvého názoru svědomitá?

1 = vůbec, 2 = trochu, 3 = středně, 4 = hodně, 5 = zcela

Manuela má dobrý smysl pro pořádek, ale někdy je v její místnosti několik dní nepořádek. Má sklon dokončovat školní úkoly těsně před termínem odevzdání.

Jak je Manuela podle tvého názoru svědomitá?

1 = vůbec, 2 = trochu, 3 = středně, 4 = hodně, 5 = zcela

Juliana je velmi pečlivá a zanáčená. Pravidelně uklízí svůj pokoj, pečlivě plní domácí úkoly a vždy je dokončí včas před termínem odevzdání.

Jak je Juliana podle tvého názoru svědomitá?

1 = vůbec, 2 = trochu, 3 = středně, 4 = hodně, 5 = zcela

3. Příloha: Nová aplikace parametrického modelu – deskriptivní statistiky

Tabulka 14 *Deskriptivní statistiky pro jednotlivé země u druhé položky k učitelově podpoře žáků (H2):* Když je třeba, učitel se nám více věnuje.

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozor- ování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
ALB	0,74	4,84	41,62	52,79	2562	3,46	2	1	6	65	63
ARE	2,49	7,59	41,89	48,03	7192	3,35	9	8	15	52	54
ARG	5,65	13,21	46,25	34,90	3553	3,10	44	37	49	10	20
AUS	3,30	11,36	47,16	38,18	9117	3,20	28	28	30	39	39
AUT	13,82	25,05	34,58	26,56	2979	2,74	68	63	68	1	1
BEL	4,58	11,76	52,45	31,21	5262	3,10	46	47	40	23	29
BGR	2,96	9,68	47,05	40,31	3285	3,25	20	20	22	42	47
BRA	3,53	11,54	46,57	38,36	11678	3,20	30	27	32	36	37
CAN	2,74	7,96	42,23	47,07	13634	3,34	13	9	18	45	51
CHE	3,24	12,09	46,27	38,41	5358	3,20	29	26	34	40	35
CHL	4,69	10,83	40,53	43,95	2828	3,24	21	13	36	21	33
COL	4,76	15,79	50,40	29,05	3389	3,04	53	55	53	20	16
CRI	11,92	22,53	38,04	27,51	2650	2,81	67	59	67	2	2
CZE	2,60	12,60	57,36	27,43	4528	3,10	47	60	33	49	36
DEU	6,67	16,91	45,23	31,19	16399	3,01	58	48	63	7	6
DNK	3,72	13,36	51,83	31,09	3133	3,10	45	50	42	35	27
ESP	2,88	10,79	50,04	36,30	5591	3,20	31	32	26	44	43
EST	6,87	13,68	49,68	29,77	2915	3,02	56	52	52	5	17
FIN	2,38	9,10	46,14	42,37	8096	3,29	17	17	19	54	50
FRA	5,63	15,8	47,98	30,59	3331	3,04	54	51	56	11	13
GBR	2,52	7,25	59,07	31,15	3004	3,19	34	49	14	51	55
GRC	4,34	17,03	49,7	28,94	3271	3,03	55	56	55	27	14
HKG	6,82	16,51	49,89	26,79	3126	2,97	61	62	59	6	10
HRV	5,57	15,15	41,69	37,59	7203	3,11	43	30	54	14	15
HUN	4,29	14,11	41,62	39,98	4324	3,17	37	21	47	28	22
IDN	0,49	4,67	60,07	34,76	3631	3,29	16	38	4	68	65
IRL	5,06	16,88	46,5	31,56	3279	3,05	51	46	57	17	12
ISL	3,04	10,14	46,9	39,92	2205	3,24	22	22	24	41	45
ISR	5,62	16,40	41,11	36,87	3103	3,09	49	31	58	12	11
ITA	5,10	13,99	47,28	33,63	20087	3,09	48	43	50	16	19
JOR	2,57	6,46	38,42	52,55	4258	3,41	4	2	13	50	56
JPN	3,78	14,30	52,18	29,75	4153	3,08	50	53	45	34	24

	Kategorie škály				Počet pozorování	Pořadí					
	Rozhodně nesouhlasím	Nesouhlasím	Souhlasím	Rozhodně souhlasím		Průměr	Průměr	Rozhodně souhlasím	(Rozhodně) souhlasím	Rozhodně nesouhlasím	(Rozhodně) nesouhlasím
KAZ	0,55	3,53	44,38	51,53	3778	3,47	1	3	1	67	68
KOR	4,51	13,59	62,46	19,43	3304	2,97	60	68	46	24	23
LIE	8,87	22,62	34,43	34,08	180	2,94	64	41	66	4	3
LTU	3,47	10,31	35,23	50,99	3003	3,34	12	5	27	37	42
LUX	11,46	20,04	39,35	29,16	3273	2,86	66	54	65	3	4
LVA	2,48	9,37	52,65	35,5	2777	3,21	25	34	21	53	48
MAC	2,30	9,50	61,15	27,05	3518	3,13	40	61	20	55	49
MEX	4,04	12,12	44,61	39,24	21912	3,19	33	25	39	31	30
MNE	4,61	12,55	48,65	34,19	2906	3,12	42	39	43	22	26
MYS	1,43	6,09	48,76	43,73	3359	3,35	10	15	7	59	62
NLD	4,91	19,22	55,09	20,78	2764	2,92	65	67	64	18	5
NOR	4,16	14,65	56,49	24,7	2832	3,02	57	66	48	30	21
NZL	2,73	12,73	49,11	35,43	2698	3,17	38	35	35	46	34
PER	2,61	10,04	52,45	34,91	3585	3,20	32	36	23	48	46
POL	6,06	17,35	51,28	25,30	2987	2,96	63	64	61	8	8
PRT	2,91	10,98	51,99	34,12	3658	3,17	36	40	29	43	40
QAT	3,45	9,77	44,08	42,70	6435	3,26	18	16	25	38	44
QCN	0,86	3,66	49,64	45,85	3449	3,40	5	11	2	64	67
QRS	1,00	7,39	51,84	39,76	1138	3,30	15	23	11	62	58
QUA	4,21	10,83	43,81	41,15	1229	3,22	24	19	31	29	38
QUB	1,52	6,47	42,81	49,21	1065	3,40	7	7	10	58	59
QUC	1,30	6,42	43,01	49,27	1103	3,40	6	6	8	60	61
ROU	4,89	12,09	47,05	35,97	3305	3,14	39	33	41	19	28
RUS	1,81	6,14	45,09	46,96	3411	3,37	8	10	9	57	60
SGP	1,02	4,45	43,42	51,11	3638	3,45	3	4	5	61	64
SRB	4,37	12,85	48,73	34,05	2939	3,12	41	42	44	26	25
SVK	5,59	17,78	51,63	24,99	2965	2,96	62	65	60	13	9
SVN	5,95	17,46	48,89	27,69	3699	2,98	59	58	62	9	7
SWE	4,01	16,00	51,78	28,22	2913	3,04	52	57	51	32	18
TAP	2,19	8,06	56,21	33,54	4005	3,21	26	44	16	56	53
THA	0,96	7,79	58,32	32,93	4366	3,23	23	45	12	63	57
TUN	5,48	8,40	41,05	45,07	2613	3,26	19	12	28	15	41
TUR	3,82	11,93	46,19	38,06	3144	3,18	35	29	37	33	32
URY	4,39	11,64	42,63	41,34	3159	3,21	27	18	38	25	31
USA	2,66	7,99	45,56	43,78	3186	3,30	14	14	17	47	52
VNM	0,66	4,13	55,67	39,54	3279	3,34	11	24	3	66	66
Celkem	3,29	10,23	49,22	37,25	306699	3,2					

Pozn. V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové studentské váhy uvedené

v datovém souboru k žakovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet studentů, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec, označeno jako *(rozhodně) souhlasím*), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec, označeno jako *(rozhodně) nesouhlasím*).

Tabulka 15 *Deskriptivní statistiky pro jednotlivé země u třetí položky k učitelově podpoře žáků (H3): Učitel nám pomáhá s učením.*

	Kategorie škály				Počet pozorování	Pořadí					
	Rozhodně nesouhlasím	Nesouhlasím	Souhlasím	Rozhodně souhlasím		Průměr	Průměr	Rozhodně souhlasím	(Rozhodně) souhlasím	Rozhodně nesouhlasím	(Rozhodně) nesouhlasím
ALB	0,74	3,22	36,45	59,59	2562	3,55	1	1	3	65	66
ARE	2,41	6,39	42,48	48,72	7192	3,38	13	10	23	36	46
ARG	2,11	6,93	43,85	47,11	3553	3,36	17	13	24	44	45
AUS	2,37	7,26	49,86	40,51	9117	3,29	30	29	30	38	39
AUT	16,96	30,56	34,30	18,19	2979	2,54	68	67	68	1	1
BEL	6,12	22,30	50,02	21,56	5262	2,87	61	62	62	9	7
BGR	2,94	11,34	46,98	38,74	3285	3,22	38	32	47	29	22
BRA	1,84	5,74	49,42	42,99	11678	3,34	22	25	16	49	53
CAN	2,40	7,51	45,37	44,71	13634	3,32	25	18	32	37	37
CHE	6,65	19,69	44,76	28,89	7203	2,96	59	49	59	6	10
CHL	2,15	7,66	43,60	46,59	4324	3,35	20	15	31	42	38
COL	1,24	5,18	47,06	46,52	5358	3,39	11	16	9	58	60
CRI	2,36	5,20	42,11	50,34	2828	3,40	9	6	15	39	54
CZE	5,38	21,95	54,15	18,51	3389	2,86	62	66	60	12	9
DEU	10,38	27,49	40,06	22,07	2650	2,74	65	61	65	3	4
DNK	1,53	7,64	61,47	29,35	4528	3,19	41	48	26	53	43
ESP	3,39	10,32	49,24	37,05	16399	3,20	40	37	44	22	25
EST	2,26	9,87	55,22	32,65	3133	3,18	42	43	39	40	30
FIN	2,59	9,55	52,78	35,08	5591	3,20	39	40	40	31	29
FRA	5,86	15,19	51,29	27,66	2915	3,01	55	54	56	10	13
GBR	1,53	5,52	50,00	42,95	8096	3,34	21	26	12	54	57
GRC	3,82	10,68	49,46	36,04	3331	3,18	43	39	49	17	20
HKG	1,90	5,78	61,01	31,31	3004	3,22	37	45	17	46	52
HRV	5,20	22,39	48,52	23,89	3271	2,91	60	59	61	13	8
HUN	4,84	15,82	51,84	27,5	3126	3,02	54	55	54	14	15
IDN	0,42	2,74	48,41	48,44	3631	3,45	3	12	1	68	68

	Kategorie škály				Počet pozorování	Pořadí					
	Rozhodně nesouhlasím	Nesouhlasím	Souhlasím	Rozhodně souhlasím		Průměr	Průměr	Rozhodně souhlasím	(Rozhodně) souhlasím	Rozhodně nesouhlasím	(Rozhodně) nesouhlasím
IRL	3,34	10,54	54,74	31,38	3279	3,14	47	44	46	25	23
ISL	1,87	5,20	48,30	44,63	2205	3,36	18	19	13	47	56
ISR	3,80	14,91	44,18	37,11	3103	3,15	46	36	52	19	17
ITA	3,95	11,21	50,02	34,82	20087	3,16	45	42	50	16	19
JOR	2,85	7,39	34,99	54,76	4258	3,42	7	3	33	30	36
JPN	3,03	12,15	56,06	28,76	4153	3,11	50	50	51	28	18
KAZ	0,58	3,03	43,37	53,02	3778	3,49	2	4	2	66	67
KOR	3,34	8,47	67,29	20,9	3304	3,06	52	63	37	26	32
LIE	9,68	21,22	42,88	26,22	180	2,86	63	57	64	4	5
LTU	2,54	8,15	34,15	55,16	3003	3,42	6	2	34	32	35
LUX	15,05	27,25	37,23	20,46	3273	2,63	66	65	66	2	3
LVA	1,90	7,56	55,48	35,06	2777	3,24	35	41	29	45	40
MAC	1,85	7,49	62,93	27,72	3518	3,17	44	53	27	48	42
MEX	1,60	5,15	43,67	49,58	21912	3,41	8	8	11	52	58
MNE	6,14	15,92	47,30	30,64	2906	3,02	53	46	57	8	12
MYS	1,15	4,98	49,83	44,04	3359	3,37	15	21	7	61	62
NLD	7,69	35,00	46,53	10,79	2764	2,60	67	68	67	5	2
NOR	2,50	9,49	59,68	28,33	2832	3,14	48	51	38	34	31
NZL	1,81	7,55	53,09	37,55	2698	3,26	33	35	28	50	41
PER	1,49	6,47	54,04	38,00	3585	3,29	29	34	20	56	49
POL	4,81	16,1	54,30	24,79	2987	2,99	57	58	55	15	14
PRT	1,51	6,68	53,77	38,05	3658	3,28	31	33	21	55	48
QAT	3,37	9,58	42,00	45,05	6435	3,29	28	17	41	23	28
QCN	1,18	5,34	50,85	42,64	3449	3,35	19	27	10	60	59
QRS	0,75	5,58	49,76	43,9	1138	3,37	14	22	8	64	61
QUA	3,13	8,42	44,84	43,61	1229	3,29	27	23	36	27	33
QUB	1,43	6,45	45,06	47,06	1065	3,38	12	14	18	57	51
QUC	1,23	6,70	43,47	48,59	1103	3,39	10	11	19	59	50
ROU	3,44	9,91	45,32	41,33	3305	3,25	34	28	42	21	27
RUS	1,05	4,33	44,39	50,23	3411	3,44	5	7	5	62	64
SGP	0,80	3,25	47,09	48,86	3638	3,44	4	9	4	63	65
SRB	5,44	18,25	48,31	28,01	2939	2,99	58	52	58	11	11
SVK	3,47	15,79	58,08	22,66	2965	3,00	56	60	53	20	16
SVN	6,57	24,30	48,53	20,60	3699	2,83	64	64	63	7	6
SWE	2,53	11,78	56,12	29,56	2913	3,13	49	47	48	33	21
TAP	1,69	5,45	56,52	36,34	4005	3,28	32	38	14	51	55
THA	0,56	5,43	55,02	38,99	4366	3,32	24	31	6	67	63
TUN	3,81	7,28	37,51	51,40	2613	3,37	16	5	35	18	34
TUR	3,36	10,04	47,53	39,06	3144	3,22	36	30	43	24	26
URY	2,23	6,42	48,30	43,04	3159	3,32	26	24	22	41	47

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozorování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
USA	2,12	6,97	46,43	44,47	3186	3,33	23	20	25	43	44
VNM	2,45	11,42	59,89	26,24	3279	3,10	51	56	45	35	24
Celkem	2,46	8,27	49,20	40,07	306699	3,27					

Pozn. V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové studentské váhy uvedené v datovém souboru k žakovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet studentů, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec, označeno jako *(rozhodně) souhlasím*), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec, označeno jako *(rozhodně) nesouhlasím*).

Tabulka 16 *Deskriptivní statistiky pro jednotlivé země u čtvrté položky k učitelově podpoře žáků (H4): Učitel nám dává příležitost vyjádřit své názory.*

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozorování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
ALB	1,67	4,26	37,06	57,00	2562	3,49	1	1	5	64	64
ARE	6,54	12,51	42,86	38,10	7192	3,13	26	17	32	23	37
ARG	5,45	10,92	46,62	37,01	3553	3,15	19	21	26	35	43
AUS	6,39	17,92	49,02	26,66	9117	2,96	47	47	49	24	20
AUT	14,67	27,95	33,88	23,49	2979	2,66	68	60	68	1	1
BEL	8,39	19,11	50,41	22,09	5262	2,86	59	62	58	11	11
BGR	4,22	9,40	44,06	42,32	3285	3,24	11	9	13	49	56
BRA	4,30	11,70	44,95	39,06	11678	3,19	17	16	23	47	46
CAN	6,79	16,39	46,76	30,06	13634	3,00	43	41	43	19	26
CHE	6,83	18,44	44,69	30,04	7203	2,98	46	42	52	18	17
CHL	5,25	13,98	43,96	36,81	4324	3,12	28	22	33	37	36
COL	2,98	8,09	44,10	44,83	5358	3,31	8	6	10	57	59
CRI	5,10	10,75	42,27	41,88	2828	3,21	15	12	21	41	48
CZE	7,32	16,87	50,32	25,48	3389	2,94	54	50	47	15	22
DEU	12,56	24,57	38,71	24,16	2650	2,74	66	57	67	3	2
DNK	4,29	17,64	56,05	22,01	4528	2,96	48	63	39	48	30

	Kategorie škály				Počet pozo- rování	Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím		Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
ESP	7,21	16,41	46,11	30,28	16399	2,99	44	39	46	17	23
EST	4,77	19,68	51,42	24,13	3133	2,95	49	58	51	43	18
FIN	6,35	20,06	51,76	21,84	5591	2,89	57	64	55	25	14
FRA	12,06	20,10	44,15	23,69	2915	2,79	61	59	63	4	6
GBR	6,60	19,87	46,25	27,27	8096	2,94	52	46	56	22	13
GRC	5,71	11,29	45,33	37,66	3331	3,15	20	18	28	31	41
HKG	3,79	11,49	59,65	25,07	3004	3,06	35	52	17	53	52
HRV	5,11	14,28	50,35	30,26	3271	3,06	37	40	35	40	34
HUN	7,58	16,76	49,14	26,53	3126	2,95	50	48	50	14	19
IDN	1,11	2,75	54,23	41,91	3631	3,37	5	11	2	65	67
IRL	9,06	20,80	45,05	25,09	3279	2,86	60	51	60	10	9
ISL	4,58	11,48	52,03	31,91	2205	3,11	29	35	24	45	45
ISR	6,78	15,18	43,45	34,59	3103	3,06	36	26	40	20	29
ITA	6,76	12,51	48,14	32,59	20087	3,07	34	30	34	21	35
JOR	5,98	7,67	36,99	49,37	4258	3,30	9	4	14	28	55
JPN	10,2	22,60	46,07	21,13	4153	2,78	63	65	64	5	5
KAZ	1,02	4,66	44,83	49,48	3778	3,43	3	3	4	66	65
KOR	7,30	22,40	57,27	13,03	3304	2,76	64	68	59	16	10
LIE	9,16	23,95	37,59	29,30	180	2,87	58	44	65	9	4
LTU	5,66	14,66	42,5	37,18	3003	3,11	30	20	37	33	32
LUX	14,51	20,72	40,59	24,18	3273	2,74	67	56	66	2	3
LVA	2,87	12,67	53,65	30,81	2777	3,12	27	38	19	58	50
MAC	3,59	12,38	59,5	24,53	3518	3,05	38	54	22	55	47
MEX	2,81	7,16	42,78	47,25	21912	3,34	6	5	8	60	61
MNE	6,16	11,56	47,52	34,76	2906	3,11	31	24	29	27	40
MYS	2,85	9,59	52,93	34,63	3359	3,19	16	25	11	59	58
NLD	8,39	22,51	54,01	15,10	2764	2,76	65	67	61	12	8
NOR	5,80	19,53	51,19	23,48	2832	2,92	56	61	53	30	16
NZL	5,56	20,70	48,94	24,80	2698	2,93	55	53	54	34	15
PER	1,94	7,23	48,89	41,94	3585	3,31	7	10	7	63	62
POL	9,28	22,31	49,02	19,39	2987	2,79	62	66	62	8	7
PRT	3,67	11,09	52,19	33,05	3658	3,15	21	29	15	54	54
QAT	9,99	14,22	39,75	36,05	6435	3,02	41	23	48	7	21
QCN	2,40	8,51	51,44	37,65	3449	3,24	12	19	9	62	60
QRS	3,87	11,61	52,42	32,10	1138	3,13	25	34	18	51	51
QUA	8,33	18,39	44,12	29,16	1229	2,94	53	45	57	13	12
QUB	4,75	17,75	45,71	31,79	1065	3,05	39	36	42	44	27
QUC	5,31	13,56	47,24	33,89	1103	3,10	32	28	30	36	39
ROU	4,81	11,46	44,36	39,37	3305	3,18	18	15	25	42	44
RUS	4,00	12,81	49,16	34,03	3411	3,13	24	27	27	50	42

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
SGP	3,28	12,28	52,22	32,21	3638	3,13	23	32	20	56	49
SRB	5,16	13,85	48,80	32,20	2939	3,08	33	33	31	39	38
SVK	6,26	17,17	52,28	24,29	2965	2,95	51	55	44	26	25
SVN	5,94	14,09	50,45	29,52	3699	3,04	40	43	36	29	33
SWE	4,41	17,95	51,54	26,10	2913	2,99	45	49	41	46	28
TAP	2,47	6,56	58,53	32,44	4005	3,21	14	31	6	61	63
THA	0,83	4,70	50,16	44,31	4366	3,38	4	8	3	67	66
TUN	10,1	10,93	34,35	44,63	2613	3,14	22	7	38	6	31
TUR	5,22	9,56	43,53	41,69	3144	3,22	13	13	16	38	53
URY	3,87	9,20	45,33	41,61	3159	3,25	10	14	12	52	57
USA	5,68	17,84	45,5	30,98	3186	3,02	42	37	45	32	24
VNM	0,76	2,71	43,93	52,59	3279	3,48	2	2	1	68	68
Celkem	5,04	12,83	47,25	34,88	306699	3,12					

Pozn. V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové studentské váhy uvedené v datovém souboru k žakovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet studentů, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec, označeno jako *(rozhodně) souhlasím*), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec, označeno jako *(rozhodně) nesouhlasím*).

Tabulka 17 *Deskriptivní statistiky pro jednotlivé země u druhé ukotvující viněty (V2):* Pan učitel Novák dává domácí úkoly z matematiky jednou za týden. Opravené je žákům vždy vrací ještě před zkoušením nebo písemkou.

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
ALB	28,34	40,58	20,93	10,15	2562	2,13	68	61	68	1	1
ARE	3,60	20,22	49,30	26,88	7192	2,99	10	6	17	50	52
ARG	3,54	15,98	50,78	29,71	3553	3,07	4	4	6	51	63
AUS	2,58	19,05	61,26	17,10	9117	2,93	20	33	12	62	57
AUT	5,83	32,55	46,21	15,41	2979	2,71	41	44	44	33	25

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
BEL	3,66	17,24	55,49	23,61	5262	2,99	11	14	10	49	59
BGR	5,08	25,44	47,92	21,57	3285	2,86	24	16	25	38	44
BRA	2,77	19,78	52,57	24,88	11678	3,00	9	11	13	58	56
CAN	4,95	29,53	50,62	14,89	13634	2,75	33	46	30	39	39
CHE	6,39	29,31	48,14	16,16	7203	2,74	36	39	33	31	36
CHL	3,35	17,17	51,88	27,60	4324	3,04	7	5	9	52	60
COL	8,14	26,89	42,67	22,30	5358	2,79	29	15	32	22	37
CRI	2,91	10,15	51,31	35,63	2828	3,20	1	1	1	57	68
CZE	2,10	15,30	62,04	20,55	3389	3,01	8	20	5	65	64
DEU	6,42	38,09	41,9	13,58	2650	2,63	48	50	52	30	17
DNK	2,36	20,36	58,64	18,64	4528	2,94	18	26	14	64	55
ESP	9,09	37,28	39,14	14,49	16399	2,59	53	48	54	19	15
EST	7,40	44,10	39,05	9,45	3133	2,51	59	63	60	25	9
FIN	7,13	40,81	43,44	8,63	5591	2,54	58	66	58	26	11
FRA	13,07	27,69	40,51	18,73	2915	2,65	46	24	46	7	23
GBR	2,74	21,58	56,98	18,7	8096	2,92	22	25	19	59	50
GRC	14,65	35,33	36,68	13,34	3331	2,49	60	51	59	5	10
HKG	8,16	34,28	47,97	9,59	3004	2,59	54	62	48	21	21
HRV	3,95	34,39	47,98	13,68	3271	2,71	40	49	43	44	26
HUN	5,23	31,24	45,46	18,08	3126	2,76	32	30	36	36	33
IDN	1,86	15,28	57,68	25,18	3631	3,06	5	9	4	66	65
IRL	6,65	36,10	47,96	9,29	3279	2,60	51	64	49	29	20
ISL	5,35	17,67	52,53	24,45	2205	2,96	14	13	15	35	54
ISR	3,94	12,00	51,42	32,64	3103	3,13	3	3	3	45	66
ITA	5,10	32,31	45,61	16,98	20087	2,74	35	35	39	37	30
JOR	13,10	21,40	40,87	24,63	4258	2,77	30	12	31	6	38
JPN	2,98	16,68	60,42	19,91	4153	2,97	12	21	7	56	62
KAZ	11,94	42,47	33,41	12,18	3778	2,46	63	55	64	10	5
KOR	8,60	27,29	56,79	7,31	3304	2,63	47	68	34	20	35
LIE	3,27	30,14	50,61	15,97	180	2,79	28	40	28	55	41
LTU	15,02	31,41	37,29	16,28	3003	2,55	57	38	55	4	14
LUX	7,05	29,73	46,13	17,08	3273	2,73	38	34	37	27	32
LVA	11,03	44,40	31,80	12,78	2777	2,46	62	54	65	13	4
MAC	9,95	43,14	38,20	8,72	3518	2,46	64	65	61	16	8
MEX	7,67	30,27	43,00	19,06	21912	2,73	37	22	41	23	28
MNE	9,23	31,81	41,34	17,62	2906	2,67	45	32	47	18	22
MYS	5,46	30,54	46,02	17,99	3359	2,77	31	31	35	34	34
NLD	1,77	21,36	58,26	18,61	2764	2,94	16	27	16	67	53
NOR	7,44	17,01	50,63	24,92	2832	2,93	19	10	20	24	49
NZL	2,43	22,17	59,57	15,83	2698	2,89	23	41	22	63	47

	Kategorie škály					Pořadí					
	Roz- hodně nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- hodně) nesou- hlasím
PER	11,17	36,17	37,12	15,55	3585	2,57	55	43	57	12	12
POL	10,64	33,38	41,07	14,91	2987	2,60	50	45	51	15	18
PRT	3,32	23,93	57,16	15,59	3658	2,85	25	42	24	53	45
QAT	6,04	20,86	47,44	25,65	6435	2,93	21	8	23	32	46
QCN	18,38	44,86	29,01	7,75	3449	2,26	67	67	67	3	2
QRS	9,34	44,63	35,25	10,77	1138	2,47	61	59	62	17	7
QUA	3,95	28,78	50,68	16,59	1229	2,80	27	36	27	43	42
QUB	4,43	33,79	49,94	11,84	1065	2,69	44	57	42	41	27
QUC	3,82	35,04	47,97	13,17	1103	2,70	43	52	45	47	24
ROU	12,98	33,37	35,14	18,51	3305	2,59	52	28	53	8	16
RUS	11,96	43,68	32,39	11,97	3411	2,44	65	56	66	9	3
SGP	4,88	25,92	58,33	10,87	3638	2,75	34	58	26	40	43
SRB	10,86	32,45	40,35	16,35	2939	2,62	49	37	50	14	19
SVK	2,66	21,92	54,38	21,03	2965	2,94	15	18	21	61	48
SVN	6,89	40,35	42,35	10,41	3699	2,56	56	60	56	28	13
SWE	3,73	11,35	51,99	32,94	2913	3,14	2	2	2	48	67
TAP	3,87	17,23	57,83	21,08	4005	2,96	13	17	11	46	58
THA	3,31	30,50	47,82	18,36	4366	2,81	26	29	29	54	40
TUN	23,11	31,14	31,02	14,72	2613	2,37	66	47	63	2	6
TUR	11,77	25,08	42,31	20,84	3144	2,72	39	19	38	11	31
URY	2,70	17,21	53,52	26,58	3159	3,04	6	7	8	60	61
USA	4,33	33,26	49,44	12,97	3186	2,71	42	53	40	42	29
VNM	1,36	22,61	57,11	18,91	3279	2,94	17	23	18	68	51
Celkem	5,65	26,73	49,06	18,55	306699	2,81					

Pozn. V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové studentské váhy uvedené v datovém souboru k žákovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet studentů, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec, označeno jako *(rozhodně) souhlasím*), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec, označeno jako *(rozhodně) nesouhlasím*).

Tabulka 18 *Deskriptivní statistiky pro jednotlivé země u třetí ukotvující viněty (V3): Paní učitelka Ježková dává domácí úkoly z matematiky jednou za týden. Opravené je žákům před zkoušením nebo písemkou nevrací.*

	Kategorie škály					Pořadí					
	Roz- hodně Nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pозо- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- Hodně) nesou- hlasím
ALB	15,11	20,56	35,09	29,24	2562	2,78	1	1	1	66	68
ARE	46,45	27,99	13,74	11,83	7192	1,91	33	19	26	27	43
ARG	40,60	29,31	17,31	12,77	3553	2,02	20	17	19	38	50
AUS	39,41	39,56	12,76	8,27	9117	1,90	34	35	42	39	27
AUT	48,27	33,46	12,45	5,82	2979	1,76	51	48	47	23	22
BEL	51,96	31,32	9,98	6,74	5262	1,72	53	42	52	16	17
BGR	23,54	37,36	23,69	15,41	3285	2,31	10	13	13	60	56
BRA	41,69	32,36	15,85	10,10	11678	1,94	30	28	25	34	44
CAN	52,42	26,82	9,94	10,81	13634	1,79	48	21	43	15	26
CHE	44,05	33,21	16,34	6,40	7203	1,85	38	44	33	31	36
CHL	33,79	34,82	17,60	13,78	4324	2,11	16	14	17	49	52
COL	17,62	32,63	27,30	22,45	5358	2,55	6	7	7	64	62
CRI	53,82	29,15	9,68	7,35	2828	1,71	55	40	50	10	19
CZE	28,48	49,44	16,65	5,42	3389	1,99	23	52	36	56	33
DEU	55,60	30,14	10,53	3,73	2650	1,62	62	64	59	8	10
DNK	37,00	47,39	12,55	3,06	4528	1,82	42	66	57	45	12
ESP	57,05	26,43	11,15	5,36	16399	1,65	59	54	54	5	15
EST	55,97	33,01	6,73	4,29	3133	1,59	65	58	65	6	4
FIN	41,02	41,65	13,10	4,23	5591	1,81	46	59	49	37	20
FRA	41,42	24,70	17,41	16,48	2915	2,09	17	11	15	36	54
GBR	41,64	33,42	11,75	13,19	8096	1,96	27	16	28	35	41
GRC	28,16	30,95	25,47	15,42	3331	2,28	11	12	10	57	59
HKG	53,67	34,69	8,98	2,66	3004	1,61	64	67	63	11	6
HRV	50,83	35,72	9,08	4,37	3271	1,67	57	57	62	18	7
HUN	43,27	38,80	12,50	5,43	3126	1,80	47	51	48	33	21
IDN	20,55	41,46	27,31	10,68	3631	2,28	12	24	14	61	55
IRL	53,05	25,65	8,05	13,25	3279	1,81	43	15	41	13	28
ISL	37,43	40,44	14,69	7,43	2205	1,92	32	38	35	44	34
ISR	44,15	39,13	11,02	5,70	3103	1,78	49	50	53	30	16
ITA	47,81	36,79	10,26	5,14	20087	1,73	52	56	58	26	11
JOR	28,04	23,73	22,29	25,94	4258	2,46	8	4	8	58	61
JPN	55,88	30,33	9,89	3,90	4153	1,62	63	63	61	7	8
KAZ	32,72	34,25	21,13	11,90	3778	2,12	15	18	16	52	53
KOR	26,71	34,02	29,61	9,67	3304	2,22	14	30	12	59	57
LIE	38,15	37,43	17,00	7,41	180	1,94	31	39	29	42	40
LTU	43,28	27,40	18,20	11,12	3003	1,97	25	20	20	32	49

	Kategorie škály					Pořadí					
	Roz- hodně Nesou- hlasím	Nesou- hlasím	Sou- hlasím	Roz- hodně sou- hlasím	Počet pozo- rování	Průměr	Průměr	Roz- hodně sou- hlasím	(Roz- hodně) sou- hlasím	Roz- hodně nesou- hlasím	(Roz- Hodně) nesou- hlasím
LUX	47,88	30,54	14,32	7,25	3273	1,81	44	41	40	24	29
LVA	37,53	35,49	17,50	9,48	2777	1,99	24	31	22	43	47
MAC	58,97	31,02	7,37	2,65	3518	1,54	67	68	68	3	1
MEX	44,34	33,99	13,69	7,98	21912	1,85	37	36	38	29	31
MNE	19,98	29,35	30,42	20,25	2906	2,51	7	9	6	63	63
MYS	38,61	34,94	17,36	9,08	3359	1,97	26	33	23	41	46
NLD	49,58	33,52	12,92	3,98	2764	1,71	54	61	51	21	18
NOR	36,40	38,25	19,34	6,01	2832	1,95	29	46	27	46	42
NZL	35,02	37,28	16,92	10,78	2698	2,03	19	22	21	47	48
PER	13,67	27,13	33,78	25,43	3585	2,71	3	5	3	67	66
POL	34,76	34,86	20,58	9,79	2987	2,05	18	29	18	48	51
PRT	44,50	35,33	12,29	7,88	3658	1,84	40	37	46	28	23
QAT	32,55	27,56	22,38	17,51	6435	2,25	13	10	11	53	58
QCN	61,02	27,92	7,14	3,92	3449	1,54	66	62	64	2	5
QRS	53,05	30,93	9,84	6,17	1138	1,69	56	45	56	14	13
QUA	47,87	28,23	13,59	10,31	1229	1,86	36	27	32	25	37
QUB	51,49	26,80	11,35	10,35	1065	1,81	45	26	37	17	32
QUC	53,32	26,03	9,89	10,77	1103	1,78	50	23	44	12	25
ROU	11,70	26,27	34,70	27,34	3305	2,78	2	2	2	68	67
RUS	57,38	28,67	7,96	6,00	3411	1,63	61	47	60	4	9
SGP	64,67	24,52	6,69	4,12	3638	1,50	68	60	66	1	3
SRB	15,60	31,25	30,95	22,21	2939	2,60	5	8	5	65	64
SVK	31,87	43,74	17,83	6,56	2965	1,99	22	43	30	54	39
SVN	54,72	29,11	10,95	5,22	3699	1,67	58	55	55	9	14
SWE	32,77	44,65	17,17	5,41	2913	1,95	28	53	34	51	35
TAP	39,10	40,61	14,53	5,77	4005	1,87	35	49	45	40	24
THA	48,61	27,05	16,05	8,29	4366	1,84	39	34	31	22	38
TUN	20,06	23,38	29,86	26,71	2613	2,63	4	3	4	62	65
TUR	28,68	24,38	24,24	22,70	3144	2,41	9	6	9	55	60
URY	33,39	40,54	16,62	9,45	3159	2,02	21	32	24	50	45
USA	49,76	28,66	11,18	10,40	3186	1,82	41	25	39	20	30
VNM	49,96	39,85	6,69	3,50	3279	1,64	60	65	67	19	2
Celkem	41,16	32,43	16,18	10,22	306699	1,95					

Poznámka: V prvních čtyřech sloupcích jsou u jednotlivých zemí uvedeny relativní četnosti kategorií škály (v procentech). K výpočtu četností jsou využity celkové studentské váhy uvedené v datovém souboru k žákovskému dotazníku PISA 2012 (OECD, 2017a). V pátém sloupci je pro jednotlivé země uveden celkový počet studentů, který je zahrnut do naší analýzy. I přes užití ordinální škály je u jednotlivých zemí pro orientaci uvedeno průměrné hodnocení (šestý sloupec) a pořadí zemí dle tohoto průměrného hodnocení (sedmý sloupec). Dále je

uvedeno pořadí zemí dle četností v následujících kategoriích: *rozhodně souhlasím* (osmý sloupec), *rozhodně souhlasím* společně se *souhlasím* (devátý sloupec, označeno jako *(rozhodně) souhlasím*), *rozhodně nesouhlasím* (desátý sloupec) a *rozhodně nesouhlasím* společně s *nesouhlasím* (jedenáctý sloupec, označeno jako *(rozhodně) nesouhlasím*).

4. Příloha: Hodnocení nucenou volbou (*forced choice assessment*) – příklad otázek

Následující otázky o strategiích učení se matematice byly využity v dotazníku PISA 2012. V manuálu ke kódování dat OECD (2017a) jsou označeny jako ST53Q01, ST53Q02, ST53Q03 a ST53Q04.

Z každé trojice tvrzení vyber jedno, které nejlépe vystihuje tvůj způsob práce v matematice.

Z trojice zaškrtni pouze jeden čtvereček.

- Když se učím na prověrku z matematiky, snažím se určit si, které části učiva jsou nejdůležitější.*
- Když se učím na prověrku z matematiky, nové pojmy v matematice se snažím pochopit tak, že si je dávám do souvislosti s tím, co už znám.*
- Když se učím na prověrku z matematiky, snažím se naučit se co nejvíc věcí zpaměti.*

Z trojice zaškrtni pouze jeden čtvereček.

- Když se učím matematiku, snažím se přijít na to, kterým pojmem jsem ještě dobře neporozuměl/a.*
- Když se učím matematiku, snažím se přijít na nové možnosti řešení.*
- Když se učím matematiku, ověřuji si, zda si pamatuji to, co jsem se už naučil/a.*

Z trojice zaškrtni pouze jeden čtvereček.

- Když se učím matematiku, snažím se dávat věci do souvislosti s tím, co jsem se naučil/a v jiných předmětech.*
- Když se učím matematiku, začnu tím, že si zjistím, co přesně se mám naučit.*
- Když se učím matematiku, řeším některé úlohy tolikrát, že mám pocit, že bych je dokázal/a vyřešit i ve spánku.*

Z trojice zaškrtni pouze jeden čtvereček.

- Abych si zapamatoval/a způsob řešení určitého typu matematických úloh, řeším stále znovu a znovu vzorové příklady.*
- Přemýšlím o tom, jak by matematická látka, kterou jsem se naučil/a, mohla být využita v každodenním životě.*
- Vždy, když něčemu v matematice nerozumím, hledám další vysvětlující informace.*

5. Příloha: Test posouzení situace (*situational judgement test*) – příklad otázek

Následující otázky k řešení problémových situací byly využity v žákovském dotazníku PISA 2012.

Situace – mobil (v manuálu ke kódování dat OECD (2017a) jsou označeny následující otázky jako ST96Q01, ST96Q02, ST96Q03 a ST96Q05)

Ted' si představ, že ze svého mobilu již několik týdnů posíláš SMS. Dnes ti to ale vůbec nejde. Chceš tento problém vyřešit.

Co uděláš? U každého návrhu vyber možnost, která tě nejlépe vystihuje.

(V každém řádku zaškrtni pouze jeden čtvereček.)

	<i>Určitě bych to udělal/a</i>	<i>Pravděpodobně bych to udělal/a</i>	<i>Pravděpodobně bych to neudělal/a</i>	<i>Určitě bych to neudělal/a</i>
<i>Zmáčknou postupně všechna tlačítka, abych zjistil/a, co nefunguje.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Přemýšlím, čím to může být způsobeno a jak to mohu vyřešit.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Přečtu si návod.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Požádám kamaráda o pomoc.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Situace – návštěva ZOO (v OECD (2017a) jsou následující otázky označeny jako ST101Q01, ST101Q02, ST101Q03 a ST101Q05)

Představ si, že chceš jít s bratrem do zoo. Nevíš ale, jak se tam dostanete.

Co uděláš? U každého návrhu vyber možnost, která tě nejlépe vystihuje.

(V každém řádku zaškrtni pouze jeden čtvereček.)

	<i>Určitě bych to udělal/a</i>	<i>Pravděpodobně bych to udělal/a</i>	<i>Pravděpodobně bych to neudělal/a</i>	<i>Určitě bych to neudělal/a</i>
<i>Podívám se, jestli je cesta popsaná v brožurce o zoo.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vyhledám nejlepší trasu na mapě.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Nechám na bratrovi, aby zjistil, jak se tam dostaneme.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vím, kde přibližně zoo je, a navrhnu tedy, abychom vyrazili.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Situace – prodej jízdenek (v OECD (2017a) jsou následující otázky označeny jako ST104Q01, ST104Q04, ST104Q05 a ST104Q06):

Představ si, že přijdeš na vlakové nádraží. Je tam automat na prodej jízdenek, který jsi ještě nikdy nepoužil/a. Chceš si koupit jízdenku.

Co uděláš? U každého návrhu vyber možnost, která tě nejlépe vystihuje.

(V každém řádku zaškrtni pouze jeden čtvereček.)

	<i>Určitě bych to udělal/a</i>	<i>Pravděpodobně bych to udělal/a</i>	<i>Pravděpodobně bych to neudělal/a</i>	<i>Určitě bych to neudělal/a</i>
<i>Zjistím, jak moc je podobný jiným automatům na jízdenky, které už jsem používal/a.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vyzkouším všechna tlačítka, abych zjistil/a, co se stane.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Požádám někoho o pomoc.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Zkusím na nádraží najít pokladnu, abych si lístek koupil/a tam.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Seznam tabulek

Tabulka 1 Korigované (sebe)hodnocení C pomocí neparametrického přístupu v případě dvou vinět v_1 a v_2 (s přirozeným řazením $v_1 < v_2$) a sebehodnocením s	65
Tabulka 2 Korigované (sebe)hodnocení C pomocí neparametrického přístupu v případě tří vinět v_1 , v_2 a v_3 (s přirozeným řazením $v_1 < v_2 < v_3$) a sebehodnocením s	67
Tabulka 3 Seznam analyzovaných zemí/regionů z výzkumu PISA 2012 včetně jejich zkratk.....	136
Tabulka 4 Hodnocení učitelovy podpory pro všechny dotázané žáky ze zemí z výzkumu PISA 2012: relativní četnosti (v %) a průměrné hodnocení	139
Tabulka 5 Deskriptivní statistiky pro jednotlivé země u první položky k učitelově podpoře žáků (H1): <i>Učitel nám říká, že musíme pilně pracovat...</i>	140
Tabulka 6 Hodnocení tří ukotvujících vinět pro všechny dotázané žáky ze zemí z výzkumu PISA 2012: relativní četnosti (v %) a průměrné hodnocení	144
Tabulka 7 Deskriptivní statistiky pro jednotlivé země u první ukotvující viněty (V1): <i>Paní učitelka Novotná dává domácí úkoly z matematiky obden. Opravené je žákům vždy vrací ještě před zkoušením nebo písemkou.</i>	145
Tabulka 8 Odhady koeficientů mezi γ^j CHOPIT modelu.....	150
Tabulka 9 Odhady mezi τ^j v CHOPIT modelu a pořadí zemí dle těchto odhadnutých mezi	151

Tabulka 10 Porovnání odhadů koeficientů β u CHOPIT modelu (korigované efekty zemí) a probitového modelu pro ordinální proměnnou (nekorigované efekty zemí)	164
Tabulka 11 Tendence k volbě určitých škálových kategorií.....	178
Tabulka 12 Signálně-detekční teorie – podnět a odpověď přítomny či nepřítomny	185
Tabulka 13 Korelace mezi motivací učit se matematiku a výsledkem v testu matematické gramotnosti po očištění o rozdílné užívání škály.....	187
Tabulka 14 Deskriptivní statistiky pro jednotlivé země u druhé položky k učitelově podpoře žáků (H2): <i>Když je třeba, učitel se nám více věnuje.</i>	233
Tabulka 15 Deskriptivní statistiky pro jednotlivé země u třetí položky k učitelově podpoře žáků (H3): <i>Učitel nám pomáhá s učením.</i>	235
Tabulka 16 Deskriptivní statistiky pro jednotlivé země u čtvrté položky k učitelově podpoře žáků (H4): <i>Učitel nám dává příležitost vyjádřit své názory.</i>	237
Tabulka 17 Deskriptivní statistiky pro jednotlivé země u druhé ukotvující viněty (V2): <i>Pan učitel Novák dává domácí úkoly z matematiky jednou za týden. Opravené je žákům vždy vrácí ještě před zkoušením nebo písemkou.</i>	239
Tabulka 18 Deskriptivní statistiky pro jednotlivé země u třetí ukotvující viněty (V3): <i>Paní učitelka Ježková dává domácí úkoly z matematiky jednou za týden. Opravené je žákům před zkoušením nebo písemkou nevrací.</i>	242

Seznam obrázků

Obrázek 1. Základní myšlenka metody ukotvujících vinět: ilustrace neparametrického přístupu pro dva respondenty	28
Obrázek 2. Základní myšlenka metody ukotvujících vinět: ilustrace parametrického modelu pro dvě země	30
Obrázek 3. Základní myšlenka metody ukotvujících vinět: ilustrace parametrického modelu s výčtem všech možností porovnání skutečných a reportovaných distribucí.....	32
Obrázek 4. Neparametrický přístup pro začínající výzkumníky: Postup získání korigovaného sebehodnocení v případě dvou vinět s hodnocením odpovídajícím přirozenému pořadí	60
Obrázek 5. Neparametrický přístup pro začínající výzkumníky: postup získání korigovaného sebehodnocení v případě dvou vinět se stejným hodnocením	61
Obrázek 6. Neparametrický přístup pro začínající výzkumníky: postup získání korigovaného sebehodnocení v případě dvou vinět s hodnocením opačným, než je přirozené pořadí	63
Obrázek 7. Ukázka zpracování dat v R: histogram korigovaného sebehodnocení pomocí neparametrického způsobu.....	106
Obrázek 8. Geografické rozdělení zemí dle odhadnuté první meze τ^1 CHOPIT modelu	157
Obrázek 9. Geografické rozdělení zemí dle odhadnuté druhé meze τ^2 CHOPIT modelu	158
Obrázek 10. Geografické rozdělení zemí dle odhadnuté třetí meze τ^3 CHOPIT modelu	159

Obrázek 11. Vztah mezi první mezí (hranice mezi *rozhodně nesouhlasím* a *nesouhlasím*) a třetí mezí (hranice mezi *souhlasím* a *rozhodně souhlasím*) .160

Obrázek 12. Rozdělení hodnocení učitelovy podpory žáků v jednotlivých zemích po korekci o rozdílné užívání škály.....169

Obrázek 13. Rozdělení hodnocení učitelovy podpory žáků v jednotlivých zemích před korekcí o rozdílné užívání škály.....170

Summary

This book focuses on the anchoring vignette method and (possibilities of) its use in educational research. The anchoring vignette method helps to correct (self-)assessments in questionnaire surveys for the heterogeneity in reporting behavior among different (groups of) respondents. Imagine that respondents are asked to rate their motivation for learning (PISA 2015 Student questionnaire): *How much do you disagree or agree with statements about yourself below? I enjoy acquiring new knowledge in science.* The rating scale is *strongly disagree, disagree, agree, strongly agree*. Though the *actual* level of motivation would be the same for two respondents, one can rate his motivation as very high and another as moderate. In these cases, the direct comparison of self-assessments would lead to incorrect results. When using the anchoring vignette method, respondents are asked to rate not only their own situation but also a situation of a hypothetical person described in a short story, so called anchoring vignette. An example of an anchoring vignette for motivation is following (PISA 2015 Student questionnaire): *Please read the description about the following three students. Based on the information provided here, how much would you disagree or agree with the statement that this student is motivated? Olivia mostly remains interested in the tasks she started and sometimes does more than what is expected from her. **Olivia is motivated.** strongly disagree, disagree, agree, strongly agree.* All respondents rate the same situation described in vignettes. The heterogeneity in evaluation of the vignettes may therefore be interpreted as the heterogeneity in respondents' reporting behavior. We then use the information about the heterogeneity in reporting behavior identified by vignettes evaluations to correct (adjust) (self-)assessments.

The use of the anchoring vignette method in educational research is not common despite the following two observations. First, there is long-term evidence about response scale differences among different groups of countries, schools and other socio-economic groups. Second, the anchoring vignette method showed its potential to correct for response scale differences in other research fields like research in health, well-being, or political efficacy. This book should help to increase the usage of the anchoring vignette method in

educational research. As far as we are aware of, there is no book about the anchoring vignette method and its use in educational research. This book is therefore a first attempt to summarize information about the method in the context of educational research.

The book is divided into seven main chapters. The first chapter describes theoretical basics of the method. It discusses methodological problems when comparing different groups, taxonomy of bias (construct bias, method bias and item bias) is described in more details. Explanation of how the anchoring vignette method could solve these methodological issues follows. The chapter concludes with the description of the main principle of the anchoring vignette method using comparison of two hypothetical countries and two individuals in case of response scale differences.

The second chapter presents key examples of using anchoring vignette method in social sciences – political efficacy (King et al., 2004), visual acuity (King et al., 2004) and work disability (Kapteyn et al., 2007). Next, applications of the method in educational research are summarized. These are divided into three groups: a) parents' evaluation of principal-parent relations, teacher-parent relations and parent-parent relationships, b) students' self-assessment of (non-)cognitive skills, school discipline and life satisfaction, c) students' evaluation of teacher classroom management and teacher support behavior.

The third chapter presents statistical approaches of the anchoring vignette method – nonparametric (simple) approach and parametric CHOPIT model. Nonparametric approach compares rankings of vignette ratings and self-assessments across groups. We present several examples how to relate vignette ratings and self-assessments so that researchers new to the method should fully understand the principle of the nonparametric approach. Concerning parametric CHOPIT model, we describe not only its standard specification but also its extensions, for example, by the time dimension. For the standard specification, we derive the maximum likelihood function.

The fourth chapter presents an example of data analysis in two software packages, R and STATA. We show an example of code which can be used for

the estimation of all basic statistics of both nonparametric approach and parametric model. For nonparametric approach, we mainly focus on calculations of relative ranks, vignette selection and plotting adjusted self-assessments in R. For parametric model, we show estimation of CHOPIT model parameters in both R and STATA. We use freely available dataset, so all readers can try to calculate the basic statistics themselves.

The fifth chapter focuses on formulation of vignettes and their position in the questionnaire. The anchoring vignette method rests upon two key assumptions, response consistency and vignette equivalence. Response consistency says that respondents use the same scale when evaluating themselves and when evaluating the vignette persons. Vignette equivalence says that different respondents interpret the same vignette in the same way. Researchers must think about the two assumptions in each step of their vignette research – formulating vignettes, questionnaire construction, data collection, data analysis and interpretation of data analysis results. We provide recommendations how to support the two assumptions and show possibilities how to test them. Next, we also focus on discriminating power of vignettes, number of vignettes and the order of vignettes and self-assessment questions in questionnaire.

The sixth chapter presents a new application of the parametric model for comparison of students' evaluation of teacher support behavior in PISA 2012 countries (68 countries, about 300,000 students). We first analyze the heterogeneity in reporting behavior among countries. Next, we compare the original students' evaluations with vignette-adjusted evaluations among countries. Finally, we compare (vignette-adjusted) evaluations of teacher support behavior with students' performance in PISA mathematics literacy test.

The seventh chapter presents four alternative approaches which can be used for the correction of the heterogeneity in response style behavior among different groups of respondents. They could be considered as alternatives to the anchoring vignette method. The first approach is based on the identification of a tendency to choose scale points regardless of the actual, underlying attitude (Buckley, 2009). The second approach is the so-called Overclaiming Technique. It is used to estimate both respondents' concept familiarity and their tendency to

overstate what they know (OECD, 2014). The third approach are the so-called Situational Judgement Tests. They present short descriptions of situation with several possible responses, and typically ask respondents to select the best option (OECD, 2014). The fourth approach is the so-called Forced-Choice Assessment. Respondents are asked to choose one out of several alternative descriptions (OECD, 2014).

The author hopes that this book will help to understand the issue of the heterogeneity in reporting behavior among different cultures, countries, schools, groups and individuals. The author strongly hopes that this book will contribute to the increase of the use of anchoring vignette method in educational research. The book can also be used by psychologists, sociologists and other social scientists who deal with the common problem of response scale differences among respondents.

Rejstřík

C

cenzorovaný probitový model pro ordinální proměnnou (*censored ordered probit model*), 69, 104
CLoSE (Czech Longitudinal Study in Education), 44, 52, 231

D

diskriminační síla vinět, 103, 126
diskriminační validita (*discriminant validity*), 46
dummy proměnná, 87, 148

E

efekt růžových brýlí, 43
efekt velké ryby v malém rybníce (*Big Fish in a Little Pond Effect*), 24
ekvivalence (*equivalence*), 19
ekvivalence vinět (*vignette equivalence*), 115, 124
entropie, 71, 102
ERA Group (Educational Research Anchors Group), 11, 43

H

heterogenita ve stylu odpovídání na otázky (*heterogeneity in reporting behavior*), 22
histogram, 68, 102, 106
histogram bez vektorových hodnot, 68, 102
histogram dle principu entropie, 71, 102
histogram s rovnoměrným rozložením prostoru, 68, 102
histogram s rozložením dle podobných respondentů, 69, 102
hodnocení nucenou volbou (*forced choice assessment*), 175, 188, 245

Ch

CHOPIT model (*compound hierarchical ordered probit model*), 73, 108, 147
chyba měření (*measurement error*), 64

I

identifikace parametrů, 80
individuální efekt, 82
intervalová hodnota, 65, 68, 101

K

Kolmogorův-Smirnovův test, 106
konzistence odpovědí (*response consistency*), 115
korigované sebehodnocení, 57, 64, 97
kriteriální validita (*criterion validity*), 46

L

Likertova škála, 13, 23, 57

M

metoda maximální věrohodnosti, 76, 149
model pro stupňovaný kredit (*partial credit model*), 73
motivace k učení, 45

N

náhodný styl odpovídání (*noncontingent responding*), 180
nekázeň, 51, 227
nekonzistence (*inconsistencies*), 66
neparametrický přístup (*nonparametric approach*), 57, 89

O

osobnostní charakteristiky, 46, 231
ostrost zraku (*visual acuity*), 36

P

panelová data, 86
parametrizace, 80
PISA (Programme for International Student Assessment), 26, 45, 53, 133, 180, 185, 188, 190, 245, 247
politická svoboda, 89, 90
pořadí vinět, 92
probitový model pro ordinální proměnnou (*ordered probit model*), 69, 163
přesmyčky (*ties*), 66
přijatelné hodnoty (*plausible values*), 45

R

referenční zkreslení (*reference bias*), 23
replikační viněta (*replica vignette*), 116
rovnice pro sebehodnocení, 73
rovnice pro viněty, 75
rozdílné fungování položky (*differential item functioning*), 21

S

sebekotvící škála (*self-anchoring scale*), 33
SHARE (Survey of Health, Ageing and Retirement in Europe), 34, 211
signálně-detekční teorie (*signal detection theory*), 184
skalár, 65, 101
software R, 89
software STATA, 112
spokojenost se životem, 52, 231
spolupráce mezi rodiči, 43, 224
spolupráce mezi rodiči a ředitelem, 43, 223
spolupráce mezi rodiči a učiteli, 43, 224

T

technika přehánění (*overclaiming technique*), 175, 184

tendence k volbě určitých škálových kategorií, 175, 178
tendence volit extrémní kategorie (*extreme response style*), 179
tendence volit nesouhlasné kategorie (*disacquiescence response style*), 178
tendence volit souhlasné kategorie (*acquiescence response style*), 178
tendence volit spíše souhlasné kategorie (*net acquiescence response style*), 179
tendence volit středové kategorie (*midpoint responding*), 179
tendence volit určitý rozsah odpovědí (*response range*), 179
test posouzení situace (*situational judgement test*), 176, 190, 247

U

učitelova podpora žáků (*teacher support behavior*), 55, 133, 226
učitelovo řízení třídy (*teacher classroom management*), 53, 225

V

vektor, 65
vnímaná politická účinnost (*political efficacy*), 33, 34, 211
vnitřní konzistence (*internal consistency*), 46
výběr vinět, 103

Z

zavedená kotvící otázka (*designated anchors*), 33
zdravotní problémy při výkonu práce (*work disability*), 38
zkreslení (*bias*), 19
zkreslení způsobené administrací (*administration bias*), 20
zkreslení způsobené agregací (*aggregation bias*), 183
zkreslení způsobené konstruktem (*construct bias*), 20, 24
zkreslení způsobené metodou (*method bias*), 20, 22

zkreslení způsobené nástrojem (*instrument bias*),
21
zkreslení způsobené obsahem položky (*item
bias*), 21, 24

zkreslení způsobené výběrem (*sample bias*), 20
znalost anglického jazyka, 49, 230
znalost ICT (Information and Communication
Technology), 47, 228

Metoda ukotvujících vinět a její využití v pedagogickém výzkumu

Hana Voňková

Vydala Univerzita Karlova — Pedagogická fakulta

Rok vydání: 2017

Počet stran: 262

Formát: B5

1. vydání

Vytiskla tiskárna Nakladatelství Karolinum

ISBN 978-80-7290-954-4

Hana Voňková (1980) se ve své vědecké a pedagogické činnosti zaměřuje především na aplikovanou statistiku a kvantitativní metodologii v pedagogickém a zdravotnickém výzkumu. Vystudovala Matematicko-fyzikální fakultu Univerzity Karlovy (2003 Mgr., 2012 RNDr.), Pedagogickou fakultu Univerzity Karlovy (2005 Mgr., 2006 PhDr., 2008 Ph.D.) a nizozemskou Tilburg University (2011 Ph.D.). V současné době působí jako akademický pracovník a výzkumník na Pedagogické fakultě Univerzity Karlovy. Zároveň spolupracuje se dvěma americkými univerzitami – Center for Economic and Social Research na University of Southern California (CESR USC) a Department of Education Reform na University of Arkansas (EDRE UARK). Dříve působila též jako výzkumný pracovník na 1. lékařské fakultě Univerzity Karlovy (2011–2015) a jako výzkumný asistent na Tilburg University (2006–2007). Je zakladatelkou výzkumné platformy Educational Research Anchors Group (ERA Group), která se zabývá aplikacemi metody ukotvujících vinět v pedagogickém výzkumu.

Monografie **Metoda ukotvujících vinět a její využití v pedagogickém výzkumu** představuje první komplexní pojednání o ukotvujících vinětách v české pedagogické literatuře. Metoda ukotvujících vinět (anglicky *anchoring vignettes method*) umožňuje *ukotvit* odpovědi na sebehodnotící otázky tak, aby je bylo možno porovnávat. Metoda tak může pomoci odstranit jednu z častých slabin kvantitativního výzkumu, totiž omezenou validitu při použití posuzovacích škál. Publikace je určena začínajícím i pokročilým výzkumníkům v pedagogických a sociálních vědách.

