

UNIVERZITA KARLOVA V PRAZE

FAKULTA SOCIÁLNÍCH VĚD

Institut sociologických studií

Diplomová práce

2012

Pavel Straka

UNIVERZITA KARLOVA V PRAZE

FAKULTA SOCIÁLNÍCH VĚD

Institut sociologických studií

Pavel Straka

**Metody testování dotazníků a možnost
zapojení výsledků MTMM studií**

Diplomová práce

Praha 2012

Autor práce: **Pavel Straka**

Vedoucí práce: **PhDr. Jiří Vinopal, Ph.D.**

Rok obhajoby: 2012

Bibliografický záznam

Straka, Pavel. *Metody testování dotazníků a možnost zapojení výsledků MTMM studií*. Praha, 2012. 79 s. Diplomová práce (Mgr.) Univerzita Karlova, Fakulta sociálních věd, Institut sociologických studií. Katedra sociologie. Vedoucí diplomové práce PhDr. Jiří Vinopal, Ph.D.

Abstrakt

Na proces evaluace otázek se dá pohlížet ze dvou úhlů. První se zabývá otázkami, jako je pochopení otázek a použitých pojmů, znění otázek, zátěž respondenta, problémy získání požadované informace z paměti, proces vytvoření odpovědi atd. Druhý se zabývá tím, jak se tyto problémy projevují kvantitativně, tzn. zabývají se odhadem chyby měření. Prvním úhlem pohledu a jeho kritickým zhodnocením jejich aplikací se zabývá první kapitola práce. Druhá kapitola se zabývá možnostmi, které nabízí druhý přístup. Statistické přístupy se v kvalitě otázek zabývají reliabilitou, validitou nebo obojím současně. Možnost vyhodnocení těchto indikátorů kvality ovšem přináší jisté nároky na realizaci a design šetření. K odhadům kvality měření se využívá několik přístupů. V české literatuře o nich pojednal Řehák (Řehák 1998) a z toho důvodu se jim tato práce věnuje pouze krátce. Naopak je pozornost zaměřena na další rozpracování MTMM designu, které dosud v české literatuře probráno není. Následné meta analýzy, pracující s výsledky MTMM experimentů se pokouší stanovit míru vlivu designu dotazu (a tedy i některých rozhodnutí výzkumníka o tom, jak daný dotaz koncipovat) na kvalitu dotazu (validitu a reliabilitu) a využít takto získané poznatky pro predikci. Předpokladem samotné predikce je jednak realizace dostatečného počtu MTMM experimentů a jejich následná analýza, která bude natolik robustní, že její výsledky budou použitelné k predikci. O tom, že je důležitým explanandem hodnot kvality dotazu jeho konkrétní znění a další parametry sekundární analýza výsledků MTMM experimentů potvrdila. Principiálně tedy jsou výsledky analýzy použitelné a můžeme na základě agregovaných zjištění z MTMM experimentů odhadovat i aktuální hodnoty odhadu kvality. Možnosti uplatnění predikčního nástroje SQP se zabývá poslední část práce.

Abstract

The process of evaluation of questions can be viewed from two point of views. The first deals with issues such as comprehension of the questions and used terms, the wording of questions, the respondent burden, problems of retrieving requested information from memory, the judgment of appropriate answer etc. The second looks at how these problems are manifested quantitatively, i.e. deal with the estimation of the measuring errors. The first angle of view and his critical assessment of their applications is analyzed in first chapter of the thesis. The second chapter deals with the possibilities offered by the second approach. Statistical approaches to quality issues are concerned to reliability, validity or both. The possibility of evaluation of these quality indicators, however, brings some requirements for realization and design of the survey. To estimate the quality of measurement several approaches are used. In the Czech literature are discussed by Řehák (Řehák 1998) and for this reason, the thesis is deals with them only briefly. On the other hand, the attention is focused on further developing of MTMM design which isn't discussed in the Czech literature so far. Subsequent meta-analyses, working with the results of MTMM experiments, attempt to determine the influence and effect of the question characteristics (and therefore also some of the researcher's decisions about how to operationalize the questions) and use the obtained findings for the prediction. Presumption of prediction itself is realization of a sufficient number of MTMM experiments and their subsequent analysis that will be robust enough that its results will be applicable to the prediction. The fact that the important explandum of quality values is its concrete wording and other characteristics was confirmed by the results of the secondary analysis of MTMM experiments. In principle the analysis results are applicable and we can, on the basis of the aggregated findings from MTMM experiments, estimate also actual/current values of the equality estimation. The last part of the thesis deals with the possibility of using of SQP prediction tool.

Klíčová slova

metody testování dotazníků, kvalita měření, MTMM design, SQP

Keywords

evaluation of survey questions, pretesting, quality measurement, MTMM design, SQP

Rozsah práce: 125 991 znaků (včetně mezer)

Prohlášení

1. Prohlašuji, že jsem předkládanou práci zpracoval/a samostatně a použil/a jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze dne 18.5.2012

Pavel Straka

Poděkování

Na tomto místě bych rád poděkoval vedoucímu práce PhDr. Jiřímu Vinopalovi, Ph.D za pomoc, podporu a odborné vedení diplomové práce. Táně a rodině za podporu a pochopení.

Námět práce

Praxe pretestování výzkumů se natrvalo usadila ve všech publikacích týkajících se praxe sociálního výzkumu, v nichž bez výjimky najdeme poukazy na její nepostradatelnost. Není se čemu divit, když se jedná o jediný způsob jak zhodnotit připravenost dotazovacích instrumentů (ale nejen jich) „nanečisto“. Tato deklarace o důležitosti se ale rozchází s tím kolik pozornosti je pak tomuto problému věnováno. Stejně tak je nelehké získat z výzkumných zpráv informace, jestli byl vůbec pilot proveden, případně jakým způsobem a s jakými výsledky. Navíc až donedávna bylo v této oblasti provedeno jen nemnoho metodologických studií.

Tradiční způsob pretestování spočívá v tom, že nemnoho tazatelů provede několik rozhovorů (maximálně v desítkách respondentů), kdy takto získané údaje slouží pro ověření dotazníku v podobě univariačních statistik a debriefingu tazatelů. A zdá se, alespoň podle metodologických statí a doporučení, že sociální vědci k tomuto přístupu chovají značnou důvěru (Sheatsley 1983, Sudman 1983). Tato metodologická víra se pravděpodobně zakládá na zkušenosti, že i malé množství uskutečněných rozhovorů dokáže odhalit problémy s chybějícími odpovědními kategoriemi, formulacemi nebo neoprávněnými předpoklady. O jaké vědecké důkazy se ovšem v tomto případě opíráme, které nám dávají jistotu, že takto odhalíme ty hlavní problémy?

Tradiční způsob pretestování stojí na předpokladu, že problémy dotazníku lze odhalit jednak přímo prostřednictvím odpovědí na konkrétní otázky (množství odmítnutí, nevím), a také skrze chování respondentů (vizuální projevy nepohody nebo váhání), o nichž může podat zprávu tazatel. Takovýto způsob pretestování je ovšem principiálně slepý k některým druhům problémů, a to zejména z toho důvodu, že některé problémy zkrátka nejsou odpozorovatelné z chování respondentů nebo odpovědí. Dohromady s tím, že pretestování probíhá většinou v nedeklarované formě (oproti formě, kdy jsou respondenti vyzváni k participaci na pretestování), pak můžeme dojít k názoru, že tradiční pretestování spíše slouží k prověření problémů, které může dotazník přinést tazatelům.

V případě, že tazatel popíše problémy, s nimiž se při dotazování setkal, tak jakými pravidly nebo návody disponuje výzkumník, aby rozhodl o jejich úplnosti, závažnosti

nebo možné revizi? Zdá se, že důvěra v práci a sensitivitu tazatelů se v tomto přístupu propojuje s víceméně intuitivním vyhodnocením chyb (Cannell a Kahn, 1953).

Tato situace, pro mnohé nespokojivá, vyústila na jedné straně k celkovému konceptu kvality dat, na straně druhé k implementaci nových metod testování.

Groves et al. v práci *Survey Methodology* nabízí tři standardy, které by výzkumné otázky měly splňovat. Jedná se o standard obsahový, kognitivní a použitelnosti. Každou z metod testování dotazníku můžeme pak chápat jako nástroj, který každý svým způsobem a do určité míry prověřuje jednotlivé standardy. Posouzení experty, skupinové diskuse, kognitivní rozhovory, kodování chování (behavior coding) nebo experimentální přístup a další metody rozšířily možné pole aplikací. K jakým výsledkům to vede? Zdá se, že tento pokrok spolu s novými poznatky a rozměry úvah přináší i některé nové otázky, specifické té které metodě testování (uvedme zde např. případ kognitivních rozhovorů nebo experimentálních přístupů) nebo i obecného charakteru.

Z prvního ranku např. problematika týkající se kognitivních rozhovorů. Ačkoliv se zdá, že přinášejí validní informace, otázkou zůstává reliabilita těchto dat, jejich hodnota pro praktické využití a vliv konkrétního způsobu provedení kognitivních rozhovorů (DeMaio a Landreth 2004, Beaty 2004, Fowler 2004, Willis, DeMaio a Harris-Kojetin 1999, Forsyth, Rothgeb a Willis 2004). Jindy výzkumníci využívají takový druh studií, kdy experimentálně srovnávají rozdílné způsoby sběru dat, rozdílné metodiky nebo odlišné verze otázek (Tourangeau 2004, Fowler 2004). Tyto metody sice dokážou změřit, jaký vliv na odpovědi má podoba otázky, jejich pořadí, metoda sběru dat atd., tj. jak rozdílné metodologické nastavení produkuje rozdílné výstupy, ale jen málokdy dokáže říci, který přístup produkuje data nejlépe.

V obecně metodologické rovině by se pak zřejmě také mělo věnovat daleko více pozornosti tomu, do jaké míry tyto jednotlivé přístupy přinášejí srovnatelné nebo lépe identické informace (Presser a Blair 1994, Forsyth, Rothgeb a Willis 2004).

Nové metody dokázaly odpovědět na některé předchozí výzvy (nejvýrazněji problém rozumění), jiné naopak vyřešit nedokázala a ještě další nově habilitovala.

Každý z těchto nových přístupů tak přináší nové poznatky, ale stejně si s sebou nese svá omezení. Opět se tak kruhem vracíme k tomu, že i přes rozšíření skupiny metod, použitelných pro pretestování dotazníku, takřka nezískáváme informace o tom, jestli jsme identifikovali důležité problémy, co a jakým způsobem revidovat ani o účinku

revizí na nově získaná data. Stále je výzkumník ponechán v zajetí své intuice, při rozhodování, jaký výstup té které hodnoty představuje již problém, který je nutno řešit nebo naopak, kdy lze identifikovaný problém s lehkou hlavou přehlédnout.

Cíl práce

Design dotazníku a statistické modelování se zdají být dvěma světy, mezi kterými bychom našli jen nemnoho společného. Naštěstí tato nevšimavost není nevyhnutelná a najdou se přístupy, které se snaží oba světy propojit. Jedním z nich je i přístup Willema Sarise, Williama van der Veld a Irmtraud Gallhofer. V 80. letech Frank Andrews aplikoval multitrait-multimethod (MTMM) měření k odhadu reliability a validity dotazníkových otázek a přišel s myšlenkou, aby takovéto výsledky byly dále použity k odhadu reliability a validity dalších výzkumných otázek. V souladu s jeho návrhem pak (mezi jinými) Saris spolu s kolegy shromáždil provedené studie a vyvinuli kódovací systém, na jehož základě lze dotazníkové položky charakterizovat vzhledem k povaze jejich obsahu, komplexitě, formátu odpovědi, pozici v dotazníku, způsobu sběru dat, typu výběru atd. Následně pomocí regresního modelu, v němž byly tyto jednotlivé charakteristiky jako nezávislé proměnné a jako proměnné závislé hodnoty MTMM identifikovali vhodný model. Hodnoty koeficientů tohoto výsledného modelu jsou pak dále využitelné pro odhady efektu těchto vstupních parametrů otázky na výsledné hodnoty validity a reliability.

Cílem práce bude detailně prozkoumat možnosti, které tento způsob testování do pole problémů přináší, jeho využitelnost i omezení. Měl by být také pokusem o prověření možností tohoto modelu v ČR. V případě vyhodnocení experimentů realizovaných v ČR by se otevírala možnost konfrontace validity predikcí tohoto modelu s výsledky MTMM studií.

Obsah

1.	ÚVOD	2
2.	VYMEZENÍ POLE ZÁJMU	2
3.	EVALUACE VÝZKUMNÉHO NÁSTROJE	6
3.1.	ÚVOD	6
3.2.	STANDARDY	7
3.3.	METODY EVALUACE	8
3.3.1.	<i>Hodnocení experty</i>	8
3.3.2.	<i>Skupinové diskuse</i>	9
3.3.3.	<i>Kognitivní rozhovory</i>	10
3.3.4.	<i>Pilotní test</i>	12
3.3.5.	<i>Experimenty</i>	15
3.3.6.	<i>Komplementarita technik</i>	16
4.	STATISTICKÉ PŘÍSTUPY	20
4.1.	KRITÉRIA KVALITY DOTAZU	21
4.2.	DESIGN PRO MĚŘENÍ KVALITY MĚŘENÍ	24
4.2.1.	<i>Quasi Simplex Model</i>	24
4.2.2.	<i>Repeated Multimethod Model (RMM)</i>	24
4.2.3.	<i>Multitrait-multimethod design (MTMM)</i>	25
4.2.4.	<i>Alternativní způsoby realizace MTMM designu</i>	26
4.3.	APLIKACE MTMM DESIGNU V ČR	31
4.3.1.	<i>ESS – Round 1</i>	31
4.3.2.	<i>ESS – Round 2</i>	37
4.3.3.	<i>ESS – Round 3</i>	43
4.3.4.	<i>ESS – Round 3</i>	43
4.4.	NÁSLEDNÁ ANALÝZA VÝSLEDKŮ MTMM EXPERIMENTŮ JAKO NÁSTROJ PRETESTOVÁNÍ	44
4.5.	VYUŽITÍ VÝSLEDKŮ A PRÁCE V SQP – OPTIMALIZACE, KOREKCE MĚŘENÍ.....	46
4.6.	SROVNÁNÍ PREDIKCE SQP S VÝSLEDKY MTMM STUDIÍ	55
4.7.	SROVNÁNÍ PREDIKCÍ SQP PRO ČR.....	55
4.7.1.	<i>Srovnání s výsledky ESS 1</i>	55
4.7.2.	<i>Srovnání s výsledky ESS2</i>	57
4.8.	VYHODNOCENÍ POUŽITELNOSTI SQP.....	60
5.	ZÁVĚR	62
6.	SUMMARY	64
7.	POUŽITÁ LITERATURA	65
8.	SEZNAM PŘÍLOH	70
9.	PŘÍLOHY	71

1. Úvod

Problematikou kvality výběrových šetření se zabývá mnoho metodologických statí. Zájem a zaměření na dílčí problémy (výběrovou chybu, atd.) se jeví čím dál tím více pro celkový výsledek jako nedostatečné a pro problematiku chyb měření se hledal obecnější rámec. Úsilí sociologie, vedeno znalostí této nekompetentnosti, směřovalo práci k nalezení principů, které by překročily parciální pohledy na zdroje možných chyb výzkumného procesu a které by koncept kvality zohlednily v celku práce sociologa.

Trend k celostnímu pojetí kvality dat pak dostal své vyjádření nejen v práci mnoha sociologických metodologů¹, ale byl transformován i do předpisů realizace výzkumné praxe v podobě oborových standardů, standardů profesních organizací nebo norem (ISO). Konkrétní vymezení managementu kvality se v rámci různých institucionálních rámců může samozřejmě mírně odlišovat (Krejčí 2008:22nn), ale důležitost těchto aplikací pro realizaci výzkumných šetření nemá smysl rozporovat.

Záměrem práce je prozkoumat kvalitu nástrojů, kterými sociologická praxe disponuje v oblasti testování dotazovacího instrumentu, tedy jedné z klíčových částí výzkumných šetření, a zabývat se jejich vazbou na kvalitu výzkumných šetření.

2. Vymezení pole zájmu

Pohled kvality výzkumných šetření samozřejmě musí obsahovat i to, co sám chce postihnout a nejlépe i měřit a to chyby. Vymezení pohledu na chyby, které provázejí jednotlivé kroky výzkumu, může být teoretické nebo zaměřené více na samotnou praxi realizace šetření. Nás zde bude zajímat více to druhé.

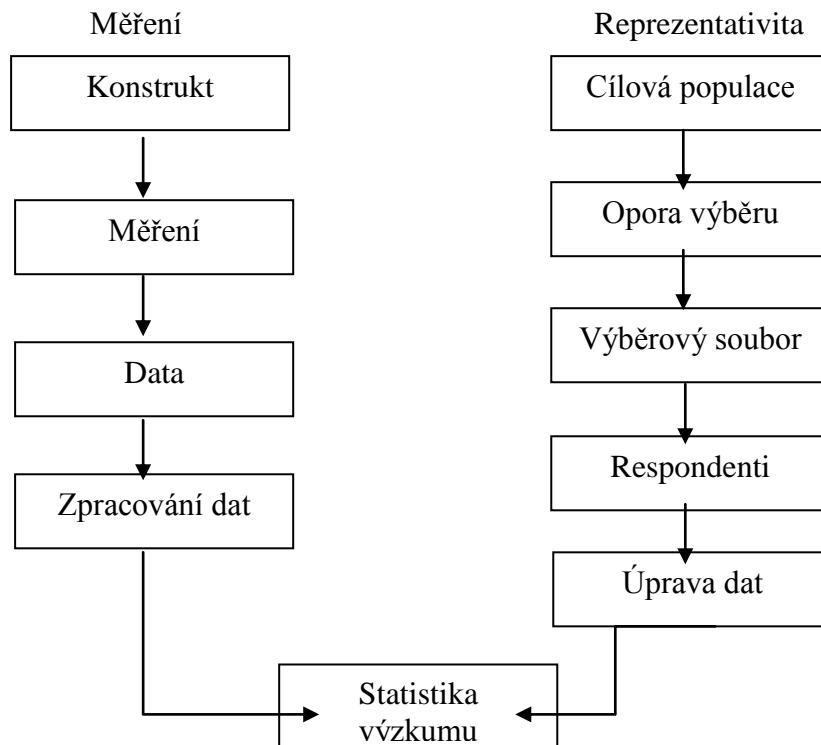
Optika procesu výzkumu skrze koncept kvality směřuje k identifikaci principů a postupů týkajících se všech částí sociologického průzkumu (design, sběr dat, zpracování,...) a jejich vztahu ke kvalitě a nákladům výstupů výzkumného šetření a procesu výzkumu. Skrze koncept celkové výzkumné chyby se pak metodologie snaží zohlednit možné nedostatky všech fází výzkumného procesu, když umožňuje nahlížet na proces výzkumu celostně a poskytuje tak vhodný analytický nástroj.

A co to znamená pro práci při realizaci výzkumných šetření? Součástí práce výzkumníků jsou rozhodnutí o konkrétním nastavení výzkumného procesu (koho a jak jej vybrat, jak dotazovat,...), které mají dopad na kvalitu výstupů dané studie. Záměrem

¹ Např. Biemer a Lyberg (2003:18).

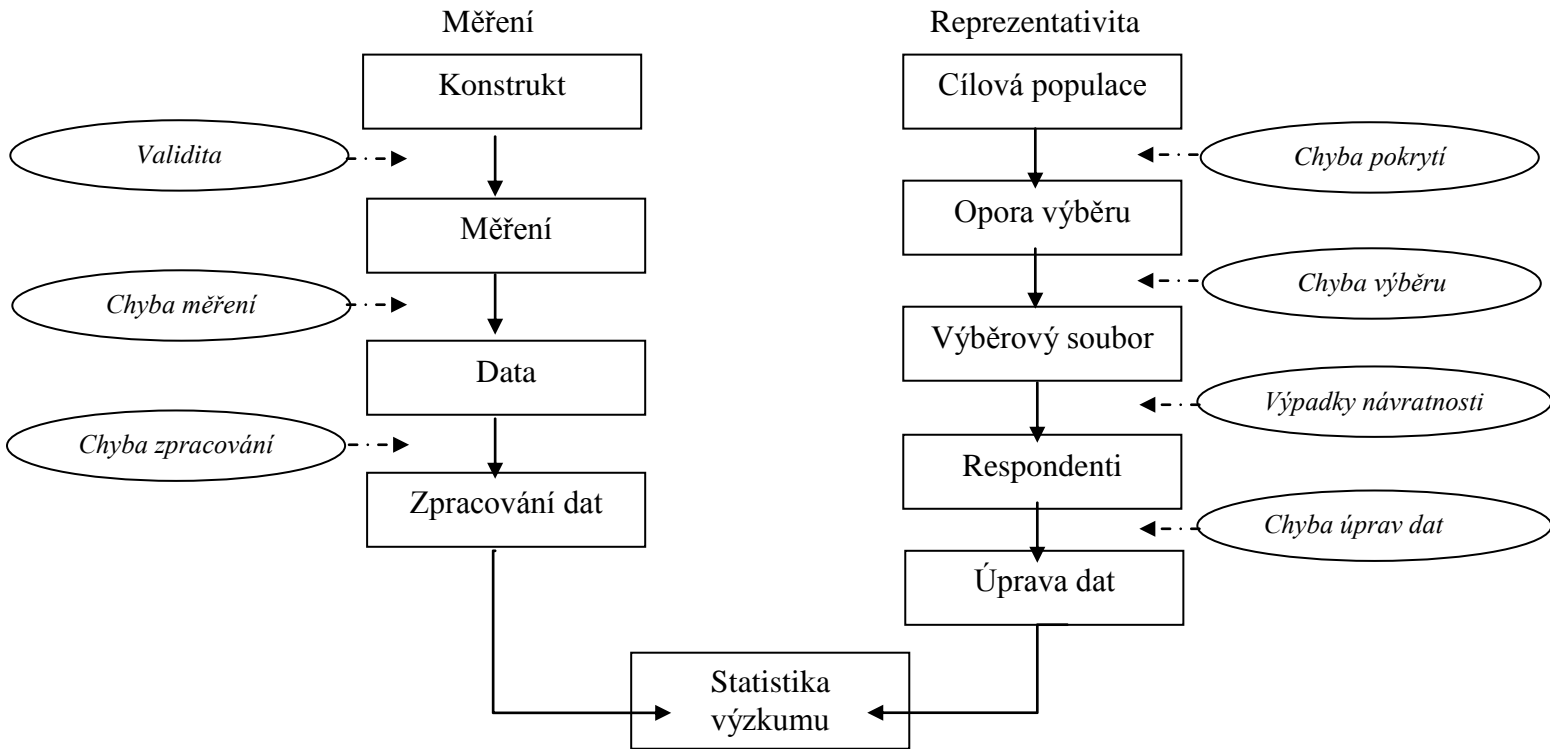
je učinit taková rozhodnutí, které nebudou mít dopad na kvalitu šetření a zároveň budou realizovatelné v alokovaných finančních zdrojích. Tolik ideál. V praxi právě zdroje a nejen ony, často i mnohé další překážky, jsou důvodem ke kompromisním rozhodnutím ohledně nastavení projektu. Schéma celkové výzkumné chyby připomíná, že výzkum je provázaným procesem a vyzdvihuje holistickou optiku. Usazuje tak jednotlivá rozhodnutí do kontextu výzkumu a ukazuje, že zaměření a třeba i alokace zdrojů pouze do jedné nebo omezené části výzkumu může mít vážný dopad na kvalitu výsledků. Výsledek totiž není o nic lepší než jeho nejhorší část.

Z tradiční metodologické literatury jsme však spíše zvyklí na jiný úhel pohledu, a to (chronologickou) perspektivu designu výzkumu a výklad „jak udělat výzkum“. Životní cyklus výzkumného procesu je tak v těchto metodologických pracích a sociologických úvodech popsán z hlediska „jak správně postupovat“ a zpravidla i chronologicky. Nabízí postupy vhodné pro realizaci šetření a dotýká se tak samozřejmě i tématu kvality, ale přinejmenším se nejedná o celostní přístup. Příkladem může být Disman (2005:120) v kapitole „Jak se to vlastně dělá“ nebo Babbie (2004:107) „Jak udělat výzkumný projekt“. Schéma takového návodu může vypadat třeba následovně²:



²Převzato Groves (2004:42).

Stejné schéma však může být použito z perspektivy kvality výzkumu. V oválech jsou přidány koncepty kvality, které jsou spojeny s jednotlivými fázemi výzkumu a upozorňují na chyby, které mohou při uplatnění postupů v dané části výzkumu vzniknout. Pro tento rámec se v literatuře používá označení celková chyba výzkumu.³



Dvě věci jsou zde důležité. A to, že každý z konceptů kvality (znázorněný oválem ve schématu výše) je nejen popsán, ale má i kvantitativní vyjádření. Za druhé, a na to upozorňuje třeba i Krejčí při výkladu konceptu celkové výzkumné chyby, chyby jako takové jsou specifické pro jednotlivé výsledky- statistiky šetření a ne pro šetření jako celek. Pro klasifikaci chyb spojených s procesem výzkumných šetření je využíváno i členění na chyby chybějících pozorování (error of non-observation) a chyby měření (errors of observation), kdy první termín se snaží obsáhnout chyby spojené s chybami pokrytí, výpadky v návratnosti a výběrovou chybou, zatímco druhý pojem v sobě obsahuje chyby týkající se práce tazatelů, respondentů, metod sběru a výzkumného nástroje.

Schéma výše poslouží pro definování dílčího prostoru zájmu práce v kontextu celkové chyby měření. V práci se chci zaměřit na oblast chyb měření, a to na část týkající se

³ Podobné schéma můžeme najít třeba i v pracích Krejčího (2008:26) nebo Vinopala (2008:15).

chyb spojených s používaným dotazovacím nástrojem. Na úvodech do sociologie tisíckrát připomínané GIGO se zarylo do paměti převážné většiny studentů. Nabídka nástrojů a postupů pro diagnostiku těchto chyb, však zdá se mi, se opomíjí. Vyhodnocení pretestů výzkumu vykazuje zajímavý paradox. Na jedné straně, pretestování představuje jediný způsob, jak v předstihu otestovat dotazovací instrument jestli nepůsobí obtíže jak respondentům, tak tazatelům. Zároveň téměř každá nebo každá metodologická práce zabývající se sociologickým výzkumem neopomene zdůraznit důležitost této práce (např. Converse, Presser 1986:51). Na druhou stranu jsou učence výzkumu ohledně tohoto tématu na informace velmi skoupé. Kapitoly s tímto tématem jsou zpravidla stručné, omezují se pouze na dlouhodobě zavedené metody a často nedokáží reflektovat další vývoj. Hlavně a to je důležité neposkytují ani návod ohledně vhodnosti použitých metod, znalosti z praxe a jak využít získané výstupy (myšleno pro zlepšení dotazníku).⁴ Až donedávna se metodologii pretestováním nevěnovalo přílišná pozornost. Obecně se tedy připouštěla neomylnost sociologů a zdůrazňovala nutnost ověřit dotazovací instrument ještě před „ostrým sběrem“. Jen málokdy je odborná veřejnost seznámena s výsledky této praxe, natož aby se výzkumník pustil do kritického zhodnocení použití a vhodnosti (použitých) metod pretestování, včetně toho, do jaké míry splnily zadaný účel a byly využitelné na zlepšení dotazníku.

⁴ Např. Converse a Presser k tomu uvádí „...nejsou žádné obecné principy dobrého testování, žádná systematizace postupů ani konsensus o účelu a jen zřídka jsou data z těchto pilotů poskytnuta dále...“ (Converse, Presser 1986:52)

3. Evaluace výzkumného nástroje

3.1. Úvod

Historie pretestování dotazníků sahá až do 40. let minulého století, kdy se začaly objevovat vědecké články na toto téma. Po dlouhou dobu byla v testování instrumentů výběrových šetření dominantní tradiční (někdy také konvenční) metoda spočívající na tom, že před samotnou realizací výzkumu byl dotazník poskytnut několika tazatelům, kteří na jeho základě provedli většinou omezený počet rozhovorů⁵. Úsilí o replikaci reálných podmínek výzkumu může mít více podob a úrovní simulace realizací a tak tazatelé mohou být na tazatelskou práci s dotazníkem vyškolení, můžeme také oslovit i respondenty, kteří jsou naší cílovou skupinou i ve skutečném výzkumu atd. Terénní sběr byl doplněn o diskusi s tazateli, kdy se spolu s výzkumníky zabývali (z jejich úhlu pohledu) možnými problémy, které dotazník představuje nebo základní tabulaci dat (rozložení proměnných, chybějící hodnoty,...). Zanedlouho potom se na základě větší či menší generalizace takovýchto zjištění začaly objevovat návody, jak udělat správně dotazník. Zajímavé také je, že v mnoha úvodech těchto praktických průvodců bychom narazili na to, že tvorba otázek a dotazníku jako celku má co do činění s uměním.

Podle Pressera a dalších (Presser 2004) vědci vkládají to tohoto přístupu nevšední důvěru. Setkáme se tak s doporučeními jako jsou: „...20-50 dotazníků je obvykle dostatečný počet pro odhalení hlavních chyb dotazníku...“ (Sudman 1983:181), Converse a Presser mluví o dostatečném počtu 25-75 dotazníků, i když připouští i kouzelné pravidlo udělej kolik je možné (Converse & Presser, 1986), Fowler mluví o realizaci 15-35 rozhovorů (Fowler 1995:115). Tato víra samozřejmě měla opodstatnění a oporu ve zkušenostech sociologů, když jim i tyto testování malého rozsahu ukázaly problémy týkající se špatného znění otázky nebo nedostatečné nabídky variant odpovědí. Vědecká podpora a zdůvodnění těchto postupů však pokulhávala. I tak však dlouhodobě tato praxe produkovala výsledky a ukazovala se jako důležitou součástí výzkumného procesu.

Přicházelo se na to, že tento způsob testování ze své podstaty není schopný identifikovat některé druhy problémů. Existují přece i problémy, které nemusí být znatelné z vizuálního projevu respondenta, a stejně tak respondenti si tyto problémy ani nemusí

⁵ Popis praxe debriefingu např. Converse a Presser (1986:51nn) nebo Fowler (1995:104-155)

uvědomovat (problémy porozumění a další). V poslední době narůstající diskuse o nedostacích tradičního přístupu a možnostech zlepšení v této oblasti sociologické práce vedla podle Pressera k dvěma vzájemně provázaným změnám. První je posun zájmu pouze od identifikace chyb dotazníku k širšímu pojetí kvality měření, druhý k etablování dalších metod testování. Předmětem zájmu se pak stává i hodnocení vhodnosti jednotlivých metod pro práci s různými typy chyb a samozřejmě také jejich komplementarita včetně komplementarity jejich zjištění.

3.2. Standardy

Na proces evaluace otázek můžeme pohlížet ze dvou, vzájemně někdy více jindy zase méně, komplementárních perspektiv.⁶ První je charakteristická snahou o redukci chyb měření a zabývá se otázkami, jako je pochopení otázek a použitých pojmů, znění otázek, zátěž respondenta, problémy získání požadované informace z paměti, procesem vytvoření odpovědi atd. Pro druhou je typické snaha změřit chybu měření, tj. zabývá se tím, do jaké míry získané odpovědi korespondují s tím, co chceme skutečně měřit, tzn. odhadem chyby měření. Pro tento účel se nabízejí dvě možnosti - porovnat získané výsledky s jinými datovými zdroji anebo zopakovat měření. Groves k tomu uvádí „srovnáním s externími datovými zdroji můžeme měřit validitu nebo odchylku měření, opakováním měření reliabilitu nebo rozptyl měření“. (Groves 2004:241) Ukážeme si, že existují přístupy a designy výzkumu, které umí adresovat oba typy kritérií. Navíc podle mého názoru díky jejich další analýze došlo k propojení obou komponent testování.

Ještě předtím než v krátkosti projdeme několik nejobvyklejších způsobů pilotování dotazníku a upozorníme na jejich slabá místa, podívejme se na standardy (za standardy svého druhu můžeme také považovat všechny předpisy a návody pro konstrukci dotazníku), které by podle Grovese měly být u otázek výzkumných šetření naplněny. Jedná se o 1) standard obsahu (tj. zdali se ptáme, na co máme), 2) kognitivní standard (rozumí respondenti otázce a jsou schopni na ni odpovědět) a 3) standard použitelnosti (jsou respondenti i tazatelé, pokud jsou zapojeni, schopni bez problémů schopni vyplnit dotazník, tak jak zamýšlel jeho tvůrce). Jednotlivé metody pretestování jsou relevantní jen pro některé z těchto standardů a jiné otázky zase na druhou stranu řešit nedokážou. O tom, jaké metody jsou vhodné pro jednotlivé standardy, odkážu čtenáře na Grovese (Groves 2004). Úmyslem zde bylo představit další způsob (a ještě širší pojetí)

⁶ Groves rozčlenil přístupy k analýze chyb měření na reducer a measurer (Groves et. al 1989:353).

uvažování o tom, co má pretestování naplňovat, na jaké otázky přinášet odpovědi a hlavně zamyslet se nad vzájemným vztahem těchto postupů.

3.3. Metody evaluace

V následující části se budu zabývat jednotlivými přístupy, které jsou v sociologii uplatňovány pro pilotní testování výběrových šetření. Úmyslem je ukázat na slabá místa jednotlivých přístupů a nabídnout alternativu, která se opírá o práci Sarise, Scherperzeel a Gallhofer. Tento autorský tým nabídl možnost spojit obě perspektivy a přístupy k testování do jedné a jak tyto výsledky a analýzy, které přicházejí po realizaci terénního sběru udělat nástrojem pretestování.

3.3.1. Hodnocení experty

Ne ve všech metodologických textech je, stejně tak to platí i pro následující přístup, uvedeno hodnocení expertní skupinou. Pro tento přístup platí, že dotazník je pro kritické zkoumání předložen skupině odborníků buď na danou problematiku, která je předmětem výzkumu anebo odborníkům na konstrukci dotazníku a dotazování. Doufám, že tento typ není eufemismem pro netestování stejně, jako to Disman uvádí v případě testování zjevné validity. Úkolem takovéto revize je prozkoumat, zda dotazník obsahuje všechny z analytického pohledu důležité indikátory, tzn., zda od respondentů získáme skutečně takové informace, které potřebujeme pro následnou analytickou práci. Může se také týkat znění otázek, zhodnocení variant odpovědí, pořadí otázek, instrukcí atd. Máme sice k dispozici mnoho návodů a principů, jak vytvořit otázky našeho dotazníku správně (např. Sudman a Bradburn 1982, Fowler 1995 nebo Converse & Presser 1986) slabým místem ovšem zůstává samotné hodnocení. Rozhodnutí o tom, zda konkrétní jednotlivá otázka splňuje ten nebo onen standard (např. je otázka jednoznačná?) je subjektivní záležitost a hodnocení odborníků se tak mohou rozcházet. Co je z jednoho pohledu otázkou bezvadnou, v jiném tomu tak být nemusí. Z toho důvodu vznikly kontrolní seznamy, které mají být návodem pro praktické provedení tohoto typu evaluace. Těchto pokusů o systemizaci, které se snaží přijít s detailním, explicitním a standardizovaným způsobem této evaluace je hned několik. Nejzajímavější je přístup Graessera a jeho kolegů, kteří šli ještě o kousek dál. Nejen že nabídli seznam problematických míst, na které by se hodnocení dotazníku mělo zaměřit⁷, ale vytvořili i počítačový program,

⁷ Podle Graessera a kolegů je třeba se při hodnocení dotazníku zabývat těmito problémy: 1) přílišná komplexnost nebo složitý syntax otázky, 2) přílišná zátěž na paměť respondenta, 3) vágní nebo

který dokáže tyto kategorie aplikovat na zadaný návrh dotazníku a získat tak automatizovaný hrubý odhad odborníků.⁸ Další systematické pojetí evaluace směřující k vyšší standardizaci hodnocení vypracovali např. Lessler a Forsyth.⁹

Ani tato zajímavá aplikace programu QUEST a možné návody ke standardizaci pole hodnocení, však nedokáže vyvrátit základní pochybnosti, jako jsou subjektivita hodnotitelů, tzn. konzistentnost hodnocení a odlišnost uplatněných kritérií jednotlivých hodnotitelů. Kromě několika dostupných studií (pokud je mi známo) jako je ta z dílny Pressera a Blaira (1994) máme jen málo metodologických textů, který by se vůbec zabývaly mírou shody těchto způsobů hodnocení (viz text dále týkající se srovnání aplikací více metod).

3.3.2. Skupinové diskuse

Dalším užívaným postupem jsou skupinové diskuse. Před nebo i v průběhu konstrukce dotazníku jsou uspořádány diskusní skupiny, do kterých jsou zváni k účasti zástupci cílové skupiny. Záleží jen na pečlivé přípravě diskusních témat a struktury diskuse výzkumníky, aby vymezili pole, k čemu chtějí získat zpětnou vazbu. Skupinové diskuse mohou výzkumníkům pomoci získat znalosti o pohledu respondentů na tematiku, jíž se chtějí věnovat, používaných pojmech, perspektivě nebo znalostech. Skupinová dynamika umožní získat různost názorů na danou problematiku, na argumentaci a sporné body. Pokud je výzkumné téma širší a pokrývá více cílových skupin doporučuje se uspořádat se zástupci každé cílové skupiny diskusi separátně. Vyhneme se tomu, že místo aby naše znalosti byly prohloubeny dojde díky nesourodosti skupin k jejich rozředění. Groves (2004:245) uvádí tři důvody, kdy a proč se vyplatí využít tuto metodu: 1) v případě, že potřebujeme získat míru znalostí nebo obeznámenosti respondentů s tématem výzkumu, včetně toho, jak je tato znalost strukturovaná, 2) skupinová diskuse je také vhodným způsobem, jak zjistit, jaké pojmy respondenti v diskusi o tématu užívají a jak jim rozumějí. Výzkumník díky tomu získá znalost o tom, jaké pojmy v otázkách použít tak, aby jim respondenti rozuměli a pokud možno byly chápány konstantně. Cílem skupinových diskusí je také zjistit (3), co může

nejednoznačná slovesa, 4) použití technických termínů, 5) vágní přísudek nebo vztažný termín, 6) zavádějící nebo chybné předpoklady, 7) nejasné odpovědní kategorie, 8) sloučení/překryv více odpovědních kategorií, 9) nesoulad mezi otázkou a nabídkou odpovědí, 10) obtížnost zhodnotit (vyvolat z paměti) požadovanou informaci, 11) nemožnost, že respondent znát odpověď a 12) nejasný záměr otázky.

⁸ Popis fungování modelu QUEST viz Greasser et al 1996:143 nn., identifikované chyby programu odpovídají klasifikaci v pozn. č.5.

respondent o výzkumném problému říci. Po položení otázky dotazníku jsou respondenti požádáni, aby popsali, co znají. Jestli se otázka dotýká jejich subjektivních stavů (pocity, názory) nebo objektivních záležitostí a zkušeností a výzkumník má možnost zjistit kontext a realie problému, na který se ptá.

K tomu je ovšem třeba přidat i omezení této metody. Účastníci skupinové diskuse nemusí nutně reprezentovat cílovou populaci, a tak zobecnění vnímání nebo znalostí je problematické. Z charakteru poznatků, které pořádání skupinové diskuse přináší, které jsou kvalitativní povahy plyne, že tyto výsledky mohou být nerealističtější, obtížně aplikovatelné a poplatné pouze respondentům, kteří danou diskusi navštívili.

3.3.3. Kognitivní rozhovory

Počátek těchto přístupů odborná literatura datuje k semináři US National Research Council v roce 1983. Výsledkem společného setkání kognitivních psychologů a sociálních metodologů bylo to, že sociální vědci se začali zabývat myšlenkou a použitím psychologických přístupů. Účelem bylo porozumět tomu, jak respondenti položeným otázkám rozumí a jak dochází ke konstrukci jejich odpovědi. Zájem o tuto oblast ještě posílil dojem ze zjištění Schumana a Pressera (1981) a Belsona (1981), která se zabývala špatným pochopením otázek.

S oporou v protokolární analýze vznikla skupina technik, které bychom mohli označit jako kognitivní rozhovory. Vinopal k tomu uvádí, že sám pojem kognitivní rozhovor nemá jednoznačnou definici a dodává, že „v současné době jde spíše o souhrnnou nálepku označující spíše určité zaměření (na kognitivní procesy), nežli konkrétní metodický postup“ (Vinopal 2008:51). Jak můžeme z vymezení kognitivních přístupů k testování kvality dotazníků usuzovat, není ustálen jediný přístup k tomu, jak kognitivní pretestování provádět.¹⁰ Cílem těchto metod je pochopení procesů, kterými respondent dospívá k odpovědi na položenou otázku včetně odhalení potenciálních zdrojů chyb (od porozumění až po vytváření odpovědi). Kognitivní rozhovor se od běžného rozhovoru liší tím, že používá techniky myšlení nahlas a slovního prověřování. Z uvedeného je jasné, že kognitivní rozhovory jsou participující technikou a respondent

⁹ Jelikož jejich metoda hodnocení má přesah i do kognitivní evaluace bude zmíněna v další kapitole. Lessler a Forsyth (1996, 259-292).

¹⁰ Taxonomii kognitivních přístupů shrnuje např. Forsyth a Lessler (1991:397). Rozdělují je na expertní hodnocení (kódování chování, hodnocení experty), rozšířené dotazování (myšlení nahlas, slovní prověřování), cílené metody (jako parafrázování, viněty) nebo skupinové metody (skupinové diskuse, dotazování nebo experimenty).

je obeznámen s účelem těchto rozhovorů. Tím ale paleta možností realizace kognitivních rozhovorů nekončí. Vzniká tak situace, kdy rozdílné organizace i jednotliví vědci užívají k tomuto účelu rozdílné techniky (myšlení nahlas nebo slovní prověřování), v rozdílné okamžiky dotazování (průběžná nebo retrospektivní forma) a rozdílný systém záznamu a analýzy. Po etablování většího počtu „kognitivních laboratoří“ došlo k pokusům o standardizaci a kodifikaci této praxe (Willis 1994, DeMaio 1993), ale ani toto nepřispělo ke shodě a dohodě vědecké komunity o tom, které nastavení kognitivního rozhovoru používat, ani jakým způsobem analyzovat získaná zjištění. Přes nepopíratelné úspěchy, které do znalostní báze tyto metody přinesly, tak zůstává praxe kognitivních rozhovorů roztříštěná. Vystávají však i další otázky, které se týkají zejména reliability zjištění konkrétních typů metod, kompatibility zjištění a přínosu těchto metod pro proces zlepšení dotazníku. Ty však zůstávají povětšinou nezodpovězené, když empirické zhodnocení přínosu těchto přístupů bylo provedeno jen v minimálním rozsahu a to nijak přesvědčivě. Uvedené obavy podporují i metodologické práce. Beatty (Beatty 2004) potvrdil vliv metody a tazatele na podobu zjištění, které kognitivní rozhovor produkuje stejně jako, že v určitém nastavení rozhovoru, kognitivní rozhovor dokáže konstruovat problémy dotazníku, které mohou být pouze problémy domnělými. V konverzačním nastavení kognitivního rozhovoru jsou respondenti méně tazateli tlačeni do využití definovaných kategorií. Vzniká tak situace, že místo odpovědi je problém prodiskutován a k využití nabízených možností odpovědi nedojde ne proto, že by na otázku nešlo odpovědět, ale proto, že respondent neodpověděl. Zatímco Beatty se zabývá oblastí vlivu chování tazatelů, včetně vlivu na výsledky testování, jiné studie se zabývají srovnáním a metodologickým zhodnocením výsledků a doporučení, které rozdílné metody kognitivních přístupů přinášejí. Conrad a Blair (2004) došli k závěru, že reliabilita výsledků (identifikace problémů) je poměrně nízká, podobný problém řešila DeMaio a Landreth. (2004) Nezávislé zhodnocení dotazníku třemi agenturami poskytlo v mnoha ohledech velmi rozdílné závěry. Co se týče počtu identifikovaných problematických otázek, nerozcházely se výsledky zásadně. Mnohem větší problém představovalo, že jen do malé míry jednotlivé testovací týmy identifikovaly u stejných otázek stejné problémy.¹¹ Ty však zřejmě nelze přiřknout pouze rozdílností v nastavení a provedení kognitivního rozhovoru, ve kterém panují velké odlišnosti (autoři srovnávají ve více než 11 dimenzích týkajících se kvalifikace

¹¹ Pro podrobnosti viz DeMaio, Landreth 2004:101nn.

tazatele až po způsob zpracování odpovědi). Nálezy jednotlivých testovacích skupin však byly jen v nevelkém překryvu. Značné rozdíly ovšem panovaly v tom, jaké otázky jsou podle jednotlivých pracovních týmů problematické a stejný problém se týkal i doporučení k nápravě a zlepšení dotazníku. Co se výstupu testování revidovaných dotazníků týče tak testování revidovaných dotazníků nepotvrdilo, že by kognitivní testování vedlo k jejich zlepšení (z hlediska počtu defektních otázek nebo problémů) a pouze jedna ze tří revidovaných verzí se z hlediska výskytu problematických míst odlišovala od dotazníku původního.

Tato zjištění ve svých závěrech potvrzuje i Vinopal, když mezi nedostatky kognitivních metod uvádí nízkou realibilitu jejich výsledků, kdy „...různí kodéři se ne zcela dostatečně shodují na identifikovaných místech a povaze problémů...“ (Vinopal 2008:53).¹² K tomu přispívá nízká standardizace techniky spojená s velmi rozdílnou úrovní individuálních tazatelů. Hlavním problémem je zde tedy validita, reliabilita zjištění a nejasná efektivita. Identifikované nedostatky nejsou příliš realibilní, neexistuje heuristika k rozlišení skutečných problémů od konstruktů, které nemají žádný nebo pouze zanedbatelný vliv (respondenti jsou přece vyzváni k úmyslnému hledání problematických sekcí) a opětovně, jako u dřívějších metod charakter a rozsah testování problematizuje jejich reprezentativita.

3.3.4. Pilotní test

Jde o tradiční způsob pilotování dotazníků spočívající v organizaci zkoušky terénního sběru dat na velmi omezeném počtu zástupců cílové skupiny ještě před samotným hlavním termínem sběru dat. Smyslem pretestu je nejen kontrola kvality dotazníku, ale také organizace sběru dat a výběru respondentů. Converse a Presser doporučují v pilotním testování uplatnit následující postupy, které mají pomoci získat co nejhodnotnější výsledky. Jedná se o: 1) poznámky ke struktuře dotazníku, 2) následný rozhovor s tazateli (debriefing), 3) písemné poznámky k jednotlivým sekcím, 4) poznámky k problematickým aspektům dotazníku, 5) terénní pozorování sběru dat, průběhu rozhovorů, 6) kódování výsledků a zobrazení rozložení proměnných (Converse a Presser, 1986:70nn). Z nabízených postupů je zřejmě nejvíce využíván debriefing tazatelů, který může být jak individuální tak skupinový a zhodnocení statistických

¹² Systematické pojetí evaluace vypracovali např. Lessler a Forsyth. V jejich přístupu je detailně vypracován systém kritérií s kterými při evaluaci otázek kodéři pracují a které jsou později základem pro další pro statistické zpracování. Nejedná se tedy o automatizovaný přístup jako v případě QUESTu, ale i tak se jedná o silný podnět ke standardizaci vyhodnocení. (Lessler a Forsyth 1996).

výsledků. Zpětná vazba od tazatelů se nemusí omezovat jen na konstatování, v které části dotazování zaznamenali nějaké obtíže, ale výzkumník může tazatele vyzvat i k tvorbě jejich vlastních návrhů, jak by bylo možné tato krizová místa vylepšit. Druhý nejčastěji užívaný postup se týká statistického zpracování dat z pilotního sběru dat. Výzkumník zde jako indikátory možných problémů může vnímat vysokou míru nezodpovězení otázek nebo nekonzistenci odpovědí s jinými otázkami. Také u položek, které mají pouze nízkou variabilitu (většina respondentů v souboru odpověděla stejným způsobem) může výzkumník tuto informaci vnímat jako podnět k reformulaci otázky, když formulace původní není schopna dostatečně diferencovat.

Tento způsob testování stojí na předpokladech, že existence možných problémů a chyb v dotazníku bude signalizována buď odpověďmi respondentů (např. varianta nevím, nechci odpovědět) anebo, v návaznosti na položené otázky, nějakým jiným způsobem (projevy) ovšem vždy takovým, aby byly potenciálně zaznamenatelné tazatelem, protože jedině tak je o nich schopný případně podat zprávu. I tak jsou do schopností tazatelů vkládána nemalá očekávání, když např. Sudman a Bradburn uvádějí, že „...důkladný pilot provedený sensitivními tazateli je nejpřímější cestou identifikace chyb...“ (Sudman, Bradburn 1982:49).

Jiná skupina vědců má s touto důvěrou problémy, když nejčastěji upozorňují, že výzkumníci nejen, že nemají žádnou kontrolu nad identifikací problémů tazatelů, ale ani pravidla pro to, jak s nimi zacházet. Jak víme, ne všechny problémy dotazování musí být zaznamenatelné tazatelem. Na tuto oblast problémů se snažily odpovědět např. kognitivní přístupy, ale ještě jedno je třeba si uvědomit. A to, že už samotné nastavení způsobu pilotního průzkumu omezovalo vlastní heuristické možnosti. Ze dvou způsobů provedení pilotního průzkumu vědci volí častěji nedeklarovaný způsob oproti participujícímu způsobu, kdy se respondentům sdělí účel rozhovorů (Converse a Presser specifikují undeclared a participating pilotního průzkumu, 1986). Vědci se tak vlastním rozhodnutím připravují o zjištění a úhel pohledu respondentů o tom, které části dotazníku nebo jednotlivé otázky jim činily potíže, čemu nerozuměli nebo jak některé pojmy interpretovali. Můžeme tedy snadno z nastavení praxe dojít k závěru, že praxe pilotního testování v deklarovaném způsobu realizace sloužila, a je také vhodná více, k identifikaci problémů, které dotazník představuje spíše pro tazatele než pro respondenty.

Další oblast nejasností otevírá, jak dále postupovat s nálezy tazatelů, když pomineme to, že názory tazatelů (zpravidla několik jedinců) jsou silně subjektivní a nereprezentativní.

Sociologie totiž nedisponuje žádnými pravidly ani návody na to, jak tyto zjištění zhodnotit. Mám tím na mysli, jak rozhodnout a odlišit problémy, které jsou závažné a vyžadují naši pozornost od těch, které si můžeme dovolit ignorovat nebo způsob, jak provést vylepšení identifikovaných problémů. Zdá se, že vědcům nezbývá v těchto případech spoléhat na nic jiného než na své zkušenosti a intuici. Jako odpověď na tuto kritiku můžeme vnímat pokusy o zapojení nových technik nebo technik z jiných vědeckých disciplín, které tradiční způsob testování doplňují. Zvukový záznam rozhovoru tazatele a respondenta, tzn. přidání systematické observace systému kladení otázek a jejich odpovídání, může poskytnout další informace o kvalitě dotazníku. Tato metoda, nazývaná kódování chování, se zabývá systematickým zhodnocením interakce tazatel-respondent a pomocí systému kódů chování analyzuje jednotlivé sekvence otázka-odpověď dotazníku.¹³ Propojení klasického pilotního průzkumu s kódováním chování nevyžaduje nic jiného než pořízení záznamu rozhovoru, který bude dále využit pro kódování definovaných aspektů chování. Výzkumník díky tomu získá data a statistické vyhodnocení o podílu tazatelů a respondentů, kterým činila daná otázka problémy. Právě v tomto spočívá specifická přidaná hodnota nad rámec běžného způsobu zhodnocení dotazníku při standardním pilotním testování. Výsledky takovýchto zhodnocení jsou systematické, objektivní a replikovatelné. (Groves 2004:249) Výzkum ukázal, že použití této metody je opravdu intersubjektivní, když dva týmy při analýze identického dotazníku dospěli k podobným statistikám chování. Potvrdilo se tak, že tato metoda může produkovat konzistentní a reliabilní výsledky (míry projevů jednotlivých druhů chování) bez ohledu na tazatele.¹⁴ Náročnost zpracování záznamů rozhovorů u této metody pravděpodobně není jedinou bariérou jejího širšího uplatnění a využití v praxi.¹⁵ Další otázky se týkají toho, že v průběhu interakce otázka-odpověď je možné kódovat rozličné aspekty chování. Zájem výzkumníků se může zaměřit na ty nebo ony projevy, důležité však je, a mělo by se to stát předmětem dalšího metodologického výzkumu, identifikovat ty prvky chování, které mohou být reliabilně kódovány a které mají dopad na kvalitu výzkumných dat.

¹³ Ukázka systému kódů např. Zouwen a Smit (2004:112) nebo Fowler a Cannell (1996:29). Jejich širší používání a přijetí by pomohlo k další standardizaci metody a případnému širšímu využití.

¹⁴ Např. Fowler a Cannell 1996:15-36.

¹⁵ V některých případech nastává situace, že problém je sice zjevný, ale identifikace příčiny problému nemusí být zjevná. Systém heuristiky krom statistického zpracování nebývá popsán. Na druhou stranu je třeba zmínit, že existují i snahy o automatické systémy kodování za pomoci PC, které pracují s přepisem rozhovorů a mají dopomoci širšímu uplatnění metody Bolton a Bronhorst (1996).

(Groves 2004:248) Limitou tohoto přístupu je také absence jakýchkoliv doporučení a návrhů k úpravě zjištěných problémů.

3.3.5. Experimenty

Experimentální přístup je, spíše než samostatným přístupem, příkladem specifického užití předchozích přístupů. Design experimentu je vždy postaven tak, aby vůči experimentální skupině proběhlo testování i na skupině kontrolní, vůči níž pak proběhne porovnání. Tento experimentální design se může týkat srovnání rozdílných metod sběru dat, rozdílných postupů nebo znění otázek. Experiment může být koncipován jako součást pilotního terénního testování. Stejně tak jsou ale i případy, kdy se jedná o samostatné metodologické studie.

Výhodou tohoto designu je kontrola podmínek testu a díky účelné manipulaci s experimentální proměnnou i možnost získat poznatky, které jsou za rovinnou informací, které poskytne respondent. Provedení experimentů mohou být velmi odlišné a pro jejich konstrukci může být použita široká škála technik.

Zastánci tohoto přístupu mají pak silnou oporu vůči ostatním metodám testování, kterým vytykají používání subjektivních přístupů, bez jakékoliv opory v testování reliability nebo validity. Upozorňují na nereprezentativní design, na studie na malých vzorcích respondentů a další aspekty, které neumožňují dostatečnou kontrolu intervenujících proměnných nebo robustní vyhodnocení. Dalším metodám nechtějí upírat právo na existenci, ale považují je spíše za metody, jak získat vědecké hypotézy. Ty by však měly být zhodnoceny objektivní vědeckou cestou, která proces tvorby dotazníku vrátí na vědeckou půdu.

K tradičním přístupům je také velmi kritický Fowler (Fowler 2004). Podle něj tradiční způsoby testování neinformují o způsobu, jak je ovlivněna kvalita dat ani o tom, zda případná úprava dotazovacího instrumentu měla požadovaný účinek a vedla ke skutečnému zlepšení. Pro zodpovězení těchto otázek je podle něj nutné nasadit experimentální design. Uplatnění tohoto designu se však nemusí omezovat jen na testování alternativních znění otázek, jak je známo z práce Schuman a Pressera (1981), ale mělo by se podle Fowlera stát standardní součástí procesu testování.

Stejná kritika je ovšem poplatná i na část konaných experimentů. Experimenty poskytují jasné důkazy o vlivu znění otázek, pořadí otázek, použité metody sběru atd. na odpovědi respondentů. Často však bohužel, ačkoli jasně ukazují, že rozdílné metodologické nastavení (instrument nebo i postup) produkuje rozdílné odpovědi, mnoho experimentů

není schopno odpovědět na otázku, která verze poskytuje lepší data. Nicméně experimenty možnost vyhodnotit dopad alternativních návrhů znění otázek na výsledná data, které další testovací přístupy neposkytují, nabízí. Jde o to tuto příležitost využít. Vhodným inspiračním materiálem týkajícím se nastavení experimentu může být text Tourangeau (Tourangeau 2004).

3.3.6. Komplementarita technik

Rozrůstání přístupů a metod pretestování přineslo i otázky týkající se jejich vzájemného vztahu. V posledních 20 letech se začaly objevovat metodologické studie, které se zabývaly právě otázkami, jako jsou: Jsou rozdílné metody vhodné k rozdílným účelům, doplňují se? Jsou jednotlivé techniky pro identifikaci problémů stejně užitečné? Jsou jejich nálezy srovnatelné? A v případě, že jsou jejich zjištění srovnatelná, která metoda je nejefektivnější (ekonomičtější)? Takové a podobné otázky vedly, jak ke srovnání funkčnosti souboru metod vzájemně mezi sebou, tak i výsledků realizací v rámci konkrétní metody. Na základě praktických zkušeností se ukazovalo, že různé metody mohou sloužit k různým cílům a přinášet rozdílné poznatky. Mezi prvními, kdo se touto otázkou zabývali, byli Presser a Blair. (Presser a Blair 1994) Tito autoři se rozhodli prozkoumat vztah a výsledky pilotního testu, kódování chování, kognitivních rozhovorů a hodnocení experty. Opakované zapojení metod ukázalo, že nejvíce reliabilní výsledky přináší metoda kódování chování. Příčinou tohoto může jistě být objektivita kritérií, které v kontextu ostatních metod tento přístup aplikuje. Nejmenší shoda naopak panovala mezi výsledky aplikace dvou pilotních testů. Z výsledků se ukázalo, že nejvíce produktivní (z hlediska odhalení chyb dotazníku) byl panel expertů. Z hlediska komplementarity metod¹⁶ se ukázalo, že kódování chování a pilotní test byly jediné metody, které byly senzitivní k zachycení problémů tazatelů, naopak kognitivní rozhovory a panel expertů diagnostikovaly sémantické problémy. Díky zapojení výzkumníků kognitivní rozhovory a panely expertů vedly i k odhalení možných analytických problémů. Rozsah diagnostikovaných problémů (nejvíce v tomto ohledu je omezená aplikace kódování chování, která se zabývá pouze chováním respondentů a problémy tazatelů) spolu s dobrými výsledky paralelních aplikací (shoda hodnocení mezi dvěma oslovenými panely) a cenou vede autory k doporučení rutinního testování dotazníků panely expertů.

¹⁶ Autoři definují následující typy problémů: sémantický (respondent), účel (respondent), chování (respondent), tazatel a analytický problém (Presser a Blair 1994:87).

Na přístup Pressera a Blaira navázal Willis (1999), který se pokusil zjistit, zda aplikované metody pretestování, zde výsledky expertního vyhodnocení, kognitivních rozhovorů a kódování chování, přináší srovnatelná zjištění. V souladu s předchozími nálezy i zde expertní hodnocení identifikovalo nejvíce problémů dotazníku, kognitivní rozhovory naopak nejméně (jednotlivé realizace metody kognitivního testování však přinesly srovnatelné výsledky). Analytická hodnota tohoto zjištění je však omezená, protože to, že jedna metoda identifikuje více problémů ještě neznamená, že je metodou lepší. Nemusí zde platit „více je lépe“ a hodnocení metod pouze na základě kritéria senzitivity bez přihlédnutí k specifitě metod může vést k tomu, že nejlepší metoda je ta, která produkuje nejvíce falešně negativních výsledků (viz dále např. nastavení QAS). Vyhodnotit je třeba nejen počet problémů, ale i míru shody mezi metodami navzájem. Korelační analýzou Willis dochází k tomu, že jednotlivé realizace stejných metod produkují velmi srovnatelné výsledky, ale ve srovnání metod navzájem nedochází k velkému překryvu. I to vnímá Willis jako pozitivní výsledek, který je dán odlišnými zaměřeními jednotlivých metod, které musí produkovat rozdílné výsledky. I negativní výsledek tak může být pozitivní a dokladem nutnosti použít více metod pretestování, aby výzkumník dostal plastičtější a komplexnější pohled.

Otázkou srovnatelnosti výsledků se zabýval i další kolektiv autorů (Rothgeb, Willis, Forsyth:2001). Zde došlo ke srovnání expertního hodnocení, formálního hodnocení pomocí systému QAS (Questionnaire Appraisal System) a kognitivních rozhovorů. Z hlediska počtu problémů expertní hodnocení a kognitivní rozhovory docházely k podobným výsledkům. Výsledky QAS byly „nadproduktivní“. Tato metoda díky nastavení prahu co je problematické (alespoň 1 z 26 kritérií nevyhovuje) vedla k velké nadprodukcii problémových situací a falešně pozitivním výsledkům. U těchto systémů, stejně jako třeba u kódování chování, je problematické nastavit práh testování tak, aby produkoval správné výsledky (u kódování chování často 15-20 %). Celkově konzistentní výsledky v tom, jak často byly otázky (ve srovnání mezi metodami) označeny za problematické, kontrastovalo s pouze nevelkou shodou nad tím, o které položky se jedná. Kvalitativní analýza pak ukázala, že nálezy metod jsou odlišné i z hlediska typu problému. Kognitivní rozhovory neidentifikují problémy tazatele, ale jsou velmi senzitivní k obsahu otázky. Expertní rozhovory vynikají v identifikaci struktury otázek nebo odpovědí. Na rozdíl od předchozích zjištění Pressera a Blaira nebylo hodnocení experty nejvíce produktivní v identifikaci chyb, ale dosahovalo srovnatelných výsledků jako kognitivní rozhovory.

Následující dvě metodologické studie si kladly i další cíle. Jejich autorům nešlo pouze o identifikaci komplementarit a rozdílností v rámci metod pretestování, ale zabývali se i vyhodnocením použitelnosti a relevance těchto výsledků pro další práci.

První práce od stejného kolektivu autorů (Forsyth Rothgeb, Willis:2004) si kladla otázky, zda metody pretestování identifikují problémy, které při terénním sběru dat nastanou a zda na základě pretestování upravený dotazník povede k lepším výsledkům. Byla tedy položena kruciólní otázka, zda činnost pretestování dotazníků vede ke skutečným výsledkům. V první fázi bylo provedeno pretestování dotazníku, v druhé fázi byl díky aplikaci split half experimentu v terénu prověřen jak dotazník v jeho původní podobě tak dotazník revidovaný, který vznikl na základě znalostí a výsledků pretestování. Ukázalo se, že techniky pretestování konzistentně indikovaly problémy, které se objevily i při následném terénním sběru (položky, které způsobovaly ve fázi pretestování problémy tazatelům a respondentům vykazovaly i vyšší míru problematičnosti při dotazování, které bylo analyzováno kódováním chování, stejně tak jako identifikovaná sensitivita otázek nebo zvýšená potřeba ujasnění otázky vyšší mírou chybějících odpovědí). Druhý cíl poskytl rozporuplné výsledky. Při srovnání původního a revidovaného dotazníku došlo k jen velmi malému snížení počtu chybějících pozorování (míra chybějících pozorování je zde brána za nepřímý indikátor kvality) ani revize položek neměla vliv na míru problémů, se kterými se tazatelé a respondenti museli při dotazování potýkat. Efektivita praxe pretestování zde zůstala otázkou, když aplikace jejich nálezů pro další práci vedla jen k nejasným výsledkům.

Účinek uplatnění a aplikace výsledků pretestování (zde kognitivního přístupu) zajímal i DeMaio a Landreth (2004). Zatímco jednotlivé způsoby kognitivního testování se významně rozcházejí v tom, kde u daných problematických otázek tkví problém, na druhou stranu poměrně odpovídali expertním hodnocením z pohledu identifikace chybných položek (otázek). Je třeba si uvědomit, že zde autoři uplatnili velmi širokou definici shody (alespoň jeden prvek hodnocení se shoduje).¹⁷ Rozdílly však stejně jako u srovnání kognitivních testů navzájem, panovaly v tom, co konkrétně na dané položce představuje problém. U méně jak poloviny otázek panovala shoda i o příčině problému mezi jednotlivými metodami. Co se výstupu testování revidovaných dotazníků týče tak testování nepotvrdilo, že by aplikace znalostí z kognitivního testování vedla ke zlepšení revidované verze (z hlediska počtu defektních otázek nebo problémů) a pouze jedna ze

¹⁷ Podporobnosti viz DeMaio, Landreth 2004:102.

tří revidovaných verzí se z hlediska výskytu problematických míst odlišovala od dotazníku původního.

Tyto nálezy vedou k doporučení výzkumníků nejen zapojit více metod pretestování, protože každá prověřuje určitou část možných problémů, ale před samotným terénním sběrem dat provést i test revidované verze, který umožní provedené změny vyhodnotit.

Na základě zjištění opírajících se o předchozí aplikace metod můžeme shrnout, že metody pretestování jsou explanatorně¹⁸ silné, když dokáží dobře identifikovat problémy dotazníku. Další práce s těmito výsledky je obtížná a vede jen k neurčitým výsledkům. Práci, zabývající se, jak zpracováním těchto výsledků a jejich využitím pro další práci je poskrovnu i proto, že metody neposkytují návod nebo doporučení, jak změny provést.

Reparatorní výzkum zabývající se dopadem a vyhodnocením změn dotazníku (provedených na základě znalostí získaných z fáze pretestování) na kvalitu výsledku výběrových šetření je ještě vzácnější. Na vyhodnocení efektivity pretestování a jeho vazby na zlepšení výsledků stále čekáme. Možnou alternativou může být podívat se na tuto problematiku z jiného úhlu pohledu.

¹⁸ Používám členění účelu pretestování na explanatorní, konfirmační a reparatorní podle Forysth, Rothgeb a Willis 2004.

4. Statistické přístupy

Nyní již snad lépe chápeme, proč texty zabývající se designem dotazníku někdy o této práci mluví v duchu, že se více tato činnost blíží práci v umění než vědě (např. Converse, Presser 1986). Jedním z důvodů, který se na tom významnou mírou podílí je i dosavadní praxe v testování dotazníků. Jsou využívány převážně metody (expertní hodnocení, kognitivní rozhovory, částečně platí i na klasický terénní pilot), které jsou neformální (konkrétní způsob realizace může nabývat mnoha odchylek), subjektivní a často neposkytují průkazné ani konzistentní výsledky. Výstupy takového testování mají pak kvalitativní povahu, kterou je obtížné zpracovat a zjednodušit do podoby, aby byly jakkoliv využitelné. Nelze se tedy divit, že výsledky založené na takových datech mohou být velmi rozdílné. Argumentace kritiků, že se do větší nebo menší míry nejedná o vědecké přístupy je tak na místě. O to více tato problematika získává na důležitosti, když si uvědomíme, v jakém kontrastu je přístup k této oblasti (pretestování) oproti pozornosti, která se věnuje statistickým úvahám designu výzkumného šetření.

V úvodu práce je vymezeno, že na proces evaluace otázek se dá pohlížet ze dvou úhlů. První se zabývá možností redukce chyb a otázkami, jako je pochopení otázek a použitých pojmů, znění otázek, zátěž respondenta, problémy získání požadované informace z paměti, proces vytvoření odpovědi atd. Druhý se zabývá tím, jak se tyto problémy projevují kvantitativně, tzn. zabývají se odhadem chyby měření. První úhel pohledu a jeho aplikace jsem se pokusit kriticky zhodnotit v předchozí kapitole. Nyní se budu věnovat tomu, jaké možnosti skýtá druhý přístup. V jakém vztahu jsou tyto jednotlivé typy přístupů bych ocitoval Vinopala, který k tomu uvádí „...je pochopitelné, že její nástroje (myšleno statistických přístupů, vloženo autorem) umožňují zkoumání jiných aspektů kvality, než přístupy kognitivní nebo experimentální, je však zřejmé, že v některých momentech se jejich zájmy protínají (a lze je tedy porovnávat), v jiných že se vzájemně doplňují (a je tedy vhodné je kombinovat)...“ (Vinopal 2008:60) Dále stejný autor k využití poznatků statistických přístupů dodává „...V současné době se zdá, že některé metody již dokážou odhalovat poměrně hluboké roviny procesu získávání informací od respondentů, a svými číselnými údaji tak mohou úspěšně přispívat ke kvalitativně orientované kognitivní analýze kvality dotazů“ (ibid.) Úryvek měl dokumentovat, že statistické přístupy nejsou nějakou odtrženou aplikací, ale její

výsledky jsou brány jako relevantní zdroj informací i pro metody zabývající se odlišnými způsoby testování kvality dotazu.

Statistické přístupy se v kvalitě otázek zabývají reliabilitou, validitou nebo obojím současně. Možnost vyhodnocení těchto indikátorů kvality ovšem přináší jisté nároky na realizaci a design šetření.

Ještě předtím než přejdu k designu výběrového šetření, které slouží k vyhodnocení kritérií kvality dotazu je třeba si tato kritéria představit a definovat.

4.1. Kritéria kvality dotazu

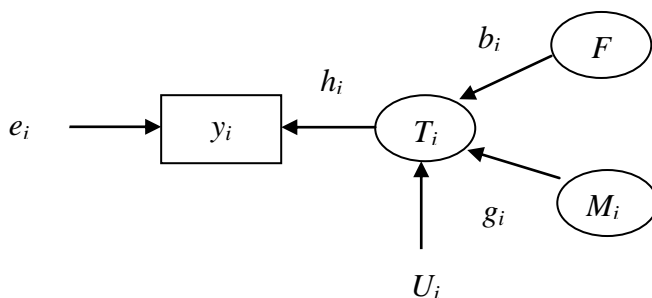
Kvalita dat se posuzuje z hlediska reprezenatativity dat a chyby měření. Konceptů pracujících s chybou měření je mnoho¹⁹, zpravidla se však užívá reliabilita a validita.

V sociálních vědách jsme při měření odkázáni k použití otázek dotazníku, kdy skrze odpovědi respondentů (indikátory) zpětně usuzujeme na latentní proměnnou (koncept), kterou otázkou zamýšlíme měřit. Je samozřejmé, že v tomto procesu dochází skrze nepochopení, použitými pojmy a zněním otázky, vlivem tazatele, chybným kódováním atd. k výskytu náhodných chyb měření. To pro výzkumníky znamená, že principiálně jsou neznámé hodnoty konceptu (nebo jeho operacionalizace do podoby konstruktů), který chceme měřit. Jediné co máme k dispozici jsou odpovědi na otázky (co jsme měřili), které je zatíženy chybou. Tento model měření se v literatuře tradičně vyjadřuje jako $y_i = F + e_i$, kde y zastupuje hodnoty, které získáme měřením, F skutečnou hodnotu konceptu, který zamýšlíme měřit a e náhodnou chybu měření. Předpoklady jsou, že chyba e není závislá na F a že průměr e při opakovaných měření je nulový. První předpoklad je realistický pokud můžeme předpokládat, že žádná další proměnná nemá vzhledem ke konceptu, který chceme měřit vliv na odpovědi respondentů. Druhý požadavek předpokládá náhodné chování chybové složky.

První odchylka od uvedeného modelu se týká převodu zjištěných informací y_i na hodnoty latentní proměnné (konceptu) F . Aby byl tento převod skutečné hodnoty na měřící škálu možný, zavádí se do rovnice koeficienty a_i a b_i (index bude záviset na použité metodě). Získáme tak nový tvar rovnice modelu měření $y_i = a_i + b_i F + e_i$. V procesu měření se však neprojevují pouze náhodné chyby, ale i vlivy, které jsou stabilní v tom smyslu, že se jejich vliv objevuje i při opakovaném měření. Za prvek, který má

takový vliv, můžete jistě považovat vliv použité metody. Za normálních okolností je měření vlivu použité metody přímou cestou nemožné, a pokud definujeme celek všech stabilních odchylek měření, jako D_i získáme novou rovnici měření ve tvaru $y_i = a_i + b_i F + D_i + e_i$. Pokud měříme pouze jednu vlastnost pak komponenty D_i a e_i není možné odlišit. Dále uvidíme, že při užití specifického typu designu to možné je. Je dokonce možné rozlišit rozptyl D_i na komponentu použité metody M_i a zbývající nevalidnost U_i . Výsledný model měření má pak tvar $y_i = a_i + b_i F + g_i M_i + U_i + e_i$, kde g_i představuje parametr metody.

V takovém modelu měření nejsme při měření jedné vlastnosti jednou metodou schopni odlišit a odhadnout náhodnou a systematickou chybu měření. Toto je možné pouze jestliže aplikujeme opakované měření. Pokud totiž zopakujeme za stejných podmínek měření stejné vlastnosti (paralelní měření) rozptyl způsobený použitou metodou M_i stejně jako U_i komponentou zůstává neměnný. Pokud nedojde ani ke změně v hodnotě měřené vlastnosti, pak jedinou intervenující proměnnou, která ovlivňuje hodnoty odpovědi respondentů je náhodná chyba měření e_i . Pokud sloučíme všechny neměnné prvky do jednoho pojmu T_i , získáme model měření ve tvaru $y_i = T_i + e_i$ a $T_i = a_i + b_i F + g_i M_i + U_i$. Parametry kvality měření jsou tradičně podávány jako standardizované koeficienty. Model měření ve standardizované formě je tedy následující: $y_i = h_i T_i + e_i$ a $T_i = b_i F + g_i M_i + U_i$.



Na základě takto definovaného obecného modelu měření můžeme definovat kriteria kvality měření.

Reliabilitou (spolehlivostí) měření se rozumí přesnost měření té vlastnosti, kterou ve skutečnosti měříme. Je definována jako míra rozptylu proměnné y_i , která se nemění při opakovaném měření stejné proměnné (reliabilita = $\text{var}(T_i) / \text{var}(y_i)$ nebo $r^2 = (T, y)$). Koeficient r^2 však není přímo zjistitelný, neboť sice známe nositele informace o T ,

¹⁹ Andrews bez nároku nabídnout kompletní rejstřík pojmů zmiňuje validitu, invaliditu, reliabilitu, preciznost, úroveň měření a další (Andrews 1991:575)

kterým je y , ale skutečnou hodnotu T zjistit většinou nelze. Z toho důvodu je nutné aplikovat opakovaná měření. Smyslem měření je pak zkonstruovat takový postup, aby y bylo co nejpřesnějším nositelem informace o neznámé a nezjistitelné skutečné hodnotě. Z vymezení reliability měření je zřejmé, že se jedná o test-retest metodu zjišťování reliability.

Validitou se rozumí přesnost toho, co ve skutečnosti měříme vzhledem k tomu, co chceme měřit. Validitu v modelu měření reprezentuje vztah mezi T a F tedy b_i a je vymezena jako mocnina efektu při přechodu od konstruktů ke konceptu, který zamýšlím měřit (validita = b_i^2). Pokud míra variability konstruktů vysvětlená latentním konstruktem odpovídá validitě, pak $1 -$ vysvětlená variabilita je rovno nevaliditě. Tuto nevaliditu můžeme ještě rozložit na komponentu použité metody (g_i^2) a U_i . Takto vymezený koncept validity se kvůli odlišení oproti odlišnému způsobu konceptualizace validity nazývá true score validity (TV). Alternativní způsob vyhodnocení validity je možný za pomoci korelace mezi měřenou proměnnou y_i a latentní proměnnou/konceptem F . Takto vymezený koncept měření validity (indikátor validity IV) odpovídá v modelu měření vztahu $IV = (h_i b_i)^2$. Rozdíl mezi takto definovanými mírami validity je zřejmý, zatímco první (TV) je čistou mírou validity nezávislou na reliabilitě, tak IV závisí na TV a hodnotě reliability. Z toho důvodu využití TV koncepce nabízí snazší interpretaci naměřených hodnot validity.

Na zopakování je třeba uvést, že reliability charakterizuje přesnost měření konstruktů pomocí jeho konkrétní operacionalizace. Validita odpovídá přesnosti operacionalizace konceptu konstruktem.

Nyní máme definována kritéria kvality otázek a vztahy mezi prvky modelu měření a těmito kritérii, které můžeme dále využít pro evaluaci otázek našeho dotazovacího instrumentu. Parametry modelu měření, které jsou předmětem našeho zájmu, však není možné odhadnout pouze na základě měření jedné proměnné. K tomuto účelu jsou využívány specifické designy dotazování.

4.2. Design pro měření kvality měření

K odhadům kvality měření se využívá několik přístupů. V české literatuře o nich pojednal Řehák (Řehák 1998) a z toho důvodu se jejich představení budu věnovat pouze krátce. Naopak se zaměřím na další rozpracování MTMM designu, které pokud vím, v české literatuře probráno není.

4.2.1. Quasi Simplex Model

Jak uvádí již Řehák, tento model byl navrhnut Heisem v situacích, kdy nejsou k dispozici baterie položek indikujících stejnou vlastnost či postoj. (Řehák 1998) V takovém případě jedinou možností, jak dospět k odhadu kvality dotazu je využít opakované měření. V tomto modelu, který vychází z formulace výše, není rozlišováno mezi konstruktem, který měříme za pomoci dotazu a latentním konceptem a je opakovaně pokládána jedna otázka za využití jediné metody. Model dovoluje evaluaci kvality dotazu, i když dochází ke změnám v hodnotě T_i . Při třech měřeních stejného dotazu je kvůli identifikovatelnosti modelu nutné zavést předpoklad, že v jednotlivých bodech měření je hodnota reliability konstantní. Je-li opakování více, můžeme odhadovat spolehlivost měření v každém jednotlivém opakování zvlášť, kromě prvního a posledního, která jsou však takto neidentifikovatelná. (Řehák 1998:198) Nevýhodou modelu je, že s jeho pomocí dokážeme odhadnout pouze reliabilitu. K odhadu validity a nevalidity by musel model využít více metod měření.

4.2.2. Repeated Multimethod Model (RMM)

Alternativou k předchozímu, pro účely hodnocení kvality měření nepříliš vhodného modelu je model, kdy sice měříme jeden koncept, ale jsou k tomuto účelu využívány nejméně dvě metody a dvě (nezávislé) měření, přičemž v každém okamžiku měření je aplikována každá operacionalizace. Opakování měření a požadavek jejich nezávislosti představuje však obtíž, se kterou se bude potýkat i následující model. Při opakovaném měření v krátkém časovém okamžiku (např. úvod a závěr dotazníku) můžeme sice oprávněně předpokládat, že v tomto časovém úseku nedošlo k změně postoje nebo názoru respondenta, na druhou stranu tato měření musí být vzdálena natolik, abychom je mohli považovat za nezávislé (problém efektu učení anebo zapamatování). Z důvodu

identifikace modelu je nutné pro měření jedné latentní koncept použít nejméně tři metody měření. Na druhou stranu model měření dovoluje stanovit odhad reliability, validity i srovnat vhodnost použitých metod. Mezi každou z metod a jejími dvěma realizacemi model předpokládána stejnou reliabilitu z důvodu, že se jedná o repliku jedné metody. Každá z metod má vlastní reliabilitu i validitu.

Nevýhodou tohoto přístupu je nemožnost dekomponovat chybu měření na efekt metody a zbývající nevaliditu a dále předpoklad, že mezi jednotlivými realizacemi nedošlo ke změně v předmětu měření.

4.2.3. Multitrait-multimethod design (MTMM)

V MTMM designu je oproti RMM každá vlastnost (konstrukt) každou uplatněnou metodou měřena pouze jedenkrát. Pro identifikovatelnost modelu je nutná realizace měření alespoň tří vlastností třemi metodami (každá vlastnost paralelně měřena stejnými metodami). Z důvodu, že každá metoda a konstrukt je měřen pouze jednou nelze v rámci tohoto modelu odlišit invaliditu U v rámci celkové chyby modelové rovnice e . Jednou variantou aplikace tohoto přístupu, která vede k identifikaci modelu, je vzdát se rozlišení na koncept (F) a konstrukt (T) a předpokládat přímý vztah konceptu k měření (F na y). Tento přístup vede k modelu, v jehož rámci je možné odhadnout validitu (IV), reliabilitu a efekt metody. Ve zbývajícím reziduálním rozptylu není možné rozlišit komponentu U a e . Druhou variantou je zavést předpoklad, že invalidita U je rovna nule (dochází k jejímu odstranění z rovnic výše).²⁰ V tomto modelu je validita rovna koeficientu bi , vliv metody koeficientu gi a reliabilita hi . Validita je zde definována (ve formě TV) jako vztah mezi konstruktem a konceptem (F na T). Výstupy obou parametrizací jsou na sebe navzájem převoditelné. Parametry rovnic tedy po odhadu komparují kvalitu metody i kvalitu vztahu mezi konceptem a operacionalizovanou dotazníkovou otázkou. Právě odhad vlivu metody, který je (za určitého předpokladu) možný (oproti RMM) pouze v MTMM designu vedl k širší aplikaci a preferenci tohoto přístupu.²¹ Řehák k tomuto modelu uvádí „...MTMM nabízí přednost v tom, že v jednom modelu separuje čistou validitu konstruktů vzhledem k vlastnosti, kterou zamýšlíme měřit, a čistý efekt metody...“ (Řehák 1998:202). Upozorňuje však i na limity tohoto přístupu, když uvádí „...největší překážkou

²⁰ Andrews k tomu uvádí, že i když tento předpoklad není možné v rámci daného modelu testovat, tak testování v rámci širšího modelu měření (multitrait-multimethod-multitime) ukázaly, že hodnoty U výrazně neovlivňují kvalitu měření (Andrews 1991:593).

konkrétní aplikace je však obtížnost realizace příslušných schémat v praxi, neboť specifická dotazová situace vyžaduje panelový přístup a vysoké nároky na respondenty násobným opakováním dotazu na totéž. Vyžaduje alespoň trojí dotaz na totéž... V experimentu je možné provést maximálně dvě formy dotazu (dvě metody) během jednoho rozhovoru (na jeho začátku a na jeho konci). Třetí dotaz musí být časově separován, a proto je panel nutný.“ (Řehák 0998:202). I z tohoto důvodu pokračovala práce na další specifikaci modelu, který by přinesl zjednodušení jeho administrace a otevřel širší možnosti pro jeho praktické využití.

4.2.4. Alternativní způsoby realizace MTMM designu

Běžně je v sociologickém výzkumu pro měření jednotlivých proměnných využívána pouze jedna metoda. Tudiž výzkumník nemá možnost odhalit do jaké míry je rozptýl proměnné způsoben náhodnou chybou měření a do jaké míry systematickým vlivem metody. Campbell a Fiske (1959) z toho důvodu navrhli opakování měření různých otázek různými metodami, aby bylo možné zjistit míru chyby těchto komponent. Svým návrhem specifikovali právě MTMM design ve své klasické podobě (měření tří proměnných třemi různými metodami), který byl zamýšlen jako instrument k posouzení konvergentní a diskriminační validity.²² Hlavním problémem jejich přístupu bylo, že se nijak systematicky nezabývali nereabilitou.²³ Další postupy se zaměřili na rozpracování možnosti analýz a použitých způsobů (od korelací k metodě strukturního modelování), odhadem vlivu rozdílných dílčích specifikací modelu (zavádění restrikcí na proměnné modelu)²⁴ a jeho statistických vlastností (identifikovatelnost, ...)²⁵.

Klíčový problém spojený s praktickou aplikovatelností modelu, na který upozorňovalo mnoho sociálních vědců, na řešení čekal. Použitelnost klasického schématu, vázaná na měření tří vlastností třemi metodami, tak byla výrazně omezena. Využití při panelové sběru dat bylo i tak sporadické, využití v běžném režimu sběru dat ještě více ojedinělé. Navíc praktické obtíže způsobené efektem zapamatování nebo učení, není možné v MTMM modelu testovat.

Nový inovativní přístup, který se toto snažit řešit navrhli Saris, Sattora a Coenders (2004). Tito autoři navrhli propojit split ballot a MTMM design do jednoho celku.

²¹ Dalším silným důvodem je možnost zkoumat i vztahy v rámci latentních proměnných (konceptů).

²² Campbell a Fiske 1959:83.

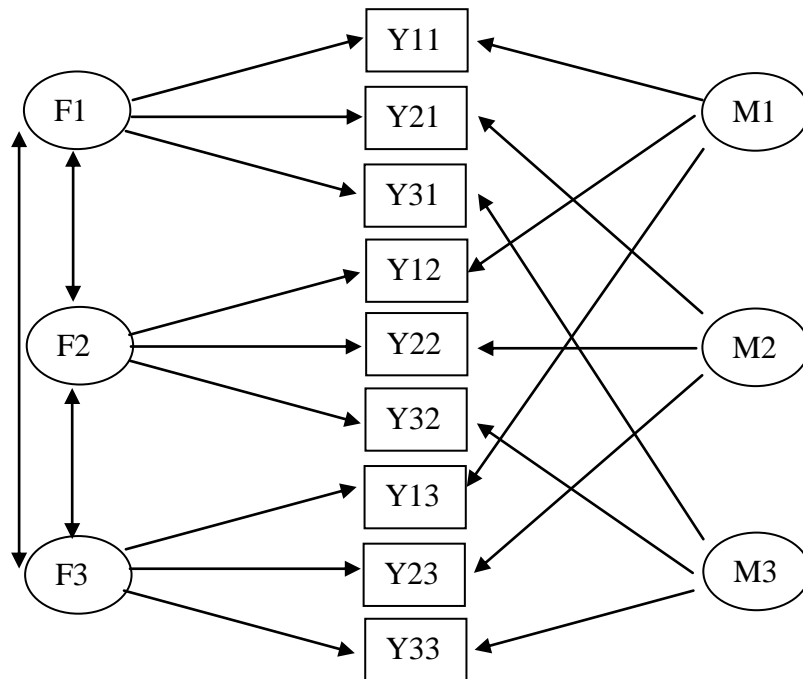
²³ Kritika přístupu např. Alwin 1974.

²⁴ Alwin nabízí hodnocení 4 typů modelů (Alwin:1974).

²⁵ Viz např. Saris, Meurs (1990).

Zatímco první je znám z použití při evaluaci vlivu alternativních metod na distribuci odpovědí respondentů, druhý dokáže odhadovat reliabilitu, validitu a vliv metody. Navržená alternativa se pokouší spojit výhody obou přístupů: využití náhodných výběrů z cílové populace a opakovaných měření stejných vlastností za pomoci více metod. Rozdíl spočívá v tom, že každý respondent (resp. jejich výběr) odpovídá pouze na kombinaci dvou metod měření a celý model je odhadován jako kdyby se skládal z více skupin (multiple-group model). Díky tomu došlo k omezení zátěže respondenta, možnému vlivu efektu zapamatování i efektu pořadí. Když se Meurs a Saris (1990) zabývali problematikou zapamatování zjistili, že respondenti si nepamatují své předchozí odpovědi zhruba po 20 minutách dotazování, a proto je nutné mezi následná měření vložit blok otázek odpovídající délky. Pokud bychom toto pravidlo aplikovali pro klasický MTMM design, dostáváme se k minimální délce dotazování v rozsahu 40 minut. Z praxe víme, že sociálně-vědní výzkumy svým rozsahem tuto časovou zátěž i překračují. I kdyby se tímto způsobem podařilo eliminovat vliv zapamatování, stejně zůstává aktuální problematika zátěže respondenta, která může ústít v nižší přesnost měření druhé, případně třetí aplikace. Dalším možným problémem se ukazovaly vyšší hodnoty korelace mezi druhým a třetím měřením než mezi jinými možnými variantami. Tyto problémy se pokouší řešit alternativní způsoby aplikace MTMM designu, které bývají také označovány SB-MTMM design (SB=split ballot).

Pro úplnost výkladu začněme 1) klasickým MTMM designem ve kterém jsou měřeny tři vlastnosti každou ze tří metod. Model by se dal znázornit v grafické podobě, která vyjadřuje výše uvedené vztahy (a respektuje systém zápisu běžný pro grafické znázornění prvků a vztahů ve strukturním modelování) následovně (každé proměnné Y náleží ještě chybová složka e , která je ve znázornění vynechána):



Na základě matice korelací dochází k odhadování parametrů modelu, který by se dal vyjádřit i tímto způsobem.

$$y_{11}=b_{11}F_1+g_{11}M_1+e_{11}$$

$$y_{21}=b_{12}F_1+g_{21}M_2+e_{21}$$

$$y_{31}=b_{13}F_1+g_{31}M_3+e_{31}$$

$$y_{12}=b_{21}F_2+g_{12}M_1+e_{12}$$

$$y_{22}=b_{22}F_2+g_{22}M_2+e_{22}$$

$$y_{32}=b_{23}F_2+g_{32}M_3+e_{32}$$

$$y_{13}=b_{31}F_3+g_{13}M_1+e_{13}$$

$$y_{23}=b_{32}F_3+g_{23}M_2+e_{23}$$

$$y_{33}=b_{33}F_3+g_{33}M_3+e_{33}$$

Jednotlivé metody nejsou vzájemně korelovány, jednotlivé faktory mohou korelovat avšak jsou na metodách nezávislé. Navíc vliv metody je mezi jejich realizacemi stejný ($g_{11}=g_{12}=g_{13}$). Při takové specifikaci modelu dochází podle Sarie jen zřídka k problémům s odhadem parametrů modelu (Saris, Gallhofer 2007). Dochází tedy k odhadu (v MTMM modelu 3 vlastností a metod) 9 koeficientů reliability, 9 koeficientů validity, 3 koeficientů vlivu metody a 3 vztahů mezi vlastnostmi.

Díky matici korelací získaných za pomoci tohoto schématu a aplikace modelu výše můžeme odhadovat kritéria kvality následujícím způsobem.

	F1	F2	F3	M1	M2	M3
Y11	V11			M11		
Y12		V21		M21		
Y13			V31	M31		
Y21	V12				M12	
Y22		V22			M22	
Y23			V32		M32	
Y31	V13					M13
Y32		V23				M23
Y33			V33			M33

Alternativní designy snižují zátěž respondenta (už není dotazován třikrát) a tuto okolnost kompenzují tím způsobem, že chybějící data (z důvodu nastavení designu) jsou nahrazena a oddotazována na jiném výběrovém podsouboru zkoumané populace. SB-MTMM design tedy aplikuje na rozličné výběrové podsoubory části MTMM designu tak, aby v každém z nich proběhlo pouze jedno opakování. Dalšími možnostmi aplikace tohoto přístupu jsou: 2) Design se dvěma skupinami, kde výběrový soubor je rozdělen do dvou skupin. Jedna skupina odpovídá na tři otázky první metodou, druhá na stejné otázky metodou dvě. Na závěr dotazování je oběma skupinám položena opětovně baterie otázek s využitím třetí metody. Odhady za pomoci tohoto modelu jsou méně efektivní, když tímto způsobem nezískáme výsledky pro vztah mezi první a druhou metodou.

	Čas 1	Čas 2
Skupina 1	Metoda 1	Metoda 3
Skupina 2	Metoda 2	Metoda 3

3) Design s třemi skupinami používá následující schéma:

	Čas 1	Čas 2
Skupina 1	Metoda 1	Metoda 3
Skupina 2	Metoda 2	Metoda 3
Skupina 3	Metoda 3	Metoda 1

Při využití tohoto designu je se všemi metodami zacházeno ekvivalentně. Všechny metody jsou nasazeny poprvé jak v prvním časovém úseku tak i v druhém. Efekt pořadí je tak v tomto designu zrušen. Zároveň můžeme odhadovat vztahy mezi všemi metodami, ale díky tomu, že ne všechny metody jsou aplikovány na všechny respondenty nemáme k dispozici srovnatelné báze za všechny respondenty

4) Za zmínku stojí ještě jeden typ designu, který díky přesné replikaci (dvou) metod dovoluje zkoumat i efekt pořadí. Výběrový soubor ale musí být rozdělen na čtyři skupiny a vyžaduje tak ještě vyšší velikost vzorku. Pokud bychom vynechali jednu se skupin 1 a 4 nebo 2 a 3 dostaneme se lehce k předchozímu designu.

	Čas 1	Čas 2
Skupina 1	Metoda 1	Metoda 1
Skupina 2	Metoda 1	Metoda 2
Skupina 3	Metoda 2	Metoda 1
Skupina 4	Metoda 2	Metoda 2

Hlavním rozdílem oproti klasickému MTMM designu tak zůstává, že ne jeden vzorek, ale více je analyzováno simultánně. Jednotlivé výběrové soubory jsou vybrány z původního jednoho souboru respondentů. Podle Sarise, Sattora a Coenderse můžeme k vyhodnocení použít obecný model MTMM designu, i když nejsou všechny metody oddotazovány v každém z podsouborů respondentů. Údaje od respondentů, které nezískáme díky nastavení designu, jsou brány jako chybějící data a pro výpočty jsou nahrazeny hodnotami 0.²⁶

Po představení možných variací aplikace MTMM designu se zaměřím na využití tohoto designu v českých výběrových šetřeních. Zabývat se budu i výsledky, které zapojení této metody přineslo.

²⁶ Pro detaily týkající se výpočtu SB-MTMM designu odkazují na Saris, Sattora a Coenders (2004:325nn).

4.3. Aplikace MTMM designu v ČR

V České republice, pokud ji mi známo, byl MTMM design využit v šetřeních Evropského sociálního výzkumu (European Social Survey, ESS). Opakované schéma dotazování bylo využito i při Výzkumu spokojenosti a hodnotových preferencí a výzkumu „Senátní volby 1996“ realizovaném SC&C. V aplikacích agenturou SC&C se autoři primárně a výhradně zaměřili na vyhodnocení reliability použitých otázek. Představme si otázky, které se v ESS šetřeních metodologicky testovaly.

Administrace experimentu v rámci ESS probíhá tak, že po vyplnění hlavního dotazníku (core) má tazatel k dispozici ještě i samostatný doplňkový dotazník (supplementary), který obsahuje různé rotující moduly. Jeho součástí je např. dlouhodobě testování hodnotových orientací metodou, kterou vyvinul Schwartz.²⁷ Druhou pro nás zajímavou okolností je, že se v tomto dodatku se objevují otázky doplňující design MTMM a umožňující cíleně vyhodnotit kvalitu některých dotazů hlavního dotazníku.

Administrace doplňkového dotazníku je možná ve třech modech: 1) přímočaré pokračování dotazování navazující na hlavní dotazník, 2) požádat respondenta o samovyplnění tohoto dotazníku, 3) případně předat respondentovi dotazník k samovyplnění s možností vrátit jej vyplněný poštou.

Podívejme se na způsoby realizace v jednotlivých šetřeních a výsledky těchto experimentů.

4.3.1. ESS – Round 1

V první vlně sociálního výzkumu, který byl realizován v roce 2002, bylo testováno celkem 16 položek dotazníku. Po hlavním dotazníku následoval doplňkový dotazník, který byl ve dvou verzích (variantnost se týkala pořadí otázek nikoliv jejich zapojení) a tazatel jej vyplňoval s respondentem.

Jednalo se o otázky:

- A1.** Kolik času celkově strávíte během průměrného všedního dne sledováním televize?
- A3.** Kolik času celkově strávíte během průměrného všedního dne posloucháním rádia?
- A5.** Kolik času celkově strávíte během průměrného všedního dne čtením novin?

²⁷ O této metodě referovala Blanka Řeháková na stránkách sociologického časopisu (Řeháková Sociologický časopis, 2006, Vol. 42, No. 1: 107–128).

*Metoda 1: 7 bodová škála, slovní vyjadřující množství času*²⁸

*Metoda 2: otevřená otázka*²⁹

*Metoda 3: 7 bodová škála*³⁰

Z důvodu použití otevřené otázky nejsou tyto otázky vyhodnoceny.³¹

A8. Obecně vzato, řekl(a) byste, že se většině lidí dá důvěřovat, nebo že člověk nemůže být při jednání s lidmi nikdy dost opatrný? Odpovězte mi s použitím stupnice 0 až 10, kde 0 znamená, že člověk nemůže být nikdy dost opatrný a 10 znamená, že většině lidí se dá důvěřovat.

HF10. Obecně vzato, řekl(a) byste že se většině lidem dá důvěřovat, nebo že člověk nemůže být při jednání s lidmi nikdy dost opatrný? Odpovězte prosím s použitím stupnice 0 až 5, kde 0 znamená, že člověk nemůže být nikdy dost opatrný, a 5 znamená, že většině lidí se dá důvěřovat.

HF28. Obecně vzato, řekl(a) byste že se většině lidem dá důvěřovat, nebo že člověk nemůže být při jednání s lidmi nikdy dost opatrný?

A9. Myslíte si, že by se Vás většina lidí snažila podvést, pokud by měli tu možnost, nebo by se snažili být poctiví?

HF11. Myslíte si, že by se Vás většina lidí snažila podvést, pokud by měli tu možnost, nebo by se snažili být poctiví?

HF29. Myslíte si, že by se Vás většina lidí snažila podvést, pokud by měli tu možnost, nebo by se snažili být poctiví?

A10. Řekl(a) byste, že se lidé většinou snaží pomáhat druhým, nebo že se většinou starají o sebe?

HF12. Řekl(a) byste, že se lidé většinou snaží pomáhat druhým, nebo že se většinou starají o sebe?

HF30. Řekl(a) byste, že se lidé většinou snaží pomáhat druhým, nebo že se většinou starají o sebe?

Pro uvedené otázky byly použity následující metody měření.

*Metoda 1: 11 bodová škála*³²

²⁸ Vůbec žádný čas, Méně než ½ hodiny, ½ hodiny až 1 hodinu, Více než 1 hodinu, ale ne více než 1,5 hodiny, Více než 1,5 hodiny, ale ne více než 2 hodiny, Více než 2 hodiny, ale ne více než 2,5 hodiny, Více než 2,5 hodiny, ale ne více než 3 hodiny, Více než 3 hodiny

²⁹ Vypište hodiny: a minuty.

³⁰ Vůbec žádný čas, Velmi málo času, Trochu času, Poměrně dost času, Docela hodně času, Hodně času, Velmi mnoho času.

Metoda 2: 5 bodová škála³³

Metoda 3: dichotomie

Vyhodnocení kvality měření je pak následující:³⁴

	Validita			Reliabilita	
	F1	F2	F3		
A8	0,986			0,697	
A9		0,984		0,65	Metoda 1
A10			0,984	0,65	
HF10	0.848			0.801	
HF11		0.805		0.787	Metoda 2
HF12			0.813	0.772	
HF28	0.787			0.609	
HF29		0.778		0.710	Metoda 3
HF30			0.787	0.569	

V dalším textu již nebudu uvádět kompletní znění všech otázek, které byly k měření konstruktů použity.

B2. Jak často se politika zdá být tak složitá, že vlastně nemůžete pochopit, co se děje?

B3. Myslíte, že byste mohl(a) hrát aktivní roli ve skupině, která se zabývá politickými otázkami?

B4. Jak je pro Vás těžké či lehké udělat si názor na politické otázky?

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 5 bodová škála užívající slovní varianty specifické měřenému konstrukt

Metoda 2: 5 bodová škála, směr negativní-pozitivní³⁵

Metoda 3: 5 bodová škála, směr pozitivní-negativní³⁶

³¹ Ve výsledcích bude dále uváděna pouze hodnota validity a reliability. Efekt metody není nutné uvádět, protože jej lze dopočítat (vliv metody)² = 1-(validita)². Výsledky čerpány ze zdroje spq.nl.

³² 11 bodová škála (0-10), slovně popsané mezní hodnoty, bipolární

³³ 5 bodová škála (0-5)

³⁴ Pro připomenutí v tomto modelu je validita rovna koeficientu *bi*, vliv metody koeficientu *gi* a reliabilita *hi*. Validita je zde definována (ve formě *TV*) jako vztah mezi konstrukt a konceptem (*F na T*).

³⁵ Rozhodně nesouhlasím, Spíše nesouhlasím, Ani nesouhlas, ani souhlas, Spíše souhlasím, Rozhodně souhlasím.

Vyhodnocení kvality měření je pak následující:

	Validita			Reliabilita	
	F1	F2	F3		
B2	0.997			0.562	
B3		0.979		0.766	Metoda 1
B4			0.997	0.642	
HF4	0.992			0.557	
HF5		0.994		0.619	Metoda 2
HF6			0.992	0.534	
HF22	0.973			0.603	
HF23		0.979		0.665	Metoda 3
HF24			0.975	0.635	

Řekněte mi prosím s použitím stupnice 0 - 10 jak moc Vy osobně důvěřujete institucím, které budu číst nahlas. 0 znamená, že instituci vůbec nedůvěřujete, 10 znamená, že jí naprosto důvěřujete

B7 Český Parlament

B8 Právní systém

B9 Policie

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 11 bodová škála, bipolární, karta

Metoda 2: 11 bodová škála, bipolární, karta

Metoda 3: 11 bodová škála, bipolární

Vyhodnocení kvality měření je pak následující:

	Validita			Reliabilita	
	F1	F2	F3		
B7	0.901			0.906	
B8		0.910		0.856	Metoda 1
B9			0.915	0.895	

³⁶ Rozhodně souhlasím, Spíše souhlasím, Ani souhlas, ani nesouhlas, Spíše nesouhlasím, Rozhodně nesouhlasím.

HF13	0.929		0.937	
HF14		0.859	0.965	Metoda 2
HF15			0.927	
HF31	0.811		0.825	
HF32		0.839	0.916	Metoda 3
HF33			0.858	

B30. Jak jste celkově spokojen(a) se současným stavem ekonomiky v České republice?

B31. Nyní se zamyslete na vládou České republiky. Jak jste spokojen(a) s tím, jak dělá svou práci?

B32. A jak jste celkově spokojen(a) s tím, jak funguje demokracie v České republice?

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 11 bodová škála, bipolární, směr negativní-pozitivní³⁷

Metoda 2: 4 bodová škála³⁸ unipolární

Metoda 3: 4 bodová škála³⁹ bipolární

Vyhodnocení kvality měření je pak následující:

	Validita			Reliabilita	
	F1	F2	F3		
B30	0.820			0.782	
B31		0.855		0.851	Metoda 1
B32			0.857	0.772	
HF7	0.761			0.667	
HF8		0.809		0.728	Metoda 2
HF9			0.951	0.612	
HF25	0.813			0.643	
HF26		0.836		0.678	Metoda 3
HF27			0.888	0.836	

³⁷ 0 naprosto nespokojený, 10 naprosto spokojený.

³⁸ Velmi nespokojen(a), Docela nespokojen(a), Docela spokojen(a), Velmi spokojen(a).

³⁹ Naprosto nespokojený(á) 0, Naprosto spokojený(á) 5.

B43. Čím méně vláda zasahuje do ekonomiky, tím je to pro Českou republiku lepší.

B44. Vláda by měla přijmout opatření, která by snižovala rozdíly v příjmech.

B45. Zaměstnanci potřebují silné odbory, aby mohli bránit svoje pracovní podmínky a platy.

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 5 bodová škála souhlasu, baterie výroků

Metoda 2: 5 bodová škála souhlasu⁴⁰

Metoda 3: 5 bodová škála užívající slovní varianty specifické měřenému konstrukt

Vyhodnocení kvality měření je pak následující:

	Validita			Reliabilita	
	F1	F2	F3		
B43	0.925			0.638	
B44		0.957		0.827	Metoda 1
B45			0.934	0.775	
HF16	0.919			0.743	
HF17		0.950		0.839	Metoda 2
HF18			0.915	0.770	
HF34	0.951			0.477	
HF35		0.981		0.841	Metoda 3
HF36			0.971	0.753	

Pokud bychom se měli pokusit o nějaké, i když velmi předčasné, zobecnění tak výsledky ukazují, že lepších výsledků lze dosáhnout použitím 11 bodových škál, dále se jako výhodný krok zdá i použití specifických variant konceptu do odpovědní škály. V jednom případě však využití 11 bodové škály zaostávala za škálou 5 bodovou v případě vyhodnocení reliability (otázky týkající se sociální důvěry). Také použití dichotomie nevykazuje nejlepší výsledky. Dále je zajímavé srovnání výsledků otázek na sociální a politickou důvěru při použití stejné metody (11 bodové škály), kdy otázky týkající se sociální důvěry vykazovaly vyšší validitu (rozdílný popis krajních možností).

⁴⁰ Rozhodně souhlasím, Spíše souhlasím, Ani souhlas, ani nesouhlas, Spíše nesouhlasím, Rozhodně nesouhlasím.

Na druhou stranu otázky zabývající se sociální důvěrou měly nižší reliabilitu než otázky zaměřující se na důvěru v politické instituce. Nejvyšší validity dosahovaly v průměru otázky zabývající se participací a postoji k politice (B2-4). Naopak nejnižší hodnoty vykazovaly otázky týkající se spokojenosti s politickými institucemi v ČR (B30-32). U reliability dosáhly nejvyšších hodnot otázky týkající se důvěry v politické instituce, nejméně reliabilní se naopak ukázal dotaz zabývající se postoji k politice (B2-4).

4.3.2. ESS – Round 2

Stejně jako v první vlně ESS i nyní doplňkový dotazník obsahoval dvě sady položek. Schwartzovu koncepci měření hodnot následovaly otázky zabývající se kvalitou měření. Každá z těchto otázek se v dotazníku vyskytuje ve třech verzích, každá z nich byla administrována na třetině vzorku (split ballot). Na rozdíl od první vlny, kde si participující země mohly vybrat způsob administrace doplňkového dotazníku byl proces nyní zjednodušen a všechny země použily stejný experimentální design (jediná volba, jenž byla ponechána otevřená a kde můžeme případně najít rozdíly se mohla týkat toho, zda se národní koordinátoři rozhodnou pro dotazování nebo samovyplnění respondentem).

Situace v terénu pak vypadala následovně. Každá země použila tři verze doplňkového dotazníku, který obsahovat identické otázky měřící hodnoty (Schwartz) a v rámci verzí odlišných 12 experimentálních otázek.

Jako v případě první vlny byl i zde využit split-ballot design. Nyní však byl vzorek rozdělen do tří podskupin, kdy každá skupina vyplňovala příslušnou verzi doplňkového dotazníku, takže žádný respondent nevyplňoval více jak 12 otázek pro měření kvality dotazu.

Jednalo se o otázky:

Řekněte mi prosím s použitím stupnice 0 - 10 jak moc Vy osobně důvěřujete institucím, které budu číst nahlas. 0 znamená, že instituci vůbec nedůvěřujete, 10 znamená, že jí naprosto důvěřujete.

B4. českému Parlamentu?

B5. právnímu systému?

B7. politikům?

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 11 bodová škála, baterie, popis krajních bodů škály, karta⁴¹

Metoda 2: 11 bodová škála, popis krajních bodů škály, karta⁴²

Metoda 3: 11 bodová škála, popis krajních bodů škály, karta⁴³

Vyhodnocení kvality měření je pak následující:

	Validita			Reliabilita	
	F1	F2	F3		
B4	0.916			0.897	
B5		0.925		0.852	Metoda 1
B7			0.905	0.841	
I25	0.857			0.920	
I26		0.870		0.903	Metoda 2
I27			0.840	0.901	
I38	0.854			0.766	
I39		0.731		0.918	Metoda 3
I40			0.733	0.762	

B25. Jak jste celkově spokojen(a) se současným stavem ekonomiky v České republice?

B26. Nyní se zamyslete nad vládou České republiky. Jak jste spokojen(a) s tím, jak dělá svou práci?

B27. A jak jste celkově spokojen(a) s tím, jak funguje demokracie v České republice?

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 11 bodová škála, popis krajních hodnot, směr nespokojen-spokojen⁴⁴

Metoda 2: 11 bodová škála, popis krajních hodnot i střední hodnoty, směr nespokojen-spokojen⁴⁵

Metoda 3: 11 bodová škála, popis krajních hodnot, směr nespokojen-spokojen⁴⁶

Vyhodnocení kvality měření je pak následující:

⁴¹ Vůbec žádná důvěra 0, Naprostá důvěra 10.

⁴² Vůbec žádná důvěra 0, Naprostá důvěra 10.

⁴³ Vůbec žádná důvěra 0, Naprostá důvěra 10.

⁴⁴ Naprosto nespokojený(á) 0, Naprosto spokojený(á) 10.

⁴⁵ Naprosto nespokojený(á) 0, ani spokojený(á) ani nespokojený(á) 5, Naprosto spokojený(á) 10.

	Validita			Reliabilita	
	F1	F2	F3		
B25	0.823			0.759	
B26		0.849		0.826	Metoda 1
B27			0.927	0.782	
I11	0.766			0.842	
I12		0.784		0.901	Metoda 2
I13			0.791	0.787	
I35	0.852			0.836	
I36		0.863		0.879	Metoda 3
I37			0.938	0.796	

D25. Lékaři neříkají pacientům celou pravdu/ Lékaři málokdy zatajují před pacienty celou pravdu./ Lékaři obvykle zatajují před pacienty celou pravdu.

D26. Praktičtí lékaři jednají s pacienty jako se sobě rovnými. / Praktičtí lékaři málokdy jednají s pacienty jako se sobě rovnými. /Praktičtí lékaři obvykle jednají s pacienty jako se sobě rovnými.

D27. Než lékaři rozhodnou o způsobu léčení, pohovoří si o tom se svým pacientem./ Než lékaři rozhodnou o způsobu léčení, málokdy si o tom pohovoří se svým pacientem./ Než lékaři rozhodnou o způsobu léčení, obvykle si o tom pohovoří se svým pacientem.
Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 5 bodová škála⁴⁷

Metoda 2: 5 bodová škála souhlasu⁴⁸

Metoda 3: 5 bodová škála souhlasu⁴⁹

Vyhodnocení kvality měření je pak následující:

⁴⁶ Velmi nespokojený(á) 0, Velmi spokojený(á) 10.

⁴⁷ Nikdy nebo téměř nikdy, někdy, Tak napůl, Většinou, Vždy nebo téměř vždy.

⁴⁸ Rozhodně souhlasím, Spíše souhlasím, Ani souhlas, ani nesouhlas, Spíše nesouhlasím, Rozhodně nesouhlasím.

⁴⁹ Rozhodně souhlasím, Spíše souhlasím, Ani souhlas, ani nesouhlas, Spíše nesouhlasím, Rozhodně nesouhlasím. V případě použití stejných metod se jedná o rozdílné znění otázky.

	Validita			Reliabilita	
	F1	F2	F3		
D25	0.740			0.335	
D26		0.912		0.723	Metoda 1
D27			0.902	0.674	
I5					Metoda 2 ⁵⁰
I6					
I7					
I28	0.956			0.555	
I29		0.958		0.490	Metoda 3
I30			0.943	0.389	

G6. Žena by měla být připravena omezit svou placenou práci kvůli rodině./ Žena by neměla být nucena omezovat placenou práci kvůli rodině./ Vyberte prosím číslo mezi 1 a 5 podle toho, jak blízký je Váš názor k výrokům uvedeným na této kartě.

G7. Muži by měli přijmout stejnou odpovědnost za domov a děti jako ženy./ Ženy by měly mít větší odpovědnost za domov a děti než muži./ Vyberte prosím číslo mezi 1 a 5 podle toho, jak blízký je Váš názor k výrokům uvedeným na této kartě.

G8. Když je málo pracovních míst, muži by měli mít větší právo na zaměstnání než ženy./ Když je málo pracovních míst, ženy by měly mít stejné právo na zaměstnání jako muži.

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 5 bodová škála souhlasu⁵¹

Metoda 2: 5 bodová škála souhlasu⁵²

Metoda 3: 5 bodová škála⁵³

Vyhodnocení kvality měření je pak následující:

⁵⁰ Výsledky této metody bohužel nejsou dostupné a ani po vyžádání u zhodnotitele The Research and Expertise Centre for Survey Methodology (RECSM) jsem je neobdržel.

⁵¹ Rozhodně souhlasím, Spíše souhlasím, Ani souhlas, ani nesouhlas, Spíše nesouhlasím, Rozhodně nesouhlasím.

⁵² Rozhodně souhlasím, Spíše souhlasím, Ani souhlas, ani nesouhlas, Spíše nesouhlasím, Rozhodně nesouhlasím.

⁵³ Žena by měla být připravena omezit placenou práci kvůli rodině1, Žena by neměla být nucena omezovat placenou práci kvůli rodině 5.

	Validita			Reliabilita	
	F1	F2	F3		
G6	0.962			0.619	
G7		0.976		0.249	Metoda 1
G8			0.972	0.629	
I8					Metoda 2 ⁵⁴
I9					
I10			0.907	0.421	
I22	0.908			0.648	
I23		0.889		0.633	Metoda 3
I24			0.930	0.734	

G22. Nyní bych rád(a) hovořil(a) o domácích pracích tak, jak jsou popsány na kartě. Domácími pracemi rozumíme činnosti, které se vykonávají doma, jako je vaření, praní, uklízení, péče o oděvy, nakupování, údržba domu. Nezapočítáváme ale mezi ně péči o děti a volnočasové aktivity. Kolik hodin celkově členové Vaší domácnosti stráví domácími

G23. A kolik z tohoto času strávíte domácími pracemi Vy osobně?

G24. A kolik z tohoto času věnuje domácím pracím Váš manžel/partner Vaše manželka/partnerka?

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: otevřená otázka

Metoda 2: otevřená otázka

Metoda 3: otevřená otázka

Není hodnoceno jako MTMM experiment.

G64. Moje práce je hodně rozmanitá.

G66. Moje práce je jistá.

G70. Moje práce ohrožuje mé zdraví nebo moji bezpečnost.

Pro uvedené otázky byly použity následující metody měření.

Metoda 1: 4 bodová škála⁵⁵

⁵⁴ ⁵⁴ Výsledky této metody bohužel nejsou dostupné a ani po vyžádání u hodnotitele The Research and Expertise Centre for Survey Methodology (RECSM) jsem je neobdržel.

⁵⁵ Není vůbec pravdivý, Málo pravdivý, Docela pravdivý, Velmi pravdivý.

Metoda 2: 4 bodová škála užívající slovní varianty specifické měřenému konstruktů⁵⁶

Metoda 3: 11 bodová škála, popis krajních hodnot.⁵⁷

Vyhodnocení kvality měření je pak následující:

	Validita			Reliabilita	
	F1	F2	F3		
G64	0.980			0.726	
G66		0.981		0.611	Metoda 1
G70			0.976	0.534	
I19	0.993			0.734	
I20		0.993		0.857	Metoda 2
I21			0.993	0.751	
I32	0.890			0.743	
I33		0.912		0.739	Metoda 3
I34			0.912	0.880	

Pokud bychom se i zde pře neúplnost výsledků měli pokusit o dílčí vyhodnocení, pak výsledky ukazují, že aplikace otázek na spokojenost s politickým systémem dosahuje velmi podobných hodnot jako v ESS 1 (validita obojí na úrovni 0,84, reliabilita v ESS 1 nižší o 0,1). Podobně to platí i v otázky zabývající se důvěrou v politické instituce. U tématu práce nabízí lepší výsledky aplikace 4 bodové škály užívající specifické varianty konceptu (podobně jako v ESS1), reliabilita je u této metody podobná jako při použití 11 bodové škály.

Nejvyšší validity dosahovaly v průměru otázky zabývající se hodnocením vlastní práce (G64, 66, 70), naopak nejnižší hodnoty vykazovaly otázky týkající se spokojenosti s politickými institucemi v ČR (B25-27). U reliability dosáhly nejvyšších hodnot otázky týkající se důvěry v politické instituce, nejméně reliabilní se naopak ukázal (i přes neúplné vyhodnocení) dotaz zabývající se chováním lékařů a postoji ke spojení rodinného a pracovního života u žen.

⁵⁶ Není vůbec rozmanitá, Je málo rozmanitá, Je celkem rozmanitá, je velmi rozmanitá.

⁵⁷ Není vůbec rozmanitá 0, Velmi rozmanitá 10.

4.3.3. ESS – Round 3

Tato vlna Evropského sociologického výzkumu v ČR nebyla realizována. Pro obsah a další informace odkazují na stránky tohoto výzkumného projektu (<http://www.europeansocialsurvey.org>).

4.3.4. ESS – Round 3

Ve čtvrté vlně, která již byla opětovně realizována i v ČR, byl použit design, který je popsán u druhé vlny.

Vyhodnocení této vlny, která proběhla v roce 2009, však ještě není dostupné a z tohoto důvodu ani predikční nástroj SQP, jehož možnosti nás zajímají primárně, s nimi nemůže pracovat. V dalším textu představený nástroj pretestování SQP se opírá (prozatím) pouze o výsledky vyhodnocení prvních tří vln (v ČR pouze dvou) MTMM experimentů a dalších studií, které proběhly v dalších zemích.

4.4. Následná analýza výsledků MTMM experimentů jako nástroj pretestování

Zatímco MTMM design a jeho užití pochází z 50. let minulého století, myšlenka následné analýzy zjištění získaných touto metodou je mnohem mladší. Autorem této idey je Andrews (1984), který s podporou Michiganského Survey Research Center provedl a dále analyzoval tehdy dostupné výsledky MTMM studií. Záměrem této meta analýzy bylo stanovit míru vlivu designu dotazu (a tedy i některých rozhodnutí výzkumníka o tom, jak daný dotaz koncipovat) na kvalitu dotazu (validitu a reliabilitu) a využít takto získané poznatky pro predikci. Andrewsovi možnosti však byly v této době limitované, když počet studií, kde byl MTMM design uplatněn byl jen velmi omezený. I tak si připomeňme některé jeho závěry i proto, že Andrewsův (1984) způsob meta analýzy principiálně následují i další autoři. Myšlenka je taková, že poznatky získané z vyhodnocení MTMM experimentů (tj. odhady validity, reliability a vlivu metody) využijeme pro další analýzu, kde je definujeme jako závislé proměnné a jako jejich prediktory (nezávislé proměnné) definujeme prvky charakterizující dotaz (např. typ a směr škály, pozice v dotazníku atd.).⁵⁸ Takto získané výsledky potom můžeme aplikovat na každou další otázku (u které definujeme identicky parametry užitě k původnímu zakódování predikátorů) a získáme tak možnost odhadnout její kvalitativní parametry. Důležitým zjištěním, které potvrdilo, že podobné analýzy mají smysl a zasluhují další pozornost bylo, že charakteristika dotazu vysvětluje velkou míru rozptylu odhadů kvality.⁵⁹ Andrews na základě analýz dospěl k tomu, že největší vliv na kvalitu dotazu má počet kategorií odpovědní škály, to jestli odpovědní kategorie obsahují i možnost neví, délka baterie a jestli dotaz užívá absolutní nebo komparativní perspektivu. Neméně zajímavé bylo i zjištění, které prvky designu dotazu naopak vliv na kvalitu nemají. Z tohoto úhlu pohledu Andrews identifikoval jako nedůležité doménu (téma) výzkumu, zda škála obsahuje střední (únikovou) možnost (typu ani ani) a metodu sběru. Posledním inspiračním bodem byl návrh pro další výzkumníky, aby tato zjištění extrapolovali na vlastní výzkumy a získali tak (s jistými omezeními) odhady míry kvality dotazů vlastního dotazníku.

Naštěstí tento průkopnický nápad neupadl v zapomnění a myšlenkou další práce s výsledky MTMM experimentů se zabývalo několik dalších autorů. Jejich snažení

⁵⁸ Konkrétní proměnné, které využil Andrews 1984: 426nn.

napomáhalo i to, že se ve svých analýzách mohli opřít o stále se rozrůstající soubor použitelných dat. Rodgers, Adrews a Herzog pracovali na datech z amerických výzkumů, Kotlinger rakouských, Scherpenzeel a Saris (1997) analyzovali holandské studie. Do jejich analýz jako závislá proměnná vstupují již odhady validity a reliability více než 600 proměnných. Jejich zjištění potvrdily nálezy Andrewse, že míra rozptylu odhadů kvality je vysvětlitelná charakteristikou dotazu. Zpětně a to je důležité to také znamená, že kvalita dotazu může být s určitou mírou jistoty predikována. Dle jejich zjištění je kvalita dotazu nejvíce ovlivněna tématem a jeho sensitivou, zvolenou škálou⁶⁰ a počtem opakování. Mírný vliv měla zvolená metoda sběru dat, stejně tak se např. ukázalo, že použití středních bodů škály má dopad pouze na odhad validity a nikoliv na realibilitu. Naopak přítomnost varianty „nevím“, symetrie škály a pozice v dotazníku⁶¹ byly prakticky bez významu na odhady kvality. Zajímavým problémem pro další zkoumání se ukázalo prověření vlivu zapamatování, když ve studiích, v nichž byly tři dotazy opakovány v průběhu jednoho rozhovoru, byly odhady kvality vyšší u druhého a třetího měření.

Komplexnější analýzou je práce Sarise, van der Velda a Gallhofer (2004), která navazuje a je rozpracováním práce Sarise a Scherpenzeel. Celkem je v této analýze použito více než 1000 měření, kdy byl aplikován MTMM design (a jsou dostupné i odhady validit a reliability) a poprvé byly do analýzy zahrnuty výsledky z více jazykových oblastí. Z toho důvodu bylo nutné unifikovat a integrovat výstupy všech MTMM experimentů tak, aby byly srovnatelné a využitelné pro další analýzu.⁶² Pro analýzu byla využita regrese s použitím dummy proměnných, kterou bychom mohli zapsat:

$$C = a + b_{11}D_{11} + b_{21}D_{21} + \dots + b_{12}D_{12} + b_{22}D_{22} + \dots + b_3Ncat + \dots + e, \text{ kde}$$

C reprezentuje kritérium kvality (hodnota reliability nebo validity), D reprezentuje dummy proměnné j -té nominální proměnné, b regresní koeficienty. Všechny dummy proměnné nabývají hodnoty 0, pokud nenastane případ, kterému daná kategorie odpovídá (referenční kategorie 0). Intercept závislé proměnné (validity anebo reliability) odpovídá případu, kdy všechny nezávislé proměnné mají hodnotu 0. Detailní výsledky např. ukázaly, že otázky týkající se chování (zejména četnosti) dosahují

⁵⁹ Design dotazu vysvětluje 66-72 % rozptylu závislé proměnné Andrews 1984:428.

⁶⁰ Zajímavý se o rozdíl mezi marking a typing vzhledem k délce škály (p. 336).

⁶¹ Je zde třeba vzít v potaz, že v tomto meta výzkumu nebyly testovány dotazníky tak rozdílných délek jako u Andrewse, viz také poznámka k diskusi.

⁶² Postup integrace dat i jazykové unifikace viz Saris, van der Veld, Gallhofer 2004:282.

horších výsledků než postojové otázky. Využití neutrální střední kategorie na odpovědní škále naopak kvalitu dotazu zlepšuje. Stejně tam má pozitivní efekt použití popisů kategorií (ovšem ne ve formě celých vět). Vliv pozice otázky v dotazníku se ukázal ambivalentní. Zatímco reliabilita narůstá s pozicí otázky v dotazníku zřejmě v souvislosti s tím, jak se respondent „učí“ zvládat odpovědní úkol, validita dosahuje vysokých hodnot již od počátku dotazování a postupně u ní dochází k mírnému poklesu.⁶³

I přes limity tohoto přístupu spočívající v omezeném počtu studií, v omezeném počtu témat (volný čas, rodina) aj. analýza v souladu s předchozími zjištěními ukázala, že nezávislé proměnné dokáží vysvětlit téměř 50 % rozptylu hodnot reliability a 60 % u validity. Stejně tak standardizovaná chyba regresních koeficientů nedosahuje vysokých hodnot, což autoři vysvětlují jako indikátor nízkých hodnot korelace mezi vysvětlujícími proměnnými modelu.

Z důvodu další práce s těmito výsledky v následujícím textu byl uveden i způsob výpočtu. Na základě tohoto postupu byl vybudován i program usnadňující práci s těmito výsledky, který autoři nazvali Survey Quality Predictor (SQP)⁶⁴.

4.5. Využití výsledků a práce v SQP – optimalizace, korekce měření

Jak je uvedeno výše, předpokladem vyhodnocení a predikce hodnot kvality dotazu je jeho určitá kategorizace. Předpokladem samotné predikce je jednak realizace dostatečného počtu MTMM experimentů a jejich následná analýza, která bude natolik robustní, že její výsledky budou použitelné k predikci. O tom, že důležitým explanandem hodnot kvality dotazu je jeho konkrétní znění a další parametry sekundární analýza výsledků MTMM experimentů potvrdila. Principiálně tedy jsou výsledky analýzy použitelné a můžeme na základě agregovaných zjištění z MTMM experimentů odhadovat i aktuální hodnoty odhadu kvality.

Postup je takový, že při analýze výsledků MTMM experimentů se analyzované dotazy podrobily určité kategorizaci parametrů, kterých mohou nabývat. Analýza pak ukázala, jaký vliv mohou tyto různé jeho individuální charakteristiky mít na míru kvality dotazu (při kontrole vlivu ostatních proměnných). Při odhadu každé další otázky

⁶³ Detailní výsledky ještě před posledním rozšířením souboru analýz jsou dostupné Saris, Gallhofer (2007).

jsou využity tyto získané koeficienty, které se aplikují na stejně provedenou schematizaci parametrů nové otázky, což dovoluje odhad jejich kvalitativních měř (validity, reliability). Důležité je tedy nyní ukázat, v jakých parametrech konkrétní otázku zakódovat (a jak byly také schematizovány již pro účely analýzy MTMM experimentů), abychom byli schopni získat odhad kvality.

Tento systém se opírá o teorii měření, kterou Saris představil v knize *Design, evaluation, and analysis of questionnaires for survey research* (2007). Tento způsob operacionalizace teoretických konceptů sociálního výzkumu do podoby měřitelných otázek výzkumného instrumentu se do jisté míry promítl i do SQP schematizace. Ponechme tuto záležitost a inspirační prvek stranou a zaměřme se na konkrétní atributy otázkem s kterými meta analýza i SQP pracuje. Jedná se o parametry (pro úplnost a ukázkou komplexity schematizace jsou uvedeny všechny parametry)⁶⁵:

doména	koncept	sensitivita	centralita	referenční období
typ formulace žádosti o odpověď	využití podnětu nebo výroku v žádosti o odpověď	absolutní nebo komparativní měřítko	typ odpovědní škály	počet kategorií
popisky kategorií	délka textu popisu kategorií	pořadí popisků	korespondence mezi popisky a číselnou hodnotou/rozsahem škály	stupnice bipolární / unipolární
počet pevných referenčních bodů	možnost nevím	instrukce pro tazatele	instrukce pro respondenta	motivace respondenta

⁶⁴ Program SQP není pouze nástrojem pro predikci kvality dotazu i když je to jeho primární účel. Zároveň se jedná o databázi výsledků měření, kde bylo využito MTMM schéma dotazování.

⁶⁵ Detailní seznam s popisem kódů včetně možností je dostupný na sqp.nl.

použití informace nebo definice	představení/úvod	počet vět v žádosti o odpověď	počet slov v žádosti o odpověď	celkový počet podstatných jmen v žádosti o odpověď
počet abstraktních podstatných jmen v žádosti o odpověď	celkový počet slov v žádosti o odpověď	počet vedlejších vět v žádosti o odpověď	počet slov v odpovědní škále	celkový počet podstatných jmen v odpovědní škále
celkový počet abstraktních podstatných jmen v odpovědní škále	použití karty	horizontální nebo vertikální škála	korespondence textu a kategorií	čísla nebo písmena před odpovědní kategorií
škála s číselnými údaji ohraničenými v kolonkách	počátek odpovědi respondenta na kartě	znění otázky na kartě	použití grafických prvků na kartě	zapojení počítače
zapojení tazatele	vizuální prezentace	pozice v dotazníku		

V analýze konkrétního dotazu se pak dá postupovat dvěma způsoby. První možnost je provést odhad na základě výstupů meta analýz (pro následující příklad budu vycházet z údajů, které jsou přetištěny v Saris, Gallhofer (2007)⁶⁶, jež pro jednotlivé možnosti parametrizace otázky uvádí hodnoty účinku, které na kritéria kvality mají.

⁶⁶ I přeš vyžádání aktuálního stavu meta analýz, které jsou implementovány v SQP, v podobě seznamu regresních koeficientů a jejich odchylek mi zástupce Research and Expertise Centre for Survey Methodology (RECSM), která má tento experiment v rámci ESS na starosti, sdělili, že přes probíhající analýzy tento seznam ještě není k dispozici.

Každý koeficient⁶⁷ pak indikuje míru změny reliability nebo validity v případě, že dojde ke změně nezávislé proměnné o jeden stupeň za předpokladu, že všechny ostatní charakteristiky zůstanou neměnné. Např. proměnná doména se týká tématu výzkumu. Z koeficientů zjistíme, že pokud se naše otázka zabývá tématem národní politiky (proměnné jsou pro účely analýzy charakterizovány jako dummy proměnné, kdy v tomto případě $D_x=1$ pro kategorii národní politika, ostatní varianty se rovnají nule) projeví se to ve zlepšení kvality dotazu konkrétně o 0,052 u reliability a 0,044 u validity. Pokud je naším tématem výzkumu nějaká z obecných otázek týkající se života (obecně), můžeme očekávat negativní efekt (-0,077 reliability, -0,016 validita). Z toho je možné lehce dovozovat, že otázky týkající se národní politiky dosahují např. o $0,052+0,077$ vyšší reliability než otázky, kdy respondent se má vyjadřovat k obecným otázkám života. U parametrů, nejméně ordinální úrovně měření vyjadřuje koeficient míru změny hodnot kvality dotazu při změně o jednu hodnotu nezávislé proměnné. Tak např. u proměnné centralita, která je zakódována do pěti kategorií může dojít k tomu, že rozdíl mezi extrémními kategoriemi představuje např. u reliability rozdíl $5 \times (-0,0172) = -0,087$.


I když se s koeficienty pracuje jako s normálními regresními koeficienty, na které je široká sociologická veřejnost zvyklá, představuje tento přístup velké náklady na lidské zdroje spojené se zakódováním otázek zájmu nebo dokonce celého dotazníku a výpočtem výsledných hodnot kvality dotazu. To vše provedeno opakovaně vzhledem k očekávaným změnám a snaze o případné zlepšení výsledných hodnot.

Druhá možnost je využít program SQP dostupný na internetové adrese sqp.nl. Program vznikl v roce 2000 a v jeho původní verzi obsahoval automatickou a manuální formu vyhodnocení dotazu. Z důvodu, že automatický způsob kódování dotazu (stačilo zapsat pouze znění otázky a program sám analyzoval a zakódoval jeho větnou strukturu do patričních kódů) představoval z mnoha důvodů neúměrnou zátěž (gramatická a sémantická struktura jazyka, rozdílnost gramatik v rámci jazyků atd.), tak v současné verzi programu je k dispozici pouze manuální způsob kódování otázky. V tomto způsobu využití výzkumník nezadává znění dotazu do textového pole, ale odpovídá na sérii otázek (jedná se o kódovací schéma výše), které jeho dotaz charakterizují. (Ukázka konkrétního zakódování otázky viz Příloha 1) Program posléze na základě těchto kódů (které reprezentují charakteristiku dotazu) postupuje dále k odhadu validity, reliability a

⁶⁷ V dalším odstavci využiji hodnoty přetištěné v Saris a Gallhofer 2007. Hodnoty v této konkrétní stati uvedené výše musí být násobeny hodnotou 1000.

celkové míry kvality dotazu.⁶⁸ Pro otázku, kterou máme kompletně zakódovanou (viz další obrázek), můžeme získat odhad její kvality.

Question Coding Complete


Coding Complete!

This question has been completely coded.

Question Quality Prediction

Get a prediction of the quality of this question based on the choices made for each characteristic.

Get Quality Prediction >

Return to the question list.

Go back to your search results for questions.

Back to Question List

Question

test1 | test1 | *Praktičtí lékaři jednají s pacienty jako se sobě rovnými.*

test Czech Republic - Czech

<p style="font-size: 0.8em; color: #4a7c9c;">Request for Answer Text:</p> <p style="font-size: 0.8em; color: #4a7c9c;">Praktičtí lékaři jednají s pacienty jako se sobě rovnými.</p>	<p style="font-size: 0.8em; color: #4a7c9c;">Answer options:</p> <ul style="list-style-type: none"> • Nikdy nebo téměř nikdy 1 • někdy 2 • Tak napůl 3 • Většinou 4 • Vždy nebo téměř vždy 5
--	---

Edit Question

Your Quality Prediction for this Question

View a prediction of the quality of this question based on the choices made for each characteristic.

View Quality Prediction >

Pro otázku, kterou máme zakódovanou (detail v příloze) získáme odhad reliability 0,647, validity 0,846 a hodnotu celkové míry kvality 0,548.⁶⁹ Jak bylo uvedeno výše, výsledky analýzy MTMM experimentů ukazují, že 50 %, resp. 60 % rozptylu reliability a validity vysvětluje právě parametrizace otázky, což indikuje, že získané predikce by měly být dobrými odhady kvality dotazu.⁷⁰ Autoři sami upozorňují na možnou chybovost v odhadech validity a reliability, která může být způsobena jednak nepřesným kódováním a jednak nepřesností modelu a nabádají k prověření výstupů SQP. Ještě předtím než se k tomu dostaneme, tak bych zmínil, jak použít výsledky, které nám program SQP nabízí.

⁶⁸ Ukázku konkrétního vyhodnocení najdeme např. u Saris, van der Velt, Gallhofer (2004).

⁶⁹ Efekt metody není nutné uvádět, protože jej lze dopočítat $(\text{vliv metody})^2 = 1 - (\text{validita})^2$, celková míra kvality = reliability x validita.

Získané výsledky můžeme využít pro zhodnocení kvality dotazu. U uvedeného příkladu, máme odhad reliability, validity a celkovou míru kvality (který se rovná součinu koeficientu reliability a validity). Mocnina této hodnoty uvádí míru rozptylu pozorované proměnné vysvětlitelnou za pomoci proměnné, kterou jsme měli v úmyslu měřit. Zde $(0,548)^2 = 0,3$. To znamená, že 30 % rozptylu měřené proměnné je vysvětleno za pomoci měřeného konstruktů. Zbývajících 70 % rozptylu pochází z chybové složky, systematická chyba z této variability představuje 12 % $(0,647)^2 \times (1 - 0,846)^2$, náhodná chyba pak 58 % $(1 - (0,647)^2)$. To ukazuje, že náhodná chyba měření je značná a kvalita dotazu není vysoká.

Získané údaje můžeme také použít k úpravě hodnot měření. Ze vztahů, které jsou v rámci modelu měření nastavené plyne, že vliv metody měření může ovlivňovat vztah mezi měřenými proměnnými v případě, kdy pro jejich měření používáme stejnou metodu. Představme si situaci, že vědce zajímá vzájemný vztah mezi otázkou na spokojenost s životem a spokojenost s politickým systémem, které jsou ve všech dalších ohledech identické. V takovém případě může korelace mezi proměnnými být uměle posílena právě o vliv metody. Za předpokladu, že obě proměnné mají stejné hodnoty validity a reliability jako v našem příkladě, dojdeme k závěru, že u naměřené hodnoty korelace mezi proměnnými můžeme z důvodu použití stejné metody očekávat zvýšení hodnoty korelace o $0,12 (0,647) \times (0,548) \cdot (0,548) \times (0,647)$. I když se v tomto případě jedná o vliv, který můžeme pokládat za zanedbatelný, přece jen tomu tak nemusí být ve všech případech.

Primárním účelem použití SQP je však pomoc při optimalizaci dotazu. V našem konkrétním případě není kvalita dotazu ideální z důvodu vysoké míry náhodné chyby měření. Z tohoto důvodu by bylo dobré pokusit se zvýšit primárně reliabilitu měření, přičemž validitu bychom rádi udrželi na stejné úrovni (nebo se také můžeme pokusit o její mírné zlepšení). Nabízí se v podstatě dvě možnosti. První je změnit ty charakteristiky dotazu, které mají zásadní negativní vliv na kvalitu měření nebo se pokusit se změnit ty parametry, které i když nyní třeba žádný (negativní) vliv nemají, mohou vést ke zlepšení. Oba přístupy samozřejmě mají své limity, když některé nastavení jsou z principu neměnné (téma výzkumu, metoda sběru, využití tazatele, ...), v jiných se naopak manévrovací prostor pro výzkumníka otevírá. Jak jsme viděli výše, lze se k podobným simulacím dostat i početně, ale po vynaložení nemalého úsilí by nás

⁷⁰ Saris, van der Velt, Gallhofer (2004:294).

čekalo rozhodování o tom, který parametr vylepšit. Navíc metodologická literatura zabývající se touto otázkou je tak rozsáhlá a fragmentovaná, že obtížně bychom hledali oporu i pro to, kterou změnu stojí zato provést a kterou nikoliv.

Program SQP automaticky nabízí možnosti změn dotazu, které by mohly vést k jeho zlepšení.

Quality Prediction

A Authorized Quality Prediction

Prediction Overview [View Quality Coefficients](#)

		Prediction
Reliability	r^2	0.647
Validity	v^2	0.846
Quality	q^2	0.548

Common Method Variance	cmv	0.100
------------------------	------------	--------------

Potential Improvement to Question Quality

View the potential for improving this question based on this quality prediction by changing individual characteristics.

[View Potential Improvements](#)

Po kliknutí na příslušnou ikonu se nám zobrazí seznam parametrů, které bychom mohli využít, a postupně dojde k propočtu změny, jež mohou pro kvalitu dotazu znamenat (číselný údaj za pruhem grafu oznamuje míru zlepšení v parametru celková kvalita měření). Nejdříve systém nabídne vyhodnocení efektu parametrů, které mají obecně nejvyšší vliv na kvalitu měření. Kliknutím si lehce můžeme nechat spočítat vliv i dalších parametrů dotazů na což nás i systém upozorňuje. U specifických otázek totiž mohou hrát roli specifické charakteristiky dotazu, které v tomto rychlém náhledu nemusí být vyhodnoceny.

Potential Improvements		
Top Variables (over all questions)		
The top variables by effect in general have been evaluated. However for this specific question, other variables may also be important.		
Evaluate all remaining variables >		
Variable	Potential Max Quality by Change in Variable	
position	0.602	+ 0.054
ncategories	0.584	+ 0.036
nsyll_total	0.569	+ 0.021
range.correspondence	0.561	+ 0.013
labels	0.556	+ 0.008
stimulus	0.556	+ 0.008
scal_neutral	0.552	+ 0.004
scale_corres	0.548	
fixrefpoints	0.548	
interviewer	0.548	
balance	0.548	
labels_order	0.548	
form_basic	0.548	
gradation	0.548	
showc_horiz	0.548	
scale_basic	0.548	
used_VWH_word	0.548	
visual	0.548	
showc_start	0.548	
avgwrdr_total	0.540	

V uvedeném příkladu vidíme, že nejvyšší potenciál pro zlepšení kvality dotazu mají změny týkající se pořadí otázky v dotazníku a počet odpovědních kategorií. Po kliknutí na ikony jednotlivých parametrů se nám ukáže i předpokládaný efekt, který změny v daném parametru přinesou (číselný údaj za pruhem grafu oznamuje míru zlepšení v parametru celková kvalita měření). V tomto konkrétním případě tak např. zjistíme, že

Selected Variable		
Position (position)		
Current choice: 120		
Choice	Average what if-prediction	
2	0.602	+ 0.054
40	0.546	- 0.003
80	0.540	- 0.008
120	0.548	
160	0.547	- 0.001
300	0.511	- 0.037
240	0.541	- 0.007
280	0.511	- 0.037

Position
The position of the item in the questionnaire. Number of questions asked before this question.

kvalitu dotazu by zlepšilo umístění v dotazníku mezi prvních čtyřicet otázek (nyní 120. otázka dotazníku) dotazníku nebo

Selected Variable		
Number of categories (ncategories)		
Current choice: 5		
Choice	Average what if-prediction	
2	0.574	+ 0.026
4	0.578	+ 0.03
5	0.548	
7	0.551	+ 0.003
11	0.584	+ 0.036

Number of categories
The code is equal to the number of categories of the category scale provided. Do not include the "don't know" category in the count. For instance, if the categories go from 0 to... more >>

kdybychom místo pětibodové škály použily čtyřbodovou nebo jedenáctibodovou. Dále by k mírnému zlepšení vedlo např. vynechání střední možnosti odpovědi atd.

Vidíme tak, že užití programu SQP přináší zajímavé možnosti pro testování. Důležitou okolností, kterou zbývá prověřit je kvalita predikce, kterou získáme k dotazům v českém jazyce.

4.6. Srovnání predikce SQP s výsledky MTMM studií

Samotní autoři SQP uvádějí že, i když lze díky míře vysvětlené variability modelem předpokládat, že predikce budou přesné, tak i přesto doporučují predikční schopnosti modelu dále testovat. Je např. možné, že predikce fungují dobře pouze na určité typy otázek, zatímco u dalších selhávají, stejně jako je možné, že omezená opora pro české odhady je bude činit neupotřebitelnými. Možnou chybovost v odhadech validity a reliability se pokusím odhalit srovnáním výsledků MTMM experimentu v ČR s predikcemi SQP. Zároveň se jako druhý datový zdroj pro externí validizaci pokusím využít hodnoty reliability, které indikoval Řehák a jeho tým ve výzkumech „Výzkum spokojenosti a hodnotových preferencí“ a výzkum „Senátní volby 1996“. Toto je velmi důležité, protože víme, že u některých statistických metod (regrese) přesnost predikce může být lepší u dat, na nichž již probíhala specifikace modelu než u jeho dalších aplikací.

4.7. Srovnání predikcí SQP pro ČR

Pro evaluaci predikčního potenciálu a použitelnosti predikcí programu SQP využiji srovnání výsledků vyhodnocení MTMM experimentů a hodnot predikce programu.

Vždy uvedu nejdříve hodnoty reliability a validity získané z analýzy dat, následně odhady validity a reliability pomocí SQP a vypočtu rozdíl mezi těmito hodnotami. Zároveň pokud se nebude odhad SQP pohybovat v 95 % intervalu spolehlivosti výsledků získaných pomocí MTMM experimentů, bude tato hodnota zvýrazněna červeně.

4.7.1. Srovnání s výsledky ESS 1

	MTMM výsledek			SQP		Rozdíl odhadu validity	Rozdíl odhadu reliability	
	Validita			Reliabilita	Odhad validity			Odhad reliability
	F1	F2	F3					
A8	0,986			0,697	0,954	0,706	0,032	-0,009
A9		0,984		0,65	0,952	0,672	0,032	-0,022
A10			0,984	0,65	0,953	0,67	0,031	-0,02

HF10	0,848			0,801	0,826	0,784	0,022	0,017
HF11		0,805		0,787	0,797	0,765	0,008	0,022
HF12			0,813	0,772	0,799	0,757	0,014	0,015
HF28	0,787			0,609	0,792	0,642	-0,005	-0,033
HF29		0,778		0,71	0,797	0,665	-0,019	0,045
HF30			0,787	0,569	0,797	0,605	-0,01	-0,036
B2	0,997			0,562	0,981	0,661	0,016	-0,099
B3		0,979		0,766	0,958	0,728	0,021	0,038
B4			0,997	0,642	0,979	0,649	0,018	-0,007
HF4	0,992			0,557	0,94	0,652	0,052	-0,095
HF5		0,994		0,619	0,944	0,673	0,05	-0,054
HF6			0,992	0,534	0,939	0,629	0,053	-0,095
HF22	0,973			0,603	0,907	0,671	0,066	-0,068
HF23		0,979		0,665	0,917	0,671	0,062	-0,006
HF24			0,975	0,635	0,909	0,654	0,066	-0,019
B7	0,901			0,906	0,89	0,861	0,011	0,045
B8		0,91		0,856	0,89	0,869	0,02	-0,013
B9			0,915	0,895	0,894	0,882	0,021	0,013
HF13	0,929			0,937	0,836	0,896	0,093	0,041
HF14		0,859		0,965	0,849	0,926	0,01	0,039
HF15			0,936	0,927	0,859	0,914	0,077	0,013
HF31	0,811			0,825	0,855	0,844	-0,044	-0,019
HF32		0,839		0,916	0,854	0,895	-0,015	0,021
HF33			0,837	0,858	0,865	0,873	-0,028	-0,015
B30	0,82			0,782	0,812	0,833	0,008	-0,051
B31		0,855		0,851	0,807	0,831	0,048	0,02
B32			0,857	0,772	0,809	0,814	0,048	-0,042
HF7	0,761			0,667	0,902	0,682	-0,141	-0,015
HF8		0,809		0,728	0,894	0,69	-0,085	0,038
HF9			0,951	0,612	0,905	0,674	0,046	-0,062
HF25	0,813			0,643	0,827	0,768	-0,014	-0,125
HF26		0,836		0,678	0,832	0,764	0,004	-0,086
HF27			0,888	0,836	0,837	0,788	0,051	0,048
B43	0,925			0,638	0,89	0,67	0,035	-0,032
B44		0,957		0,827	0,83	0,803	0,127	0,024
B45			0,934	0,775	0,853	0,74	0,081	0,035
HF16	0,919			0,743	0,908	0,693	0,011	0,05
HF17		0,95		0,839	0,862	0,824	0,088	0,015
HF18			0,915	0,77	0,867	0,784	0,048	-0,014

HF34	0,951			0,477	0,923	0,668	0,028	-0,191
HF35		0,981		0,841	0,902	0,838	0,079	0,003
HF36			0,971	0,753	0,903	0,759	0,068	-0,006

Z výsledků je patrné, že přesnější odhady SQP produkuje u hodnot reliability (odhady reliability byly ve 14 případech mimo 95 % interval spolehlivosti, u validity to bylo dokonce 24krát ze 45 možností). Nejvyšší rozdíl mezi výsledkem a predikcí je u validity 0,141, u reliability dokonce 0,191. Průměr odchylek ukazuje, že zatímco hodnoty validity predikce spíše nadhodnocuje (průměr 0,026), tak hodnoty reliability jsou spíše mírně podhodnocené (průměr -0,015). Průměr z absolutních hodnot rozdílů ukazuje, že bez ohledu na pod nebo nadhodnocení výsledků je průměrná odchylka 0,042 u validity a 0,039 u reliability. Nehledě na to, že máme v případě reliability přibližně 31 % šanci $((14/45)*100)$ a u validity 53 % šanci $((25/45)*100)$, že se odhad SQP nebude pohybovat v rámci 95 % intervalu spolehlivosti, průměrně se jedná o odchylky velmi nízké.

4.7.2. Srovnání s výsledky ESS2

	MTMM výsledek			SQP		Rozdíl odhadu validity	Rozdíl odhadu reliability	
	Validita			Reliabilita	Odhad validity			Odhad reliability
	F1	F2	F3					
B4	0,916			0,897	0,909	0,838	0,007	0,059
B5		0,925		0,852	0,93	0,86	-0,005	-0,008
B7			0,905	0,841	0,919	0,853	-0,014	-0,012
I25	0,857			0,92	0,822	0,865	0,035	0,055
I26		0,87		0,903	0,819	0,882	0,051	0,021
I27			0,84	0,901	0,82	0,861	0,02	0,04
I38	0,854			0,766	0,822	0,852	0,032	-0,086
I39		0,731		0,918	0,807	0,883	-0,076	0,035
I40			0,733	0,762	0,813	0,853	-0,08	-0,091
B25	0,823			0,759	0,873	0,831	-0,05	-0,072
B26		0,849		0,826	0,871	0,826	-0,022	0
B27			0,927	0,782	0,888	0,808	0,039	-0,026

I11	0,766			0,842	0,816	0,837	-0,05	0,005
I12		0,784		0,901	0,813	0,851	-0,029	0,05
I13			0,791	0,787	0,822	0,813	-0,031	-0,026
I35	0,852			0,836	0,839	0,839	0,013	-0,003
I36		0,863		0,879	0,835	0,847	0,028	0,032
I37			0,938	0,796	0,847	0,818	0,091	-0,022
D25	0,74			0,335	0,784	0,403	-0,044	-0,068
D26		0,912		0,723	0,846	0,647	0,066	0,076
D27			0,902	0,674	0,828	0,62	0,074	0,054
I5								
I6								
I7								
I28	0,956			0,555	0,88	0,495	0,076	0,06
I29		0,958		0,49	0,876	0,529	0,082	-0,039
I30			0,943	0,389	0,862	0,475	0,081	-0,086
G6	0,962			0,619	0,829	0,702	0,133	-0,083
G7		0,976		0,249	0,833	0,708	0,143	-0,459
G8			0,972	0,629	0,861	0,654	0,111	-0,025
I8								
I9								
I10			0,907	0,421	0,854	0,533	0,053	-0,112
I22	0,908			0,648	0,838	0,742	0,07	-0,094
I23		0,889		0,633	0,831	0,732	0,058	-0,099
I24			0,93	0,734	0,893	0,713	0,037	0,021
G64	0,98			0,726	0,981	0,696	-0,001	0,03
G66		0,981		0,611	0,982	0,65	-0,001	-0,039
G70			0,976	0,534	0,981	0,635	-0,005	-0,101
I19	0,993			0,734	0,991	0,75	0,002	-0,016
I20		0,993		0,857	0,991	0,807	0,002	0,05
I21			0,993	0,751	0,991	0,772	0,002	-0,021
I32	0,89			0,743	0,918	0,752	-0,028	-0,009
I33		0,912		0,739	0,918	0,754	-0,006	-0,015
I34			0,912	0,88	0,929	0,831	-0,017	0,049

Při vyhodnocení odhadů této vlny, SQP produkuje přesnější hodnoty pro validitu než pro reliabilitu (odhady reliability byly v 19 případech mimo 95 % interval spolehlivosti, u validity to bylo nyní pouze 14krát ze 40 možností). Nejvyšší rozdíl mezi výsledkem a predikcí je u validity 0,143, u reliability dokonce 0,459. Pokud pomínu tuto extrémní hodnotu, která naznačuje, že pro některé typy dotazu SQP nemusí být vhodným

nástrojem, druhý nejvyšší rozdíl hodnot je 0,112. Průměr odchylek opětovně ukazuje, že zatímco hodnoty validity predikce spíše nadhodnocuje (průměr 0,021), tak hodnoty reliability jsou spíše mírně podhodnocené (průměr -0,024). Průměr z absolutních hodnot rozdílů ukazuje, že bez ohledu na podhodnocení nebo nadhodnocení výsledků je průměrná odchylka 0,044 u validity (v ESS 1 hodnota 0,042) a 0,15 u reliability (v ESS 1 hodnota 0,039).

Vidíme tak, že odhady vypočtené programem SQP se v případě validity velmi přibližovaly skutečným hodnotám, predikce reliability byly lepší v první vlně. V ESS 2 došlo u predikovaných hodnot reliability k výraznému poklesu průměrné přesnosti.

Pro další validizaci a ověření fungování SQP jsem se pokusil najít další česká data, u nichž by byly k dispozici odhady kvality měření. Publikace kvality měření není častá (už pro způsob vyhodnocení) a z tohoto důvodu jsem se nakonec rozhodl využít údajů z výzkumu týmu Řeháka. I když se pro záměr práce nejednalo o data úplně vhodná, vzhledem k dostupnosti hodnot reliability nebo validity v české literatuře jsem se rozhodl zmíněný datový zdroj využít.

Primárním cílem autorského týmu bylo testování reliability. Odhady validity nebyly provedeny a autoři pro to měli objektivní důvody (nemožnost variace otázek,...). Bohužel pro naše účely nejsou v textu publikovány ani přesné znění otázek, které byly respondentům položeny, resp. celý dotazník (kontext), ve kterém se otázky nacházely. Z tohoto důvodu bude větší či menší část parametrů a způsob zakódování otázky pouze odhadnut (detailní údaje o zakódování budou uvedené v příloze).

První otázkou, u které byla vyhodnocena reliability, byla otázka na politické postoje. Autoři ji charakterizují jako subjektivní sebezařazování: dotaz na politickou pozici (sedmibodová škála, 1 extrémní pravice, 7 extrémní levice). Hodnota reliability u tohoto dotazu dosahovala průměrné úrovně 0,77 (0,74 v souboru A, 0,8 v souboru B). Pokud se na základě těchto informací pokusíme získat odhad kvality dotazu na základě SQP získáme hodnotu 0,711. Znovu upozorňuji, že se jedná pouze o hrubý odhad, kdy postrádáme některé informace pro vstup a vyhodnocení SQP. Vidíme však, že zde se odhad poměrně přibližuje získané kvalitě dotazu (rozdíl 0,059).

Dále byla testována reliability u dvoupólové preferenční škály opět zaměřené na politické téma. Respondentům byla položena otázka *Jaké vlastnosti má mít Váš senátor? Označte, čemu dáváte přednost!* Jako odpovědní škála byla využita u baterie pětibodová škála, bipolární s plně popsanými variantami, specifickými pro měřený koncept. Hodnoty reliability zde dosahovaly průměrně hodnoty 0,457 při rozpětí 0,066.

SPQ predikce u tohoto typu otázky predikovala hodnotu 0,699. Vidíme, že v tomto případě SQP silně nadhodnocuje získané výsledky. Musím ale nezauatě přiznat, že míra objektivní přesnosti zadání parametrů dotazu se oproti předchozímu příkladu snížila. Mám zde na mysli zejména její zařazení do baterie dotazů.

Pro třetí pokus jsem vybral otázku na spokojenost s politickým vývojem. Zde není v textu práce uvedeno ani znění dotazu, takže vše, kromě konceptu a použité škály je ponecháno na libovůli (opětovně se snížila přesnost zadání parametrů dotazu). Pro zadané kódování jsme z SQP získali odhad 0,810, zatímco analýza dat ukázala hodnotu 0,681. V tomto případě můžeme konstatovat, že se jedná o výsledek poměrně uspokojivý, když s rozdílem v odhadu na úrovni 0,129 (0,810-0,681) jsme se mohli setkat i při predikci dotazů ESS.

Hlavním omezením u tohoto způsobu vyhodnocení a srovnání s výsledky studie je, že nejsme schopni s dostatečnou mírou přesnosti rozlišit mezi chybou vzniklou na základě rozdílného zadání kódování (z důvodu nedostatku informací o použitých dotazech) a chybou predikce SQP. Vidíme, že u prvního dotazu pro který byly naše informace nejvíce ucelené a přesné, a kde jsme mohli vyjít z přesného znění otázky atd., byl odhad poměrně přesný. Pro další situaci bylo k dispozici méně údajů a jednalo se svým způsobem o značně benevolentní zadání. Navíc ve spojení s užitím dotazu v širší baterii dotazů se predikce ukázala značně rozdílnou. V posledním případě, kdy bylo zadání opět značně benevolentní, získané výsledky poměrně odpovídaly hodnotám predikovaným.

4.8. Vyhodnocení použitelnosti SQP

Program SPQ jde využít jednak jako zdroj informací o realizovaných MTMM studiích a odhadech kvality použitých dotazů, jednak nabízí cestu, jak snadno tyto nashromážděné znalosti týkající se efektu jednotlivých charakteristik dotazu na jeho kvalitu využít. SPQ tak poskytuje možnost otestovat kvalitu dotazu ještě předtím, než se rozhodneme jej využít pro terénní sběr. Představuje tak zajímavou možnost pretestování instrumentů výběrových šetření. Evaluace kvality dotazu je navíc doplněna o nabídku alternativních návrhů, které mohou vést k zlepšení jeho parametrů.

Možnost aplikace SQP je tak novým nástrojem, který se výzkumníkům nabízí. „Nový“ i z toho důvodu, že zatímco dříve byly sociologické znalosti týkající se kvality dotazu k dispozici nejen v omezené míře, ale jejich další a hlavní nevýhodou byla roztroušenost

po nejrůznějších odborných knihách a periodikách. SQP tak sjednocuje dosavadní znalosti a vytváří tak důležité předpoklady nejen pro využití a vytěžení těchto výsledků, ale i výzkumnou práci jednotlivých badatelů.

Srovnání predikce SQP a výsledků MTMM experimentů ukázalo možnou využitelnost tohoto nástroje i v českém prostředí, když průměrná odchylka u validity měla hodnotu 0,043, u reliability pak 0,09. Vyhodnocení vhodnosti predikčního modelu pro další data bylo nepřesvědčivé. Kvůli nemožnosti přesného zadání dotazu do kódovacího schématu SQP nelze přesvědčivě odlišit chybu predikce od chyby zadání. Extenzivnější test predikcí SQP v českém prostředí se mi zdá nutný pro definitivní zhodnocení aplikovatelnosti SQP. Prozatímně, ze současného úhlu pohledu a na základě dosavadních zjištění, se aplikace SQP ukazuje jako slibnou. Je třeba vzít v úvahu, že SQP je ve své současné verzi (2.0) v provozu teprve několik měsíců a poprvé umí nabídnout pomoc i českým vědcům (zkrátka a jednoduše umí i česky).

Je však důležité si uvědomit, že tento přístup má i své limity. Omezené jsou např. možnosti predikce, co se týče témat výzkumu, užití některých parametrů otázek (např. otevřené otázky) nebo metod sběru dat atd. Tento systém samozřejmě není uzavřený a je třeba mít na paměti, že se jedná o stálé a „kontinuální provizorium“, které bude postupně zdokonalováno. Zároveň se jedná o nejlepší provizorium, které máme v současnosti k dispozici.

Postupně lze navíc očekávat, že analýza dalších experimentů a postupná integrace těchto výsledků do predikčního modelu dále povede ke zpřesňování predikcí kvality dotazu (a rekalkulaci současných hodnot vlivu parametrů na kvalitu dotazu).

Uplatnění nástroje pro účely pretestování tak bude postupně ještě přesnější a snad i pro vědce aktuální.

5. Závěr

Po dlouhou dobu byla v testování instrumentů výběrových šetření dominantní tradiční (někdy také konvenční) metoda debriefingu. Narůstající diskuse o nedostacích tradičního přístupu a možnostech zlepšení vedla k posunu zájmu k širšímu pojetí kvality měření a k etablování dalších metod testování. První část práce se pokusila prozkoumat a zhodnotit možnosti a komplementaritu etablovaného souboru metod pretestování.

Na základě zjištění opírajících se o metodologické práce můžeme shrnout, že metody pretestování jsou explanatorně silné, když dokáží dobře identifikovat problémy dotazníku (jednotlivé metody jsou silné v měření rozdílných typů problémů).

Další práce s těmito výsledky je ale obtížná a vede jen k neurčitým výsledkům. Jednou z možných příčin je i to, že jsou využívány převážně metody, které jsou neformální, subjektivní a často neposkytují průkazné ani konzistentní výsledky. Výstupy takového testování mají pak kvalitativní povahu, kterou je obtížné zpracovat a zjednodušit do podoby, aby byly jakkoliv využitelné. Výzkumníci se tak při aplikaci metod musí vypořádat se skutečností, že nalezení problému v průběhu testování ještě neznamená odhalení jeho významu pro terénní šetření. Výsledek pretestování často neposkytuje informace o tom, jak problém napravit nebo zda bude upravená verze lepší než verze původní. Právě systematická konfrontace původního a revidovaného znění dotazníku se ukazuje pro vyhodnocení efektivity pretestování jako klíčovou.

Snahou práce bylo představit možnosti alternativního přístupu. Zatímco první přístup se zabývá možnostmi redukce chyb a otázkami, jako je pochopení otázek a použitých pojmů, znění otázek, zátěž respondenta, druhý se zabývá tím, jak se tyto problémy projevují kvantitativně, tzn., zabývají se odhadem chyby měření. Možnost vyhodnocení indikátorů kvality ovšem přináší jisté nároky na realizaci a design šetření. Jejich největší překážkou je tak právě zejména obtížnost realizace v praxi.

Pro kontext práce je důležitá možnost využít takto získaných poznatků jako bázi pro etablování specifického způsobu pretestování, který se práce pokusila představit. Ta je opřena o závěry meta analýz, jejichž smyslem bylo stanovit míru vlivu designu dotazu na kvalitu dotazu (validitu a reliabilitu). Analýzy potvrdily, že míra rozptylu odhadů kvality je vysvětlitelná charakteristikou dotazu, a tedy i s určitou mírou jistoty predikovatelná. Samotné výsledky analýz jsou pak velmi užitečným materiálem, který výzkumníkům osvětluje vliv použitých metod a dalších prvků, které jsou využívány při konstrukci dotazníku.

System pro odhad kvality dotazu opřený o takto získané výsledky je nyní implementován v programu SQP. Samotní autoři SQP uvádějí, že i když lze díky míře vysvětlené variability modelem předpokládat, že predikce budou přesné, tak i přesto doporučují predikční schopnosti modelu dále testovat. Toto se stalo cílem další části práce, která se zabývala možnostmi, kterou tento program českým vědcům přináší. Zaměřil jsem se na chybovost v odhadech validity a reliability za pomoci srovnání výsledků MTMM experimentu v ČR s predikcemi SQP. Srovnání predikce a MTMM dat ukázalo možnou využitelnost tohoto nástroje i v českém prostředí. Vyhodnocení vhodnosti predikčního modelu pro další datový zdroj, ale bylo nepřesvědčivé. Při aplikaci tohoto přístupu se však ukázaly i konkrétní omezení SQP jako je tematické zaměření, otevřená otázka atd.

Extenzivnější test predikcí SQP v českém prostředí se mi zdá nutný pro definitivní zhodnocení aplikovatelnosti SQP. Je tak stále např. možné, aby se ukázalo, že predikce fungují dobře pouze pro určité typy otázek, zatímco u dalších (v českém kontextu) selhávají. Prozatímně, ze současného úhlu pohledu a na základě dosavadních zjištění, se aplikace SQP ukazuje jako nadějnou.

6. Summary

Campbell a Fiske navrhli a Andrews poprvé využil MTMM design nejen k odhadům kvality dat (validity a reliability), ale i následné analýze, která se zabývala vlivem charakteristiky dotazu na získané hodnoty kvality a možnost jejich predikce. Po skromných začátcích, zapříčiněných zejména obtížností realizace MTMM designu, jsou v současné době tyto experimenty zahrnuty do šetření Evropského sociálního výzkumu. Můžeme tak říci, že realizace MTMM experimentů a jeho možné přínosy zcela jistě nestojí mimo pozornost odborné sociologické veřejnosti. MTMM experimenty jsou díky tomu realizovány na široké celoevropské platformě a takto získané údaje jsou využívány pro následné analýzy zabývající se vlivem charakteru dotazu na jeho kvalitu. Konkrétním vyjádřením této snahy o metodologické studium sociologického instrumentaria a využití takto získaných znalostí je i program SQP. Srovnání predikce SQP a výsledků MTMM experimentů v ČR ukázalo možnou využitelnost tohoto nástroje i v českém prostředí. Rozsáhlejší testování predikcí SQP v českém prostředí je však nutné pro definitivní zhodnocení aplikovatelnosti SQP. Prozatímně, ze současného úhlu pohledu a na základě dosavadních zjištění se aplikace SQP ukazuje jako slibnou.

7. Použitá literatura

ALWIN, D. F. 1974. Approaches to the interpretation of realtionships in the multi-trait-multimethod matrix. in Costner, H. ed. Sociological Methodology 1973-1974.

ANDREWS, F. M. 1984. Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach: The Public Opinion Quarterly, Vol. 48, No. 2, 409-442.

BABBIE, E. 2001. The Practise of Social Research. Belmont: Wadsworth.

BEATTY, P. The Dynamics of Cognitive Interviewing, in Presser et. all. Methods for Testing and Evaluating Survey Questionnaires, Wiley 2004, 45-67.

BELSON, W. A. 1983. The design and understanding of survey questions, Lexington Books.

BIEMER, P. P., LYBERG, L. 2003. Introduction to Survey Quality, John Wiley. NJ.

BOLTON, R. M., BRONHORST, T. M.. Questionnaire Pretesting, Computer-Assisted Cosiny of Concurrent Protocols, in Schwarz, Sudman: Answering Qestions, Josey-Bass, 1996, 1996:36-64.

CAMPBELL, D. T., FISKE, D. W. 1959. Convergent and discriminant validation by the multimethod-multitrait matrix. Psychological Bulletin 56:81-105.

CONVERSE, J., PRESSER, S. 1986 Survey questions: Handcrafting The Standardized Questionnaire. Newbury Park, London, New Delhi : Sage Publications.

COX, E. (1980). The optimal number of response alternatives for a scale: A review. Journal of Marketing research, 17, 407- 422.

DeMAIO, T. J., LANDRETH, A. Do Different Cognitive Interview Techniques Produce Different Results?, in Presser et. all. *Methods for Trstiny and Evaluating Survey Questionnaires*, Wiley 2004, 89-109.

DISMAN, M. 2002. *Jak se vyrábí sociologická znalost*, Praha, Carolinum.

FINK, A. 2003. *The Survey Handbook*. Newbury Park, London, New Delhi: Sage Publications.

FORSYTH, B. H., LESSLER, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. Biemer, R., Groves, L. Lyberg, N. Mathiowetz, and S. Sudman, (Eds.), *Measurement error in surveys*. New York: Wiley.

FORSYTH, B. H., ROTHGEB, J. M., WILLIS, G. B. Does Pretesting Make a Diference, An Experimental Test, in Presser et. all. *Methods for Trstiny and Evaluating Survey Questionnaires*, Wiley 2004, 525-547.

FOWLER, F. J., CANNELL, CH. F. Using Behavioral Cosiny to Identify Cognitive Problems with Survey Questions, in Schwarz, Sudman: *Answering Qestions*, Josey-Bass, 1996,1996:15-36.

FOWLER, F. J. The Case for More Split-Sample Experiments in Developing Survey instruments, in Presser et. all. *Methods for Trstiny and Evaluating Survey Questionnaires*, Wiley 2004, 173-189.

FOWLER, F. J. 1995. *Improving Survey Questions Design and Evaluation*. Sage Publication.

GRAESSER, A. C. et all. Integrating Questionnaire Design with a Cognitive Computation Model of Human Question Answerin, in Schwarz, Sudman: *Answering Qestions*, Josey-Bass, 1996, 143-174.

GROVES et all. (2004). *Survey Methodology*, Wiley, Hoboken, New Jersey.

GROVES et all. (1989). Survey Errors and Survey Costs, Wiley, Hoboken, New Jersey.

KREJČÍ, J. 2009. Kvalita sociálněvědních výběrových šetření v České republice. Praha, Slon.

LESSLER, J. T., KALSBECK, W. D. Nonsampling Error in Surveys, John Wiley&Sons, New York 1992.

LESSLER, J. T., FORSYTH, B. H. A Coding System for Appraising Questionnaire, in Schwarz, Sudman: Answering Questions, Josey-Bass, 1996, 259-292.

OPPENHEIM, A. N. 2001 Questionnaire Design, Interviewing and Attitude Measurement. Continuum: London & N. Y.

PRESSER, S., BLAIR, J. Survey pretesting: Do different methods produce different results? in Marsden, P. ed: Sociology Methodology 24, 73-104.

ŘEHÁK, J. 1998a. „Kvalita dat I. Klasický model měření reliability a jeho praktický aplikační význam.“ Sociologický časopis 34: 51-60.

ŘEHÁK, J. 1998b. „Kvalita dat II. Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur.“ Sociologický časopis 34: 195-204.

ŘEHÁK, J, BÁRTOVÁ, I., HAMANOVÁ, J. Kvalita dat III., Empirické výsledky měření reliability pro vybrané míry a stupnice. Sociologický časopis, 1998, 34:363-372.

ROTHGEB, J. M., WILLIS, G. B, FORSYTH, B. H. Questionnaire Pretesting Methods: Do Different Techniques and Different Organizations Produce Similar Results? Annual Conference of American Association for Public Opinion Research Montreal, May 2001.

SARIS, W. E. 1990. „Models for evaluation of measurement instruments.“ Pp. 52-80 in Evaluation of measurement instruments by meta-analysis of multitrait multimethod studies, ed. by W. E. Saris and A. van Meurs. Amsterdam: North-Holland.

SARIS, W. E. 1990. „The Choice of a model for evaluation of measurement instruments.“ Pp. 118-129 in Evaluation of measurement instruments by meta-analysis of multitrait multimethod studies, ed. by W. E. Saris and A. van Meurs. Amsterdam: North-Holland.

SARIS, W. E., MEURS, A. 1990. „Memory effects in MTMM studies.“ Pp. 134-146 in Evaluation of measurement instruments by meta-analysis of multitrait multimethod studies, ed. by W. E. Saris and A. van Meurs. Amsterdam: North-Holland.

SARIS, W. E., MEURS, A. ed. Evaluation of Measurement Instruments by Meta-analysis of Multitrait-Multimethod Studies, Amsterdam, New Holland 1990.

SARIS, W. E., SATORRA A., COENDERS, G. (2004). A New Approach for Evaluating the Quality of Measurement Instruments: Split Ballot MTMM Design, Sociological Methodology, 34 pp. 311-347.

SARIS, W. E., van der VELD, W., GALLHOFER, I. Development and Improvement of Questionnaires Using Predictions of Reliability and Validity, in Presser et. all. Methods for Testing and Evaluating Survey Questionnaires, Wiley 2004, 275-2299.

SARIS, W. E., GALLHOFER, I. Design, evaluation, and analysis of questionnaires for survey research. Hoboken, N.J. : Wiley-Interscience, 2007.

SARIS, W. E, GALLHOFER, I. Estimation of the effects of measurement characteristics on the duality of survey questions Survey Research Methods (2007), Vol. 1, No. 1, 29-43, European Survey Research Association.

SARIS, W., E., GALLHOFER, I. 2011. The results of the MTMM experiments in round 2 (of the European Social Survey). RECSM Working Paper 23.

SCHERPENZEEL, A. C, SARIS, W. E. 1997. The Validity and Reliability of Survey Questions, A Meta-Analysis of MTMM Studie, Sociological Methods&Research 1997, vol 25, No. 3, 341-379.

SCHUMAN, H., PRESSER, S. 1996. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context, Sage.

SCHWARTZ, N., SUDMAN, S. (eds.). 1996. Answering questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass Publisher.

TOURANGEA, R. Experimental Design Considerations for Testing and Evaluating Questionnaires, in Presser et. al. Methods for Testing and Evaluating Survey Questionnaires, Wiley 2004, 209-225.

TOURANGEA, R., RIPS, L., J., RASINSKI, K. The Psychology of Survey Response, Cambridge University Press, 2000.

De VAUS, D. A. 1991. Surveys in Social Research. London: UCL Press, Allen & Unwin.

VINOPAL, J. 2008. Kognitivní přístupy v metodologii výzkumných šetření: Metoda okamžité validizace. Praha: Sociologický ústav AV ČR, v.v.i.

Willis, G. B. (2001). A Comparison of Survey Pretesting Methods: What do Cognitive Interviewing, Expert Review and Behavior Coding Tell Us?

ZOUWEN, J. (2000). An assessment of the difficulty of questions used in the issp questionnaires, the clarity of their wording and the comparability of the responses.

<http://www.europeansocialsurvey.org>

<http://www.sqp.nl>

8. Seznam příloh

Příloha č. 1: Ukázka kodování pomoci SQP

Příloha č. 2: Kodování otázky na subjektivní sebezařazování-dotaz na politickou pozici

Příloha č. 3: Kodování otázky na subjektivní sebezařazování-dotaz na politickou pozici

Příloha č. 4: Kodování otázky na spokojenost s politickým vývojem

9. Přílohy

Příloha č. 1: Ukázka kodování pomoci SQP

D26 / dctreq1 / Regular general practitioner/doctor treat patients
ESS Round 2 Czech Republic - Czech

Request for Answer Text:

Praktičtí lékaři jednají s pacienty jako se sobě rovnými.

Answer options:

- Nikdy nebo téměř nikdy 1
- někdy 2
- Tak napůl 3
- Většinou 4
- Vždy nebo téměř vždy 5

Characteristic	Choice	Code
Domain	Health	3
Domain: health	Doctor's treatment	9
Concept	All other simple concepts	7
Concept: other simple concepts	Evaluation	3
Social Desirability	A bit	1
Centrality	A bit central	1
Reference period	Present	2
Formulation of the request for an answer: basic choice	No request present (e.g. not the first item of battery)	0
Use of stimulus or statement in the request	Stimulus or statement is present	1
Absolute or comparative judgment	An absolute judgement	0
Response scale: basic choice	Categories	0
Number of categories	5	5
Labels of categories	Fully labelled	3
Labels with long or short text	Short text	0
Order of the labels	First label negative or not applicable	1
Correspondence between labels and numbers of the scale	Medium correspondence	2

Characteristic	Choice	Code
Theoretical range of the scale bipolar/unipolar	Theoretically unipolar	0
Number of fixed reference points	2	2
Don't know option	DK option only registered	2
Interviewer instruction	Absent	0
Respondent instruction	Absent	0
Extra motivation, info or definition available?	Absent	0
Introduction available?	Not available	0
Number of sentences in the request	1	1
Number of words in request	9	9
Total number of nouns in request for an answer	2	2
Total number of abstract nouns in request for an answer	0	0
Total number of syllables in request	21	21
Number of subordinate clauses in request	0	0
Number of syllables in answer scale	22	22
Total number of nouns in answer scale	0	0
Total number of abstract nouns in answer scale	0	0
Show card used	Showcard used	1
Horizontal or vertical scale	Vertical	0
Overlap of text and categories?	Text clearly connected to category	1
Numbers or letters before the answer categories	Numbers	1
Scale with numbers or numbers in boxes	Only numbers	0
Start of the response sentence on the showcard	No	0
Question on the showcard	No	0
Picture on the card provided?	No picture provided	0
Computer assisted	No	0
Interviewer	Yes	1
Visual presentation	Visual	1
Position	120	120

Příloha č. 2: Kódování otázky na subjektivní sebezařazování-dotaz na politickou pozici

Pro znění otázky jsem využil často uváděnou verzi dotazu ve formě: „*V politice lidé někdy hovoří o levici a pravici. Kam byste se sám zařadil na této stupnici?*“ Odpovědní škála : 1extrémní levice, 7extrémní pravice

Detailní zakódování:

Domain	National politics	1
Domain: national politics	Political parties	5
Concept	Evaluative belief	1
Social Desirability	A bit	1
Centrality	Rather central	2
Reference period	Present	2
Formulation of the request for an answer: basic choice	Direct request	2
WH word used in the request	WH word used	1
'WH' word	Where PLACE	5
Request for an answer type	Interrogative	1
Use of gradation	No gradation used	0
Balance of the request	Balanced or not applicable	0
Presence of encouragement to answer	No particular encouragement present	0
Emphasis on subjective opinion in request	No emphasis on opinion present	0
Information about the opinion of other people	No information about opinions of others	0
Use of stimulus or statement in the request	No stimulus or statement	0
Absolute or comparative judgment	An absolute judgement	0
Response scale: basic choice Categories		0
Number of categories	7	7
Labels of categories	Partially labelled	2
Labels with long or short text	Short text	0
Order of the labels	First label negative or not applicable	1
Correspondence between labels and numbers of the scale	Medium correspondence	

Theoretical range of the scale bipolar/unipolar			Theoretically bipolar	1
Range of the used scale bipolar/unipolar	Bipolar			1
Symmetry of response scale	Symmetric			1
Neutral category	Present			1
Number of fixed reference points	2			2
Don't know option	DK option not present			3
Interviewer instruction	Absent			0
Respondent instruction	Present			1
Extra motivation, info or definition available?		Absent		0
Introduction available?	Available			1
Number of sentences in introduction		1		1
Number of words in introduction	5			5
Number of subordinated clauses in introduction		0		0
Request present in the introduction	Request not present			0
Number of sentences in the request		2		2
Number of words in request	17			17
Total number of nouns in request for an answer		4		4
Total number of abstract nouns in request for an answer		3		3
Total number of syllables in request	34			34
Number of subordinate clauses in request		0		0
Number of syllables in answer scale	17			17
Total number of nouns in answer scale		2		2
Total number of abstract nouns in answer scale		2		2
Show card used	Showcard used			1
Horizontal or vertical scale	Horizontal			1
Overlap of text and categories?		Text clearly connected to category		1
Numbers or letters before the answer categories		Numbers		1
Scale with numbers or numbers in boxes		Only numbers		0
Start of the response sentence on the showcard		No		0
Question on the showcard	No			0
Picture on the card provided?		No picture provided		0
Computer assisted	No			0
Interviewer	Yes			1
Visual presentation	Visual			1

Position 50 50

Příloha č. 3: Kódování otázky na subjektivní sebezařazování-dotaz na politickou pozici

Pro znění otázky jsem využil často uváděnou verzi dotazu ve formě: „*Jaké vlastnosti má mít Váš senátor? Označte, čemu dáváte přednost!?*“ Odpovědní škála : pět položek s pětibodovou škálou reprezentovalo dvojice pólů: (1) chová se nezávisle-prosazuje zájmy své strany, (2) zkušený politik-nezávislý odborník, (3) přizpůsobivý-zásadový, (4) klade důraz na místní problémy-klade důraz na celostátní problémy, (5) prosazuje rovnost všech lidí-prosazuje svobodu pro každého jednotlivce.

Detailní zakódování:

Domain	National politics	1
Domain: national politics	Prominent persons (ministers, members of parliament, etc.)	19
Concept	Complex concepts	6
Concept: complex concept	Other	3
Social Desirability	Not present	0
Centrality	Not at all central/salient	0
Reference period	Present	2
Formulation of the request for an answer: basic choice	Direct request	2
WH word used in the request	WH word used	1
'WH' word	Which	2
Request for an answer type	Interrogative	1
Use of gradation	No gradation used	0
Balance of the request	Balanced or not applicable	0
Presence of encouragement to answer	No particular encouragement present	0
Emphasis on subjective opinion in request	No emphasis on opinion present	0
Information about the opinion of other people	No information about opinions of others	0
Use of stimulus or statement in the request	No stimulus or statement	0

Absolute or comparative judgment	An absolute judgement	0
Response scale: basic choice Categories		0
Number of categories	5 5	
Labels of categories	Fully labelled	3
Labels with long or short text	Complete sentences	1
Order of the labels	First label negative or not applicable	1
Correspondence between labels and numbers of the scale	Not applicable	0
Theoretical range of the scale bipolar/unipolar	Theoretically bipolar	1
Range of the used scale bipolar/unipolar	Bipolar	1
Symmetry of response scale	Symmetric	1
Neutral category	Not present	3
Number of fixed reference points	5 5	
Don't know option	DK option present	1
Interviewer instruction	Present	1
Respondent instruction	Present	1
Extra motivation, info or definition available?	Absent	0
Introduction available?	Available	1
Number of sentences in introduction	0 0	
Number of words in introduction	0 0	
Number of subordinated clauses in introduction	0 0	
Request present in the introduction	Request not present	0
Number of sentences in the request	2 2	
Number of words in request	10 10	
Total number of nouns in request for an answer	2 2	
Total number of abstract nouns in request for an answer	1 1	
Total number of syllables in request	20 20	
Number of subordinate clauses in request	0 0	
Number of syllables in answer scale	85 85	
Total number of nouns in answer scale	9 9	
Total number of abstract nouns in answer scale	5 5	
Show card used	Showcard used	1
Horizontal or vertical scale	Horizontal	1
Overlap of text and categories?	Text clearly connected to category	1
Numbers or letters before the answer categories	Numbers	1

Scale with numbers or numbers in boxes	Only numbers	0
Start of the response sentence on the showcard	No	0
Question on the showcard	No	0
Picture on the card provided?	No picture provided	0
Computer assisted	No	0
Interviewer	Yes	1
Visual presentation	Visual	1
Position	50	50

Příloha č. 4: Kódování otázky na spokojenost s politickým vývojem

Do jaké míry jste spokojen s politickým vývojem. Škála 1-9, 1 zcela nespokojen, 9 zcela spokojen

Domain	National politics	1
Domain: national politics	National institutions (ministries, parliament, etc.)	3
Concept	All other simple concepts	7
Concept: other simple concepts	Policy	6
Social Desirability	A bit	1
Centrality	Rather central	2
Reference period	Present	2
Formulation of the request for an answer: basic choice	Direct request	2
WH word used in the request	WH word used	1
'WH' word	What	3
Request for an answer type	Interrogative	1
Use of gradation	Gradation used	1
Balance of the request	Balanced or not applicable	0
Presence of encouragement to answer	No particular encouragement present	0
Emphasis on subjective opinion in request	No emphasis on opinion present	0
Information about the opinion of other people	No information about opinions of others	0
Use of stimulus or statement in the request	No stimulus or statement	0
Absolute or comparative judgment	A comparative judgement	1
Response scale: basic choice Categories		0

Number of categories	9	9		
Labels of categories	Partially labelled	2		
Labels with long or short text	Short text	0		
Order of the labels	First label negative or not applicable	1		
Correspondence between labels and numbers of the scale	Medium correspondence	2		
Theoretical range of the scale bipolar/unipolar	Theoretically bipolar	1		
Range of the used scale bipolar/unipolar	Bipolar	1		
Symmetry of response scale	Symmetric	1		
Neutral category	Present	1		
Number of fixed reference points	2	2		
Don't know option	DK option only registered	2		
Interviewer instruction	Present	1		
Respondent instruction	Absent	0		
Extra motivation, info or definition available?	Absent	0		
Introduction available?	Not available	0		
Number of sentences in the request	1	1		
Number of words in request	8	8		
Total number of nouns in request for an answer	2	2		
Total number of abstract nouns in request for an answer	1	1		
Total number of syllables in request	17	17		
Number of subordinate clauses in request	0	0		
Number of syllables in answer scale	20	20		
Total number of nouns in answer scale	1	1		
Total number of abstract nouns in answer scale	1	1		
Show card used	Showcard used	1		
Horizontal or vertical scale	Horizontal	1		
Overlap of text and categories?	Text clearly connected to category	1		
Numbers or letters before the answer categories	Numbers	1		
Scale with numbers or numbers in boxes	Only numbers	0		
Start of the response sentence on the showcard	No	0		
Question on the showcard	No	0		
Picture on the card provided?	No picture provided	0		
Computer assisted	No	0		

Interviewer	Yes	1
Visual presentation	Visual	1
Position	50	50