

Metoda ukotvujících vinět a možnosti využití v pedagogice*

27

Hana Voňková

Univerzita Karlova v Praze, Pedagogická fakulta

Abstrakt: Ve výzkumných šetřeních se často využívají sebehodnotící otázky. Pří-
mé porovnání zemí či různých sociálních skupin na základě odpovědí na tyto otázky však může vést
k chybnému závěru o objektivním stavu respondentů. Důvodem může být, že respondenti z odlišných
skupin využívají dostupnou škálu odpovědí odlišným způsobem, např. Američané mohou častěji
hodnotit situace jako „vynikající“ ve srovnání s jinými národy. Metoda ukotvujících vinět řeší tento
problém tak, že nechává respondenty vedle vlastní situace hodnotit i hypotetické situace jiných
lidí (ukotvující viněty). Vzhledem k tomu, že všichni respondenti hodnotí jednu a tutéž hypotetic-
kou situaci, může být rozdílnost v jejich hodnocení těchto situací interpretována jako rozdílnost
v jejich využívání škály. To je pak využito pro korekci (ukotvení) jejich sebehodnocení. Metoda
ukotvujících vinět nebyla v české odborné literatuře dosud diskutována. Tento článek představuje
základní myšlenku této metody intuitivně a poté i s využitím matematického aparátu a prezentuje
několik příkladů využití této metody v sociálních vědách. V neposlední řadě uvádí možnosti využití
této metody v pedagogice.

Klíčová slova: výzkumné šetření, sebehodnotící otázky, neporovnatelnost odpovědí respondentů,
metoda ukotvujících vinět, neparametrický přístup, parametrický model CHOPIT, konzistence od-
povědí, ekvivalence vinět, pedagogika, zdraví, politika

Anchoring Vignettes Method and Possibilities of its Use in Educational Science

Abstract: Self-assessment questions are used often in survey research. A direct
comparison of countries or different socioeconomic groups based on answers to such questions may
lead to an incorrect conclusion about the respondents' actual situation. One of the reasons may
be that respondents from different groups use the response scale in different ways, for example
Americans may evaluate a situation as “excellent” more often compared to other nationalities. The
anchoring vignette method solves the problem by letting respondents evaluate both their situation
and hypothetical situations of other people (anchoring vignettes). Since the hypothetical situation
is invariant over respondents, the differences in their evaluations of the hypothetical situation may
be interpreted as the differences in their use of response scale. This is then used to adjust (anchor)
their self-assessments. The anchoring vignettes method has not yet been discussed in Czech lan-
guage academic literature. This paper introduces the basic idea of the method both intuitively and
using mathematics and presents several applications of the method in social sciences. Least but not
last, it suggests how to use this method in educational science.

Keywords: survey research, self-assessment questions, incomparability of respondents' answers,
anchoring vignettes method, nonparametric approach, parametric model CHOPIT, response consis-
tency, vignette equivalence, education, health, policy

* Článek vznikl za podpory projektu „The relationships between skills, schooling and labor market
outcomes: A longitudinal study“ (No. P402/12/G130) financovaného Grantovou agenturou České
republiky.

Ve výzkumných šetřeních se často využívají tzv. sebehodnotící otázky (*self-assessment questions*). Příkladem může být otázka *Jak byste zhodnotili své zdraví?* s pěti-bodovou škálou od *nedostatečné* až po *vynikající*. K hlavním přednostem těchto otázek patří nízké náklady (oproti objektivnímu měření např. zdravotního stavu).

Přímé porovnání zemí či různých skupin na základě odpovědí na tyto otázky však může vést k chybnému závěru o objektivním stavu těchto zemí či skupin. Různí respondenti totiž mohou pro hodnocení určitého skutečného stavu využívat různé kategorie škály, interpretovat kategorie škály odlišným způsobem. Tento fenomén je v literatuře označován jako *differential item functioning* (zkratka DIF).¹

Řešení problému heterogenity ve stylu odpovídání na otázky ve výzkumných šetřeních nabízí metoda ukotvujících vinět² (*anchoring vignette method*). Jejím cílem je pomocí přímého změření rozdílnosti využívání škály různými respondenty „očistit“ (přizpůsobit, korigovat) sebehodnocení respondentů tak, aby bylo porovnatelné. Základní myšlenkou metody ukotvujících vinět je nechat respondenty mimo sebehodnotící otázky hodnotit v dané oblasti i hypotetickou osobu popsanou v krátkém příběhu (ukotvující viněti). Vzhledem k tomu, že všichni respondenti hodnotí jednu a tutéž hypotetickou situaci, může být rozdílnost hodnocení ukotvujících vinět interpretována jako rozdílnost v jejich využívání škály. Pomocí takto odhadnuté rozdílnosti můžeme provést „očistění“ sebehodnocení a následně je mezi skupinami respondentů porovnat.

Pro objasnění diskutovaného problému nejprve uvedeme několik příkladů využití metody ukotvujících vinět v sociálních vědách (část 2). Dále představíme základní myšlenku této metody intuitivně, bez použití matematického aparátu (část 3). Následně shrneme formální statistické přístupy využívané při analýze konkrétních dat – neparametrický přístup a parametrický model CHOPIT (část 4). Heterogenita ve stylu používání škál v sebehodnotících otázkách je (potenciálně) relevantní i v pedagogice. V části 5 uvedeme (možné) využití metody ukotvujících vinět v různých oblastech pedagogiky. V závěrečné části 6 mimo jiné stručně shrneme současný výzkum zaměřený na validizaci metody ukotvujících vinět.

¹ Termín *differential item functioning* (zkratka DIF) pochází z literatury o teorii testů. Otázka v testu vykazuje DIF, pokud testovaní z různých skupin se stejnou schopností odpoví na otázku správně s různou pravděpodobností. V teorii odpovědí na položku (*Item response theory*) je této problematice věnována značná pozornost (např. Holland & Wainer, 1993).

² Termín metoda ukotvujících vinět (ani aplikace metody samotné) nebyl v české literatuře doposud nikdy použit. Tento článek zavádí překlad anglického názvu *anchoring vignettes method* jako metoda ukotvujících vinět. Překlad ukotvující je volen z důvodu hlavní funkce vinět – jejich úkolem je ukotvit odpovědi na sebehodnotící otázky tak, aby je bylo možné v případě existence DIF porovnat. Viněti tedy ukotvují sebehodnocení, nejsou však ukotvené ve smyslu, který využívá zakotvená teorie (*grounded theory*). Jako vhodná alternativa k ukotvující viněti by mohla být zakotvující viněti. Pro odlišení se zakotvenou teorií je zvolena ukotvující viněti.

2 Příklady využití v sociálních vědách

29

Metodu ukotvujících vinět zavedl King et al. (2004). V článku se zabývá možnostmi obyvatel Číny a Mexika ovlivňovat politická rozhodnutí. V obou zemích bylo provedeno dotazníkové šetření za podpory Světové zdravotnické organizace. Sběr dat byl ukončen v červnu 2002. Celkově byla sebrána data od 430 respondentů v Číně a 551 v Mexiku. Respondentům byla kladena následující sebehodnotící otázka: *Do jaké míry máte možnost ovlivňovat vládu v záležitostech, které se Vás týkají?* se škálou 1) *vůbec*, 2) *málo*, 3) *středně*, 4) *velmi* a 5) *neomezeně*. Srovnáním relativních frekvencí jednotlivých kategorií škály došel King et al. (2004) k překvapujícímu závěru. V Mexiku volila více než polovina respondentů odpověď *vůbec*, zatímco v Číně se jednalo jen o přibližně čtvrtinu respondentů. Oproti tomu každou z kategorií *středně*, *velmi* a *neomezeně* volilo podstatně větší procento Číňanů než Mexičanů (v průměru za tyto tři kategorie se jedná přibližně o 50 procent Číňanů a přibližně 20 procent Mexičanů). Respondenti v Číně tedy uvedli podstatně větší možnosti ovlivňovat rozhodnutí vlády než v Mexiku. Jak autoři uvádějí, politická situace v době sběru dat však nasvědčovala opačnému závěru. K vysvětlení paradoxu mezi aktuální situací a výpověďmi respondentů použil metodu ukotvujících vinět. Respondenti byli požádáni, aby mimo uvedené otázky ohodnotili na stejné škále pět ukotvujících vinět týkajících se možnosti občana ovlivňovat rozhodnutí vlády. Jedna z ukotvujících vinět je následující: *[Imelda] nemá dostatek pitné vody. Společně se svými sousedy se snaží na tento problém upozornit pomocí sběru podpisů petice. Plánují ji předložit všem politickým stranám před nadcházejícími volbami. Do jaké míry má [Imelda] možnost ovlivňovat vládu v záležitostech, které se jí týkají?...* 1) *vůbec*, 2) *málo*, 3) *středně*, 4) *velmi* a 5) *neomezeně*. Z hodnocení ukotvujících vinět vyplynulo, že Číňané mají oproti Mexičanům nižší standard pro hodnocení míry možnosti ovlivňovat vládu, tj. systematicky volí vyšší kategorie škály (*středně*, *velmi* a *neomezeně*) při hodnocení situace popsané v ukotvující viněti. Tímto bylo identifikováno rozdílné využití škály respondentů z Číny a Mexika. Tento rozdíl následně vyjádřili pomocí matematického aparátu (základní myšlenka metody je diskutována v části 3, statistické přístupy jsou prezentovány v části 4). Identifikovaná heterogenita v použití škály byla následně využita k „očistění“ výpovědí respondentů na sebehodnotící otázku. Na základě toho dospěli k závěru, který byl v souladu s aktuální politickou situací. Pokud by pro srovnání zemí nebyly použité „očistěné“ výpovědi respondentů o heterogenitu v použití škály, byl by závěr o možnostech obyvatel ovlivňovat vládní rozhodnutí zavádějící.

Popsaný problém heterogenity ve stylu používání škály v (sebe)hodnotících otázkách se netýká pouze výzkumných šetření v oblasti politiky. Metoda ukotvujících vinět je využita i v mnoha studiích zaměřených na komparaci zdraví obyvatel různých zemí a socioekonomických skupin (např. Salomon, Tandon, & Murray, 2004; Bago d'Uva, O'Donell, & van Doorslaer, 2008; Voňková & Hullegie, 2011). Respondentům je typicky kladena sebehodnotící otázka *Jak byste zhodnotili své zdraví?* s pětibodovou škálou *nedostatečné*, *dostatečné*, *dobré*, *velmi dobré* a *vynikající*. Analogicky jsou

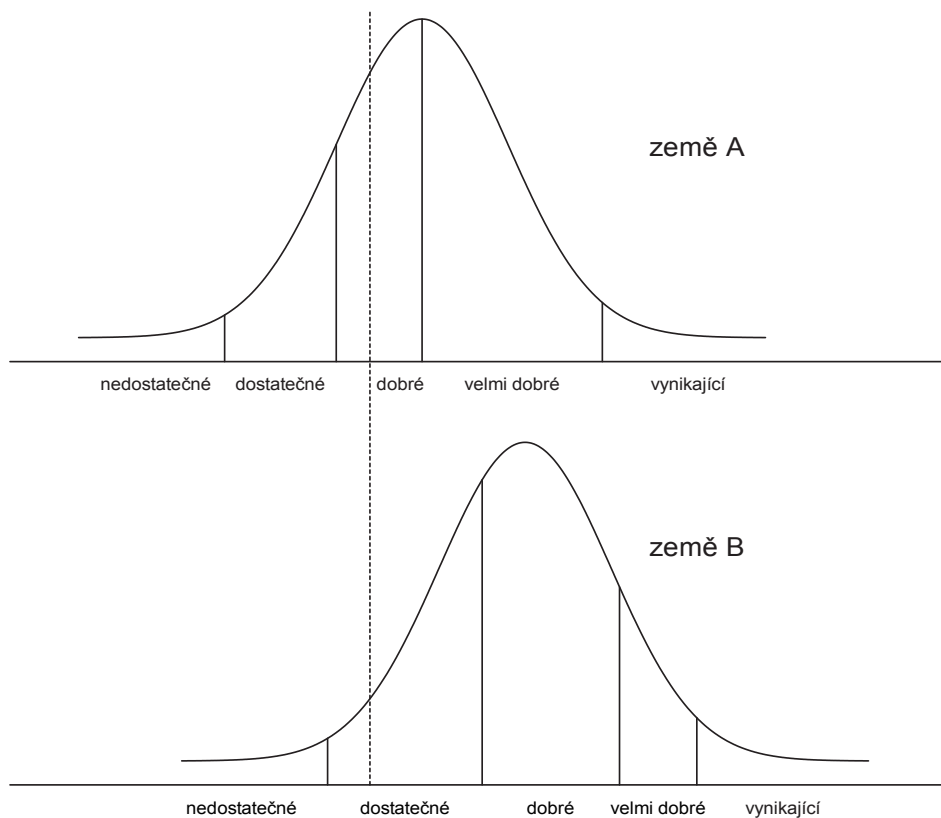
30 formulovány i sebehodnotící otázky týkající se specifických oblastí zdraví jako jsou problémy se soustředěním a pamětí, spánkem, mobilitou, depresí, bolestí a dýcháním. Např. pro oblast soustředění a paměti je používána otázka *Jak velké problémy jste měl/a v posledních třiceti dnech se soustředěním a zapamatováním věcí? se škálou žádné, mírné, střední, velké, extrémní*. Ke každé z těchto oblastí jsou následně formulovány i ukotvující viněty, které pomáhají identifikovat rozdílnost v používání škál. Příkladem ukotvující viněty pro soustředění a zapamatování je: *[Eva] se nadšeně učí nové recepty, ale velmi často v nich dělá chyby a musí si je několikrát znovu přečíst, aby je mohla bezchybně použít. Jak velké problémy má [Eva] se soustředěním a zapamatováním?* se stejnou škálou, která je využita v sebehodnotící poloze.

Při porovnávání zdraví se často ukazují rozdíly v používání škály napříč zeměmi. Obyvatelé v jižní Evropě mají nižší standard než v severní Evropě, tj. shledávají určitý problém jako zanedbatelný, zatímco obyvatelé v severní Evropě shledávají tentýž problém jako více vážný (např. Voňková, 2011). V mezinárodních srovnáních mohou mít navíc jednotlivé kategorie škály různé konotace. Jürges (2007, s. 164) uvádí, že například *excellent* je v anglosaských zemích používáno v každodenní hantýrce, zatímco Němci by slovo *ausgezeichnet* často považovali za ironickou nadsázku, a to speciálně v kontextu posuzování zdraví. Rozdíly v používání škál jsou shledávány i napříč socioekonomickými skupinami. Bago d'Uva et al. (2008) shledává, že například lidé s vyšším vzděláním mají vyšší standard než lidé s nižším vzděláním, což může být ovlivněno jejich větší informovaností o možnostech léčby.

S dalšími příklady využití metody ukotvujících vinět se setkáváme napříč různými oblastmi sociálních věd. Mimo oblasti politiky a hodnocení zdraví se jedná o hodnocení pracovní neschopnosti (Kapteyn, Smith, & van Soest, 2007) a spokojenosti se zdravotním systémem (Murray et al., 2003; Sirven, Santos-Eggimann, & Spagnoli, 2012). Na webových stránkách Gary Kinga věnovaných vinětám (<http://gking.harvard.edu/vign>) jsou uvedeny příklady ukotvujících vinět v různých oblastech sociálních věd. V pedagogice je metoda ukotvujících vinět využita pouze v jedné oblasti. V části 5 shrneme závěry této studie a uvedeme možnosti dalšího využití této metody v pedagogice.

3 Základní myšlenka metody ukotvujících vinět

Základní myšlenka metody ukotvujících vinět bez použití matematického aparátu bývá vysvětlována následujícím způsobem (např. Kapteyn et al., 2007; Voňková & Hullegie, 2011). Představme si, že chceme porovnat zdraví obyvatel ve dvou zemích A a B na základě jejich odpovědi na sebehodnotící otázku *Jak byste zhodnotili své zdraví?* s pětibodovou škálou *nedostatečné, dostatečné, dobré, velmi dobré a vynikající*. Skutečná (a výzkumníkům ovšem neznámá) rozdělení zdravotního stavu v zemích A a B, která chceme porovnat, jsou zakreslena na obrázku 1. Rozdělení v zemi A je posunuto ve srovnání s rozdělením v zemi B více vlevo, což znamená, že zdraví obyvatel v zemi A je horší než v zemi B. K porovnání zemí však nemáme k dispozici objektivní míru zdraví. Bylo by časově i finančně náročné ji získat.



Obrázek 1 Porovnání zdraví pomocí sebehodnotících otázek v případě existence DIF

Místo objektivní míry však máme k dispozici sebehodnocení vlastního zdraví obyvatel na pětibodové sebehodnotící škále *nedostatečné*, *dostatečné*, *dobré*, *velmi dobré* a *vynikající*. Použití škály se však v zemích A a B odlišuje. V zemi A jsou lidé optimističtější při hodnocení vlastního zdraví. Vertikální přerušovaná čára na obrázku 1 reprezentuje osobu s daným objektivním zdravotním stavem. V zemi A hodnotí taková osoba své zdraví jako dobré. Osoba v zemi B se stejným zdravotním stavem však hodnotí své zdraví jen jako dostatečné. Pokud bychom při srovnávání zemí vycházeli ze zdraví měřeného na sebehodnotící škále, podhodnotili bychom rozdíl v průměrném zdravotním stavu mezi zeměmi A a B, nebo bychom dokonce došli k chybnému závěru, že je průměrné zdraví v A lepší než v B.

Metoda ukotvujících vinět umožňuje tento problém řešit. Mimo hodnocení vlastního zdraví jsou respondenti požádáni hodnotit na téže škále i zdraví hypotetických osob popsané v krátkých příbězích, tzv. ukotvujících vinětách. Předpokládejme, že ukotvující viněta popisuje zdravotní stav hypotetické osoby znázorněný vertikální čarou. Vzhledem k popsanému rozdílnému využití škály v uvedených zemích by respondenti v zemi A hodnotili zdraví hypotetické osoby jako dobré, zatímco re-

- 32 spondenti v zemi B by ho hodnotili jako dostatečné. Vycházíme-li z toho, že v obou zemích byla hodnocena tatáž viněta, tj. bylo hodnoceno zdraví téže hypotetické osoby, lze rozdíly v hodnocení připsat rozdílnému využití bodů škály. Hodnocení zdraví hypotetické osoby v ukotvující vinětě nám tudíž pomáhá identifikovat rozdíly v použití škály v daných zemích. S využitím této informace pak můžeme „očistit“ hodnocení vlastního zdraví obyvatel zemí A a B. Pomocí škály jedné z uvedených zemí můžeme vyjádřit zdraví obyvatel v obou zemích. Nezáleží na tom, jakou zemi zvolíme. Podstatou je, že vše vyjadřujeme na škále jedné ze zemí, a tudíž můžeme provádět jejich komparaci. Ať využijeme k vyjádření zdraví škálu země A či B, dospějeme ke správnému závěru, že zdraví obyvatel země B je lepší než země A.

4 Statistické metody

Využití metody ukotvujících vinět při zpracování dat z velkých výzkumných šetření spočívá v aplikaci buď neparametrického přístupu (*nonparametric (single) approach*) a nebo parametrického modelu CHOPIT (*Compound Hierarchical Ordered Probit Model*).

V následujícím textu si oba statistické přístupy představíme. Vzhledem k tomu, že neparametrický přístup je relativně intuitivní, je nabízen jako alternativa pro vložení základního principu ukotvujících vinět uvedeného v části 3.

4.1 Neparametrický přístup

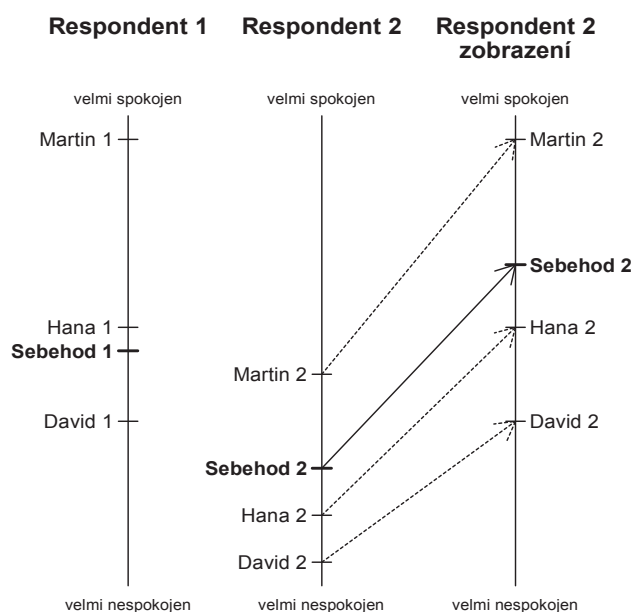
Zavedme si neparametrický přístup na příkladu porovnávání obecné spokojenosti se životem v různých skupinách (Kapteyn, Smith, & van Soest, 2010). Respondentům je kladena následující otázka: *Jak jste obecně spokojen/a se svým životem?* s pětibodovou škálou *velmi spokojen, spokojen, ani spokojen ani nespokojen, nespokojen, velmi nespokojen*. Následně jsou dotázáni, aby zhodnotili, jak spokojené se životem jsou hypotetické osoby v ukotvujících vinětech, které vzniknou podle následujícího vzoru: *[Jméno] je 42 let, je šťastně ženat, má dvě děti, které se dobře učí a vychází dobře se svými rodiči. Příjem jeho rodiny je [xxx]. Má rád svou práci, ačkoli je někdy stresující. [Jméno] trpí relativně vážnými bolestmi zad, které ho přibližně jednou týdně nenechají spát. Žádné další vážné zdravotní problémy nemá. Jak spokojen je dle vašeho názoru [jméno] se svým životem?* se škálou *velmi spokojen, spokojen, ani spokojen ani nespokojen, nespokojen, velmi nespokojen*.

V dotazníku pro respondenty je [jméno] nahrazeno nějakým konkrétním jménem. V případě ženského jména by v českém jazyce byla viněta jazykově upravena (např. je šťastně vdaná). Místo [xxx] je doplněn konkrétní příjem. Vznikne tedy sada vinět, které jsou stejné až na jména a částky příjmu rodiny. Významnou vlastností této sady je, že viněty mají tzv. „přirozené řazení“, viněty s vyšším rodinným příjmem by měly být všemi respondenty hodnoceny minimálně tak dobře, jako viněty s nižším

příjmem. Obecně by bylo možné se stejným efektem měnit i úroveň spokojenosti s rodinou a zdravotních problémů.

Neparametrický přístup je založen na relativním porovnání hodnocení vlastní situace s hodnocením situace popsané v ukotvujících vinětách.

Například si představme, že dvěma respondentům jsou předloženy společně s otázkou na zhodnocení vlastní životní spokojenosti i tři ukotvující viněty (viněty popisující Davida, Hanu a Martina) s přirozeným řazením. Na obrázku 2 je znázorněno jejich hodnocení na škále velmi nespokojen až spokojen (pro výstižnější popis situace je zvolena místo pětibodové diskrétní škály škála spojitá). Respondent A umístuje hodnocení vlastní životní spokojenosti přibližně doprostřed škály. Zároveň hodnotí spokojenost hypotetického Davida jako horší a spokojenost hypotetických osob Hany a Martina naopak jako lepší, než je ta jeho (levá část obrázku 2). Respondent B hodnotí svou spokojenost v dolní části škály, tj. u pólu velmi nespokojen. Zároveň však hodnotí spokojenost dvou hypotetických osob Davida a Hany jako horší a jen Martina jako lepší (prostřední část obrázku 2). Porovnáme-li hodnocení tří vinět respondentů A a B, je zřejmé, že respondent B je se situací daných hypotetických osob vždy méně spokojen než respondent A. Při hodnocení určité situace je pesimističtější.



Obrázek 2 Znázornění neparametrického přístupu metody ukotvujících vinět

Za předpokladu tzv. konzistence odpovědí (*response consistency*, každý respondent používá škálu pro hodnocení sám sebe stejným způsobem jako pro hodnocení hypotetických osob ve vinětách) a tzv. ekvivalence vinět (*vignette equivalence*, různí respondenti interpretují dané viněty stejným způsobem) můžeme zobrazit

- 34 hodnocení respondenta B na škálu respondenta A (pravá část obrázku 2). Hodnocení třech ukotvujících vinět respondenta B zobrazíme do úrovně hodnocení respondenta A (viz čárkované šipky z prostřední části obrázku 2 do jeho pravé části). Sebehodnocení respondenta B pak umístíme mezi jeho zobrazené hodnocení vinět (viz plná šipka na obrázku 2). Na základě vyjádření hodnocení respondenta B na škále respondenta A můžeme konstatovat, že spokojenost respondenta B je vyšší než u respondenta A. Analogickým způsobem by také šlo vyjádřit hodnocení respondenta A na škále respondenta B. Závěr by byl totožný. Pokud bychom však nepoužili tohoto neparametrického přístupu a porovnali pouze sebehodnocení respondentů, dospěli bychom k chybnému (opačnému) závěru.

Sebehodnocení respondenta nabude jednu ze 7 relativních pozic vůči hodnocení 3 vinět. Sebehodnocení může být (1) menší než hodnocení viněty o hypotetické osobě Davidovi (viněta popisující nejmenší spokojenost), (2) rovno hodnocení viněty o Davidovi, (3) mezi hodnocením viněty o Davidovi a Haně, (4) rovno hodnocení viněty o Haně, (5) mezi hodnocením viněty o Haně a Martinovi, (6) rovno hodnocení viněty o Martinovi a (7) větší než hodnocení viněty o Martinovi (viněta popisující největší spokojenost). V neparametrickém přístupu je tedy sebehodnocení „očistěné“ o heterogenitu v používání škály (C) v otázce o životní spokojenosti vyjádřeno na uvedené ordinální sedmibodové škále.³

Obecně při využití K ordinálně seřazených hodnocení vinět může sebehodnocení na diskretní ordinální škále v určité otázce v dotazníku nabýt $2K + 1$ pozic, tj. C může nabýt hodnot 1, 2, ..., $2K + 1$. Reálně však dochází k situacím, kdy respondenti ohodnotí situaci ve více vinětách stejným způsobem, či ohodnotí situaci ve vinětách v jiném pořadí než přirozeném. V takovýchto případech mluvíme o tzv. přesmyčkách (*ties*). Předpokládejme, že hodnocení tří vinět v_1, v_2, v_3 má následující přirozené pořadí: $Y_{v_1} < Y_{v_2} < Y_{v_3}$. Respondent však hodnotí viněty následujícím způsobem: $Y_{v_2} < Y_{v_1} = Y_{v_3}$, tj. situaci v druhé vinětě v_2 považuje za nejhorší, situace v první a třetí vinětě v_1 a v_3 za totožné a zároveň lepší než ve vinětě v_2 . Své sebehodnocení, Y_s , umístí mezi hodnocení druhé a první viněty: $Y_{v_2} < Y_s < Y_{v_1} = Y_{v_3}$. V tomto případě se jeho sebehodnocení vyjádřené na korigované škále v neparametrickém přístupu nachází v intervalu 1 až 5 ($C = [1,5]$).

K tomuto závěru dojdeme postupným prověřením všech možností relativního umístění sebehodnocení mezi vinětami. V tomto případě nejprve prověříme, zda je sebehodnocení menší než hodnocení v první vinětě ($Y_s < Y_{v_1}$). Pokud ano, zaznamenáme 1. V našem příkladu toto platí. Dále prověříme možnost rovnosti sebehodnocení a hodnocení první viněty $Y_s = Y_{v_1}$. Pokud tato rovnost platí, zaznamenáme 2. V našem příkladu toto neplatí. Dále prověříme, zda-li je sebehodnocení mezi hodnocením

³ Sebehodnocení respondenta může nabýt s nenulovou pravděpodobností uvedených 7 relativních pozic vůči hodnocení 3 vinět v případě ordinální škály (např. pětibodová ordinální škála spokojenosti), a nikoli v případě spojité škály. Pravděpodobnost, že se dva body na spojité škále rovnají, je teoreticky nulová – C rovno 2, 4 či 6 by nastalo s nulovou pravděpodobností. Navíc záleží na počtu bodů ordinální škály. Při malém počtu bodů by nebylo možné nabýt uvedených 7 pozic. Např. v případě dvoubodové ordinální škály (např. nespokojen, spokojen) není možné, aby sebehodnocení bylo mezi hodnocením libovolných dvou vinět. Vždy se bude alespoň jedné z nich rovnat.

první a druhé viněty, tj. $Y_{v1} < Y_s < Y_{v2}$. Pokud ano, zaznamenáme 3. V našem příkladu toto neplatí, pořadí hodnocení první a druhé viněty je obrácené. Dále prověříme, zda $Y_s = Y_{v2}$, a tak dále až po ověření, zda $Y_{v3} < Y_s$. Sebehodnocení vyjádřené na korigované škále v neparametrickém přístupu se pak nachází v intervalu nejmenšího a největšího zaznamenaného čísla. V našem příkladu je nejmenší zaznamenané číslo 1 ($Y_s < Y_{v1}$) a největší zaznamenané číslo 5 ($Y_{v2} < Y_s < Y_{v3}$).

Pro porovnání skupin respondentů je nutné jejich sebehodnocení vyjádřené na korigované škále C určitým způsobem agregovat. Můžeme např. sestavit tabulku relativních četností jednotlivých kategorií C a k tomu příslušný sloupcový graf. Při této agregaci se musíme určitým způsobem vypořádat s přesmyčkami (*ties*), tj. intervalovým vyjádřením C (viz náš příklad $C = [1,5]$). V literatuře je uvedeno několik možností, jako je vynechání intervalových hodnot, rovnoměrné rozdělení jedné jednotky v rámci jednotlivých kategorií intervalu atd. (více viz Wand, King, & Lau, 2011).

Je zjevné, že pro úspěšné nasazení neparametrické metody je zásadní mít dostatečný počet dobře přirozeně řazených vinět, které co nejrovnoměrněji pokryjí co největší část škály, na které respondenti odpovídají, zejména ty její oblasti, ve kterých se odpovědi na sebehodnotící otázky koncentrují.

4.2 Parametrický model

CHOPIT model (*Compound Hierarchical Ordered Probit Model*) je rozšířením běžně využívaného tzv. *ordered probit modelu*. Základní rozdíl spočívá v tom, že meze pro rozlišení jednotlivých kategorií ordinální škály mohou být v případě CHOPIT modelu variabilní v závislosti na pozorovaných (a případně i nepozorovaných) charakteristikách respondenta. Tím je umožněno modelovat rozdílné použití škály respondenty.

Parametrický model CHOPIT se skládá ze dvou částí. V první části modelujeme odpověď respondenta na (sebe)hodnotící otázku v rámci určité domény (možnosti ovlivnění vládních rozhodnutí, zdraví atd.). V druhé části pak jeho hodnocení viněty (či více vinět) popisující situaci hypotetické osoby v dané doméně.

První část modelu CHOPIT – (sebe)hodnocení respondenta
(Sebe)hodnocení respondenta i je modelováno následovně:

$$\begin{aligned} Y_{si}^* &= X_i' \beta_s + \varepsilon_{si} \quad i = 1, \dots, I \\ Y_{si} = j &\Leftrightarrow \tau_i^{j-1} < Y_{si}^* \leq \tau_i^j \quad j = 1, \dots, J \\ \tau_i^1 &= X_i' \gamma^1 \\ \tau_i^j &= \tau_i^{j-1} + e^{X_i' \gamma^j} \quad j = 2, \dots, J-1 \\ \tau_i^0 &= -\infty; \quad \tau_i^J = \infty, \end{aligned}$$

Y_{si}^* je latentní (nepozorovaná) proměnná popisující objektivní úroveň respondenta $i = 1, \dots, I$ v rámci dané domény (např. skutečná úroveň jeho zdraví). Je

- 36** modelována jako lineární kombinace vysvětlujících proměnných X_i a nepozorované komponenty ε_{si} . Jako vysvětlující proměnné X_i jsou mimo specifických relevantních proměnných pro zkoumanou oblast běžně používány proměnné jako pohlaví, věk, vzdělání, etnicita a země. Nepozorovaná komponenta ε_{si} může zahrnovat vliv nepozorovaných proměnných a/nebo náhodnou chybu.

Latentní proměnnou nemáme v aplikacích CHOPIT modelu k dispozici. Místo ní však máme k dispozici (sebe)hodnocení respondenta na ordinální škále označované jako Y_{si}^* . Pozorované (sebe)hodnocení nabude hodnoty $j = 1, \dots, J$,⁴ pokud je latentní proměnná Y_{si}^* mezi mezemi τ_i^{j-1} a τ_i^j . Meze τ_i^j $j = 2, \dots, J - 1$ neznáme. Jsou modelovány jako (lineární) kombinace pozorovaných charakteristik respondenta. Tímto je zachycena potenciální variabilita v používání škály různými respondenty.

V této části modelu vztažené k sebehodnocení respondenta není možné bez další informace identifikovat všechny jeho parametry. Jedná se o parametry β_s a γ^1 . Tyto dva parametry jsou však jedny z nejdůležitějších. β_s hraje primární roli při vysvětlení reálné úrovně respondenta v rámci dané domény (např. jeho skutečného zdraví). γ^1 vystihuje variabilitu v používání škály jednotlivými respondenty. S využitím pouze této části modelu je možné identifikovat pouze jejich rozdíl. Intuitivně to lze vysvětlit tak, že v jednom (sebe)hodnocení respondenta jsou obsaženy dvě složky – jeho objektivní (skutečná) úroveň v rámci dané domény a jeho způsob využívání škály (více viz část 3).

Druhá část modelu CHOPIT – hodnocení ukotvujících vinět

Pro identifikaci těchto parametrů využijeme hodnocení ukotvujících vinět v rámci dané domény. Pro každého respondenta i je modelováno následujícím způsobem:

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

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

Y_{vi}^* označuje pro respondenta $i = 1, \dots, I$ latentní (nepozorovanou) proměnnou popisující v rámci dané domény úroveň hypotetické osoby v ukotvující viněti v . Je součtem konstanty θ_v a náhodné veličiny ε_{vi} . Oproti první části modelu vztažené k sebehodnocení respondenta není latentní proměnná modelována v závislosti na respondentových charakteristikách (místo θ_v neuvádíme $X_i' \theta_v$). Důvodem je to, že respondentům je předložena k hodnocení určitá ukotvující viněta, a tudíž respondenti hodnotí tutéž situaci popsanou v dané viněti. V tomto je zároveň skryt předpoklad tzv. *ekvivalence vinět* (*vignette equivalence*), tj. že respondenti interpretují situaci v dané viněti stejným způsobem.

Y_{vi} je pozorované hodnocení ukotvující viněty v respondentem i na stejné ordinální škále, jaká byla použita při sebehodnocení respondenta. Y_{vi} je rovno $j = 1, \dots, J$,

⁴ Např. při využití pětibodové ordinální škály spokojenosti je $J = 5$, přičemž na škále vzestupně míry spokojenosti označuje $j = 1$ kategorii velmi nespokojen $j = 2$ nespokojen, $j = 3$ ani spokojen, ani nespokojen, $j = 4$ spokojen a $j = 5$ velmi spokojen.

pokud je latentní proměnná Y_{vi}^* mezi mezemi τ_i^{j-1} a τ_i^j . Vzhledem k předpokladu tzv. konzistence odpovědí (*response consistency*) jsou meze τ_i^j $j = 1, \dots, J$ modelovány stejným způsobem jako meze pro sebehodnocení respondenta.

S využitím druhé části modelu je umožněna identifikace parametrů γ^1 a β_s . Intuitivně to lze vysvětlit tak, že hodnocení vinět nám umožňuje identifikovat variabilitu v používání škály u jednotlivých respondentů, tj. parametr γ^1 . V první části modelu bylo možné bez využití hodnocení vinět identifikovat rozdíl $\beta_s - \gamma^1$. S využitím identifikovaného parametru γ^1 z druhé části modelu můžeme již tedy identifikovat i parametr β_s , což umožní pro každého respondenta odhadnout jeho nepozorované zdraví. Pomocí všech identifikovaných parametrů můžeme produkovat výsledky podobné výsledkům neparametrické metody.

O náhodných složkách modelu, ε_{sj} a ε_{vj} , $v = 1, \dots, K$, se předpokládá, že jsou nezávislé jak vzájemně, tak i na vysvětlujících proměnných X_i a že mají normální rozdělení se střední hodnotou nula a rozptyly σ_s^2 a σ_v^2 (pro různé viněty se obvykle předpokládá stejný rozptyl).

CHOPIT model je nejčastěji odhadován metodou maximální věrohodnosti. Při odhadu se většinou využívá normalizace parametrů $\beta_{s,1} = 0$ a $\sigma_s = 1$.

Zjevnou nevýhodou CHOPIT modelu je jeho relativní technická náročnost. Přesto je v praxi nasazován častěji než neparametrický přístup. Hlavní předností parametrického modelu je, že k identifikaci parametrů stačí mít jen jedinou vinětu. V aplikacích neparametrického přístupu jsou obvykle využívány alespoň tři ukotvující viněty, které jsou společně se sebehodnotící otázkou předloženy každému respondentovi (např. King et al., 2004; Voňková, 2011). Popsané nižší požadavky na množství dat jsou v parametrickém modelu pochopitelně kompenzovány větším počtem matematicko-statistických předpokladů.

V konkrétních aplikacích neparametrického přístupu i parametrického modelu CHOPIT je možné využít předprogramované balíčky ve statistických softwarech. Neparametrický přístup je implementován v statistickém softwaru R (balíček *anchors*). Parametrický model CHOPIT je implementován v R (stejný balíček *anchors*) jako pro neparametrický přístup a statistickém softwaru STATA (balíček *gllamm*).

5 (Možné) využití v pedagogice

Přestože možnosti relevantního využití metody ukotvujících vinět v pedagogice jsou velké, reálné využití je doposud malé. Autorka je si vědoma pouze jedné publikované studie. Jedná se o knihu Buckley a Schneider (2007), která se zabývá porovnáním hodnocení školy rodiči žáků navštěvujících *charter schools*⁵ a tradiční veřejnou školu ve Washingtonu D.C.

⁵ Charter schools jsou veřejně financované školy v USA, kterým je ponechána relativně velká autonomie v jejich řízení a při tvorbě kurikula výměnou za větší zodpovědnost v dosažení určitých výsledků, které jsou formulovány ve stanovách (*charter*) každé školy. Stanovy jsou vždy platné na určité pevně dané období (běžně 3–5 let). Po tomto období se hodnotí výsledky školy a rozhoduje, zda bude škola nadále podporována. Charter schools tedy nejsou soukromé školy, jsou

38 Rodiče žáků *charter schools* hodnotí školu svých dětí lépe a jsou více spokojeni s jejími různými charakteristikami ve srovnání s rodiči žáků navštěvujících tradiční školy. Vzhledem k tomu, že si rodiče volí školu žáků sami, mohou vnímat objektivní stav na škole pozitivněji, tj. dívají se na školu svého dítěte „růžovými brýlemi“ (více o *rose-colored glasses effect* např. Erikson, 1982). Pokud však chceme využít hodnocení spokojenosti rodičů k hodnocení skutečné kvality těchto škol, je otázkou, jakou roli zde hraje „efekt růžových brýlí“. Pro analýzu tohoto problému zvolil Buckley, Schneider (2007) parametrický model CHOPIT metody ukotvujících vinět.⁶

Výzkum byl proveden mezi 374 rodiči žáků ve Washingtonu D.C., přičemž 200 z nich byli rodiče žáků *charter schools* a 174 z nich rodiče žáků tradičních škol. Rodiče hodnotili tři dimenze školy svých dětí. Jednalo se o míru spolupráce (a) rodičů a ředitele, (b) rodiče a učitele jeho dítěte a (c) rodičů navzájem. Ke každé dimenzi byla formulována jedna sebehodnotící otázka a tři ukotvující viněty. Sebehodnotící otázka na míru spolupráce rodičů a ředitele např. je: *Jak byste zhodnotili míru spolupráce mezi ředitelem a rodiči na škole Vašeho dítěte?* na škále A, ... , F.⁷ Příkladem ukotvující viněty k této otázce je: *Představte si následující situaci: ve třídě je dítě, které pravidelně vyrušuje při vyučování a je učitelem pravidelně posíláno k řediteli. Ředitel této školy pošle rodičům upozornění o problémovém chování jejich dítěte, ale situaci již dále nijak neřeší. Jak byste zhodnotili míru spolupráce („vztah“) mezi ředitelem a rodiči na této škole?* Předložená škála je opět A, ... , F.

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

Jak již bylo uvedeno, možnosti využití metody ukotvujících vinět v pedagogice jsou však velké. Např. autorka článku (spolu)pracuje na výzkumných šetřeních týkajících se porovnání kázně žáků na školách a porovnání hodnocení školního prostředí žáky, jejich rodiči, učiteli a řediteli na víceletých gymnáziích a v odpovídajících ročnících základních škol. Jedná se o první využití metody ukotvujících vinět v populaci žáků, a nikoli dospělých.

6 Závěr

Odpovědi na (sebe)hodnotící otázky ve výzkumných šetřeních mohou být závislé nejen na skutečné (objektivní) situaci respondenta, ale i na jeho individuálním po-

veřejně financované školy nabízející však alternativu k běžným veřejně financovaným školám. První *charter school* byla zavedena v Minnesotě v roce 1991. V dubnu 2006 fungovalo v USA 3500 takových škol s celkem více než jedním miliónem žáků (Buckley & Schneider, 2007 s odkazem na webové stránky Center for Education Reform, duben 2006).

⁶ Jako vysvětlující proměnné X_i jsou pro modelování mezi použity následující proměnné: typ školy (*charter school* či tradiční škola), rasa, vzdělání, postoj k veřejným školám a pohlaví.

⁷ Tato stupnice hodnocení se na amerických školách běžně využívá. Rodiče ji tedy dobře znají.

užívání škály. Pokud je cílem porovnat skutečnou situaci nezatíženou subjektivním vyžíváním škály, může metoda ukotvujících vinět hrát zásadní roli. Pomocí ní je možné identifikovat rozdílné využití škály v různých zemích a/nebo různých socio-ekonomických skupinách. V článku je představena základní myšlenka ukotvujících vinět na příkladu v oblasti hodnocení zdraví, který je často využíván v mezinárodních výzkumech. Jsou diskutovány i dva statistické přístupy, které se využívají při analyzování sebraných dat. Jedná se o neparametrický přístup, pomocí něhož je také objasňována základní myšlenka této metody. Dále se jedná o parametrický model CHOPIT, který se při analýze dat využívá častěji než neparametrický přístup. Je uvedeno i konkrétní využití této metody v různých oblastech sociálních věd, např. možnost ovlivňování politických rozhodnutí občanů v Číně a Mexiku a porovnávání zdravotního stavu obyvatel v různých zemích.

V oblasti pedagogiky prozatím neexistuje mnoho aplikací. Výjimkou je analýza Buckley a Schneider (2007) týkající se porovnání míry spolupráce (a) rodičů a ředitele, (b) rodiče a učitele jeho dítěte a (c) rodičů navzájem na tzv. *charter schools* a tradičních veřejných školách ve Washingtonu, D.C. Možnosti využití metody ukotvujících vinět v pedagogice jsou však velké. Např. autorka článku (spolu)pracuje na výzkumných šetřeních týkajících se porovnání kázně žáků na školách a porovnání hodnocení školního prostředí žáky, jejich rodiči, učiteli a řediteli na víceletých gymnáziích a odpovídajících ročnících základních škol.

Jednu z hlavních linií výzkumu ukotvujících vinět reprezentují v současné době metodologické studie zabývající se validizací této metody. Jde především o testování dvou základních předpokladů metody ukotvujících vinět – konzistence odpovědí (*response consistency*) a ekvivalence vinět (*vignette equivalence*). Statistický test předpokladu konzistence odpovědí pomocí objektivní míry daného konceptu byl zaveden ve studii van Soest et al. (2011). Test předpokladu ekvivalence vinět založeného na předpokladu konzistence odpovědí při hodnocení různých vinět je diskutován v článku Bago d'Uva et al. (2011). Experimentální přístup k testování předpokladů je představen ve studii Kapteyn et al. (2011). Závislost korigovaného sebehodnocení na formulaci viněty je testována v článku Voňková a Hullegie (2011). Dále je analyzován i vliv pořadí sebehodnotících otázek a vinět v dotazníku na jejich hodnocení respondenty (Buckley, 2008; Hopkins & King, 2010) a adekvátnost statistických předpokladů v parametrickém modelu CHOPIT (Voňková, 2011). Podrobná diskuze validizace metody ukotvujících vinět je nad rámec tohoto článku. V každém případě by však měly být nově formulované viněty podrobovány testům jejich adekvátnosti pro korekci sebehodnocení respondentů. Toto se týká (i nově připravovaných) výzkumů s využitím metody ukotvujících vinět v pedagogice.

Literatura

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

- 40 Bago d'Uva, T., Lindeboom, M., O'Donnell, O., & van Doorslaer, E. (2011). Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. *Journal of Human Resources*, 46(4), 872–903.
- Buckley, J., & Schneider, M. (2007). *Charter schools: Hope or hype?* Princeton, New Jersey: Princeton University Press.
- Buckley, J. (2008). *Survey context effects in anchoring vignettes. The Society for Political Methodology working paper*. St. Luis, Washington University, č. 822. Dostupné z <http://polmeth.wustl.edu/media/Paper/surveyartifacts.pdf>.
- Erikson, D. (1982). *The British Columbia story: Antecedents and consequences of aid to private schools*. Los Angeles: Institute for the Study of Private Schools.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Hopkins, D. J., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, 74(2), 201–222.
- Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics*, 16(2), 163–178.
- Kapteyn, A., Smith, J. P., & van Soest, A. (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economic Review*, 97(1), 461–473.
- Kapteyn, A., Smith, J. P., & van Soest, A. (2010). Comparing life satisfaction. In E. Diener, J. F. Helliwell, & D. Kahneman (Eds.), *International differences in well-being* (s. 70–105). New York: Oxford University Press.
- Kapteyn, A., Smith, J. P., van Soest, A., & Voňková, H. (2011). Anchoring vignettes and response consistency. RAND Labor and Population working paper. Santa Monica, RAND Corporation, č. WR-840. Dostupné z http://www.rand.org/pubs/working_papers/WR840.html.
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 567–83.
- Murray, C. J. L., Özalın, E., Tandon, A., Salomon, J. A., Sadana, R., & Chatterji, S. (2003). Empirical evaluation of the anchoring vignette approach in health surveys. In C. J. L. Murray & D. Evans (Eds.), *Health systems performance assessments: Debates, methods and empiricism* (s. 369–399). Geneva: World Health Organization.
- Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *British Medical Journal*, 328(7434), 258–260.
- Sirven, N., Santos-Eggimann, B., & Spagnoli, J. (2012). Comparability of health care responsiveness in Europe. *Social Indicators Research*, 105(2), 255–271.
- van Soest, A., Delaney, C., Harmon, A., Kapteyn, A., & Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society: Series A*, 174(3), 575–595.
- Voňková, H., & Hullegie, P. (2011). Is the anchoring vignettes method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society: Series A*, 174(3), 597–620.
- Voňková, H. (2011). *The use of subjective survey data: Anchoring vignettes and stated preference methods* (Disertační práce). Tilburg: Tilburg University, Center. Dostupné z <http://www.tilburguniversity.edu/research/institutes-and-research-groups/center/graduate-school/thesis/vonkova.html>.
- Wand, J., King, G., & Lau, O. (2011). Anchors: Software for anchoring vignette data. *Journal of Statistical Software*, 42(3), 1–25.

PhDr. Mgr. Hana Voňková, Ph.D. et Ph.D., Ústav výzkumu a rozvoje vzdělávání,
Pedagogická fakulta, Univerzita Karlova v Praze
hana.vonkova@pdf.cuni.cz