

Spôľahlivosť otvoreného kódovania v kvalitatívnom výskume a možnosti jej zvyšovania

Lucia Gálová, Ľubor Pilárik

Katedra psychologických vied, FSVaZ, UKF, Nitra
galova.lucia@pobox.sk, lpilarik@ukf.sk

Abstrakt

Cieľom príspevku je vymedzenie reliability v rámci kvalitatívneho výskumu a následné zhodnotenie spoľahlivosti kódovania medzi kódermi z hľadiska ich blízkosti v identifikovaní významových jednotiek a ich podobnosti, overenie efektu výcviku formou diskusie a zároveň tréningu kódovania u začínajúcich kvalitatívnych výskumníkov. Ako metóda zberu dát boli použité pološtruktúrované fokusné skupiny, metódou analýzy bolo otvorené kódovanie a index reliability je vyjadrovaný Cohenovou kappou a Krippendorffovou alphou. Výsledky indikujú k tomu, že diskusia posudzovateľov počas analýzy, po rozdelení textu na časti, následne vedie k zvýšeniu reliability v rámci kvalitatívneho výskumu, pričom vyšší počet stretnutí nezaručuje aj vyššiu zhodu medzi posudzovateľmi, a krátkodobý tréning kódovania nevedie k zvýšeniu zhody medzi začínajúcimi kódermi pri identifikovaní významových jednotiek.

Kľúčové slová: kvalitatívny výskum, kritériá kvality, reliability, verifikačné stratégie, triangulácia, zhoda medzi posudzovateľmi, otvorené kódovanie, Cohenova kappa, Krippendorffova alpha

Úvod

Výskum ako taký musí spĺňať určité kritériá. Inak sa stáva fikciou, je bezcenný a stráca užitočnosť. Veľký dôraz sa preto kladie na kritériá kvality, ktorých definovanie a „zaistovanie“ je odlišné v závislosti na type výskumu, ale podstata tohto konceptu je rovnaká: bez zaistenia kvality vo výskume naše získané údaje nebudú dobré a použiteľné.

V praxi sa stretávame s dvoma typmi výskumov: kvantitatívnym a kvalitatívnym. V rámci kvantitatívneho sú kritériá kvality jasne vymedzené a uznávané v radoch metodológov, ktorý sa venujú takémuto typu výskumu. Ide menovite o validitu (platnosť), reliability (spoľahlivosť) a objektivitu, ktoré Kvale (1994, In: Švaříček, Šed'ová, 2007) označuje ako „svätú trojicu“ sociálno-vedného výskumu (angl. holy trinity).

V kvalitatívnom type výskumu to nie je také jednoznačné, pretože sa stretáme s rozličným prístupom výskumníkov: od úplného odmietania týchto kritérií, až po ich prijatie, či redefinovanie (Goodwin & Goodwin, 1984, In: Harris, Pryor, Adams, 1999), pričom vo všeobecnosti sa tento typ výskumu považuje za validnejší ako výskum kvantitatívny (pozri napr. Disman, 2002). Reliability je tak často opomínaná. Je však známe, že bez stanovenia reliability nie je možné, aby bol výskum aj validný. Kirk a Miller (1986, s. 76), ktorých môžeme zaradiť do skupiny výskumníkov, ktorí kritériá pozitivistickej tradície preberajú, pričom sú upravené pre potreby kvalitatívneho výskumu hovoria, že: „Kvalitatívni výskumníci si už nemôžu dovoliť odsúvať otázku reliability. Zatiaľ čo sila terénneho výskumu bude vždy spočívať v jeho spôsobilosti zabezpečiť validitu tvrdení, jeho výsledky sa (odôvodnene) nebudú brať na vedomie, ak sa nebude venovať pozornosť reliability.“

Pre kvalitatívnych výskumníkov reliability znamená presnosť a zrozumiteľnosť zozbieraných dát. Inak povedané, reliability je presnosť či správnosť, s akou výskumník dokumentuje to, čo pozoroval, počul, cítil a skúsil. Podľa Hammersleyho (1990, In: Silverman, 2005, s. 188) je reliability „stupeň konzistencie, s akou konkrétne prípady zaraďujú do rovnakej kategórie rôzni pozorovatelia alebo ten istý pozorovateľ pri rozličných príležitostiach.“

Hoci reliabilita v kvalitatívnom výskume nie je taká vysoká ako reliabilita v kvantitatívnom výskume prevedená podľa tradičných štandardov, môže byť dosiahnutá cez rôzne verifikačné stratégie, pričom verifikáciu vymedzujeme ako „proces kontroly, schválenia, uistenia sa a istoty“ (Morse, 2002).

Jednou z najpoužívanejších stratégií pre overovanie kvality výskumu je triangulácia (viď napr. Merriam, 1995), pričom kvalitatívne orientovaní výskumníci ju často používajú na zaisťovanie reliability vo výskume. Janice Morse (1991, In: Mason, 2001) vymedzuje trianguláciu ako metódu získavania komplementárnych zistení, ktoré posilňujú výsledky výskumu a prispievajú k teórii a vývoju poznatkov.

Triangulácia môže zahŕňať množstvo odlišných zdrojov dát, početné teoretické pohľady na interpretáciu dát, viaceré metódy k štúdiu jedného problému, a viacerých výskumníkov alebo hodnotiteľov (Plichtová, 2002). Táto posledná forma triangulácie naznačuje, že rozdiel medzi výskumníkmi môže byť použitý ako metóda na podporu lepšieho porozumenia. V literatúre sa stretáme s pomenovaním zhoda medzi posudzovateľmi (angl. interrater reliability). Ide o určenie stupňa súhlasu medzi rozličným utváraním zdroja dát vzťahovaných k posudzovateľom (Gwet, 2001).

V rámci kvantitatívneho výskumu existuje množstvo indexov, ktoré sa používajú na stanovenie zhody medzi posudzovateľmi. Ich použitie je možné aj v rámci kvalitatívneho prístupu a to ich vhodnou aplikáciou.

Klasickým ukazovateľom je percento súhlasu. Ďalej môžeme použiť Holstiho koeficient reliability, Scottove π , Cohenovu kappu, Krippendorffovu alphu, Spearmanove rho, Pearsonov korelačný koeficient, Linov korelačný koeficient zhody, Kupper-Hafnerov index a veľa ďalších (viď: Shoemaker, 2003, Wever a kol., 2005). Pri výbere neexistuje všeobecná zhoda, ktorý index by mal byť použitý, nakoľko každý má svoje výhody a nevýhody.

Reálne triangulácia výskumníkov vyzerá tak, že výskumníci samostatne kódujú rovnaké dáta a potom sa stretnú a porovnajú si kódy a na základe zvoleného indexu môžu vyrátať skóre reliability. Miles and Huberman (1994, In: Cooper, 2001) odporúčajú stretávať sa viac krát a porovnávať si kódy, pokiaľ reliabilita nedosiahne uspokojujúci stupeň, pričom podľa nich zhoda medzi kódami môže eventuálne dosiahnuť 90 % rozsah v mnohých prípadoch. Tento „design“ preberá mnoho výskumníkov (napr. Harris, Pryor, Adams, 1999) a aj my sme zvolili tento postup a predpokladáme, že diskusia posudzovateľov pri kvalitatívnej analýze dát bude viesť k zvýšeniu zhody medzi posudzovateľmi a že posun v reliabilite bude väčší po stretnutiach dvoch ako po stretnutí jednom.

ŠTÚDIA I: Metódy

Výskumnú vzorku tvorilo päť nezávislých kóderov, ktorí boli zoskupení do dvojíc, pričom jeden kóder sa zúčastnil dvoch kódovaní. Traja kóderi boli študentmi psychológie na UKF a kódovania sa zúčastnil aj jeden pedagóg z Katedry psychologických vied.

Ako metóda zberu dát boli použité pološtruktúrované ohniskové skupiny, ktoré využívajú práve dynamiku sociálnej skupiny pre zber kvalitatívnych dát. V dvoch prípadoch samotní kóderi viedli skupinové interview, ktoré bolo zvukovo zaznamenávané a následne prepísané pre potreby kódovania, a v jednom prípade bol prepis interview poskytnutý pedagógom z Fakulty sociálnych vied a zdravotníctva.

Pre analýzu získaných prepisov bolo vybrané otvorené kódovanie, ktoré tvorí časť analýzy zakotvenej teórie (Strauss, Corbin, 1999), pričom práve otvorené kódovanie je pomerne jednoduchý spôsob ako naštartovať analýzu dát.

Prvá dvojica kóderov kodovala prepisy samostatne a potom sa stretla a porovnala si kódy, druhá dvojica mala text rozdelený na dve časti: prvú časť kodovali kóderi samostatne, potom sa stretli a

porovnali si kódy a následne kódovali druhú časť opäť samostatne, a znovu sa stretli a porovnali si kódy. Tretia dvojica mala prepis rozdelený na tri časti, a postupovala zhodne ako dvojica druhá. Za každým kódovaním (stretnutím) bola vypočítaná Cohenova kappa pre jednotlivé časti prepisov pomocou štatistického programu SPSS.

Výsledky

V rámci skúmania početnosti súhlasu a nesúhlasu, resp. zhody a nezhody v identifikácii kódov medzi kódermi, sa za súhlas považovala prítomnosť ako aj neprítomnosť významovo rovnakého kódu u oboch kóderov. Za nesúhlas sa považovala situácia neprítomnosti významovo rovnakého kódu u jedného z kóderov.

Výsledok zisťovania zhody u prvej dvojice poukazuje na skutočnosť, že ich podobnosť v kódovaní sa nelíšila od náhodných súhlasov, pričom vzájomná zhoda dosahovala 16.7 %, u druhej dvojice zhodu v 22.2 % kódov po prvej časti a zhodu 41.8 % po stretnutí druhom, pričom ich podobnosť v kódovaní nebola náhodná. Posun v tomto prípade bol o 19,6 %. V rámci tretej dvojice sme zaznamenali významný nesúhlas medzi kódermi, ktorý dosahoval 29.1 % po stretnutí prvom. V rámci ich druhého stretnutia ich vzájomná nezhoda dosiahla 9.6 % a po stretnutí treťom sme zachytili už zhodu 16,3 %, pričom v posledných dvoch prípadoch ich podobnosť v kódovaní sa nelíšila od náhodných súhlasov. Posun u tretej dvojice bol o 45,4 % (viď *Tab. 1*).

Tabuľka 1 Hodnota Cohenovej kappy pre všetky dvojice kóderov

	1 stretnutie	2 stretnutia	3 stretnutia
	Cohenova kappa	Cohenova kappa	Cohenova kappa
1. dvojica	0,167 (p>0,05)		
2. dvojica	0,222 (p<0,01)	0,418 (p<0,001)	
	Posun o 19,6 %		
3. dvojica	- 0,291 (p<0,01)	- 0,096 (p>0,05)	0,163 (p>0,05)
	Posun o 45,4 %		

ŠTÚDIA II

Výsledky zo štúdie I podporujú predpoklad, že diskusia nezávislých posudzovateľov pri analýze kvalitatívnych dát vedie k zvýšeniu zhody medzi posudzovateľmi v následnej analýze. Posun v zhode medzi posudzovateľmi je väčší pri dvoch stretnutiach ako pri stretnutí jednom, ale vyšší počet stretnutí ešte nezaručuje aj vyššiu reliabilitu.

Jedným z dôvodov môže byť aj rozdielna miera skúseností u výskumníkov, na ktorú sme sa zamerali v rámci štúdie II, nakoľko sme si všimli, že v rámci výsledkov štúdie I, kóderi, ktorí mali viac skúsenosti s kódovaním kvalitatívnych dát, dosahovali vyššiu zhodu v rámci kódovania (2 dvojica). Predpokladali sme, že krátkodobý tréning v otvorenom kódovaní zvýši zhodu medzi posudzovateľmi, čo znamená, že experimentálna skupina bude dosahovať v rámci skupiny vyššiu zhodu medzi posudzovateľmi ako kontrolná skupina v rámci skupiny.

Metódy

Výskumnú vzorku tvorilo 20 študentov 4. ročníka odboru psychológia, ktorí v školskom roku 2007/2008 mali v letnom semestri zapísaný výberový predmet Metódy analýzy dát v kvalitatívnom výskume II. Študenti boli náhodným výberom (ťahanie mien z klobúka) rozdelení do kontrolnej a experimentálnej skupiny.

Ako metóda zberu dát boli použité pološtruktúrované ohniskové skupiny, pričom študenti samotní dáta nezberali, ale prepisy z pološtruktúrovaných ohniskových skupín im boli poskytnuté, tak ako v jednom prípade v štúdiu I.

Pre analýzu prepisov bolo opäť použité otvorené kódovanie.

V rámci výskumného plánu sme použili následné pozorovanie na dvoch skupinách (podľa Dismana, 2002), ktoré je podobné klasickému experimentu s jednou dôležitou výnimkou: predbežné pozorovanie je vypustené, nakoľko výsledky následného pozorovania v kontrolnej skupine nám podávajú informáciu aj o výsledkoch predbežného pozorovania v oboch skupinách.

Nezávislou premennou v experimente bol krátky workshop (2 x 2 vyučovacie hodiny, spolu 2 x 90 min.), ktorý bol zameraný na analýzu dát v kvalitatívnom výskume, presnejšie sa venoval kódom, kódovaniu a tomu, ako kódovať, pričom 2/3 času boli venované praktickému nácviku. Závislou premennou v experimente bola potom zhoda medzi posudzovateľmi.

Po ukončení sme mali k dispozícii okódované prepisy od experimentálnej aj kontrolnej skupiny s tým, že experimentálna skupina kodovala prepisy po absolvovaní prvej časti workshopu, a kontrolná kodovala bez absolvovania workshopu. Získané dáta sme podrobili deskripcii a následne boli pre obe skupiny vypočítaná zhoda medzi posudzovateľmi, a to konkrétne Cohenova kappa a zároveň aj Krippendorffova alpha, pretože zvládne väčší počet výskumníkov ako dva na rozdiel od Cohenovej kappy.

Výsledky

Analýza výsledkov pozostávala z identifikovania prítomnosti a neprítomnosti významovo rovnakých kódov v experimentálnej aj kontrolnej skupine tak ako v štúdiu I.

Pre obe skupiny bola vypočítaná Cohenova kappa pre všetky kombinácie dvojíc kóderov (spolu 45 kombinácií). Následne bol vypočítaný rozdiel medzi oboma skupinami pomocou t-testu pre dva závislé výbery v štatistickom programe SPSS a z Tab. 2 vidíme, že neexistuje štatisticky významný rozdiel ($p > 0,05$) medzi experimentálnou (AM= 25,8378) a kontrolnou (AM=22,5622) skupinou.

Tabuľka 2 Porovnanie Cohenovej kappy medzi kombináciami kóderov v experimentálnej a kontrolnej skupine

	Experimentálna skupina	Kontrolná skupina	t	p
	AM	AM		
Hodnota Cohenovej kappy	25.8378	22.5622	1,093	,281

Legenda: AM- aritmetický priemer, t – hodnota t-testu pre dva závislé výbery, p- signifikancia

Pre výpočet zhody sme použili aj Krippendorffovu alfu, k prednostiam ktorej patrí aj fakt, že je použiteľná aj pri viacerých výskumníkoch. Výpočet prebehol vložením makra¹ do SPSS, nakoľko tento štatistický program neobsahuje Krippendorffovu alfu vo svojom základnom balíku.

Z výsledkov vidíme (Tab. 3), že v rámci experimentálnej aj kontrolnej skupiny bolo identifikovaných 6570 kódovaných párov, čo je číslo všetkých kombinácií kóderov (v našom prípade 45) násobené počtom identifikovaných jednotiek. Krippendorffova alpha v experimentálnej skupine dosiahla hodnotu $\alpha = 0,257$, pričom spodný limit hodnoty alphy je $\alpha_{\min} = 0,2341$ a horný $\alpha_{\max} = 0,2806$, čo znamená, že hodnota alphy sa v rámci tejto skupiny pohybuje v tomto rozpätí.

V rámci kontrolnej skupiny hodnota Krippendorffovej alphy bola $\alpha = 0,2303$, jej spodný limit $\alpha_{\min} = 0,2003$ a horný limit $\alpha_{\max} = 0,2591$.

¹ http://www.comm.ohio-state.edu/ahayes/SPSS_programs/kalpha.htm

Rozdiel medzi hodnotami Krippendorffovej alphy je medzi experimentálnou a kontrolnou skupinou 0,0269, čo sa približne rovná rozdielu medzi aritmetickým priemerom Cohenovej kappy v experimentálnej a kontrolnej skupine (0,032756), ak ju neprevádzame na percentuálne vyjadrenie.

Tabuľka 3 Hodnota Krippendorffovej alphy pre experimentálnu skupinu a kontrolnú skupinu

	Alpha	Spodný limit; 95% zhoda	Horný limit; 95% zhoda	Jednotky	Kóderi	Páry
Experimentálna skupina	0,2572	0,2341	0,2806	146	10	6570
Kontrolná skupina	0,2303	0,2003	0,2591	146	10	6570

Záver

Z výsledkov oboch štúdií vidíme, že diskusia posudzovateľov počas analýzy kvalitatívnych dát vedie k zvýšeniu reliability v rámci kvalitatívneho výskumu, pričom vyšší počet stretnutí nezaručuje aj vyššiu zhodu medzi posudzovateľmi, a krátkodobý tréning kódovania nevedie k zvýšeniu zhody medzi začínajúcimi kódermi pri identifikovaní významových jednotiek.

Diskusia

V rámci prezentovaných zistení u oboch štúdií a najmä počas ich dôkladnej analýzy sme identifikovali niekoľko problematických okruhov, ktoré môžu vplývať na veľkosť zhody medzi posudzovateľmi. Zaradili sme sem osobnosť výskumníka, jeho mieru skúseností a aj povahu dát, čím máme na mysli, že rozdiel medzi zhodou v kódovaní v rámci dvojíc sa líšil aj vzhľadom na to, či si kóderi sami zbierali dáta (t.j. tzv. živé dáta), alebo či kodovali prepisy, ktoré im boli poskytnuté (tzv. mŕtve dáta). Doplniť ako problémový okruh môžeme tiež motiváciu výskumníka, ktorú sme zaznamenali v rámci krátkej kvalitatívnej analýzy spätnej väzby probantov, ktorí sa zúčastnili experimentu (Gálová, 2008).

Literatúra

COOPER, Cecily D. 2001. Not Just a Number Things: Tactics for Improving Reliability an Validity in Qualitative Research [online]. [cit. 01-03-2008]. Dostupné na WWW: <http://division.aonline.org/rm/2001forum/methods_article_with_refs.pdf>.

DISMAN, Miroslav. 2002. Jak se vyrábí sociologická znalost. Praha: Karolinum, 2002. ISBN 80-246-0139-7.

GÁLOVÁ, Lucia. 2008. Zhoda medzi posudzovateľmi ako verifikačná stratégia reliability v kvalitatívnom výskume (kvantitatívna štúdia). Diplomová práca. FSVaZ, UKF Nitra, 2008. s.60.

GWET, Kilem. 2001. Handbook of Inter-Rater Reliability. How to Estimate the Level of Agreement Between Two or Multiple Raters [online]. [16-02-206]. Dostupné na WWW: <http://www.stataxis.com/files/3chapters_handbook.pdf>.

HARRIS, Judith , PRYOR, Jeffry, ADAMS, Sharon. 1999. The Challenge of Intercoder Agreement in Qualitative Inquiry [online]. [cit. 24-11-2005]. Dostupné na WWW: <<http://emissary.wm.edu/templates/content/publications/intercoder-agreement.pdf>>.

KIRK, Jerome, MILLER Marc L. 1986. Reliability and validity in qualitative research. Sage Publications, Inc.: London, 1986. ISBN 0-8039-2470-4 80s.

- MASON, Oliver J., 2001. The Application of Mindfulness Meditation in Mental Health: Can Protocol Analysis Help Triangulate a Grounded Theory Approach? *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [online], 3(1). [cit. 30-10-2005]. Dostupné na WWW: <<http://www.qualitative-research.net/fqs-texte/1-02/1-02mason-e.htm>>.
- MERRIAM, Sharan B. 1995. "What Can You Tell From an N of 1?": Issues of Validity and Reliability in Qualitative Research. *AACE Journal of Lifelong Learning*; v4 p51 60 1995. Dostupné na WWW: <<http://etd1.library.duq.edu/theses/available/etd-06302003-145932/unrestricted/09Chap03.pdf>>.
- MORSE, Janice M., BARRETT, Michael, MAYAN, Maria, OLSON, Karin, SPIERS, Jude. 2002. Verification strategies for establishing reliability and validity in qualitative reseach. *International Journal of Qualitative Methods* 1 (2), Article 2. Dostupné na WWW: <http://www.ualberta.ca/~iiqm/backissues/1_2Final/pdf/morseetal.pdf>.
- PLICHTOVÁ, Jana. 2002. *Metódy sociálnej psychológie zblízka. Kvalitatívne a kvantitatívne skúmanie sociálnych reprezentácií*. Bratislava: MÉDIA, 2002. ISBN 80-967525-5-3.
- SHOEMAKER, Pamela J. 2003. Intercoder reliability [online]. [cit. 26-01-2006]. Dostupné na WWW: <http://web.syr.edu/~snowshoe/content_analysis/intercoder_reliability.doc>.
- SILVERMAN, David. 2005. *Ako robiť kvalitatívny výskum*. Bratislava: Ikar, 2005. ISBN 80-551-0904-4.
- STRAUSS, Anselm, CORBIN, Juliet. 1999. *Základy kvalitatívneho výzkumu. Postupy a techniky metódy zakotvené teórie*. Boskovice: ALBERT, 1999. ISBN 80-85834-60-X.
- ŠVARÍČEK, Roman, ŠEĎOVÁ, Klára a kol.. 2007. *Kvalitatívny výzkum v pedagogických viedach*. Praha: Portál, 2007. ISBN 978-80-7367-313-0.
- WEVER, B. De, SCHELLENS, T., VALCKE, M., KEER, H. Van. 2005. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education* (2005), 11 April 2005.[cit. 16-2-2006]. Dostupné na WWW: <http://allserv.ugent.be/~mvalcke/CV/content_analysis_review.pdf>.