DO WE KNOW WHAT WE ARE ASKING?
INDIVIDUAL AND GROUP COGNITIVE INTERVIEWS[1]

MIROSLAV POPPER AND MAGDA PETRJÁNOŠOVÁ

Abstract: The paper deals with cognitive interview, a method for pre-testing survey questions that is used in pilot testing to develop new measures and/or adapt ones in foreign languages. The aim is to explore the usefulness of the method by looking at two questionnaires measuring anti-Roma prejudice. The first, the Stereotype Content Model (SCM), contains questions that are dominantly used to test two dimensions of social perceptions of various groups: warmth and competence. The second, Interventions for Reducing Prejudice against Stigmatized Minorities (INTERMIN) consists of the items most frequently used in contact research to measure attitudes, social distance, anxiety, trust and behavioural intentions towards outgroups. Two rounds of cognitive interviews were held on both questionnaires to verbally evaluate participants' understanding and/or interpretation of the draft questions. The first round was attended by university students, while the second round (with improved versions of the questionnaires) was done with high school students, as they are the target group for planned interventions based on the contact paradigm. The paper explains the problems/difficulties the participants had answering some of the questions and our attempts at improving the questionnaires. The problems can be grouped around six issues: The first two deal with the strategies participants used to answer our questions – whom exactly did they have in mind when answering the questionnaires and whose viewpoint did they represent in their answers. The next four problems are around nuances in the formulations of our questions and generally have to do with how the participants interpreted our questions – they concern assumptions that distinct items were logically interconnected, the period of time and locality referred to in our questions, translation and transferability of meanings from one language to another and double negation.

Key words: cognitive interview; questionnaire testing; SCM; INTERMIN; measuring stereotypes and prejudice.

1. Introduction

There is a relatively long tradition of using questionnaires in social psychology in spite of persistent criticism that questionnaire scales and items are unclear, vague and open to various interpretations (see e.g. Furnham & Steele, 1993 for a critique of questionnaires measuring locus of control, Gutek, Murphy, & Douma, 2004 for a critique of the Sexual

---

Experiences Questionnaire (SEQ) and Gray and Durrheim, 2013 for a critique of the Right-Wing Authoritarianism scale). Questionnaires that have not been standardized but created ad hoc have serious limitations. The paper deals with two instruments for measuring prejudice against stigmatized minorities: the SCM (Stereotype Content Model, Fiske et al., 2002) and INTERMIN (Interventions for Reducing Prejudice against Stigmatized Minorities, for more details, see Lášticová & Findor, 2016, this issue). These will be used to measure the effectiveness of direct and indirect contact interventions designed to reduce ethnic prejudice[2].

Since there are no Slovak versions of the standardized instruments used to measure prejudice, which would allow for international comparison and testing of the impact of this kind of intervention, our aim was firstly to create a comprehensive questionnaire (INTERMIN) covering the areas most relevant to contact research (see Lášticová & Findor, 2016, this issue), and secondly to adapt the Stereotype Content Model for use in Slovakia (SCM, Fiske et al., 2002). The first versions of these two instruments were than verified in a quantitative pilot study, which also included qualitative cognitive interviews about the instruments. Standard statistical analyses of scale reliability and factor structure were conducted in order to eliminate items that did not fit the scales well. The second version of the instrument was tested consecutively in a new (second) round of cognitive interviews held to test items that had been changed, added and were generally problematic.

Two different methods were therefore used to improve the reliability and validity of the instruments for measuring prejudice—quantitative and qualitative—although this is not common practice in social psychology. Many researchers advocate the use of these two methods on the basis of several reasons. For instance, Ryan et al. (2012, p. 415) argue that despite quantitative methods being commonly used to analyse the relationship between the questionnaire items and the scales, "the process of testing survey question with cognitive interviews, as part of questionnaire design and refinement, can lead to better informed judgements about the potential quality of survey evidence". Bradburn (2006) explains that many comprehension problems occur because of lexical or structural ambiguity. He gives some simple examples: the word "table" can mean a piece of furniture for placing things on or data arranged in rows and columns, i.e. lexical ambiguity means that one word can have different meanings. Structural ambiguity occurs in sentences like "Flying planes can be dangerous", where "flying" can be read as a verb or adjective and thus the meaning of the whole sentence alters. This can create significant problems when creating items for questionnaires. When designing a question, the researcher's aim is to "get respondents to understand the question in the same way that the researcher does", which is quite problematic

---

[2] The aim of the project is to test various kinds of intervention based on the intergroup contact paradigm in order to ascertain which can be best applied in schools to reduce prejudice against the stigmatized Roma minority (for more details see Lášticová, Findor, 2016, this issue). The members of the research team are: Barbara Lášticová (principal investigator), Magda Petrjánošová, Simona Andraščiková (all from the Institute for Research in Social Communication, Slovak Academy of Sciences), Andrej Findor, Martin Kanovský, Miroslav Popper, Peter Dráľ (all from the Faculty of Social and Economic Sciences, Comenius University in Bratislava), and Miroslav Sirota from the University of Essex, UK. This paper looks primarily at the advantages of pilot testing questionnaires using cognitive interviews rather than the challenges of analysing anti-Roma prejudice in Slovakia. Questions regarding anti-Roma prejudice will follow in the later stages of the project.

"because of the many subtleties and ambiguities of language" (Tourengeau & Bradburn, 2010, p. 318). Of course, these problems can also emerge when trying to translate and/or transfer a questionnaire into another language and/or culture. When they occur as part of an attempt at something complicated, e.g. translating a group of synonyms that have specific connotations, they can be more difficult to spot and to avoid. The aim of the paper is to demonstrate the usefulness of cognitive interviews for psychologists or the instruments they routinely use so they can be sure they really measure what they ought to measure and/or are culturally universal (like the SCM).

Cognitive interviewing can be characterised as a method for pre-testing survey questions or questionnaires[3]. It focuses on questions that may be vague or not easy to understand or leave too much scope for unintended subjective interpretation. Beatty and Willis (2007, p. 287) define cognitive interviewing "as the administration of draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help to determine whether the question is generating the information that its authors intend". Blair and Brick (2010) warn that while this method has been widely used since the late 1980s, there is still no generally accepted procedure for using cognitive techniques and interpreting the data obtained. Equally there is no agreement on the optimal sample size, the number of interviews, interviewers or rounds of interviewing (Beatty & Willis, 2007). Furthermore, Presser et al. (2004) stress that there is no consensus on best practices regarding techniques, procedures and evaluating the results. For example, we have yet to resolve the questions of which of the two basic techniques—*thinking-aloud interviewing* and *verbal probing*—is better to use in what circumstances. The first technique involves the participants actively verbalizing their thoughts whilst answering the survey questions (e.g. "*How Roma are viewed by Slovaks in general? Please tell me everything that comes into your mind*"). When using the second technique the interviewer asks specific probing questions either after each survey question or at the end, once the entire questionnaire has been completed, in order to obtain additional information to help in understanding and/or interpreting the survey questions (e.g. *"What does the term 'in general' mean to you?"*). The advantages and disadvantages of these alternative techniques can be found in full in Willis (1999), and Willis and Artino (2013). Since in this research we use verbal probing we only cover the pros and cons of using this technique.

If it is to be performed appropriately and without bias, verbal probing demands some skills or training, to avoid, for instance, devoting too much attention to or speculating over questions which can be quite easily understood without probing. This technique allows the researcher to focus on more problematic cognitive processes such as understanding the particular wording/phrasing of more complex questions. It can also help in cross-cultural comparisons of questionnaire data in ascertaining whether the items are understood or

---

[3] To be clear on the terminology, the origins of cognitive interviewing are to be found in sociology, where it is a relatively commonplace technique. The vast majority of the literature uses the term *survey* to describe the main sociological research method. However, in social psychology the term *questionnaire* as a quantitative data collection tool is preferred over *survey*. As the authors' aim is to highlight the use of this method for social psychologists, we will be using the term *questionnaire* throughout, except where sociological resources are directly cited.

interpreted in the same way in different cultures. At the same time it is a flexible procedure, so any problems and/or difficulties that emerge during the interview can be addressed and explained.

When designing surveys the perceived seriousness and extent of the problems associated with cognitive interviews depend mainly on the aims of the cognitive interviewing. For someone seeking to create the ideal questionnaire, it may become a never-ending process of constantly refining the wording of questions. However, abandoning the requirement to achieve the ideal and settling for having an understanding of the pros and cons of the alternative phrasing of the items can lead to relative satisfaction with the improvements to the questionnaire. In both cases, systematic and rigorous cognitive interviewing may significantly enhance the design of questionnaire items. Our aim is closer to the second, i.e. to learn about the ambiguities of the questionnaire items and either improve the wording or replace or eliminate the "troubling" item. In another words, our aim was to obtain useful information in order to improve the questionnaire items to ensure they were unambiguous, i.e. respondents would know what we are asking. Our aim is to enhance the validity of the questionnaires rather than produce "ideal" items.

We pre-tested two questionnaires. The first is based on questions that feature most frequently in studies using and testing the Stereotype Content Model (SCM) (Cuddy, Fiske, & Glick, 2007; Fiske et al., 2002; Cuddy et al., 2009; Fiske & North, 2014). These questions were translated into Slovak from the original English. The SCM measures two basic dimensions of social perception: warmth and competence. These two universal dimensions of social judgments relate closely to stereotypes and/or prejudices towards different outgroups or minority groups (Cuddy, Fiske, & Glick, 2008). Different groups can be placed in one of four possible quadrants specified/defined by these bipolar scales. However, according to the SCM, we frequently find two mixed clusters of ambivalent stereotypes on many outgroups: high warmth together with low competence or high competence combined with low warmth (Fiske et al., 2002). Stereotypically high status groups are viewed as competent, while low warmth groups are perceived as competitive (Cuddy et al., 2009). The Slovak version of the SCM also included scales to measure the emotions of contempt, admiration, pity and envy in addition to protective vs. aversive behaviour towards the minority group and frequency of different types of direct contact with the minority group in accordance with the BIAS map proposed by Cuddy, Fiske, and Glick (2007).

It should be noted that there is no definitive version of a questionnaire for investigating the SCM. However, there is a clearly defined pool of items employed in the various versions and the four-item and two-item versions are most used to measure each of the two basic dimensions of social perception: warmth and competence. Translating the items from the English versions into Slovak caused us substantial problems as the number of synonyms and their connotations differ between languages. Other difficulties resulted from cultural transferability, as different social practices prevail in different societies and if the questions are translated literally, they can have very different meanings from the original. The aim of course is to find a translation that captures the spirit of the original rather than a literal one, but when the goal is to use a specific measure, e. g. the SCM questionnaire, finding a good translation – one that is precise and meaningful in the other country's context) is often far from easy and a slow process.

The second questionnaire – INTERMIN – was compiled by our team, based on the items that are most frequently used in contact research. It has subscales for social anxiety, social trust, behavioural intentions and contact in general because these are important in intergroup relationships (Turner, West, & Christie, 2013). We adapted the questionnaire for schools and plan to use it later on a large sample. It is designed for use in contact research on outgroups, but at the time of writing we are mainly interested in the Roma minority, which is heavily stigmatized in Slovakia, and we pre-tested it (along with the SCM questionnaire) by asking specifically about the Roma outgroup. The first draft discussed in this article consisted of 26 items plus demographic questions about the participants (for more details, see Lášticová & Findor, 2016, this issue).

2. Cognitive interview research sample

We conducted two rounds of cognitive interviews[4] with both the questionnaires in order 1) to learn how respondents understand or interpret the meanings of the items, and 2) to pre-test whether the translation of the items from the English instruments used in contact research is suitable for use in Slovakia. The first round (see Table 1) was administered to university students, while the second round (using improved versions of the questionnaires) was carried out with high school students, as they are the target group of planned experiments based on the contact paradigm. The university student sample was a convenience sample; we were attempting to gain access to bigger groups of students from different educational establishments as well as of different specializations and from different years of study. The gender proportion was skewed as more women attend social sciences courses.

The cognitive interviews conducted differ slightly from the standard type, as many of them were held not with one respondent, but with groups of students. The main reason for

Table 1: Participants in the first round of cognitive interviews

	Type of cognitive int.	Locality	Education establishment	Year of study	Specialization	Numbers of participants
SCM	group disc.	Bratislava	university	1st	psychology	4
SCM	group disc.	Trnava	university	4th	pedagogy	9
SCM	group disc.	Bratislava	university	3rd	psychology	17
SCM	group disc.	Bratislava	university	4th	psychology	17
SCM	group disc.	Bratislava	university	5th	psychology	30
SCM	group disc.	Bratislava	university	5th	anthropology	15
Intermin	group disc.	Trnava	university	1st	pedagogy	6
Intermin	group disc.	Bratislava	university	1st	psychology	24
Intermin	individual int.	Bratislava	university	3rd	psychology	1
Intermin	group disc.	Bratislava	university	3rd	anthropology	12

---

[4] We will be discussing the first and second drafts of the questionnaires.

this was that group interviews lead to more heated discussion and argument. On the other hand the disadvantages of group cognitive interviews are that many participants tend to adopt the dominant view and that opinions may become polarized in a way that would not arise in individual cognitive interviews.

We also tested a simplified version in which the students first completed the questionnaire pilot and then wrote comments on each question after they had been asked probing questions formulated by the research team (N=100, Trnava, university, 1st year, pedagogy; and N=50, Bratislava, university, psychology, 1st year). We used this data as an additional source of information.

In the second round of cognitive interviews (see Table 2), the research sample was much more specific. It consisted of 24 participants, half of whom were from Prešov, Slovakia's third biggest city, located in the east, which has a relatively numerous Roma population (the proportion of Roma residents is 10.06%) and half were based in Bratislava, the capital, where the proportion of Roma residents is 2.12% (Mušinka & Matlovičová, 2015). In each of the cities, six participants were involved in pre-testing and discussing the SCM and six in pre-testing and discussing the INTERMIN questionnaire. Women and men were equally represented in these samples.

Table 2: Participants in the second round of cognitive interviews

	Type of cognitive int.	Locality	Education establishment	Year of study	Numbers
SCM	individual	Bratislava	High school	3rd	6
SCM	individual	Prešov	High school	3rd	6
INTERMIN	individual	Bratislava	High school	3rd	6
INTERMIN	individual	Prešov	High school	3rd	6

In the group cognitive interviews we used verbal probing once the participants had completed the questionnaire. In some cognitive interviews two researchers were involved –one asked the probe questions while the other took written notes on each item discussed. In some cases only one researcher was involved. Other interviews were recorded and then analyzed. We did not transcribe the interviews verbatim (as is standard in qualitative interviewing for example) as we were not interested in the linguistic or discursive nuances; instead we simply analyzed the information they provided.

3. Results

We will not describe the two questionnaires in full but will present categories of typical problems (concerning one or both questionnaires) that triggered relatively extensive and passionate discussion among the participants and researchers. The problems can be grouped around six issues. The first two deal with the strategies the participants used to answer our questions – what representations of the category Roma they used and whom were they

thinking of when completing the questionnaires (see 3.1), and what "optics" were used for the opinions they provided and whose viewpoint did they give in their answers (see 3.2). The next four problems are around nuances in the way the questions were formulated and generally have to do with how the participants understood our questions—they concern the assumption that distinct items were logically interconnected (see 3.3); the period of time and locality referred to in our questions (see 3.4); translation and transferability of meanings from one language to another (see 3.5); and double negation (3.6).

Our results are arranged such that in the first instance we introduce the original item, then the comments of the participants on the wording of the item (whether it was clear or vague, how it was understood, what was disruptive, etc.) and finally we consider whether, how, why and to what extent the original wording was changed.

3.1 Whom did the participants have in mind when answering the questionnaires?

The main problem related not to a specific question but to both questionnaires, and it emerged when we later asked the participants in a cognitive interview whom they had been thinking of. This of course is a problem from the viewpoint of the researchers because we cannot be sure which representations of the category of Roma the participants were using and what they would do if they had access to more than one representation, especially when those are contradictory.

For example, the first draft of the INTERMIN questionnaire included the following question measuring social distance (inspired by Bogardus (1925) but adapted for school context where the data will be gathered in future): *To what extent would the following situations be acceptable or inacceptable for you?* The five items were: *If a Roma (female or male) was a pupil at your school/ was your classmate/ sat next to you on a school bench/ spent free time with you as part of your group of friends/ was your neighbour.* The participants answered using a 7-point Likert scale, 1 – that would be completely inacceptable, 7 – that would be perfectly acceptable.

When we asked the participants whom they had been thinking of, some reported they had only thought about "good Roma who make an effort"; some included the impact of media and general opinions in Slovakia and had not just considered their opinion but were attempting to assess the situation in Slovakia objectively; some had pictured a person they knew; some had also answered in relation to a specific person but then "caught" themselves and then tried to produce an "average" opinion based on different people (some of whom they viewed negatively, some neutrally, and some positively); some had tried using "averages" as a strategy from the beginning (because they had wanted to be as objective as possible); and finally some had tried to use "averages", but realized they were thinking of extremely positive and negative experiences at the same time and therefore experienced problems coming to a conclusion. It appeared that different participants used very different strategies to construct a representation of the "Roma" social category, and, moreover, the representation the participant used sometimes changed from question to question.

We did not find an easy way to test this variation in the second draft of the questionnaire (and in the second round of cognitive interviews). We simply realized how important it was to ask about this point explicitly – and we did that via a question at the end of the

questionnaire (used in the first draft) that asked: "*When you were filling in the questionnaire, who did you picture when you read the word ´Roma´? If you pictured different persons/ groups, please say who all of them were.*"

A very similar problem arose in the SCM questionnaire, for example with the question measuring competence: "*As viewed by most Slovaks, how prestigious are the jobs generally held by Roma?*" The participants answered on a 5-point Likert scale, 1 – not at all, 5 – extremely.

The problem with this question was that for some participants thinking about the jobs triggered a stereotypical image/idea of two basic groups of Roma people, one representing the majority of Roma who work as manual labourers, and the second, depicting Roma as artists (namely musicians). Participants used one of two strategies: they either decided on the basis of numbers and rated the jobs generally held by Roma as very low in prestige or they averaged the two groups (manual labourers and artists) and assigned a middle value reflecting how prestigious the jobs generally held by Roma are conceived to be. Our research team did not consider the problem to be a major issue since it relates to the questionnaire ratings of all kinds of groups, like teachers, students, parents, etc. People will probably always have a stereotypical or main idea when rating any larger group of people, which cannot be avoided[5]. Stereotyping Roma as either blue collar workers or musicians might be specific to Slovakia (see e.g. Mann & Ruppeldtová, 2009) and in other cultures different stereotypes may prevail; however, there would still be a problem as to whether the middle point on the Likert scale reflected uncertainty of judgment (concerning Roma) or the average of two distinct judgments of different subgroups (of Roma), which is a problem generally when using these kinds of self-report measures. Moreover, as in the INTERMIN questionnaire, we have extra information from the question at the end of the questionnaire: "*When you were filling in the questionnaire, who did you picturing when you read the word ´Roma´? If you pictured different persons/ groups, please say who all of them were.*"

3.2 Whose viewpoint should be represented in the answers?

The SCM is based on the assumption that by asking participants not to express their personal beliefs about (minority) groups but think about how these groups are viewed by others reduces socially desirable answers and captures perceived cultural stereotypes (Fiske et al., 2002). The original item was "Consider how [group, e.g. the elderly] are viewed by Americans in general" (Cuddy et al., 2007, p. 648). The Slovak version used the same wording: *"Consider how Roma are viewed by Slovaks in general"*. However, in cognitive interviews, some participants suggested that it was quite difficult for them to establish what "in general" meant for the following reasons: (1) a contradiction between their own (positive) experience and the negative messages concerning Roma from the media; (2) a contradiction between personal opinion and the opinion of the majority; (3) uncertainty over whether to reflect the opinions of those close to them or the negative media messages that

---

[5] As stressed by Fiske (1998, p. 367) "Categorization and stereotyping save cognitive resources, which help explain why their relative automaticity is functional, but the form of automaticity may vary, depending on people's goals and the form of contact".

have greater influence/impact/power; (4) a difficulty depersonalizing their own view; (5) implicitly at least the phrasing of the question suggests that Slovaks consider themselves to be different from Roma, as if the Roma constituted a group to which Slovaks do not belong[6]. Ultimately, on the basis of these comments from the cognitive interviews, we changed the phrasing "*Consider how Roma are viewed by Slovaks in general*" to "*Consider how Roma are viewed by the majority of people in Slovakia*". This still left the problem concerning the key role of the media in influencing the thinking unresolved, since we could not satisfactorily solve the problem of how to "suppress" the dominant influence of media in cases where the participants and those close to them view Roma differently than the dominant media discourse. So the question *"Consider how Roma are viewed by Slovaks in general"* is often answered as if it were *"Consider how Roma are viewed by the Slovak media in general"*. This issue is not addressed in the original questionnaire by Fiske et al. (2002).

3.3 Assumption that distinct items are logically interconnected

We were surprised to find that, in some cases the participants had unexpected problems answering some questions which they assumed were interconnected and which we considered to be distinct items. This occurred on both the general level and where something is considered to have a cause-consequence relation.

A) Interconnectedness in general: For example in the INTERMIN questionnaire, we used a broader question with six subquestions to measure intergroup anxiety and these subquestions were intended as six distinct questions that tapped into different facets of anxiety; however, some participants reported that they had looked for a logical connection between them. The introductory sentence was: *If a new Roma pupil came to your class, how would you feel?* Six distinct statements followed, and these were divided visually and came with a separate scale on which to answer: *I would feel ok/ I would feel awkward/ I would be cautious/ I would feel agreeable/ I would be worried/ it would be no problem for me.* The answers were again given on a 7-point scale, but this time, 1 – does not correspond to my feelings at all and 7 – corresponds precisely to my feelings.

Answering the question became much more complicated for those participants who were looking for a logical connection between the items (e.g. they argued that the items did not exclude each other, which we had never intended, or that they might feel differently at the beginning of the interaction from during the interaction, or first cautious and then agreeable, etc.).

We found no easy solution to this issue (and had to satisfy ourselves with the idea that only some participants had reported this and only when they had been explicitly instructed to find problems, so perhaps they reported minor ones as well…). However, our team statistician Martin Kanovský conducted a statistical analysis of scale reliability and factor structure (see

---

[6] This problem is quite complex. In Slovak, different words are used to refer to state citizenship and nationality/ethnicity, whereas in English nationality and citizenship are often considered to be the same thing. Our aim was to ascertain the nature of the relationship between the majority (Slovak) population and the minority (ethnic Roma); however, some participants interpreted the item as meaning that Roma are not citizens of Slovakia (i.e. Slovaks). (For more details on Slovak national identity see e.g. Lášticová, 2006.)

discussion and for more details Kanovský, 2016), and subsequently in the second draft of the questionnaire we excluded two "weak" items, shortening the list of six items to four, hoping this would cause fewer problems relating to interconnectedness.

B) Connecting items because of a perceived cause-consequence relation: In the cognitive interviews another problem emerged with the way in which items logically interconnected, related to a perceived cause-consequence relation. This can be seen in one of the questions in the INTERMIN questionnaire that measures approach and avoidance behavioural intentions (largely inspired by measures used in Turner, West, & Christie, 2013). In the first draft questionnaire we used the question: *If a new Roma pupil came to your class, how would you react?* Six statements followed: a) I would like to learn more about him/her; b) I would like to have a chat with him/her; c) I would like to spend some time with him/her; d) I would like to avoid him/her; e) I would like to keep far away from him/her; f) I wouldn´t want to have anything to do with him/her. The answers were again given on a 7-point scale, where 1 – does not correspond to my feelings at all and 7 – corresponds perfectly to my feelings, and each item was rated separately on a scale provided separately.

Again, the intention was that these items should remain distinct and independent of each other. The approach tendencies were arranged together (a-c) and the avoidance tendencies together (d-f). The problem was, once again, that in the cognitive interviews several participants reported that they had considered whether there was a logical cause-consequence relation between one item and the others and this made the whole question more complicated for them to answer. For example, some claimed that the information learnt in situation a) "I would like to learn more about him/her" would make a big difference to them and would be very relevant to their decision in situations c) "I would like to spend some time with him/her" and e) "I would like to keep far away from him/her".

In an attempt to solve this we first looked at the results of the statistical analysis of scale reliability and factor structure and found that the last three items measuring avoidance behavioural tendencies (d-f) were rather "weak", so we were able to omit them in the second draft of the questionnaire, making the whole question shorter and easier. Also, according to this analysis, our participants apparently had not dealt with the questions on approach and avoidance behavioural tendencies as if some questions were part of a scale for approach and some of a scale for avoidance (compare Tam et al., 2009), but as if there was one scale only, measuring the tendency to approach or to avoid. Thus, the new shortened version for the second draft with only three questions will focus only on approach tendencies and should still accurately measure what we intended it to measure.

3.4 What is the period of time and locality referred to in our questions?

In both questionnaires we decided to measure the reported direct contact between the participants and Roma, as this may strongly influence perceptions of Roma. In the first drafts of the questionnaires, we asked for example the following question about contact: *"How often do you come into contact with Roma in public space?"* The participants answered on a scale of 1 – never, 2 – less than once per month, 3 – once per month, 4 – several times per month, 5 – once a week, 6 – several times per week, 7 – every day. The question raised three issues. The first problem related to the period of time as some participants could have

come into contact with Roma in the past, but no longer did so in the present (or vice versa). The second problem was the locality since people who study or work outside their place of primary residence had difficulties deciding in which of these two places they should assess the frequency of their contact with Roma (particularly where there were significant differences between the localities). The third problem concerned how they should interpret the term "contact".

The solution attempted in the second drafts of the two questionnaires was to rewrite the question entirely and replace it with more specific ones: "How often do you see Roma in everyday life?" (to assess mere exposure to Roma); "How often do you communicate with Roma in everyday life?" (to assess direct contact); "How often do you spend time with Roma?" (to assess a more personal and intimate form of direct contact); "How many of your friends have Roma friends you know about?" (to assess extended contact via cross group friendships); "How often do you encounter media news about Roma (e.g. newspapers, TV, the internet)?" (to assess vicarious contact via the media). The reason for this relatively substantial change to the original question was that contact with Roma can differ in nature and quality. The research team therefore decided to make the questions on the different forms of contact as specific as possible, since these differences may play a substantive role in the formation of stereotypes on Roma.

3.5 Translation and transferability of meanings from one language to another

When using several items that are close in meaning to obtain a full picture of the phenomenon being investigated there is a danger that the translation may lead to confusion. For example in the English version of the SCM questionnaire there are three questions on the competence, capability and skilfulness of the perceived groups. This led to a non-trivial problem in Slovak, as the meanings of these words are interchangeable and so can be viewed as synonyms. These words are of course synonyms in English as well; however, it is hard to assess and/or compare the measure of synonyms in the two languages. This was for example the case with the original item: *"As viewed by most Slovaks, how capable are Roma?"* with answers on a 5 point Likert scale, 1 – not at all, 5 – extremely. Participants interpreted the word "capable" in several ways: as artists, as the potential to assert oneself in life, and as the ability to be inventive in an unexpected, nonstandard or novel situation[7]. The participants were uncertain as to whether "capability" meant talent or competence at work. It also evoked the negative connotation – "capable thieves". Our research team eventually came to the conclusion that all of these interpretations describe the full range of meanings that the word "capable" has in Slovak and so the question stayed unchanged.

---

[7] Just as a reminder, the original idea behind the SCM was that "people want to know others' intent (i.e., warmth) and capability to pursue it (i.e. competence)" (Fiske et al., 2002, p. 878). The competence scale contained the following items: competent, confident, independent, competitive, inteligent and later two new competence traits were added: capable and skillful (Fiske et al., 2002).

3.6 Double negation

The last problematic issue we want to explore in this paper concerns a well-known problem – double negations (cf. Helfrich, 1986). This can be illustrated using the question about intergroup anxiety above (see 3.3, part A), or more specifically using one of its subitems – some participants reported problems with the double negation occurring in the item "*it would be no problem for me*" and the end of the scale that used negation "*1 – does not correspond to my feelings*". Here it is important to say that double negation is frequently used in Slovak and does not change the meaning into a positive one as it does in English, e.g. in Slovak the English sentence "I have nothing" is "Nemám nič" [lit. "I don't have nothing"]; however, the double negation can be confusing when it involves the combination of several statements or questions and answers (especially short ones) (where the logic dictates that, as in mathematics, two minuses mean a plus). Therefore some participants reported having to "work out" what answer would express their opinion – if they wanted to express that "it would be no problem for them", should they choose the positive or the negative end of the scale if they did not want to mathematically negate their statement?

In order to solve this problem, in the second draft we used a combination of several changes. First, in all items we used exclusively a grammatically positive formulation (in this case – "it would be a problem for me"). Second, following the statistical analysis of scale reliability and factor structure (see discussion) we eliminated the "weak" items, shortened the list of six items to four, and thus got around our original problem[8] with the whole question in the first draft of the questionnaire.

4. Second round of cognitive interviews

After modifying the two questionnaires on the basis of the results of the statistical analyses and the first round of individual and group cognitive interviews, a second round of cognitive interviews was held with high school students. This showed that the modifications to the original questions made in light of the comments and suggestions from the first round of individual and group cognitive interviews had improved the way our participants understood the instruments. The instructions were now clear to participants and the wording/phrasing of questions was not vague. The items concerning the various types of contact with Roma were now comprehensible and respondents could answer them without problems. Nevertheless, some other problems emerged, but these seem easy to improve. For example in the SCM questionnaire in relation to the question "How much does special treatment (e.g. occupation selection privilege) given to Roma make things more difficult for other people" it was suggested that we replace our example in parenthesis with a more specific example, like housing, education, meal allowances, or welfare privileges. Another kind of problem that several of the participants in the second round of cognitive interviews had with the SCM

---

[8] In the first draft we had six items translated from the English, but they were too similar in Slovak and so to make the choices more different, we included the grammatically negative formulations and unfortunately, this led to the whole problem with the double negation which would otherwise have been obvious (see also the translation problems mentioned above).

questionnaire concerned the scales to measure emotions. They consist of pairs of items which have quite similar meanings, so distinguishing between them can be tricky. Specifically, the participants have to consider the feelings of the majority towards Roma regarding contempt (contempt, disgust), admiration (admire, proud), pity (pity, sympathy) and envy (envious, jealous). Surprisingly, we subsequently realized that it had also been mentioned in the first round of cognitive interviews, but only marginally, so we had paid insufficient attention to it. After the second round of cognitive interviews it was suddenly clear that quite a lot of the participants had had problems differentiating between feelings of envy/jealousy and contempt/disgust. In order not to confuse participants by putting them in the position of having to consider the degree of similarity our research team decided that instead of placing the two relatively similar questions one after another it would be better to mix them up. So, for example the item "To what extent do people tend to feel envy toward Roma" would no longer be followed by "To what extent do people tend to feel jealousy toward Roma" but by "To what extent do people tend to feel contempt toward Roma".

The second tested draft of the INTERMIN questionnaire proved even less problematic, as we quite often asked whether a different (usually more specific) wording of the question would be better and generally the participants thought both variations were perfectly understandable. Also we have to stress that when we shortened the lists of items used in line with the results of the statistical analysis of the scale reliability a lot of the earlier problems regarding excessively similar synonyms and perceived interconnectedness disappeared.

For an overview of the six problematic areas, the original versions of the questions/items, and the changes following the first and second round of the cognitive interviews, see Table 3.

5. Discussion and conclusions

To conclude, according to our findings the participants in the cognitive interviews generally expressed difficulties in understanding and/or interpreting questions that seemed clear to us. However, in the context of the cognitive interviews we have to take into consideration the fact that we were also actively problematizing the measure and sometimes explicitly asking about existing problems and difficulties – therefore the participants may have focused on finding issues and so perhaps also reported minor ones.

With this in mind, we learned (or rather, we already knew this in theory, but our experiences during the cognitive interviews made this a much more urgent concern) that firstly, even seemingly very banal, easy and/or straightforward questions or items can be understood /interpreted quite differently by different participants. This concerns the content and clarity of the questions (c.f. Willis & Artino, 2013). Obviously, the solutions to this include formulating the items more clearly so they are easier to understand, not using foreign or complicated words, stating clearly the time frame or locality being asked about, paying immense attention whilst translating from one language to another, not using double negation, giving repeated information on unusual scales, etc.

Secondly, we have learned that sometimes different participants will adopt very different strategies when answering the same questions. For example, the questions may be answered from different viewpoints or using very different representations of the outgroup being investigated. In the SCM questionnaire, we wanted participants to answer them from a

Table 3: Sample question changes following two rounds of cognitive interviews

	Original question in the first draft	Changes after 1st round of cognitive interview	Changes after 2nd round of cognitive interviews
Problem 1: Whom did the participants have in mind when answering the questionnaires?	*"When you were filling in the questionnaire, who did you picture when you read the word 'Roma'? If you pictured different persons/ groups, please say who they were."*	The question remained unchanged; we just learnt how important it is to have this information.	The question remained unchanged; we just learnt how important it is to have this information.
Problem 2: Whose viewpoint should be represented in the answers?	*"Consider how Roma are viewed by Slovaks in general".*	*"Consider how Roma are viewed by the majority of people in Slovakia"*	No further changes needed. No problems understanding question now.
Problem 3: Assumed distinct items were logically interconnected	*"If a new Roma pupil came to your class, how would you feel? I would feel ok I would feel awkward I would be cautious I would feel agreeable I would be worried it would be no problem for me"*	*"If a new Roma pupil came to your class, how would you feel? I would feel ok I would feel awkward I would be worried it would be a problem for me"*	No further changes needed.
Problem 4: What is the period of time and locality referred to in our questions?	*"How often do you come into contact with Roma in public space?"*	Original draft question replaced by several more specific ones *"How often do you see Roma in everyday life?" "How often do you communicate with Roma in everyday life?" "How often do you spend time with Roma?" "How many of your friends have Roma friends you know about?" "How often do you encounter media news about Roma (e.g. newspapers, TV, the internet)?"*	No further changes needed. No problems understanding questions now.

Table 3: Continued

	Original question in the first draft	Changes after 1st round of cognitive interview	Changes after 2nd round of cognitive interviews
Problem 5: Translation and transferability of meanings from one language to another	*"To what extent do people in Slovakia tend to feel envy toward Roma?"* *"To what extent do people in Slovakia tend to feel jealousy toward Roma?"* *"To what extent do people in Slovakia tend to feel contempt toward Roma?"* *"To what extent do people in Slovakia tend to feel disgust toward Roma?"*	No change as comments on similarity of meaning between the pairs of questions were marginal	Order of questions changed : *"To what extent do people in Slovakia tend to feel envy toward Roma?"* *"To what extent do people in Slovakia tend to feel contempt toward Roma?"* *"To what extent do people in Slovakia tend to feel jealousy toward Roma?"* *"To what extent do people in Slovakia tend to feel disgust toward Roma?"*
Problem 6: Double negation	Item *"it would be no problem for me"* together with the grammatically negative end of the scale "1 – does not correspond to my feelings"*	Positive reformulation of the item: *"it would be a problem for me"*	No further changes needed.

267

specific viewpoint (that of the majority in Slovakia), and so we solved this by reformulating and clarifying the question, despite it not simply being a banal issue of unclear wording. In order to deal with sometimes very different representations of the outgroup, we came up with the solution of including an explicit "category representation question" at the end of the questionnaire to improve the researchers' understanding, e.g. in this case asking participants how exactly they were constructing the category of the Roma person or persons they were thinking of when answering the question.

Thirdly, the logic underpinning the construction of the questionnaire can become problematized if items that the researchers intended to be part of a subscale, e.g. for anxiety, are understood by the participants to measure something else or are simply not considered as part of the subscale. As the main part of this article was about how to mitigate the problems summarized in the first and second points, we did not dwell on the problems with scales reliability in any depth. We just mentioned them briefly when discussing changes to the second drafts of the questionnaires, but again, we would like to stress that a thorough statistical analysis of scale reliability and factor structure can reveal internal problems with the consistency of subscales. In other words, it can clearly show which items participants do not consider fit into the scale (even where researchers think they fit perfectly) and badly affect coherence. Based on this analysis we were able to discard several problematic items in the two questionnaires, and the new versions are more consistent. However, this method does not suffice alone (without being combined with cognitive interviews). A scale reliability analysis shows which items are weak, which cause problems for the (sub)scales and which can/should be removed. Statistics cannot be used to establish exactly why some of these items are weak and how they should be improved where we want to measure what we are interested in, but cognitive interviews can be used to achieve this.

In general, using cognitive interviews helped us to refine the questionnaire design and thus improve the quality of the questionnaire. We are fully aware that questionnaires are still not ideal, as all instruments of this kind contain language subtleties and ambiguities and although it is possible to enhance the validity of the questionnaires, producing "ideal" items is an unrealistic expectation.

To conclude, we think that if researchers use a questionnaire without having conducted cognitive interviews, they have only a very small degree of control over what they are "really asking" – what the participants understand the questions to mean and what they are answering. With cognitive interviews, the researchers can see where the main problems are and try to improve the questionnaire. However, at least for some areas of research, other data gathering methods seem to provide a more fine-tuned option, e.g. the qualitative interview, where researchers can explain a question to the participants when it is clear they do not understand the question or have interpreted it differently and the researchers can ask deeper questions if they do not understand. However, if we want to measure the effect of different interventions e.g. in dealing with bias against a stigmatized outgroup, we need to compare data about post-intervention attitudes with the control group data and for this we need to use quantitative measures. Thus, in our project we have created new versions (modified for Slovakia, for the Roma minority and for schools) of the two questionnaires measuring contact, attitudes, emotions, social anxiety, social distance and behavioural intentions. The cognitive interviews fulfilled their assigned function enabling us to improve these measures

to the extent that we now at least partially understand what we are "really" asking and what our participants are answering about.

References

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287-311.

Blair, J., & Brick, P. D. (2010). *Methods for the analysis of cognitive interviews.* Section on survey research methods – JSM 2010.

Bogardus, E. S. (1925). Measuring social distances. *Journal of Applied Sociology, 9*, 299-308.

Bradburn, N. M. (2006). Understanding the question-answer process. *Statistics Canada, 30*(1), 5-15.

Cuddy, A. J. C, Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*(4), 631-648.

Cuddy, A. J. C, Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS Map. *Advances in Experimental Social Psychology, 40*, 61-149.

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.P., ..., Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1-33.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357-411). New York: McGraw-Hill.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902.

Fiske, S. T., & North, M. S. (2014). Measures of stereotyping and prejudice: Barometers of bias. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 684-716). London: Academic Press.

Furnham, A., & Steele, H. (1993). Measuring locus of control: A critique of general, children's, health- and work-related locus of control questionnaires. *British Journal of Psychology,* 84(4), 443-479.

Gray, D., & Durrheim, K. (2013). Collective rights and personal freedoms: A discursive analysis of participant accounts of authoritarianism. *Political Psychology, 34*(4), 631-648.

Gutek, B. A, Murphy, R. O., & Douma, B. (2004). A review and critique of the sexual experiences questionnaire (SEQ). *Law and Human Behavior, 28*(4), 457-482.

Helfrich, H. (1986). On linguistic variables influencing the understanding of questionnaire items. In A. Angleitner, & J. S. Wiggins (Eds.), *Personality Assessment via Questionnaires* (pp. 178-188). Berlin – Heidelberg: Springer.

Kanovský, M. (2016). Robustné štatistické metódy v sociálnych vedách [Robust statistical methods in social sciences]. Bratislava: Slovenská asociácia sociálnej antropológie.

Lášticová, B. (2006). Identification with large scale social categories: A social psychology perspective. *Sociológia – Slovak Sociological Review, 38*(6), 546-561.

Lášticová, B., & Findor, A. (2016). Developing explicit measures of stereotypes and anti-Roma prejudice in Slovakia: Conceptual and methodological challenges. *Human Affairs, 26*(3), 233-252.

Mann, A., & Ruppeldtová, S. (2009). Pohodlné stereotypy o Rómoch [Indolent Roma stereotypes]. *Britské listy*, 23.02.2009 http://blisty.cz/art/45488.html#sthash.QWVQAOrb.dpuf

Mušinka, A., & Matlovičová, K. (2015). Atlas rómskych komunít na Slovensku 2013 ako pramenná databáza pre analýzu situácie Rómov na Slovensku a jeho potenciál pre ďalšie výskumy a analýzy. In T. Podolinská, & T. Hrustič (Eds.), *Čierno-biele svety: Rómovia v majoritnej spoločnosti na Slovensku* [Atlas of Roma Communities in Slovakia 2013 as a source database for analysis of Roma

situation in Slovakia and its potential for further research and analysis]. Bratislava: VEDA, Ústav etnológie Slovenskej akadémie vied.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly, 68*(1), 109-130.

Ryan, K., Gannon-Slater, N., & Culberstone, M. J. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation*, *33*(3), 414-430.

Tam, T., Hewstone, M., Kenworthy, J., & Cairns, E. (2009). Intergroup trust in Northern Ireland. *Personality and Social Psychology Bulletin*, *35*(1), 45-59.

Tourengeau, R., & Bradburn, N. M. (2010). The psychology of survey response. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research* (pp. 315-346).West Yorkshire: Emerald Group.

Turner, R. N., West, K., & Christie, Z. (2013). Out-group trust, intergroup anxiety, and out-group attitude as mediators of the effect of imagined intergroup contact on intergroup behavioural tendencies. *Journal of Applied Social Psychology, 43*(S2), 196-205.

Willis, G. B. (1999). *Cognitive interviewing*. A "how to" guide, from the short course "Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys". Paper presented at the meeting of the American Statistical Association.

Willis, G. B., & Artino, A. R. (2013). What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. *Journal of Graduate Medical Education,* September, 353-356.

Institute of Applied Psychology,
Faculty of Social and Economic Sciences,
Comenius University in Bratislava,
Mlynské luhy 4,
821 05 Bratislava,
Slovakia
E-mail: miroslav.popper@fses.uniba.sk

Institute for Research in Social Communication,
Slovak Academy of Sciences,
Dúbravská cesta 9,
841 04 Bratislava 4,
Slovakia
E-mail:magda.petrjanosova@savba.sk