UNIVERZITA KARLOVA V PRAZE

FAKULTA SOCIÁLNÍCH VĚD

Institut sociologických studií, katedra sociologie

Michal Kunc

Testing the van der Veer Et al. Fear-based Xenophobia scale

Bakalářská práce

Praha 2016

Autor práce: **Michal Kunc**

Vedoucí práce: **PhDr. Martin Buchtík**

Rok obhajoby: 2016

Bibliografický záznam

Abstrakt

Práce poskytuje stručnou teoretickou reflexi pojmu xenofobie a kvantitativního výzkumu tohoto konceptu. Používá analýzu reliability, konfirmační faktorovou analýzu a metodu Mokkenova škálování k testování reliability a vnitřní validity devíti položkové van der Veerovy škály xenofobie na vzorku reprezentativním pro ČR. Následně používá škálu jako index ke konfirmaci několika hypotéz pro zjištění vnější validity škály. Dochází k závěru, že škála je spolehlivá i platná pro použití v ČR, nicméně její výsledky jsou velmi silně ovlivnitelné pořadím odpovědí, a proto by měl z ní vzniklý index být používaný pouze jako ukazatel pro analýzu jiných proměnných, a nikoliv jako samostatná informace.

Abstract

The paper offers a brief theoretical reflection of the concept of xenophobia and a brief review of quantitative research on the topic. It uses reliability analysis, Confirmatory factor analysis and Mokken scale analysis to test the reliability and internal validity of the 9- tem van der Veer Et al. Fear-based xenophobia scale on a sample representative for the Czech Republic. It then uses the scale´s results to forma n index which is then used to confirm several hypotheses to examine the construct´s external validity. It reaches the conclusion that the scale is reliable and valid and suitable for use on the Czech population, however its results are too easily affected by the ordering of item options, and its results should therefore be used as an index to analyze other variables, not as a standalone source of information.

Klíčová slova

xenofobie, kvantitativní analýza, konfirmační faktorová analýza, Mokkenovo škálování, reliabilita, validita

Keywords

Rozsah práce: 65 802 znaků včetně mezer

Prohlášení

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

V Praze dne 13.5.2016 Michal Kunc

Institut sociologických studií
Projekt bakalářské práce

Prague, 10th April 2015

keywords: xenophobia, quantitative analysis, confirmatory factor analysis, reliability, validity

DISSERTATION TOPIC

The issue of xenophobia is and has been a topical one. Currently, with the media being substantially filled with news and commentary on immigration-related incidents and trends, it is arguably useful to use good quality research instruments to learn about the public opinion on such issues. Instruments that offer relevant data and are related to the essence of the matter of xenophobia – fear-based attitudes towards foreigners, and subsequently, immigrants. In my dissertation I would therefore like test one such instrument.

In the first section of the proposed paper, I plan to offer a comprehensive theoretical reflection of the concept of xenophobia and its meanings in social science and sociology in particular. After examining different definitions of the term (Yakushko 2009) I will present the results of my broad literature review of quantitative research instruments aimed to measure the concept of xenophobia. I am going to deal with instruments created as secondary analyses from big survey data (De Master and Le Roy 2000), instruments aimed at measuring specific subtypes of xenophobia (Watts 1996), or even xenophobia directed at specific social groups. (Lee et al. 2009) The description of these instruments will also serve as grounds to comparing individual conceptualizations and operationalizations of xenophobia. The result of this comparison should be the justification for choosing a particular scale and using it for collecting data and testing. This will be followed by its detailed description.

The scale chosen for analysis is a scale developed by van der Veer et al. In order to measure a single aspect of xenophobia, *"namely the fear, that 'the other' (i.e. the immigrants) can cause personal and societal harm"*. (van der Veer et al. 2011) The scale was initially developed by a team of experts as a pool of thirty items. These were then

reduced into a 14-item version, keeping only items directly related to fear-based attitudes. These 14 items were, through maximizing reliability (represented by Cronbach´s alpha) downsized into a 9-item Likert-type scale. This scale was then subject to Mokken scaling procedure[1], which yielded the indexes of scalability for each item and also a 5-item scale consisting of the 5 "best" items and ordered by their descending "difficulty" of positive (xenophobic) response. (van der Veer et al. 2011:34). The testing was done on three distinct samples, namely undergraduate students from the USA (608 respondents) and Netherlands (193 respondents). The third sample consisted of 303 Norwegian students enrolled in the psychology program at their university.

In my paper I am going to test the above mentioned 9-item scale (and its reduced 5-item subset), the data for which were gathered in the Czech Public Opinion Research Center´s "Our Society 10-15" survey (representative for the Czech Republic, N=1045) in October 2015. I will also describe some basic frequencies of the scales´ responses, followed by some of the more interesting correlations with the respondents´ demographics or other general opinions. After that I will briefly describe the methods of statistical analysis I will be using. I will not go into unnecessary detail regarding the rather conventional exploratory methods used, such as correlation analysis and principal component analysis (PCA). Instead I will describe the methods which are of essence for this paper – confirmatory factor analysis and scale reliability analysis.

Confirmatory factor analysis (CFA) allows us to test the goodness of fit of a proposed structural model to the observed data. In my paper I intend to use the AMOS software (Byrne 2010) to apply CFA to two models implied in the van der Veer Et al. Article (2011), which are based on the 9-item and 5-item scale respectively, and the explicitly stated intent for the scale to measure a single aspect of xenophobia, which I will reflect in constructing them as single-factor models. For both of these scales, I will also examine Cronbach´s alpha, to learn about their reliability, the *"accuracy of the measurement of concept using a particular operationalization"*. (Řehák 1998:51, translation MK)

---

[1] An IRT-based method examining the relationship between the presence of a latent trait and the probability of the respondent´s positive response to an item.

In the final section, I will describe the findings of the above listed analyses. In the discussion, I will, given the model comes out inadequate, propose possible causes of this fact. In case the model is confirmed, I will propose aspects of the findings which deserve further analysis.

PROPOSED OBJECTIVE

Test the van der Veer et al. fear-based xenophobia scale using new representative data.

METHODS

- Confirmatory factor analysis
- Scale reliability analysis

EXPECTED CONTENTS

1. Introduction
2. Context
   a. Theoretical and conceptual reflection
   b. Research Instrument Literature Review
   c. The Veer Et al. Fear-based Xenophobia Scale
3. Data
4. Scale Testing
5. Results
6. Discussion
7. Summary

REFERENCES

Byrne, Barbara M. 2010. Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.

Lee, Sherman a., Jeffrey a. Gibbons, John M. Thompson, and Hussam S. Timani. 2009. "The Islamophobia Scale: Instrument Development and Initial Validation." International Journal for the Psychology of Religion 19(2):92–105.

De Master, Sara and Michael K. Le Roy. 2000. "Xenophobia and the European Union." Comparative Politics 32(4):419–36.

Řehák, Jan. 1998. "Kvalita Dat I. Klasický Model Měření Reliability a Jeho Praktický Aplikační Význam." Sociologický časopis 34(1):51–60.

van der Veer, Kees, Oksana Yakushko, Reidar Ommundsen, and Laurens Higler. 2011. "Cross-National Measure of Fear-Based Xenophobia: Development of a Cumulative Scale 1." Psychological Reports 109(1):27–42.

Watts, Meredith W. 1996. "Political Xenophobia in the Transition from Socialism : Racism and Ideology among East German Youth." Political Psychology 17(1):97–126.

Yakushko, Oksana. 2009. "Xenophobia : Understanding the Roots and Consequences of Negative Attitudes toward Immigrants." The Counselling Psychologist 37(1):36–66.

TABLE OF CONTENTS

1 INTRODUCTION

Even if not mentioned explicitly, xenophobia, the fear of the foreign, and the attitudes, statements, policies or even important events resulting from it, have always been a topical issue. In European society, this topic has taken a new meaning in recent months (years, even) in connection to the so-called migrant crisis. Xenophobia and related topics headline the news and public discourse in general. Experts, laymen, and intellectuals of all sorts have been pointing out xenophobia in various forms, either criticizing it as unacceptable, or defending what was being called xenophobic as natural, rational, or otherwise legitimate.

Regardless of which of these sides prevails, xenophobia has in some respect been an important issue, and is one – nowadays especially. This fact should then allow us to argue, that it is not only pertinent to rigorously study and examine xenophobia as a social phenomenon, but also measure it using the appropriate research instruments. To assure an instrument is fit for such purposes and its findings are useful, and ideally internationally comparable, it is also necessary to test such instruments – and that is the purpose of this paper – to test the reliability and validity of a cross-national measure of fear-based xenophobia on a nationally representative sample.

In order to find an instrument to be tested, I first did a broad literature review of quantitative research on xenophobia and its instruments, which I describe in chapter 3 of this paper, along with some of the theoretical discussion of xenophobia as a concept. Having looked more closely at the research instruments, I then explain choosing the fear-based xenophobia scale (van der Veer et al. 2011) and describe the instrument in more detail.

In order to perform the tests, gathering appropriate data was necessary. This process, along with the particularities of the used research instrument are described in the Methodology chapter, chapter 4. Its second half then consists of the explanations of the measured concepts, reliability and validity, along with the particular methods, both from Classical Test Theory and Item Response Theory, which are used in the analysis.

The following chapter deals with analyzing the outputs of the particular analyses. The reliability measure Cronbach´s alpha is interpreted, along with two internal validity

analysis methods – confirmatory factor analysis and Mokken scale analysis, the latter of which is also compared to the results of the Mokken scale analysis performed in (van der Veer et al. 2011). The analysis is concluded by utilizing the scale as a fear-based xenophobia index to confirm hypotheses regarding other data from the respective survey, confirming its external validity.

The ultimate chapter of the paper connects, compares and discusses the most important findings from all chapters dealing with methodological and issues stemming from the instrument´s conceptualization and operationalization as well ethical concerns and general conclusions resulting from the results of the performed analysis. This is followed by a short summary of all findings.

2 THE AIM OF THIS PAPER

The goal of this paper, as has been stated above is to measure the validity and reliability of the fear-based xenophobia scale (van der Veer et al. 2011) on a Czech representative sample. This then leads to the following research question:

Is the Van der Veer et al. fear-based xenophobia scale applicable in the Czech Republic?

The particular sub-questions leading to the main research question and the related theoretical and/or methodological issues will be discussed in the following chapters – the particular sub-questions will be stated and discussed in their respective chapters.

3 THEORETICAL BACKGROUND

In this chapter I would like to discuss the term xenophobia itself with the ultimate goal to provide rationale for choosing one particular scale for analysis and also comparing other instruments, which aim to measure xenophobia, and show their varying levels of similarity to the tested scale. After generally introducing the term itself, this will be done by providing a brief account of approaches to measuring xenophobia in quantitative research and their inherent definitions of xenophobia.

3.1 DEFINING XENOPHOBIA

The term xenophobia is comprised of two Greek words: *xenos* (foreign, stranger) and *phobos* (fear). Probably due to its frequent usage in contemporary society, and the media in particular, its etymology is widely known. It is also apparent from the issues the term is usually adapted to describe, that it is used in a slightly more specific meaning, than simply the fear of foreign or strangers, but rather negative or hostile attitudes toward other cultures and their bearers.

Instead of taking that as a given, let us now have a look at how xenophobia has been described in works both theoretical and empirical. For the term itself, there has been a long standing vagueness in its definition (Canetti-Nisim and Pedahzur 2003), which then of course varies based on context, be it geographical, or that of a scientific discipline.

Quite a comprehensive account of the different aspects the term xenophobia has encompassed and encompasses can be found in (Yakushko 2009). In her paper she clearly states that xenophobia is a prejudice. And as such, it is constituted by three dimension: attitudinal, affective and behavioral. She also points out, that even while maintaining this framework, xenophobia can differ in its particularities – it is specific to a particular community and its relationship to foreigners (or the foreign in general).

The specifics of xenophobia as a term can be very well described in comparison to some of the other terms close to it. The first to be mentioned here is ethnocentrism. Since both xenophobia and ethnocentrism are based on a general preference to what an individual may see as their "own" culture, it comes quite intuitive to view them as two sides of the same coin. However, (Cashdan 2001) claims, that this is not the case. She points out, that *"in-group favoritism is not a necessary concomitant of out-group hostility"*. (Cashdan 2001:760) A presence of a perceived threat is then mentioned as an important factor for the social configuration to result into xenophobia.

Another term that is sometimes used interchangeably with xenophobia, but is not identical in meaning, is racism. As described in (Yakushko 2009), despite being often used to describe or characterize the same phenomena, unlike racism, xenophobia lacks the racial framework. The notion of being foreign is enough for xenophobia to be applicable – the perceived biologically inherited difference need not be present. Also, whereas racist conceptions are often used to describe mechanisms that are applicable worldwide,

xenophobia does not have to be applicable through geographical boundaries – as has been mentioned above, it is the particular relationship of the community in question to the foreign (although not necessarily) in its local context.

It is also pointed out in Yakushko´s paper, that xenophobia and racism as terms come from different historical contexts. On the one hand, racism comes from environments of subordination, such as colonialism or slavery, where the social relationship of superiority is determinant. On the other hand, xenophobia comes from environments of instability associated with massive migration, where then the confrontation with foreigners under such conditions constitutes the term. Also from a research perspective, there are different contexts to the uses of xenophobia and racism. In the USA, racism has been used to describe relationships and events on a more local level, however in Europe, racism has, throughout history, reached the levels of state politics. This situation renders xenophobia as a more acceptable term for the European scientist (Canetti-Nisim and Pedahzur 2003:320). This then transfers to common speech as well, where for example in Britain, the term *xenophobia* is used in similar quantities to the term *racism*, both to describe quite similar phenomena. In Germany, the word *racism* is so closely tied to Nazism, that *xenophobia* is used predominantly. (Banton 1996)

Since we have now briefly covered some of the aspects the term of xenophobia has, or does not, let us now look at two different theoretical models, which do not directly define it, but serve as a useful framework to better understand some of its inner specifics. The first is the Stereotype Content Model (SCM), which assert that out-groups and their members are rated by individuals along the dimensions of perceived "warmth" and "competence" (in implementing their intentions towards the in-group). (van der Veer et al. 2011:28) Based on this model a xenophobic attitude would be based on rating foreigners as un-warm and competent at the same time. The second model in question is the Integrated Threat Theory model (ITT). (Stephan and Stephan 2008:25) Despite not looking similar at first, it shares several common points with SCM. Describing a theory of prejudice (which xenophobia is, as has been established) based on perceived threats from the out-group, the authors discern two kinds of threats – symbolic (to an individual´s culture, way of life) and realistic (to their health, economic situation, etc.). To emphasize the similarities between these two models, I would like to point out, that this notion of threat is inherently based both on the mention un-warmness (hence the perceived ill wish from the

outgroup), and competence (to implement the threat, rendering it relevant) as to be found as one of the possible configurations in SCM.

It could be said, that all the descriptions of the term mentioned so far approached it as quite static, describing attitudes or beliefs. The work of Meredith Watts, however, adds a very interesting dynamic to xenophobia as a term, mostly thanks to the fact that having focused on her own concept of political xenophobia (as will be mentioned further on in this bachelor´s thesis), she tried to emphasize xenophobia´s instrumentality, bringing forth some of its more interesting sociological aspects – treating xenophobia as discriminatory potential. (Watts 1996) Watts does not dispute the above mentioned characterization of xenophobia as a prejudice. However, she identifies it as a force, which resides inside a culture. This force can potentially gradually corrode the culture it is present in, or erupt into an extreme event, it does not, however, do so on its own. In itself, it does not possess the ability to be an operational factor. For this, Watts explains, it must be given a political focus. This happens through a phase Watts calls targeting. This essentially consist of a group adopting the belief, that a certain foreign minority (or several foreign minorities) is responsible for a problem, or more abstractly speaking, a state of affairs that is perceived as undesirable (or its various aspects with varying sources and degrees legitimacy – immoral, illegal, unjust, etc.). This targeting then links, using mostly ideology or promoting the perceived threat, the, so far more or less distant, minority to the personal beliefs and needs of those targeting it, thus creating a demand for rectifying a problem presented as if caused by the targeted minority.

Watts then makes a very important observation, that since it is problems that the xenophobic group wants to be solved, since it is needs they want saturated, the negative and destructive extent, to which the presence of xenophobia in a society may lead, depends on whether its people believe in already established processes for institutions to deal with problems. A society, where institutions are perceived as being able to deal with problems, would then be quite resistant to the corrosive or eruptive presence of xenophobia. Also, analogically, even a small amount of xenophobia in a society which perceives its institutions as lacking competence, can have devastating effects. Watts therefore leads us from a definition of **xenophobia as a prejudice** having attitudinal, affective and behavioral aspects, to a one that is enriched by an important social dynamic – **xenophobia**

as a discriminatory potential, a dormant force, the actual consequences of which are defined by other social conditions.

With xenophobia being a term that is both ancient in etymology and contemporary in usage, there are many definitions to be used in research and theoretical work. The concept I use in this bachelor´s thesis is fear-based xenophobia, defined as *"namely the fear, that ´the other´ (i.e. the immigrants) can cause personal and societal harm"*. (van der Veer et al. 2011:29) Since this paper aims to test a quantitative research instrument, it simply adapts the conceptualization and operationalization used by said scale. To discuss its conceptual characteristics, I will use the following sub-chapter to provide a brief account of the types of conceptualization and operationalization of xenophobia used in quantitative research, to serve as a reference point to the tested instrument, and also partly as rationale for its choice.

3.2 XENOPHOBIA IN QUANTITATIVE RESEARCH

In quantitative research on the topic of xenophobia, I believe there are two main kinds of defining the concept for measurement to be encountered, both being, quite straightforwardly, shaped by the nature of their data. There are, of course other criteria that could be applied to these instruments, but based on the aim of this paper the criterion of the process of the instruments´ design is appropriate.

3.2.1 Scales derived from bigger surveys (secondary analyses)

The first kind are secondary analyses with xenophobia defined by a set of items from, typically a large-scale, survey. The common denominator of such measurements is that they were extracted ex-post from a larger piece of research. They usually were not created to directly measure xenophobia, but are treated as such. Therefore, they do not have necessarily have an a priori conceptualization of xenophobia other than being comprised of the extracted items, which may, or may not, have been designed to be related.

A general note: when looking at these different instruments, it is apparent, that xenophobia in quantitative research is most often measured indirectly through attitudes towards immigrants (and to a lesser extent foreigners). The broader context of xenophobia as simply fear of the strange or foreign is (if it is) represented by items asking about attitudes toward the immigrants´ cultures – both in particular and in general. Choosing the right attitudinal items to represent xenophobia is difficult, especially due to strong

negative connotations of the term, and therefore the presence of a social desirability bias. (De Master and Le Roy 2000)

To mention some of the instruments that fall in the current category – secondary analyses of xenophobia draw from large-scale survey such as ISSP, for example in (Hjerm 2001), where four items were used as a sufficient measure of xenophobia – those items deal with the attitudes to immigrants´ effect on crime rates, the country´s economy, job market and it becoming more open to new ideas. Another large-scale survey serving this purposes is Eurobarometer (De Master and Le Roy 2000) , where seven items were extracted. First two examined the general attitudes toward immigrants and asylum seekers – whether their numbers were too high and whether their presence was a big problem or not. The remaining five items asked about blaming immigrants for particular social issues, such as exploiting social welfare, increase in unemployment, causing problems for neighbors, violent crime and failed marriages.

The most common reason for not choosing such an instrument for analysis was it being too complex. A good example of such a case is the Special Eurobarometer Survey (different from Standard Eurobarometer) conducted in 1997 and 2001 that was dedicated specifically to racism and xenophobia (Thalammer et al. 2001). A large set of questions was subject to Exploratory Factor Analysis, which yielded seven different dimensions – blaming minorities, policies improving social coexistence, restrictive acceptance of immigrants, disturbance, multicultural optimism, conditional repatriation and cultural assimilation. Even though the article is no arbitrary choice of several items which resemble xenophobia, but quite a sophisticated approach, it still is a good example of the two main reasons why none of the instruments of the first kind was chosen. The first, minor one – the number of items was simply too high. Second, major one – like the other examples, the scale was extracted from data, not designed based on theory and then tested. I therefore kept looking for a scale that would be both compact and designed based on an a priori conceptualization.

3.2.2 Xenophobia-specific instruments

The choosing process therefore led me to instruments that were specifically created to measure xenophobia in general. My assumption was such instruments would be ideal to draw from, also based on the concerns mentioned in the previous subchapter.

Nevertheless, these concerns remained present – if an instrument fit such criteria, it remained too complex.

An example of such a too complex instrument was (Chtouris et al. 2014). By measuring attitudes of political xenophobia (a concept described above), authoritarian and anti-authoritarian attitudes, and the level of inter-group contacts of respondents, the researchers were later able to combine these data in a correspondence analysis, which allowed them to identify different groups among Greeks with xenophobic attitudes. If the Eurobarometer Special was too large, this instrument was unacceptable in the complexity of its method.

Looking for a scale to test, I abandoned the expectations of finding and instrument to measure xenophobia as a whole, and instead turned to more specific instruments for a possible solution. Despite having established earlier, that xenophobia in quantitative research focuses on attitudes towards immigrants (or foreigners, to some extent), such demarcation is still quite broad. A substantial number of academics therefore strove to use quantitative analysis to explore some of its more specific aspects. Two examples of different instruments of such quality will be described in this section – rather than focusing on the measurements themselves, I will try to point out their conceptual specifics. The first is the work on the measurement of *political xenophobia* (Watts 1996) – a conceptually specific instrument. The term of political xenophobia (also explained above) refers to the "*desire or willingness to use public policy to discriminate against foreigners*". (Watts 1996:97) According to Watts, such conceptualization is important, because, unlike xenophobia in general, it is operationable, having the backing in public policy, and therefore state-supported reinforcement, as opposed to xenophobia in general, which does not have this direct link to political power. It was possible to use this instrument, but it was decided not to, since it was based on comparing attitudes to a multitude of different particular nationalities, therefore being quite large as well.

The second example, the *Islamophobia Scale*, is specific in the objects of xenophobia it analyzes. (Lee et al. 2009) The instrument was developed to measure xenophobic attitudes towards Muslims. Therefore, it contains items, that are not transferable to other groups of people – other possible objects of xenophobia. Such items are for example supporting hypothetical policies to ban the building of mosques or opinions on Islam as a religion (it being the defining aspect of the analyzed out-group).

These two examples clearly show that xenophobia is a multi-faceted concept, to say the least, and that whole instruments deserve to be developed only to explore some of its more detailed aspects. However, due to their size and their high level of specificity, such instruments were also not suitable to be used in the here tested survey, as the intention to get as close as possible to measuring xenophobia in general remained.

Eventually, the instrument chosen for the survey and subsequent testing was the 9-item scale of fear-based xenophobia (van der Veer et al. 2011) *"defined by the irrational fear of immigrants and foreigners",* which will be described in detail in the following section. However, to compare its conceptual characteristics and to summarize on the findings of this chapter – its notion of xenophobia focuses on the negativity of attitudes towards immigrants and foreigners through asking about attitudes towards the immigrant´s potential attitudes, actions or effects on the in-group. In this, it is mostly in line with both the Stereotype Content Model (which it is based on), essentially looking for whether the respondent finds immigrant generally un-warm and competent, and with the model of Integrated Threat Theory as well, having items that deal both with symbolic and realistic threats. The aim of measuring xenophobia in general was not necessarily achieved, but a generally important, and theoretically justifiable aspect has been chosen. Given its reasonable size of 9 items, the instrument is well suited for the analysis in this paper.

At the end of this chapter I find it appropriate to mention some of my concerns. I do indeed realize that the list of instruments described is not exhaustive, although not too short either, and that choosing an instrument to test (along with its conceptual implications) using the above described criteria and rationale does not on its own ensure the instrument´s quality. In fact, some of the issues I tried to avoid (such as items being arbitrarily extracted) can still be present nonetheless – the possibility of a purposely designed scale also being an arbitrarily assorted set of questions and nothing more cannot be ruled out. I believe, however, that I made sure I do not have to take these issues as a given, but only as a risk, and after all, the whole purpose of the remainder of this bachelor´s thesis will be to find out, whether these choices were sound or not.

3.3 VAN DER VEER (ET AL.) FEAR-BASED XENOPHOBIA SCALE

The scale chosen for analysis is a scale developed by van der Veer et al. In order to measure a single aspect of xenophobia, *"namely the fear, that ´the other´ (i.e. the immigrants) can cause personal and societal harm"*. (van der Veer et al. 2011:29) This unidimensionality – operationalizing a single aspect of the wider concept of xenophobia using this scale – is the main assumption they later tested using MSA, and the testing of which is also described in this paper.

The scale was initially developed by a team of experts as a pool of thirty items. The items in general are statements about attitudes and/or beliefs related to immigrants and the effect immigration and the immigrants' presence in a country may and/or will have – either on the respondent as a person or their particular society. To these statements the respondents then answered with a particular level of either agreement or disagreement. The initial pool was then reduced into a 14-item version, keeping only items directly related to fear-based attitudes (see Figure 1).

Figure 1, the initial 14-item scale

ITEMS EXPRESSING XENOPHOBIA

1. Immigration in this country is out of control.
2. Borders should be made more secure to prevent immigrants from entering this country.
3. Immigrants cause increase in crimes.
4. Immigrants take jobs from people who are here already.
5. Interacting with immigrants makes me uneasy.
6. I enjoy interacting with immigrants (R).
7. It is OK for immigrants to stay close to their cultural roots (R).
8. I would welcome interaction with immigrants (R).
9. I worry that immigrants may spread unusual diseases.
10. I am afraid that in case of war or political tension, immigrants will be loyal to their country of origin.
11. I trust immigrants will support my country in times of crisis (R).
12. With increased immigration I fear that our way of life will change for the worse.
13. I doubt that immigrants will put the interest of this country first.
14. I am afraid that our own culture will be lost with increase in immigration.

Note.—Answer format was 1: Disagree strongly, 2: Disagree somewhat, 3: Disagree slightly, 4: Agree slightly, 5: Agree somewhat, 6: Agree strongly. (R) means the item was reverse-scored for analysis.

These 14 items were, through maximizing reliability (represented by Cronbach´s alpha) downsized into a 9-item Likert-type scale (see Figure 2). This is also the version of the scale that served as the basis for the research instrument utilized in this paper. Note, that all of the reversed items (marked R), were removed in this step of the item reduction

process. From then on the scale consisted only of statements framed negatively towards immigrants and their potential effect on the members of the majority and/or the majority itself. This may have been a problematic step in the creation of the scale resulting in a possible weakness, since, even though reversed items lower the value of the measure of reliability, it is the ordering of their options which may affect the scores of these items (Friedman, Herskovitz, and Pollack 1994), and thus could have served as a possible control in the instrument (as is also indirectly discussed in the final chapter of this paper).

Figure 2, the 9-item version of the fear-based xenophobia scale

ITEMS OF SUMMATED RATING SCALE TO MEASURE XENOPHOBIA

1. Immigration in this country is out of control.
2. Immigrants cause increase in crimes.
3. Immigrants take jobs from people who are here already.
4. Interacting with immigrants makes me uneasy.
5. I worry that immigrants may spread unusual diseases.
6. I am afraid that in case of war or political tension, immigrants will be loyal to their country of origin.
7. With increased immigration I fear that our way of life will change for the worse.
8. I doubt that immigrants will put the interest of this country first.
9. I am afraid that our own culture will be lost with increase in immigration.

Note.—Cronbach's coefficients alpha: USA = .77; Norway = .87; The Netherlands = .86; Scale ratings ranged from 9: Not xenophobic at all to 54: Extremely xenophobic.

This scale was then subject to Mokken scaling analysis, which yielded the indices of scalability for each item and also a 5-item scale consisting of the 5 "best" items and ordered by their descending "difficulty" of positive (xenophobic) response (van der Veer et al. 2011:34), also see Figure 3. The testing was done on three distinct samples, namely undergraduate students from the USA (608 respondents) and Netherlands (193 respondents). The third sample consisted of 303 Norwegian students enrolled in the psychology program at their university.

Figure 3, the final 5-item instrument derived from the results of a Mokken scale analysis

1. Interacting with immigrants makes me uneasy.
2. With increased immigration I fear that our way of life will change for the worse.
3. I am afraid that our own culture will be lost with increase in immigration.
4. Immigration in this country is out of control.
5. I doubt that immigrants will put the interest of this country first.

It would also be pertinent to mention, that its authors have performed a qualitative validation of the 5-item instrument – using the Three-Step Test-Interview method. The results of this qualitative validation were that there is not a general consensus on the understanding of the term "immigrant" and that the respondents mostly identify a difference between economic and political immigrants. (van der Veer et al. 2013:1436) This paper will not, however, examine this part of the original validation.

As has been mentioned already, the scale aims to measure a single aspect of xenophobia, which is also explicitly stated by its authors. It could then be argued that this should be evident from the scale´s name – I believe the attribute "fear-based" does not explain its actual focus on the fear of potential harm coming from immigrants. Since the conceptualization does not hide this though, this becomes more of a formal-detail issue than a one of conceptual debate.

In the following chapters I will describe how the step between a 9-item and 5-item version was, in a way, replicated on a sample representative for the Czech Republic. In order to offer more sensitivity of potential difference between the findings on the three student samples mentioned above and the Czech sample, the 9-item scale was used in a survey and a reliability analysis and quantitative validation were performed. To describe this process in detail, let us start with the characteristics of the data.

4 METHODOLOGY

4.1 SCALE

In this section I would like to describe the specifics of the particular instrument that was used for gathering the data used in the analyses in this paper, which despite being a translation of the above described scale, was not an identical copy of it. In the second part I give a description of the data file followed by basic frequencies of the scale´s item´s responses.

The 9-item scale (containing its 5-item subset) was translated into Czech by the Public Opinion Research Center team with the items being kept in the order of the original 9-item scale (see Figure 4). I believe the meanings of individual items were well preserved. However, the biggest differences are to be found in the offered responses. As should be apparent from Figure 4, the items´ options are in reverse order relative to the

original 9-item scale, with disagreement having the highest number and the value of agreement being number 1. There is also a *Do not know* (DK) option in addition to the agreement scale, which is represented by the number nine, as this is common practice in the survey the scale was a part of. It is, of course, possible that this will affect the results in some manner and this fact will be taken into account.

Figure 4, the Czech version of the 9-item xenophobia scale

POKYN: PODEJTE DOTÁZANÉMU KARTU OV.205							
OV.205 „Do jaké míry souhlasíte nebo nesouhlasíte s následujícími výroky?							
ROZHODNĚ SOUHLASÍ	SOUHLASÍ	SPÍŠE SOUHLASÍ	SPÍŠE NESOUHLASÍ	NESOUHLASÍ	ROZHODNĚ NESOUHLASÍ	NEVÍ	
1	2	3	4	5	6	9	
a) Imigrace v naší zemi se vymkla kontrole.				1 2 3 4 5 6 9			
b) Imigranti způsobují nárůst trestné činnosti.				1 2 3 4 5 6 9			
c) Imigranti berou práci lidem, kteří zde již žijí.				1 2 3 4 5 6 9			
d) Stýkat se s imigranty je pro mě nepříjemné.				1 2 3 4 5 6 9			
e) Mám strach, že imigranti mohou šířit neobvyklé nemoci.				1 2 3 4 5 6 9			
f) Obávám se, že v případě války nebo politického napětí budou imigranti stát na straně země svého původu.				1 2 3 4 5 6 9			
g) Obávám se, že se zvýšenou imigrací se zhorší náš způsob života.				1 2 3 4 5 6 9			
h) Pochybuji, že imigranti budou klást zájem naší země na první místo.				1 2 3 4 5 6 9			
i) Obávám se, že s nárůstem imigrace bude naše vlastní kultura ztracena."				1 2 3 4 5 6 9			

It is not clear from the original paper, whether the respondents from the student samples received the scale translated into the language of their country, or whether all of the data was gathered using English-language scales, although the former seems more intuitive. Even though a possible loss of validity or change of meaning is possible when translating scale items, in the case of this paper´s data, whether to translate was not a matter of debate, as the regular survey the scale was a part of is in Czech by default. Despite not following any rigorous scale translation procedure, the translation was performed by a team of experts and subject to multiple discussions until it was decided to keep this particular version. There is another ambiguous point in the original article, and

that is the number of the values of the scale. Even though a 6-point scale is described in Figure 2, a 5-point scale is mentioned in the text of the source article. (van der Veer et al. 2011:31) It was decided to use the 6-point scale, since it is more easily dichotomizable, which is needed to perform an MSA comparable the original article.

4.2 DATA

The data used in this analysis were collected by the Public Opinion Research Center at the Institute of Sociology of the Academy of Sciences of the Czech Republic. The scale was a part of the "Our Society 10-15" survey (N=1045), which was conducted in October 2015 of a representative quota sample of respondents of 15 years of age and more, with the quotas utilized being gender, age, educational attainment, place of residence and its size, along with NUTS-3 regions.[2]

If we look at the basic relative frequencies of the individual items of the scale, we see that the respondents have tended to agree with all of the xenophobic statements quite strongly, as opposed to the frequency of disagreement.

---

[2] This information was kindly provided to me by the Public Opinion Research Centre itself.

Figure 5, basic relative frequencies of the scale´s items

	Immigration in this country is out of control	Immigrants cause increase in crimes	Immigrants také jobs from people who are here already	Interacting with immigrants makes me uneasy	I worry that immigrants may spread unusual diseases
Strongly agree	23.4%	29.1%	22.9%	24.3%	39.3%
Agree	21.8%	27.0%	22.5%	22.3%	25.0%
Rather Agree	20.1%	28.2%	24.5%	25.1%	23.3%
Ratther Disagree	24.5%	12.0%	21.8%	19.6%	8.2%
Disagree	7.6%	2.3%	5.6%	6.6%	2.8%
Strongly disagree	2.5%	1.5%	2.7%	2.0%	1.5%

	I am afraid that in case of war or political tension, immigrants will be loyal to their country of origin.	With increased immigration I fear that our way of life will change for the worse.	I doubt that immigrants will put the interest of this country first.	I am afraid our won culture will be lost with the increase of immigration
Strongly agree	48.9%	39.0%	49.7%	30.8%
Agree	26.0%	29.6%	29.4%	26.6%
Rather Agree	18.9%	18.7%	15.3%	20.8%
Ratther Disagree	4.2%	9.2%	3.7%	13.9%
Disagree	1.2%	2.4%	1.3%	4.8%
Strongly disagree	.7%	1.1%	.5%	3.2%

The most frequent strong agreement was observed for the two items regarding the political loyalty of the immigrants. It is also important to mention that the amount of DK ranges from only above 30 cases to more than 150 in the case of the item regarding uneasiness during contact with immigrants. These frequencies, however, are not crucial to this analysis, as they only serve to allow the reader to get to know the response to the individual items and their data better.

4.3 METHODS OF ANALYSIS

In this subchapter I am going to deal with some of the matters connected to testing the scale – finding the answers to the main questions posed in this bachelor´s thesis. In order to do that, I will first explain the concepts of reliability and validity, and also the way in which they are central to this analysis, and then proceed to describe the particular

methods and procedures I will use to examine them. This chapter is not expected to be an exhaustive study of statistical methods of analysis, instead it aims to provide a description comprehensive enough to allow a person previously unfamiliar with the particular methods to understand the processes involved in this analysis and the ways in which its results are presented and interpreted.

I would like to make clear, that there are of course other methods that I have used during my research, but since they are not directly related to the aim of this bachelor´s thesis, or are quite notoriously known, they will be not described in detail, and only mentioned instead. Such method includes correlation analysis, two-sample t-tests (see 5.4), or principal component analysis (PCA). PCA was used to quickly explore the initial hypothesis, that the 9 items in the scale indeed form a single factor. Since it supported this expectation by putting all of the items in a single factor, which explained over 60% of the data´s total variance, it was decided to test this hypothesis with more detailed, testable methods.

In my analysis I will examine the whole 9-item scale as a baseline to be tested on the data, and also, when stated, the 5-item scale selected by MSA in the original article, in order to provide information about how this 5-item measure behaves in the Czech context.

4.3.1 Reliability and validity

The two main concepts around which scale testing revolves here, reliability and validity, are important concepts for research in general. In this subchapter, I would like to briefly explain them, in order to be able to later link their most important characteristics to the findings in my particular analyses. The level of detail I will go into should only be sufficient to serve this purpose, there is no ambition to provide an exhaustive or mathematical explanation of these terms. Every research instrument, however, every scale or index – every construct – should be arguably reliable and valid, in order for researchers to be able to justify its potential use and eliminate most possible doubts about their data and interpretation. Therefore, it is important to examine and test research instruments to measure their reliability and validity, which essentially sums up the aims of this paper.

The difference between reliability and validity may seem quite subtle, but it could not be more substantial. Reliability deals with the question of the accuracy and consistency of the measurement itself, to put it more exactly, the *"accuracy of the measurement of a concept using a particular operationalization"*. (Řehák 1998:51, translation MK) There are of course multiple ways to measure a scale´s reliability. In this paper the decision was to go with one of the most commonly used methods, and that is the calculation of Cronbach´s alpha. If a measurement is reliable in its measurements – we can expect it to give us reliable information about our data, it does not necessarily mean it measures what we have designed it to measure. This is where the concept of validity becomes useful.

Validity does not only deal with the measurement itself, but connects it to wider configurations, to put it briefly, it tells us to which extent we can safely assume that it is valid to operationalize a particular concept with a particular construct. It deals with the relationship of what we have measured, and what we wanted to measure in the first place. Therefore, the question of validity is not only one of calculation, but of theoretical background as well.

Validity is usually divided into two related and somewhat complementary concepts – internal validity and external validity. Internal validity refers to whether the instrument in itself (hence internal) really works in the ways we have intended it to – this can be done in multiple ways. In this analysis they are confirmatory factor analysis and Mokken scale analysis. Since our presumed model of measurement is a single-aspect one, we are effectively looking for the scale´s homogeneity. Despite the name, even the concept of internal validity is relational to external criteria (Řehák 1998:51), in this case the standards of the utilized analytical methods. External validity, where this relational character is explicit, refers to whether the instrument measures to what we have modelled it to measure relatively to things *outside* of the instrument itself, and is therefore whether it can be generalized into different instances as well. (Calder, Phillips, and Tybout 1982) In this analysis I do that by confirming hypotheses we make about the rest of the survey´s data and the way the scale explains it when calculated into an index.

4.3.2   The aims of the analysis

When it comes to the analysis itself, the goal set in the 2nd chapter of this paper can be further divided into the following sub goals:

- Analyze the reliability of the 9-item and 5-item versions of the scale using Cronbach´s alpha (see 5.1)

- Analyze the homogeneity of the 9-item and 5-item versions of the scale using CFA (see 5.2)

- Analyze the homogeneity of the 9-item version of the scale using MSA (see 0)

- Compare the MSA results regarding scalability of items and their ordering by difficulty to the results of the original article (see 0)

- Test the external validity of the scale by using its score to predict some of the relationships between its value and various other related variables from the survey (see 5.4)

4.3.3 Classical test theory

In this paper I am using methods coming both from the so called Classical test theory (CTT) and Item Response Theory (IRT), to come to conclusions regarding similar, if not identical questions. As is apparent from its name, the main focus of IRT is on the single item, rather the test itself. However, the main difference between these approaches is, that, unlike CTT, IRT does not assume that the items in a scale are all of the same difficulty, or "parallel" (van Schuur 2003). Instead, it treats the analyzed unidimensional scale as, in the case of MSA, cumulative in being representative of a single latent trait (in our case xenophobia). The representatives of CTT in this paper are Cronbach´s alpha and CFA, whereas IRT is represented by MSA.

4.3.3.1 *Reliability analysis*

The method used to answer the question of reliability of the utilized scale in this paper is Cronbach´s alpha. One of the easiest ways to describe this measure is a correlation between a summative index created from the items in a scale and an index of the same number of items which measures the same concept. (Řehák 1998:58) This then leads us back to the basic description of the concept of reliability, namely by assuming that this correlation between two indices is indicative of the extent the constructs results are determined by the measured concept, and not error in measurement. In this analysis I calculate Cronbach´s alpha using SPSS, along with the table of potential Cronbach alpha values after individual items have been removed from the scale. I will provide a Cronbach´s alpha for each of the models tested using confirmatory factor analysis (explained below)

and compare it to the threshold of .80, which is considered adequate for applied experimental settings, ideally reaching the .90 level, indicative for possible utilization in more general subsequent decisions. (Lance 2006:205)

4.3.3.2   Confirmatory factor analysis (CFA)

Confirmatory factor analysis belongs into the family of structural models. Despite sharing the name with the much more commonly used exploratory factor analysis, which is very similar in its usage (not calculations), to the previously described PCA, and serves as a means of exploring data to identify possible ways of dimension reduction, often by replacing sets of items by different factors, confirmatory factor analysis works in a different way. In CFA, we put in a previously designed structural model of how we presume a structure of factors to affect the observed data. The analysis then tells us via a set of statistics, how well the model proposed by us fits the data we have observed.

Of course to test the model fit, we should have a model that is somehow theoretically supported, ideally by more than a set of questionnaire of items being arbitrarily put together. (Byrne 2010:6) In this paper I use CFA (among others) to assess the claim that the scale measures a single aspect of xenophobia – this means putting all the items in the scale into a single-factor model. As the original study mentions a 9-item scale that was tested, and a 5-item scale that came out of this testing as a better version of the 9-item instrument, I will test both of these as single-factor models. In order to test items similar to those tested in MSA, the CFA was also done using dichotomized versions of the items (see 4.3.4).

I have created both of the models, as well as performed the confirmatory factor analyses, using SPSS AMOS, generally drawing from one of the most widely used textbooks on the matter. (Byrne 2010) In order to read the models properly, distinguishing between the separate involved variables´ shapes is important. A rectangular-shaped variable is a one that was in fact observed in the data – in our case this would mean the nine or five scale items. A round-shaped variable indicates an unobserved variable, one that was not directly addressed in then data, but is assumed to be a part of the structural model. This would of course be the latent trait of fear-based xenophobia. However, there is also a separate unobserved variable for each of the scale items. These are the presumed "errors", simply those sets of factors which influence the items´ values, but are not a part of the measured latent trait.
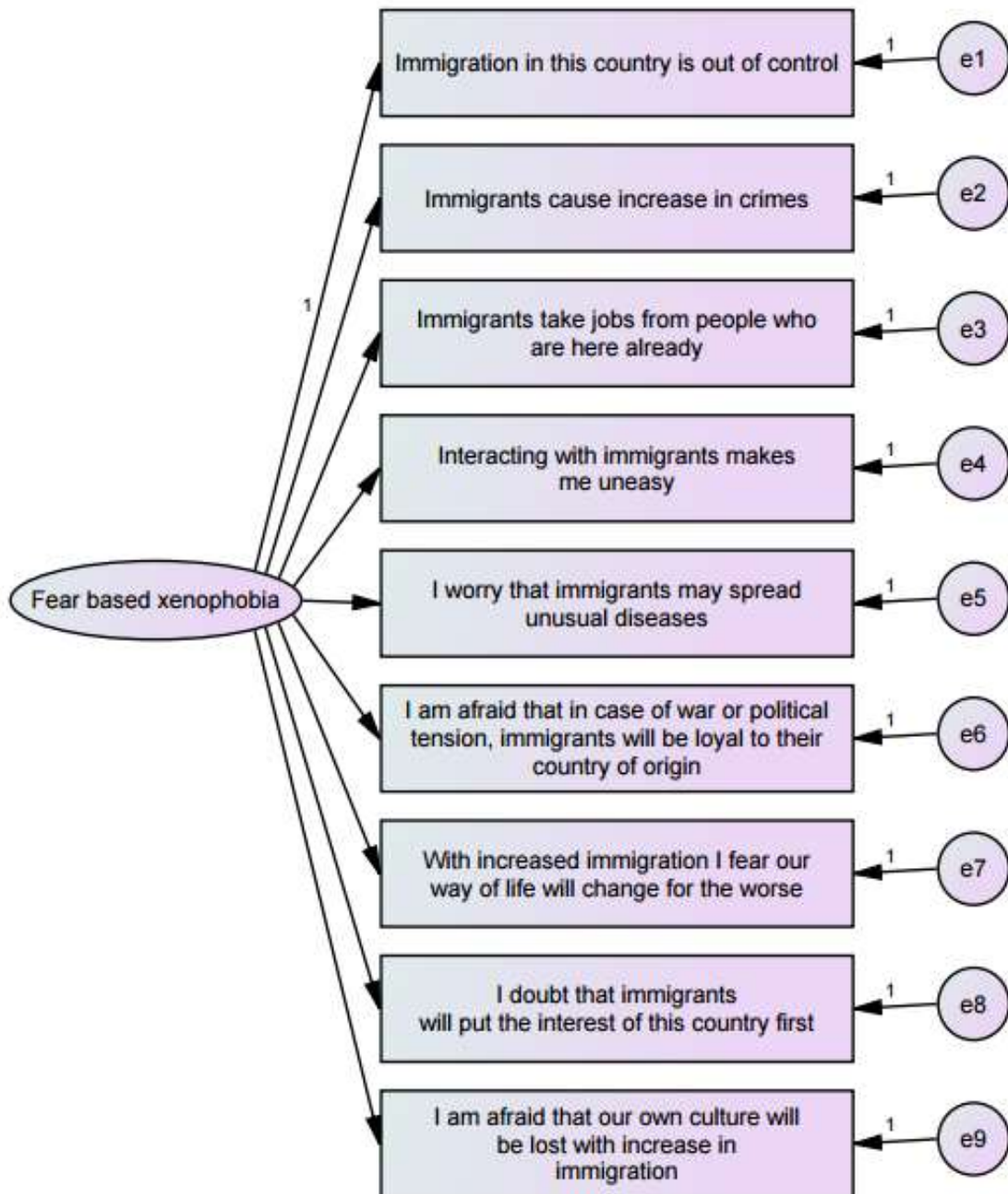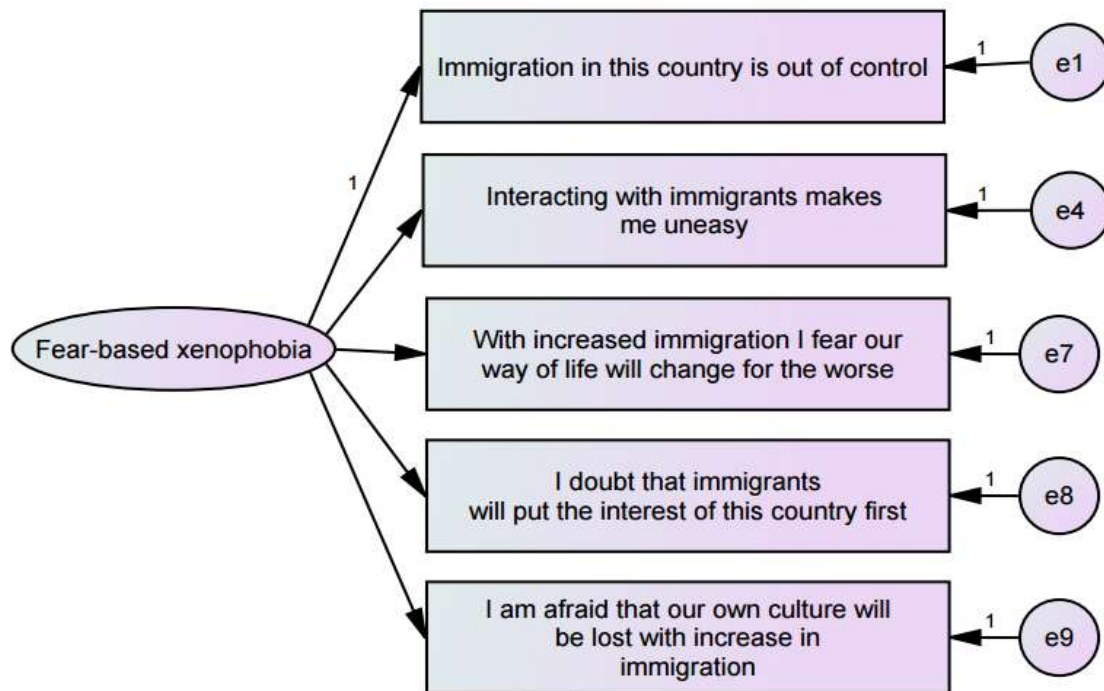
Figure 6, the 9-item single-factor model

Figure 7, the 5-item single-factor model

After the models have been built in AMOS and their estimates were calculated, I will mainly focus on model fit, various measures which tell us about how well the above constructed structural model fits our observed data. In this structure of interpretation and the choice of measurements for the results of the analysis I am generally following the example of (Buchtík 2012). As his paper quite clearly states, the number of model fit measures AMOS provides us with is far too large to be considered in its entirety, and I will therefore focus only on four of the most commonly used measures, and use this chapter to briefly explain the general rules of their interpretation, based mostly on (Arbuckle 2013).

- Chi-square – chi-square is quite commonly used in statistics. However, if we are using it to assess a model´s fit to the data, the logic is reverse to what we might usually encounter, e.g. as a test in cross tables. Since its null hypothesis states there being no significant difference between two elements (in our case the data and the model), not being able to reject it indicates a good model fit. We are therefore looking for p values as high as possible, ideally above the 0.05 level. However, with complex models that manifest a higher number of degrees of freedom, or have larger samples, expecting the p value to be over this threshold is unrealistic and a bad value does not necessarily indicate bad model fit. (Byrne 2010:76)

- RMSEA – a measure best explained by its full name – root mean square of approximation. The value of this statistic should always be below 0,1 for a well-fitting model, ideally even below 0.05. It is also accompanied by a Closeness-of-fit p-value, which should ideally be over 0.05.

- CFI – is one of the baseline comparison measures, meaning that it compares the analyzed model to a saturated (the proposed factors have the biggest effect possible) model of the same structure. Its values range from 0 to 1 tell us about the extent to which our model is similar to a saturated one, therefore the bigger the value, the better the fit, ideally reaching over 0.90, or even 0.95.

- AIC – The Aikake information criterion is a measure of model fit based on the concept of parsimony, the simplest possible configuration. This measure is defined as to decrease with goodness of fit, and therefore has us look for the lowest values in our analyzed models, rendering itself useful when comparing several models.

Also, if a model does not fit ideally, AMOS allows us to perform several types of post-hoc analysis, one of which is determined by Modification indices. These tell the amount of change in the models´ estimates caused by redefining some of the fixed parameters as free. This information can be then used to improve the model fit to a certain extent. Even though this does not necessarily fit the idea of confirmation (being post-hoc), modification indices will be utilized here as they do not overpower the whole factor analysis, and their fine-tuned models will only be used as additional information.

4.3.4 Mokken scale analysis

The second portion of scale testing is done using Mokken Scale Analysis (MSA), which belongs under Item Response Theory (IRT). As is common for IRT models, Mokken scale provides us with a response function for each item (IRF) which is indicative of the relationship of an item´s probability of an affirmative response and the presence of the analyzed latent trait (see Figure 8). This then serves both as a measure of reliability, identifying whether respondents with lower trait values are separated from those with relatively high trait values, and by ensuring that the separate items actually all refer to the same latent trait, validity as well. (Sijtsma 2005:875)

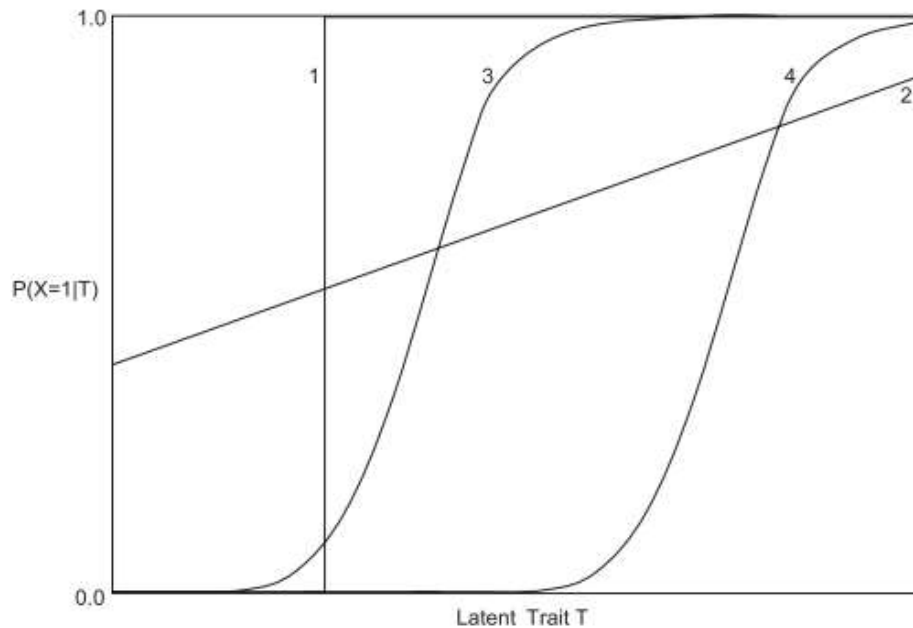*Figure 8, (van der Veer et al. 2011:33)*

FIG. 1.   Four item response functions for dichotomous items. Function 1 is deterministic; Function 2 is probabilistic but does not discriminate well between subjects with a large and those with a small amount of T; Functions 3 and 4 are probabilistic, and both differentiate well between subjects with different amounts of T.

The reason for including MSA in this paper despite not being mentioned in the project is being able to provide comparison with (van der Veer et al. 2011), which used MSA exclusively to test the characteristics of their scale. In this sense the analysis performed here can be viewed as a sort of a replication of that research on a nationally representative sample, but of course with the limitations and differences discussed in the Data chapter.

In this analysis, MSA provides us (among others) with the following information:

- A list of items that do form a cumulative Mokken scale (see the assumptions below)
- Their order in terms of difficulty
- The predictive power of the scale – what percentage of the respondents´ behavior can be explained by the scale

In the remainder of this section I would like to deal with the particularities which will eventually lead us to these results. MSA, being a nonparametric method, is designed for ordinal, and namely dichotomous, variables. (Sijtsma 2005:875) It is important to

note, that for the purposes of MSA, the data was recoded to resemble the replicated analysis (van der Veer et al. 2011:34), with the first three affirmative options being recoded into 1 (xenophobic) and the three negative responses into 0 (non-xenophobic), thus creating a dichotomous variable of a xenophobic response to the particular item.

The new results of MSA presented in this paper were computed using the open-source R statistical software, the *Mokken* package in particular. (Van der Ark 2015) All of the eventual results described below will be compared to those of the original analysis. This will be done by the means of a table, plots of different IRFs will not be provided, due to the limitations of the current R package, which does not offer the same visualization options as the commercial MSP software used by Van der Veer and their associates.

Let us now briefly walk through the separate steps of the analysis. Mokken scale analysis works under three main assumptions (Sijtsma 2005:876):

- Unidimensionality, an assumption meaning that the relationship between the different items of the scale can be explained by one shared latent trait (this is in fact the characteristic we are testing with CFA as well)
- Local independence, meaning that a particular respondent´s score on one of the items is not determined by the scores on the other items, but of the latent trait instead
- The Item Response Functions (IRF) in a Mokken Scale are monotonous and non-decreasing, meaning that "If all members of a set of items measure the same latent trait, then the ordering of the subjects by their probability of a positive response should be the same for all items." (van Schuur 2003:145)

To be able to analyze our items as a Mokken scale, we must first decide, which of the proposed items belong into a Mokken scale. This is determined by the Loevinger´s homogeneity (H) – the coefficient of scalability. The simplest way to view this coefficient would be to treat is as a ratio of the observed number of deviations from the assumed model and the number of such deviations that would have been expected if the items were not related, which is then subtracted from the number 1. It is common practice to treat items with an H lower than 0. 4 as weak for the respective scale (van der Veer et al. 2011:39) with items with a higher value being acceptable. In the original analysis, items with the H value as low as 0. 36 were also considered acceptable. In this analysis, the

scale is constructed by the Mokken´s automatic item selection process (aisp) (Van der Ark 2015:6), which by default includes all items with a H above 0. 3, displaying their scalability coefficients. The items were then ordered by hand in an ascending order by their mean value, with a higher mean value indicating a less difficult item.

Mokken scale works on a probabilistic, rather than a deterministic basis. (van der Veer et al. 2013) The last piece of analysis to be done related to MSA is identifying the percentage of respondents, who answered to the questions in accordance with the difficulty pattern.

5 ANALYSIS

5.1 RELIABILITY ANALYSIS

The results of the reliability analysis were positive for the scale. The values for Cronbach´s alpha (0.925 for the 9-item scale and 0.866 for the 5-item version) have not only reached the recommended value of 0.8, but for the 9-item version of the scale, Cronbach's alpha also exceeded the 0.9 value which indicates an instrument that can be considered generally reliable. (see 4.3.3.1) It is also indicative of the scale´s reliability, that removing each of the items would only lower the general reliability, and so there seems to be no items, which should be removed at this point.

5.2 CONFIRMATORY FACTOR ANALYSIS

In order to see, how well the two examined single-factor models work, let us first look at their diagrams with standardized estimates (see Figure 9 and Figure 10). From the factor loadings (numbers next to the arrows on the left) we can discern, that the items themselves are not particularly strong, but not too weak e, with the majority of them being over 0.6. As generally high loadings are desirable on variables we have assumed to be important in a factor, (Byrne 2010:6) this serves us to assert that none of the individual items´ relationship with the factor are particularly negligible in the model.

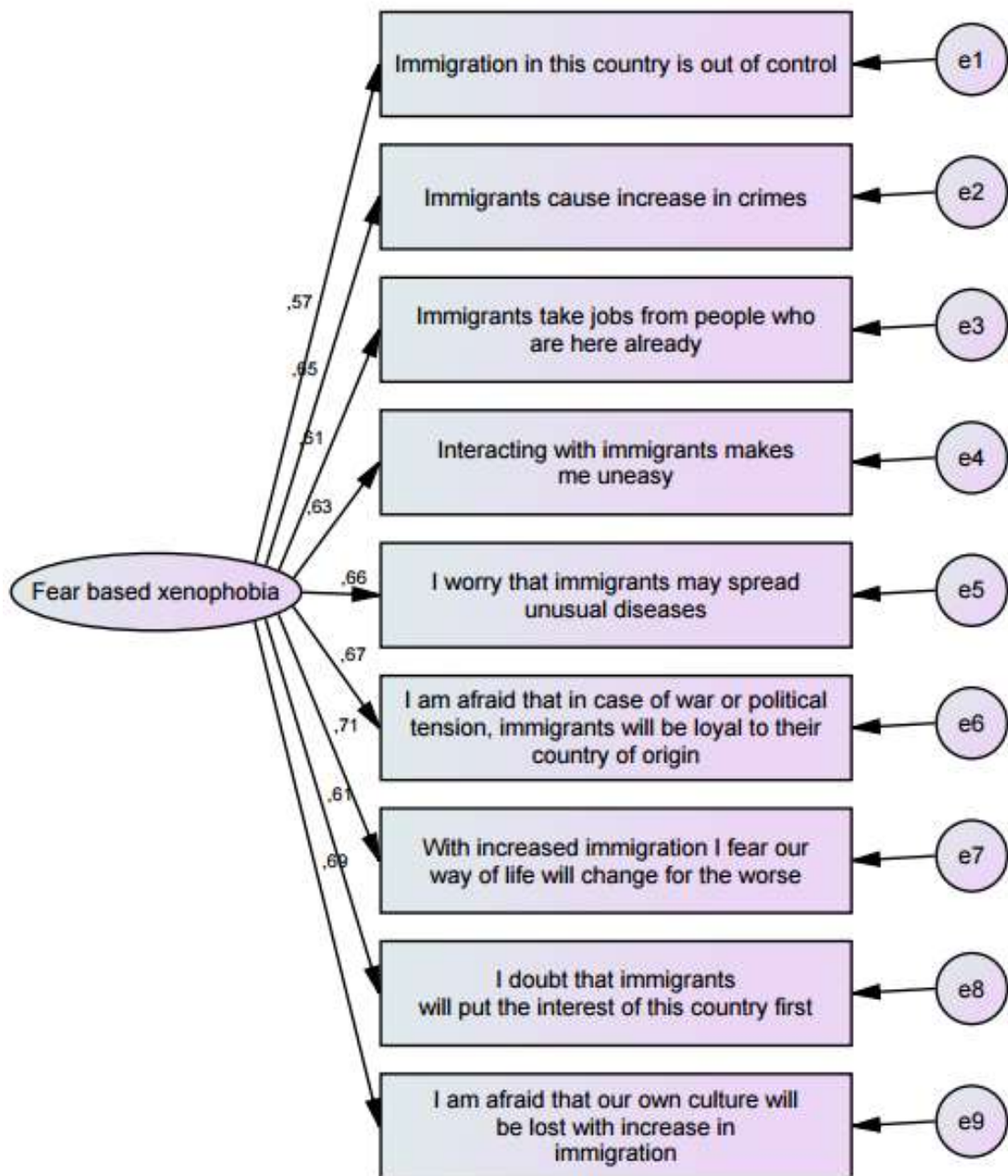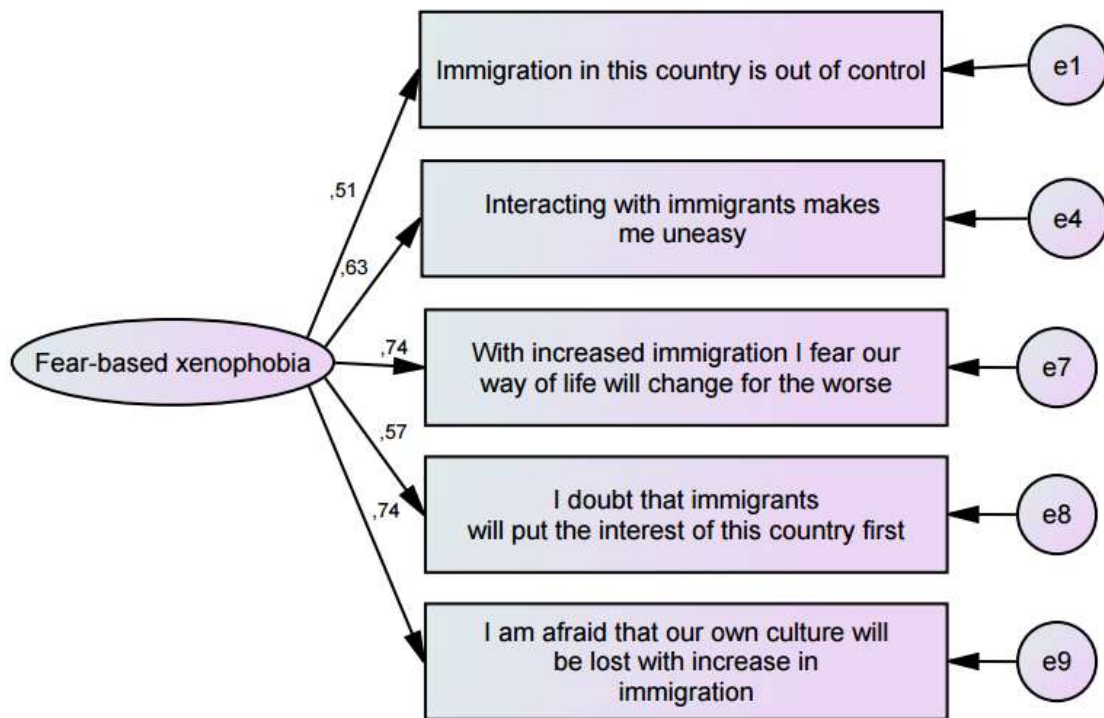*Figure 9, the 9-item single factor model with standardized estimates*

Figure 10, the 5-item single-factor model with standardized estimates

To conclude this section, I will interpret the model fit statistics that were calcu-lated for both of the models based on (Byrne 2010:76–82). For both of the models, the p-value for the chi-square test was far below the desired 0.05 threshold, indicating a less-than-ideal model fit. This is well in line with the statement mentioned earlier in 4.3.3.2, that expecting such values is most often unrealistic. The focus should be therefore shifted to other measurements. Not much revealing information also came from the AIC statistic, since the number of items and therefore also degrees of freedom differ greatly between the models. The finding that a lower relative value of AIC indicated the 5-item model to be better in terms of parsimony than the 9-item model, is also not surprising.

Interesting findings are therefore concentrated around the two remaining statistics – RMSEA and CFI, for which the two models will be now first described separately and then compared. In case of the 9-item model, the RMSEA was mediocre (0.099), however combined with a less-than ideal p-value (0.000), which was the case for the p-values of all the models mentioned here as well. The CFI statistic for this model was moderately good (0.916). Both RMSEA and CFI got better after the post-hoc modification (connec-tions were made between e1-e8 and e3-e8), with the statistics´ value shifting to 0.89 and 0.937 respectively. The model fit for the 9-item single-factor model remains mediocre at

best. The results for the 5-item single-factor model were quite different. The model itself started out with a relatively worse and generally bad RMSEA (0.114), but a reasonably good CFI value (0.948), almost reaching the 0.95 level. These two shifted substantially during the post-hoc analysis (modification indices indicated the two most effective connections to be made between e1-e8 and e7-e8). After the modification we were left with a reasonably good RMSEA (0.78) and a very good CFI (0.985). In the end, the 5-point model fared better in the analysis, especially under the assumption that applying the two modification indices was justified, than the 9-point model. Thus partially justifying the 5- point subset of the original scale and leaving the whole scale´s homogeneity as a 9-item scale uncertain.

Table 1, the results of MSA

New analysis				Van der Veer et al. Analysis**		
order	item	H	mean	H	mean	order
1	Immigration in this country is out of control	0.57	0.70	0.48	0.53	5
2	Immigrants také jobs from people who are here already	0.56	0.74	NA	NA	NA
3	Interacting with immigrants makes me uneasy	0.57	0.75	0.45	0.21	1
4	I am afraid our won culture will be lost with the increase of immigration	0.58	0.80	0.52	0.36	3
5	Immigrants cause increase in crimes	0.58	0.86	NA	NA	NA
6	I worry that immigrants may spread unusual diseases	0.58	0.88	NA	NA	NA
7	With increased immigration I fear that our way of life will change for the worse.	0.61	0.88	0.56	0.34	2
8	I am afraid that in case of war or political tension, immigrants will be loyal to their country of origin.	0.70	0.93	NA	NA	NA
9	I doubt that immigrants will put the interest of this country first.	0.65	0.94	0.49	0.49	4

*Items are ordered by decreasing difficulty of an affirmative response
**NA indicates that this item was not described in the original analysis, because its low H did not allow it to be a part of the creat Mokken Scale

5.3 MOKKEN SCALE ANALYSIS

Regarding Mokken scale analysis, the focus of this paper was on two main indicators: 1) the H scalability coefficient, indicative for which items fit the Mokken scale assumptions well enough to be a part of it 2) the items´ mean scores, the implication of which is the ordering by descending difficulty – from the item with the lowest mean score

to the item with the highest average value. These were both to be compared to findings from (van der Veer et al. 2011). The results of this analysis are to be found in.

As is apparent from Table 1, the scalability of the items in the data were generally substantially higher than those of the original analysis. This is best emphasized by the fact that none of the items' in fact went below the 0.5 value, making them not only all a part of the Mokken scale, but also indicate that the scale can be considered strong (van der Veer et al. 2011:39). The ordering of items by difficulty also differs from the original article. Of course, almost half of the items cannot be compared due to not being fit for the original scale, but from what we see, only three items (items 3,7, and 9 in the new ordering) at least maintained their original ordering.

Table 2, response patterns adherent to the cumulative model from Table 1

item*	1	2	3	4	5	6	7	8	9
pattern 1	0	0	0	0	0	0	0	0	0
pattern 2	0	0	0	0	0	0	0	0	1
pattern 3	0	0	0	0	0	0	0	1	1
pattern 4	0	0	0	0	0	0	1	1	1
pattern 5	0	0	0	0	0	1	1	1	1
pattern 6	0	0	0	0	1	1	1	1	1
pattern 7	0	0	0	1	1	1	1	1	1
pattern 8	0	0	1	1	1	1	1	1	1
pattern 9	0	1	1	1	1	1	1	1	1
pattern 10	1	1	1	1	1	1	1	1	1

Items are ordered according to decreasing difficulty of affirmative response

When looking at the resulting Mokken scale´s explanatory power, we want to calculate how many items fit in the ordering of items by their difficulty. This is done by first identifying all the possible response patterns that adhere to the cumulative model. (van der Veer and Higler 2013:90) There are ten (K+1 with K being the number of items) different patterns, in which the respondent may answer in accordance with the cumulative model (see Table 2). In (van der Veer et al. 2011:39) the percentages of such respondents were 56%, 67%, 58% for the Dutch, Norwegian and U.S. samples respectively. In this

analysis the discussed percentage was 64%, making the explanatory power of the model comparably useful to the previous studies.[3]

5.4 EXTERNAL VALIDITY

In order to be able to make some informed conclusions about whether the instrument in question actually provides us with relevant information about what can be considered a respondent´s value of what is considered fear-based xenophobia in this paper also related to measurements other than the scale itself – meaning if it is not only homogenous, but also externally valid.

The approach I have taken in this paper tries to make the most of the fact that the scale was administered as a part of a representative public opinion survey, and therefore also provides us with valuable information regarding the respondents´ socio-demographical characteristics or their attitudes regarding political and social issues. I will therefore make several basic hypotheses about said data and try to confirm them using the fear-based xenophobia as a quantitative index – a quite straightforward process since the scale contains no turned items.

To do this I have used SPSS to compute a mean index – a mean of all the items in the scale, since our MSA has deemed all of the 9 items as scalable, and also standardizing it to a range of 0-100 (instead of the previous 1-6 range), resulting in a continuous variable, seemingly in the form of a percentage score. This index is then used to test the following hypotheses.

H1: Respondents with a higher xenophobia score tend to have lower levels of educational attainment than those with a lower xenophobia scale.

H2: Respondents who agree that other nationalities are a problem in the whole Czech Republic tend to have a higher xenophobia score than those who do not.

---

[3] The percentage of respondents which adhered to the cumulative model was calculated in SPSS using a series of IF commands resulting in a relative frequency of a dichotomous variable of adherence to the model. This process is in itself quite straightforward and I believe it does not require its own methodological section.,

H3: Respondents who declare trust in Tomio Okamura tend to have a higher xenophobia score than those who do not.

H4: Respondents who declare trust in Jiří Dienstbier tend to have a lower xenophobia score than those who do not.

The choice of these hypotheses draws on various sources for its legitimacy, it is, however, not based on mere whim. Barring the inevitable fact, that the number of hypotheses that it was possible to test from this dataset is in fact much higher (and should be given their own separate analysis), the rationale of the choice are as follows: Hypothesis 1 comes from the fact backed by academic research, that the extent to which individuals manifest xenophobic attitudes decreases with ascending levels of educational attainment. (Hjerm 2001) Hypothesis 2 goes for the obvious assumption, that a questionnaire item, which is not a part of the scale, but carries similar meaning, should also have similar characteristics in its response. Hypotheses 3 and 4 are based on the current Czech political climate, where Tomio Okamura is a political party leader explicitly opposed to immigration and defending attitudes which are labeled as xenophobic in the public discourse, whereas Jiří Dienstbier is the Czech minister for human rights, who has been explicitly criticizing what are considered to be xenophobic attitudes and in favor of more open immigration policies. It is therefore assumed the relationship between the index´s score and the attitudes towards these politicians should be the direct opposite.

The statistical methods used to confirm these hypotheses were based on the particular character of the tested variables. With the educational variable being an ordinal one, I decided to use Spearman´s nonparametric correlation coefficient. The rest of the hypotheses regard dichotomous variables that represent either the agreement with a statement (alternatively the trust in a politician), or the lack thereof. These were then used as grouping variables for the standardized index in an independent two sample t-test.

The results of the correlation analysis and statistical test speak largely in favor of the fear-based xenophobia index. Higher levels of educational attainment correlate negatively with the index with a correlation coefficient of -.172, being statistically significant

at the three-star level[4], making its null hypothesis of no association between the two variables easy to reject. The case of statistical significance was similar (three stars) for all of the hypotheses subject to t-tests, with the exception of H4, where the difference was statistically significant only at the one-star level. Otherwise

If the hypotheses chosen to be confirmed regarding the external characteristics of the fear-based xenophobia index were chosen well, then in this representative survey it would appear externally valid, a useful index to discern between the responses of the more and the less xenophobic respondents. The mean differences in the standardized mean index for the respondents who wither agree with other nationalities being a problem in the Czech Republic or trust the politician in question are 20.7, 9.3 and -3.5 for H2, H3 and H4 respectively. This leads us to the somewhat cheerful conclusion that all of the four hypotheses have been confirmed in the data, rendering our instrument externally valid.

6 DISCUSSION

6.1 DISCUSSING THE RESULTS

In this final chapter I would like to connect, compare and discuss the findings I have made throughout this paper, especially in the analysis chapter, with the ultimate goal of answering the original research question of this paper and stating its most important issues. As was established in the previous chapter, the results of the analysis were good enough to consider the scale reliable on a Czech representative sample. When it came to internal validity, however, CFA left the 9-item scale on quite a shaky ground and spoke more in favor, although not entirely, for the 5-item version. However, the results from MSA provided us with a different story – all of the items (not only the originally chosen 5) were deemed scalable enough to be part of a unidimensional cumulative Mokken scale, and the original 5 items did not even manifest the best scalability, let alone maintain their ordering. Therefore, whereas CFA supported the original choice of items, the same analysis that chose them, led us to a different, and more sound in its statistics, conclusion. The obvious reason for this difference is the different approach of the methods, offering the possible conclusion, that the scale is more valid, if the items are not treated as parallel

---

[4]I adapt the common significance marking system of *** p<0.001; **p<0.01; *p<0.05

(CFA and CTT in general), but cumulative instead. By using MSA to back the question-able results of CFA for the 9-point scale, I would argue using the 9-point version of the scale is more justifiable. However, there are still some things to be addressed before the final conclusion is made.

6.2 METHODOLOGICAL CONCERNS

There is a difference in the mean response to all of the items, when compared to the three student samples that deserves to be addressed, since it might have affected the analyses as well. The first explanation at hand would be that simply the three student samples were too different in their characteristics as a sample from the general population sample, and therefore the original validation´s results were not applicable. This was how-ever addressed by the authors and argued not to be a concern. (van der Veer et al. 2011:29)

Another way to interpret these differences would be as stemming from a DK op-tion being added to the instrument, resulting into a substantial number of missing values, especially in the item regarding uneasiness when interacting with immigrants, which then, due to list wise deletion being used in the methods of analysis, effectively lowered the sample size by more than a hundred. Nevertheless, looking at the responses of those who chose a DK option in one of the items, there was no visible pattern in the response to the other items (a rigorous approach would, however, require its own analysis).

A second, and a much more arguable possible cause to the big difference in both mean values and eventually scalability is the switching of the order of items, which, by putting the agreement to the front of the scale, makes it easier for respondents to agree with the statements, and thus influences the result. (Friedman et al. 1994) As the numbers suggest, public opinion in the Czech Republic is definitely more xenophobic than not and this difference might have magnified this tendency.

6.3 SOCIAL-CONTEXTUAL CONCERNS

Methodological issues are definitely not the only arguable way in which to inter-pret the analysis results in question. It is possible that the origin of the difference is not methodological, and simply reflects the situation in the Czech Republic. It could of course be, that respondents from the Czech Republic are more xenophobic than from all of the three countries from the original analysis, but I believe it to be more probable, that time

is a factor here. The situation regarding immigration in Europe has changed, at least in the public discourse, substantially between 2011 a 2015 with immigration as an issue becoming one of, if not the main topic in the media and public debate. Therefore, the scale and the meaning of the tested concept may be taking a more tangible form than ever, possibly even acquiring a new meaning since the original analysis. All of these possibilities should be taken into account. They cannot, however, be verified in this data.

7 CONCLUSION

Taking into account the results from all the analytical methods combined, I would like to conclude, that the 9-item version of the scale is a reliable and valid instrument to measure what has been defined as fear-based xenophobia in the Czech population. There is, however a major additional imperative to this statement.

Since the scale´s results have shown themselves to be strongly influenced by the ordering of the items´ options, it becomes clear, that the results can be easily altered into one direction or the other. This, due to the possible political implications, constitutes an ethical problem in the scale. It is therefore advisable, to use the scale only in the form of a continuous variable, a cardinal index, used to discern between the more and the less xenophobic in relation to other characteristics, not as a characteristic itself.

8 SHRNUTÍ

V práci se podařilo provést stručnou teoretickou reflexi pojmu xenofobie, stejně jako popsat rešerši jeho zkoumání v kvantitativním výzkumu. Ta byla zároveň použitá jako vysvětlení volby konkrétní škály k testování. Van der Veerova devíti položková škála xenofobie byla detailně popsána, a po popisu pro ČR reprezentativního výběrového souboru byla testována její reliabilita a vnitřní validita za pomocí Cronbachovy alfa, konfirmační faktorové analýzy a Mokkenova škálování. Tyto metody a jejich interpretace byly v práci rovněž popsány. Výstupem interpretace byl poznatek, že nástroj je spolehlivý, nicméně údaje o jeho vnitřní platnosti se liší. To bylo, vzhledem k odlišným přístupům metod, očekávatelné a po kombinaci výsledků obou analýz byla škála označena jako vnitřně validní. Její vnější validita byla ověřena úspěšnou konfirmací 4 hypotéz. Po diskusi možných metodologických a sociálně kontextuálních příčin výsledků analýz byl učiněn závěr, že škála je vhodná pro užití na české populaci, nicméně že její výsledky jsou příliš snadno ovlivnitelné pořadím odpovědí, a její výsledky by tedy měly být používány jako kardinální index pro popisování a analýzu jiných proměnných a nikoliv jako samostatně platná informace.

9 REFERENCES

Arbuckle, James L. 2013. *IBM SPSS Amos 22 User's Guide.*

Van der Ark, L. Andries. 2015. "Mokken Scale Analysis in R." *R* 1–35. Retrieved May 1, 2016 (https://cran.r-project.org/web/packages/mokken/mokken.pdf).

Banton, Michael. 1996. "The Cultural Determinants of Xenophobia." *Anthropology Today* 12(2):8–12.

Buchtík, Martin. 2012. "Měření Soudržnosti Třídního Kolektivu: Interdisciplinární Přístup." *Pedagogika* 62(2):147–66. Retrieved (http://userweb.pedf.cuni.cz/wp/pedagogika/files/2013/08/Pedag_13_2_Měření_Buchtík.pdf).

Byrne, Barbara M. 2010. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.*

Calder, Bobby J., Lynn W. Phillips, and Alice M. Tybout. 1982. "The Concept of External Validity." *Journal of Consumer Research* 9(3):220–44.

Canetti-Nisim, Daphna and Ami Pedahzur. 2003. "Contributory Factors to Political Xenophobia in a Multi-Cultural Society: The Case of Israel." *International Journal of Intercultural Relations* 27(3):307–33.

Cashdan, Elizabeth. 2001. "Ethnocentrism and Xenophobia: A Cross-Cultural Study." *Current Anthropology* 42(5):760–65. Retrieved (http://www.jstor.org/stable/10.1086/323821?origin=JSTOR-pdf).

Friedman, Hershey H., Paul J. Herskovitz, and Simcha Pollack. 1994. "The Biasing Effects of Scale-Checking Styles on Response to a Likert Scale." *Proceedings of the American statistical association annual conference: survey research methods* (1972):792–95.

Hjerm, Mikael. 2001. "Education, Xenophobia and Nationalism: A Comparative Analysis." *Journal of Ethnic and Migration Studies* 27(1):37–60.

Chtouris, Sotiris, Anastasia Zissi, George Stalidis, and Kostas Rontos. 2014.

"Understanding Xenophobia in Greece: A Correspondence Analysis." *European Journal of Sociology* 55(01):107–33. Retrieved (http://www.journals.cambridge.org/abstract_S0003975614000058).

Lance, C. E. 2006. "The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say?" *Organizational Research Methods* 9(2):202–20.

Lee, Sherman a., Jeffrey a. Gibbons, John M. Thompson, and Hussam S. Timani. 2009. "The Islamophobia Scale: Instrument Development and Initial Validation." *International Journal for the Psychology of Religion* 19(2):92–105.

De Master, Sara and Michael K. Le Roy. 2000. "Xenophobia and the European Union." *Comparative Politics* 32(4):419–36.

Řehák, Jan. 1998. "Kvalita Dat I. Klasický Model Měření Reliability a Jeho Praktický Aplikační Význam." *Sociologický časopis* 34(1):51–60.

van Schuur, Wijbrandt H. 2003. "Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory." *Political Analysis* 11(2):139–63. Retrieved (http://pan.oupjournals.org/cgi/doi/10.1093/pan/mpg002).

Sijtsma, Klaas. 2005. "Nonparametric Item Response Theory in Action: An Overview of the Special Issue." Pp. 875–82 in *Encyclopedia of Social Measurement*, vol. 25, edited by K. Kempf-Leonard. New York. Retrieved (http://apm.sagepub.com/cgi/doi/10.1177/014662210122032028).

Stephan, Walter S. and Cookie White Stephan. 2008. "An Integrated Threat Theory of Prejudice." Pp. 23–47 in *Reducing Prejudice and Discrimination*, edited by S. Oskamp. Psychology Press.

Thalammer, Eva, Vlasta Zucha, Edith Enzenhofer, Brigitte Salfinger, and Günther Ogris. 2001. *Attitudes towards Minority Groups in the European Union*. Vienna.

van der Veer, Kees et al. 2013. "Psychometrically and Qualitatively Validating a Cross-National Cumulative Measure of Fear-Based Xenophobia." *Quality & Quantity* 47(3):1429–44. Retrieved (http://link.springer.com/10.1007/s11135-011-9599-6).

van der Veer, Kees and Laurens Higler. 2013. "Developing Cumulative Scales." Pp. 78–97 in *Advances in International Psychology*.

van der Veer, Kees, Oksana Yakushko, Reidar Ommundsen, and Laurens Higler. 2011. "Cross-National Measure of Fear-Based Xenophobia: Development of a Cumulative Scale." *Psychological Reports* 109(1):27–42. Retrieved (http://www.amsciepub.com/doi/abs/10.2466/07.17.PR0.109.4.27-42).

Watts, Meredith W. 1996. "Political Xenophobia in the Transition from Socialism : Racism and Ideology among East German Youth." *Political Psychology* 17(1):97–126.

Yakushko, Oksana. 2009. "Xenophobia : Understanding the Roots and Consequences of Negative Attitudes toward Immigrants." *The Counselling Psychologist* 37(1):36–66.

10   ANNEX