

Suitability in law enforcement

Assessing multifaceted selection criteria

Miguel Inzunza



Department of Applied Educational Science
Educational Measurement
Umeå University
No. 8 · 2015

This work is protected by the Swedish Copyright Legislation (Act 1960:729)

ISBN: 978-91-7601-373-1

ISSN: 1652-9650

Cover art and design by Björn Sigurdsson

Electronic version available at <http://umu.diva-portal.org/>

Printed by Print & Media, Umeå University

Umeå, Sweden

October 2015

© Miguel Inzunza

Abstract

When selecting applicants for professions with job descriptions involving complex human interaction, identifying suitable candidates is essential. This strongly applies in police selection, because police officers must act appropriately at all times, deal with difficult and sometimes dangerous work situations, and come into contact with people who are in highly emotional states, such as anxiety, fear, excitement, and shock. However, suitability is a broad and vague concept that cannot be even defined without understanding the value of various personal skills and qualities for specific duties, ways that such skills are manifested during life as an officer (potentially both on- and off-duty), and appropriate ways to assess them.

The overall aim of this thesis was to explore the personal skills and qualities that are relevant to suitability for the police profession with a special focus on the concept of empathy, and to investigate how it can be conceptualized and measured in the context of selection. Two of the four studies focused on different perspectives regarding the personal skills and qualities of a good police officer, while the other two focused on theoretical aspects and practical measurement of empathy.

Mixed methods were used, in acknowledgement of the value of using different forms of data collection, material, and analytical methods to achieve valid information.

The results and findings support several of the personal skills and qualities that have been identified in previous research and also can be found as criteria in police selection. Empathy should be considered in the selection process, but it is also important to ensure that there is an alignment between the definition of the concept and practical outcomes in police work. The findings indicate that empathy could be conceptualized in a relevant way using a social cognitive neuroscience (SCN) approach. In particular, a modified, Swedish 4-factor version of the Empathy Assessment Index (based on recent theoretical advances in SCN) appears to have considerable potential for further development.

Future studies will focus on areas identified as requiring further research, such as the development of this instrument by adding complementary measures focusing on manifestations of empathy and other relevant qualities in actual work-related situations. Other aspects that require further attention include the definition, assessment, and impact of social desirability in selection contexts.

Keywords: selection criteria, measurement, empathy, self-report instrument, validity, interpersonal qualities

Acknowledgements

Writing this thesis has been a valuable experience for me. I have learnt a great deal about the world of research, and the northern parts of Sweden (more specifically Umeå). Both experiences have enriched me and many people have contributed to this positive experience. First, I would like to express my gratitude to my supervisor Christina Wikström for all her support in my research. Thank you for sharing your knowledge so generously, giving insightful feedback on my work and valuable comments on my English. I would also like to thank my assistant supervisor Tova Stenlund, for providing guidance on structure and helpful feedback to improve my research. Both of you are also motivational people, who I think bring the best out of any person you guide. I also express my thanks to Widar Henriksson as an assistant supervisor for wise suggestions and comments concerning measurement aspects in the first two years of my studies.

I would also like to thank my closest PhD student colleagues and friends: Jonathan Wedman, my lunch companion, who always has interesting stories and experiences to share, and Eva Knekta, for sharing her knowledge about the “nature life” in the northern parts of Sweden. I will also think of the valuable discussions we all had in our previous office, where we considered problems and solutions to issues ranging from complex research questions to aspects of life as PhD students. I also feel lucky to have been introduced to the whole measurement unit at the Department of Applied Educational Science, where all topics are discussed.

As a researcher from another part of the country with criminological experience but no connections to the police field in the northern parts, I am grateful to several persons for opening doors regarding my doctoral work. I am thankful to Lars-Erik Lauritz from the police unit in Umeå University for being so open to research. For the involvement with the RECPOL project, I thank Staffan Karp and Simon Wolming from Umeå University. In this project, I got to know representatives from Catalonia, particularly Lola Vallès who became a valuable contact person there. Another person who has been valuable for my research is Stefan Anell from the Swedish Defence Recruitment Agency (SDRA). I am thankful for our inspirational discussions and for sharing all kinds of challenges, such as trying to make organizations more open to the benefits of research. In those attempts, I thank you Stefan for introducing me to Rose-Marie Lindgren from the SDRA, who I also would like to thank for her openness to research. Special thanks are also due to all the reviewers and editors for providing feedback and suggestions to improve my work.

Furthermore, I would like to express my gratitude to all participants in my studies and to all of you who have in some form contributed to my research.

I would also like to thank my former teachers, and later colleagues and friends from Sociology and Criminology at Kristianstad University for preparing me with valuable theoretical perspectives in social sciences. For the practical help and layout of the thesis, I thank Björn Sigurdsson.

Finally, I would like to express my gratitude to my family. My parents I thank for always believing in me. My father for being such a positive person, regardless of the situations confronted with and my mother who will always be in my heart. I will especially think of her whenever I do something good as she is the reason for it. I am grateful to my sister Cecilia who is in my heart and my brother Luis for always being there, providing support in whatever I set my mind to. I wish you and your family the best. Lastly, but most of all, I thank my dear Anette for being supportive and for having enriched my life in many positive ways, especially with my two beloved sons Dario and Marlo, you mean the world to me.

Miguel Inzunza

Umeå, October 2015

Articles

This thesis is based on the following articles, which are referred to in the text by the corresponding Roman numerals:

- I. Inzunza, M. (2015) *Selection practitioners' views on recruitment criteria for the profile of police officers: A comparison between two police organizations*. Manuscript submitted for publication.
- II. Inzunza, M. (2015) *Validating a measure of the 'successful/good police officer' using data from recruits in two different police systems*. Manuscript submitted for publication.
- III. Inzunza, M. (2015) Empathy from a police work perspective. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 16(1), 60-75. doi: 10.1080/14043858.2014.987518
- IV. Inzunza, M. (2014) Adaptation and development of the Empathy Assessment Index (EAI). *International Journal of Comparative and Applied Criminal Justice*, 39(3), 239-255. doi: 10.1080/01924036.2014.989245

All published articles are reprinted with the permission of the respective copyright holders.

Table of Contents

| | |
|---|------------|
| Abstract | i |
| Acknowledgements | ii |
| Articles | v |
| Table of Contents | vii |
| 1. Introduction | 1 |
| Aims and research questions | 2 |
| Disposition of the thesis | 3 |
| Philosophy of science | 4 |
| Terminology | 4 |
| Demarcations | 5 |
| 2. Selection | 6 |
| Logics of selection in general | 6 |
| The Swedish police selection system | 7 |
| The Swedish selection system in a wider perspective | 8 |
| 3. Suitability within the police context | 9 |
| Aspects relevant over time and context | 9 |
| The Swedish context | 11 |
| 4. Concurrent views and the measurement of empathy | 12 |
| Views of empathy | 12 |
| The measurement of empathy | 14 |
| 5. Validity theory | 15 |
| The classical view of validity | 15 |
| The modern view of validity | 16 |
| Practical considerations | 17 |
| Reflections on validity | 18 |
| 6. Methods | 19 |
| Validity considerations | 20 |
| Data and participants | 21 |
| Literature search | 21 |

| | |
|--|-----------|
| Focus groups | 22 |
| Common factor modelling | 23 |
| Ethical considerations | 26 |
| Reflections concerning the methodological choices | 26 |
| 7. Summary of studies | 27 |
| Study I | 27 |
| Study II | 28 |
| Study III | 29 |
| Study IV | 30 |
| 8. Discussion | 31 |
| Discussion concerning the overall aim | 34 |
| Validity considerations for measurements of constructs applied in police selection | 36 |
| Conclusions | 38 |
| Limitations and future research | 39 |
| References | 42 |
| Appendix | 51 |

1. Introduction

When selecting applicants to join an education programme for a profession, the process of deciding who to select and the grounds for decisions are extremely important. In modern democracies fairness and validity are critical issues in this context (Gilliland, 1993; Messick, 1980; Wolming, 1999). A first step towards establishing a fair and valid selection process is to construct a profile that matches descriptions of the requirements to be successful, or to perform well in the intended context. This will provide robust foundations for the selection criteria. In most cases, the focus will be on finding applicants who will perform well in the education programme, and the main criterion will be academic achievement. Furthermore, in most cases appropriate preparation for the job, i.e. the vocational education, is assumed to be fulfilled if an individual passes the courses and other requirements in the education programme. However, in some cases the selection process must also take into consideration each applicant's suitability, i.e. whether he or she has the personal qualities required to become a member of the profession, such as a teacher, social worker, police officer or other "street-level bureaucrat" (Lipsky, 1980). These professionals make decisions that may profoundly affect people's lives in daily contact situations with the public. Thus, their behaviour and actions may potentially either resolve or inflame difficult situations. Some responses (positive or negative) are associated with the organizations and specific constraints that can be typical, such as the shortage of resources or the culture and work climate. However, others may be related to individual professionals' approaches to solving problems and behaviour in specific situations. Then suitability is critical, but it is often difficult to define. It may be possible to teach applicants at least some elements of suitability, or develop them through appropriate training or education. However, some elements may be innate and much more difficult (or impossible) to change, especially in the limited timeframes of most educational programmes. Therefore, personal suitability is an important aspect to consider during the selection, education, training, professional development and working lives of such professionals.

Personal suitability is particularly important in law enforcement professions, for several reasons. Firstly, educating a person who will drop out or fail is costly, in terms of both time and money (DeCoster-Martin, Weiss, Davis, & Rostow, 2004). Secondly, candidates' entry to the professions has other costs and consequences that may be difficult to predict (Hibler & Kurke, 1995), but social costs may be very high if inappropriate applicants are selected, or successful candidates are inappropriately educated and trained. No other profession is entrusted with so much authority to use force, if necessary, in contact situations with citizens

(Cochrane, Tett, & Vandecreek, 2003; Hibler & Kurke, 1995). Hence, the educational preparation for this profession is only one part of the requirements for the job. In Sweden, the number of students admitted to the basic training program for police officers matches the number of police officers required, so if they pass their courses they are guaranteed positions as police officers. Thus, to be eligible for the education and profession there are both formal requirements (such as a sufficient level of previous education and possession of a valid driver's license), and requirements based on the personal skills and qualities considered relevant for a good police officer. This approach is supported by research indicating that personal suitability is an important factor for a successful career as a police officer (Dantzker, 2011).

However, including requirements related to personal skills and qualities is not straightforward since they involve conceptual constructs (hereafter constructs) for which perceptions and definitions may be highly time-, situation-, culture- and context-dependent. Another issue is whether the key constructs should identify desired or undesirable characteristics of the applicants. In this respect, it is worth noting that previous research have recommended not only screening out undesirable applicants but also looking for individuals with desirable characteristics (Burbeck & Furnham, 1984; Guffey et al., 2007; Sanders, 2008). Given the importance of fair and valid selection, it is important to ensure that any characteristics used are relevant, well-defined and measured in a valid and reliable way.

Aims and research questions

The overall aim of this thesis was to explore the personal skills and qualities that are relevant to suitability for the police profession, focusing particularly on empathy, its conceptualization and its measurement in selection contexts. More specific aims were to address two related research questions. The first aim concerned the nature of suitability for positions as law enforcement officers, and the question posed was *What personal skills and qualities are relevant to suitability from different perspectives?* The second aim was to conceptualize and measure a quality that appears to be relevant for suitability in this context, empathy, and the question posed was *How can empathy be conceptualized and measured in a relevant manner for the police profession?* These aims and research questions were addressed in four studies outlined below.

Study I investigated important requirements and constructs by analysing official documents guiding selection procedures of police organizations in Sweden and Catalonia to address questions such as *What are officially valued as important criteria for police officers, and what selection procedures do each organization apply?* The information in the official documents or statements was then complemented with subjective

perspectives from selection practitioners. A major objective was to illuminate similarities and differences between the two organizations regarding the qualities deemed essential for the police profession.

Study II investigated recruits' attitudes regarding a good police officer, based on data from the European Longitudinal Study of Recruitment, Education and Careers in the Police (RECPOL). The study included validation of part of an instrument intended to measure attitudes towards competencies relevant for a successful/good police officer, in order to make valid comparisons between two of the participating organizations. A second aim was to investigate possible differences and similarities between them.

Study III investigated how to conceptualize a model of empathy relevant for a police context, based on the main aspects of empathy as defined historically and currently. An objective of this study was to include recent advances in empathy research and assess their value in a police context. It was concluded that social cognitive neuroscience (SCN) theory provides a valuable approach.

Study IV focused on ways to measure empathy practically, and included the translation and adaptation of an empathy-measuring instrument, the Empathy Assessment Index (EAI; Gerdes, Lietz, & Segal, 2011). Modifications including the addition of items designed to assess interpersonal components of empathy identified as being potentially important in Study III. The study also addressed empirical validity issues related to use of the modified instrument in a Swedish context and the convergent validity of the constructs with other conceptually valid measures. Research questions were: *Which measurement model is most appropriate for the translation and adaptation of the EAI for use with the chosen groups of participants?*

A unifying element of the studies was the common concern to consider pertinent validity aspects, either explicitly or implicitly, since validity is one of the most important aspects when investigating theoretical constructs. This approach is further discussed in the validity chapter of the thesis.

Disposition of the thesis

The thesis consists of a summary and four studies. The following chapters will provide background information to the different studies and put them in a context. Chapter 2 briefly describes the logic and issues relevant to selection in general, and then specific circumstances related to the Swedish process for selecting police applicants. Chapter 3 discusses suitability in a police context, from both international and Swedish perspectives. Chapter 4 describes the development of empathy theory and how the concept has been measured in other contexts. Chapter 5 discusses validity, including the classical perspective, the modern view, and practical considerations of validity relevant to Studies I-IV. Chapter 6 describes the methods used and

choices made when carrying out the studies. Summaries of the four studies are presented in Chapter 7. Chapter 8 discusses and interprets the findings from the studies and their value for future research. Chapter 9 sequentially presents the four studies.

Philosophy of science

Producing knowledge is an important part of research. However, “knowledge” is not a straightforward concept because its value in analyses of real phenomena depends on how closely it reflects reality. Thus, particularly given the complexities of the constructs involved in the research reported here, regarding the nature of pertinent skills or qualities in selection contexts, the nature of reality (ontology) should be considered first (Berger & Luckmann, 1966; Jacquette, 2002). Should the world be treated as something solid and constant, or is it important to understand variations in perceptions of a phenomenon? This is especially relevant when applying the models of theoretical constructs frequently used in social sciences. For example, loyalty could be defined and perceived in different ways by people from different times and places due to cultural differences in the importance attached to various aspects of loyalty. A combined approach has been adopted in the different studies of this thesis; conducting research to produce knowledge, in terms of both identifying existent problems (for instance in selection, where real decisions have to be made on a daily basis) and how different concepts representing skills and qualities are perceived from different perspectives.

Terminology

The terminology used in this thesis can vary somewhat, due to variations in the ways concepts describing relevant skills or qualities have been presented in the literature. This section attempts to clarify some of this variation. The research presented in this thesis deals with qualities such as empathy and honesty that are valuable in selection, but not directly observable. This is because behavioural patterns or responses to a specific stimulus may be observable, but not their causes, so such internal causes or attributes are hypothetical constructs (MacCorquodale & Meehl, 1948). Studying phenomena that are not directly observable is a central activity in the social sciences. However, the terminology adopted can vary between disciplines and may also depend on whether phenomena are being considered on a theoretical or at a measurement level. For instance, at a theoretical level, such phenomena can be described as unobservable concepts, unobservable variables or latent constructs, where the use of construct refers to construction in the minds of the researchers (Lee & Lings, 2008). Use of the term concept is appropriate when trying to name a specific phenomenon, or class of phenomena, and a concept may become a formalized concept

following descriptions of what it includes (Lee & Lings, 2008). This distinction is not always very clear in the literature, but valuable to reflect upon. When moving to the measurement level, the commonly used terminology is usually latent variables or factors (Bollen, 2002).

Another issue associated with the terminology used is that qualities can be described in different ways. A frequently recurring discussion on whether certain qualities should be described as traits or abilities conveys some of the associated difficulties. For example, empathy is sometimes described as a trait, implying that it is something that is innate. This is consistent with some recognized aspects or dimensions of empathy, such as “mirroring”: the ability to simulate another person’s emotional state, mediated by the mirror neuron system (Iacoboni, 2008). However, it is not consistent with some other dimensions, such as the ability to take another person’s perspective, because (according to a common assumption at least) a person with experience of the observed person’s situation will have greater ability to do this than someone who lacks such experience. This implies that empathy is something that can be trained, or at least is dependent on cognitive development (Roan et al., 2009). Hence, in social psychology literature, for instance, there is often a dynamic view of the construct as an ability that can be increased by training or experience (Holm, 2001). This is important to reflect upon when referring to qualities that contribute to what is regarded as suitability in a selection context, because if an important quality is considered to be innate it might be essential to assess its strength in candidates early in the selection process, while if it is an ability it might be possible to train them to display it after entering the organization.

Demarcations

The research this thesis is based upon has focused mainly on the Swedish police context, although Studies I and II included perspectives from other police organizations, in order to obtain indications of the generality and constancy of recurrent concepts in police contexts in both literature and practice. Including comparative analysis of other organizations also broadened the cultural contextualization. In this respect it should also be noted that most research conducted globally is published in English, and a large proportion of police-related research is conducted in and refers to conditions in Anglo-Saxon countries (particularly the USA). This is important to keep in mind, as there are major differences between police forces in the USA and European countries (particularly countries that do not have Anglo-Saxon traditions). Some of these differences are pertinent in selection contexts, for example there are shortages of qualified applicants to fill positions in American police departments (White & Escobar, 2008), while there are more qualified applicants than positions in Sweden (see The Swedish police selection section). However, police work involving intimate

contact with citizens (street-level policing, handling witnesses, contact with victims, etc.) has often been seen as a similar endeavour in all modern police organizations, indicating that a similar competence profile could be applicable in most contexts.

2. Selection

Logics of selection in general

As already mentioned, it is widely accepted in modern democracies that selection procedures for higher education and official positions should be fair and efficient. However, fairness in this context is not straightforward and hence is a frequently debated issue, as illuminated by Wolming (1999), who discusses selection for higher education from egalitarian, utilitarian, and meritocratic perspectives. He argues that fairness in selection is not absolute and that it requires reflection on the demands and context.

Every selection system follows some logical rules or principles. Egalitarian, utilitarian and meritocratic principles can be described as follows. Egalitarian fairness has generally been described as striving to ensure that all participants have equal opportunities. This means that in competitive procedures for the allocation of scarce resources no irrelevant criteria such as sexual orientation, ethnicity, or other characteristics of individuals should matter, and only criteria that are strictly relevant should be considered, to allow equality of opportunity (Jacobs, 2003). Utilitarian distribution of resources is based on maximising utility, in other words, maximizing human happiness (Mulgan, 2007), or (in this context) society's interest. In a meritocratic system the best or most merited candidates are selected (Brezis & Crouzet, 2006). In contrast to the egalitarian approach, which promotes use of external, non-personal attributions, the meritocratic approach emphasizes the abilities and virtues of the individual (Kornai, Rothstein, & Rose-Ackerman, 2004). In the meritocratic ideal, the rights of the individual are paramount. Selection in this model is based solely on merit, rather than parentage (as in many previous inheritance-based systems). This idea rose with the emergence of liberal democracies and was applied in most selection systems that replaced nepotism and family connections with concepts such as personal effort and ability (Zimdars, 2007). It is important to note that this model allows competition among applicants for places.

Decision theory is also relevant here, as it concerns the methods used to gather and analyse information, and the subsequent strategies applied for making decisions (Urbina, 2014). In practice, it concerns estimations of the value of correct decisions versus costs of incorrect decisions, which are often considered in monetary terms. The foundation of utility analysis exists in the

premise that measures will lead to more rational and productive choices regarding people and that the measures are important for getting support when presented to decision takers (Boudreau & Ramstad, 2003).

Further parameters to consider include the selection ratio (the number of applicants per position) and base rate, the proportion of recruits who succeed in the profession (Urbina, 2014). Having many more applicants than study positions (a small selection ratio) is beneficial for the profession, since an instrument with reasonable predictive validity can be used to help ensure that suitable candidates are selected (Urbina, 2014). A basic model can then be further developed to include multiple predictors in different combinations, each providing indications of the degree to which applicants have (or lack) a specific quality needed in the profession (Boudreau, Sturman, & Judge, 1994). Ideally a set of selection methods should also be used (Le, Oh, Shaffer, & Schmidt, 2007), hence most complex professions require successful applicants to display a set of qualities deemed appropriate in a multiple hurdle procedure, as seen in selection processes for the police.

The Swedish police selection system

A centralized process for selecting applicants to join the Swedish police was introduced in January 1998, in conjunction with the launch of a new police education programme (SOU, 2008:39). This selection process admits students to the education programme, and hence profession (see the description of the situation in Sweden in the Introduction) twice a year. There is a large number of applicants and few study places, thus there is a small selection ratio. Indeed the ratio has fallen recently from approximately 10 to one (SOU, 2008:39) to 9015 applicants for just 367 study places in 2011 (<https://polisen.se>). A new element of this centralized selection process was a pronounced emphasis on the importance of personal suitability for the police profession (Ds, 1996:11). The selection process adopted more of the internationally recognized principles of police selection, where personal suitability and psychological concepts are key criteria (Lough & Von Treuer, 2013). The National Swedish Police Board contracted out the admissions process to the Swedish National Service Administration in efforts to ensure that the centralized selection process was more structured and efficient than the previous system. This is a government agency responsible for carrying out recruitments for national services, recently renamed (from 1 January 2011) the Swedish Defence Recruitment Agency (SDRA).

As described by Annell (2012), the selection process is based on several steps, each with a different purpose. First, there is a rough screening based on the documented merits of the applicants. Finer screening is then conducted, based on medical exams, physical tests, a language test, a cognitive ability test, and a psychological interview, all used to rate the applicants. A last, competitive selection step for those who have not been screened out includes

extended interviews with both police representatives and psychologists, who then decide who they think are most suitable. A criminal background check, based on records and a specific interview, is also included before the final decision is made by the National Swedish Police Board.

The process has recently (2014) been subject to some changes, but it follows the same rationale as before, using several methods for evaluating the various qualities that are considered valuable for the profession. Requirements for eligibility are listed, and include a minimum age of 18 years, Swedish citizenship, swimming ability, a driver's licence class B (car), and previous education to upper secondary level with specific minimum accomplishments in certain subjects, (<http://www.rekryteringsmyndigheten.se>). The applicants submit the required information and answer questions regarding their background through a Web portal. Thereafter all the applicants are ranked according to their previous school merits and their responses to the background questions. The top-ranked applicants are invited to take further tests at centres in Stockholm, Gothenburg or Kristianstad. These tests include evaluations of physical ability, medical condition, cognitive ability, and psychological aspects based on a specified profile. The information generated from these tests is then evaluated through a psychological interview, which is most relevant for the final decision. Appraisals of the qualities addressed in the multiple-hurdle procedure are summarized in a competence profile under the headings personal maturity, responsiveness and flexibility, commitment and responsibility, patience and thoroughness, and communicativeness (RPS, 2012).

The Swedish selection system in a wider perspective

When viewing the Swedish police selection model from a theoretical perspective, it can be concluded that the present model embodies elements of more than one selection principle, but primarily follows a meritocratic principle of choosing the best or most merited people for the profession (Brezis & Crouzet, 2006). This is consistent with legal guidelines and expectations of the applicants. Using the multiple hurdle procedures in the selection process is expected to provide the information needed to choose the most qualified and suitable applicants. The procedure should also meet egalitarian goals of striving to provide equal opportunities and only applying relevant criteria (Jacobs, 2003). Utilitarian principles are also involved, as the organization has a history of guidance by social objectives. A general societal expectation of the Swedish selection system is that recruits will represent the society they serve in terms of gender, ethnicity and cultural diversity (SOU 2007:39). This idea dates back in time, although the demographic constitution (and hence expected proportional representation) was very different. For instance, a 1948 governmental commission recommended that police should be recruited from diverse social and

occupational strata (Furuhagen, 2015). Demands for recruitment of women and other groups became stronger later, to boost the organization's legitimacy and efficiency (Stensöta, 2004).

Viewing the police selection process from a historical perspective may also be valuable for understanding the potential importance of the utilitarian principle of seeking to do what is good for all (or as many as possible), which is highly relevant for the police. For such understanding of the police's role in society, social contract theory is helpful. Briefly, the theory describes the citizen's agreement to abide by certain rules in return for receiving protection (Locke, 2000). In the modern state, the role of the police is then to secure the order required for society to function and enable government of a particular territory (Loader & Walker, 2001). A good relationship between the police and the citizens is therefore essential, because otherwise trust will fall, with negative consequences at both individual and societal levels. The individuals representing the police in contacts with the citizens will clearly play a key role in maintaining the good relationship, emphasising the importance of recruiting suitable people (Cochrane, Tett, & Vandecreek, 2003). Carefully considering the work descriptions of a police officer, including the significant degrees of autonomy and discretion, is also clearly critical (MacVean & Neyroud, 2012). In summary, recruiting suitable individuals as police officers is beneficial for everyone.

3. Suitability within the police context

Having established the need for a rigorous police selection process, the next issue to address is what to look for. As noted by Burkhart (1980), before asking how "the best" or most suitable applicants should be chosen it is necessary to address the question "Best at what?", i.e. to establish the key criteria. Clearly they should include the personal suitability of the applicants as future police officers, but there is no universal consensus regarding the personal skills and qualities to include, and (more importantly) the connections between these qualities, their descriptions and their value in actual police work need to be more thoroughly elaborated.

Aspects relevant over time and context

The qualities seen as important for a police officer might depend on time, culture and other contextual elements related to the mission and procedures of the law enforcement organization. For instance, closeness to the citizens is a generally recognized feature or aim of police organizations rooted in Anglo-Saxon traditions, e.g. in US contexts, hence there is a relatively strong focus on valuable qualities for maintaining good relations with the citizens compared to other traditions. In the US context, this can be seen in the

history of attempts to recruit civilians representing groups and geographical areas where the police are active, with the intention to minimise gaps between the police and citizens. A different trend has been seen in police organizations more related to a military tradition. For example, in Germany or France recruitment has predominantly been from former soldiers but also from people from the outside, for instance from the countryside to the towns, with the purpose to maintain a distance to the group they were policing (Furuhagen, 2015).

Who the police are supposed to be serving is also highly relevant to suitability. For instance, police working in organizations following the Anglo-Saxon model has had its legitimacy in local communities in need of citizens consent to enforce the law and exert authority (McGloin, 2003). In contrast, the legitimacy of colonial police organizations derived from the colonial masters rather than the people the police were policing. Public consent was not expected from the citizens and activities were more focused on maintaining public order (Mawby, 1999; McGloin, 2003). In these contexts, different qualities would be prioritized in selection.

The discussion regarding desired qualities is not new, and most of the time similar qualities are in focus, although the rank ordering may be unclear and the “best at what” changes (Burkhart, 1980). In the US, Bain (1939, p. 454) analysed the research field and concluded that a policeman should have a background in college, and be of “superior mental and physical endowment”, and this is a view that seem to be coherent in police requirements in general, although the exact meaning of these qualities may be interpreted differently. But the discussion of educational merits and practical abilities has been a recurrent issue (Roberg & Bonn, 2004). According to Bain (1939) the qualities looked for were physical courage, disciplined temper, good judgement, alertness of observation and knowledge of law and procedure. All these qualities were exemplified with situations typical from the time era. Furthermore, physical courage and moral stamina were not seen as sufficient if a police officer lacked tact. He meant that there are situations that are not so straightforward such as delinquent children or handicapped defectives, which requires other qualities. Bain (1939) also portrayed how police officers of the time should interpret their mission using dichotomies: the policeman should see himself as a preventer of crime rather than a catcher of criminals, as a community servant rather than a watchdog and symbol of fear, and as an expert in handling complex human relations rather than a simple enforcer of laws. Almost fifty years later, Pugh (1986) presented qualities that are sought in applicants and are also measurable. He found that a need directly associated with the job was to react quickly and effectively to problematic situations. Another job-specific quality was the ability to apply mature judgment when deciding whether to arrest someone, issue a warning or use force. Qualities directly associated with the

personality of the future police officer were abilities to take initiative, solve problems, and apply sufficient judgment (and imagination) to deal with complex situations. Other were to maintain a balanced perspective and tolerate stress, these were considered critical, given the constant exposure in the profession to the dark side of humanity.

The Swedish context

In a Swedish historical context, similar qualities to those mentioned above have been seen as relevant, but the list has expanded over time. In the early descriptions there were also normative descriptions, stating how the person should be when off-duty, reflecting a common belief that policing is not solely a profession but also a way of life.

In 1948, when requirements were more precise, applicants had to be at least 21 years old and have good physical fitness. The personal qualities sought were trustworthiness, a disciplined lifestyle, good judgment, a peaceful temperament, a civic attitude and reasonable general knowledge. Other implicit directions to the recruiters were to select individuals who were reliable in a civic sense, holding democratic values (Furuhagen, 2015).

The most recent Swedish document describing such a competence profile (RPS, 2012) lists several desired qualities for an ideal applicant, influenced by the suggestions in Ds, 1996:11. It states that the applicant should show maturity and self-confidence when handling stressful situations, and interact positively in encounters with others. The applicant should also be able to communicate effectively both orally and in writing, and more importantly have the ability to take another person's perspective (i.e., be empathic). In addition, an applicant should be able to view situations objectively, to avoid disproportionate involvement, and co-operate with others. He or she should also be self-motivated and contribute constructive and creative solutions to diverse problems. Furthermore, he/she should be responsible, maintain good ethical and moral standards in stressful situations, resist pressure, and avoid giving in to negative influences from co-workers (Ds, 1996:11). These qualities are all probably valuable, but they are also complex and thus difficult to measure, a major drawback as ideal criteria should be measurable (Pugh, 1986).

In conclusion some suitability criteria have been expressed quite consistently, including skills and qualities that are valuable in: interactions with citizens, coping in difficult situations, and helping to ensure that an individual functions well in the organization. It should be noted, however, that the prioritization of these criteria has varied over time and context.

To summarize, the abilities required to succeed in the complex police profession can be portrayed in a competence profile that includes a number of required or desired qualities, which are not always easy to define. This may not change the criteria included, but may affect their prioritisation.

Emotional stability and understanding others are clearly important, and some formulations of certain required qualities are clearly related to empathy. Thus, this thesis focuses on empathy as a personal quality that is widely considered important for suitable police officers. It is hypothesised that advances in understanding of the nature of empathy, its measurement, and its importance in police work should be valuable for future police officers, the citizens in contact with them, and police organizations, through improvements in officers' preparation for a profession dealing with complex human relations embedded in an increasingly complex society.

4. Concurrent views and the measurement of empathy

This section presents a short theoretical description of the empathy concept, how it has been defined in different disciplines, and the approaches that have been adopted in attempts to measure it.

Views of empathy

Empathy has been defined, and classified, in various ways (Gerdes, 2011; Jolliffe & Farrington, 2004). Classifications include: predictive empathy, the ability to predict how another person is going to react; situational empathy, the ability to perceive emotions aroused in a specific situation; and trait empathy, which defines empathy as something innate. Empathic understanding refers to a person cognitively understanding another person and empathic communication concerns specific behaviour in patterns of interaction towards another person (Holm, 2001). The different definitions of empathy highlight a lack of consensus that, according to Johansson (2004), is not necessarily problematic because it may allow researchers from different disciplines to provide different and valuable perspectives that collectively provide a comprehensive, unified view. More important is understanding that empathy is a complex construct, and that different disciplines have focused on different parts of a commonly agreed division between affective, cognitive and behavioural components (Batson, 2009). The affective component is the emotional, to have the affective state of the other, and the cognitive component of empathy is often referred to as the conscious part of empathy. The behavioural is how a person acts after being in contact with the other.

The following paragraphs describe eight phenomena or processes that Batson (2009) claims capture processes that are often used in literature or research when referring to empathy, in order to facilitate understanding of what the concept entails. Based on a situation where a person is in contact

with another person who has a problem of some kind, Batson (2009, p. 3) poses two key questions: “How can one know what another person is thinking and feeling?” and “What leads one person to respond with sensitivity and care to the suffering of the other?” For simplicity, the person encountering the other with a problem is called the “observer” and the person with the problem is called the “target” (Davis, 1996). The eight phenomena are presented below, with suggestions for classifying them in parenthesis, whether they are regarded as cognitive or affective, and examples of their use in published literature.

The six first phenomena are related to the first question posed by Batson (2009). The observer gets to know about the target’s internal state, i.e. thoughts and feelings (phenomenon 1: cognitive, empathic accuracy). When the observer has this information, he/she can adopt the target’s expression to a certain degree (phenomenon 2: affective, motor mimicry, imitation, neural response) or come to feel the target’s feelings (phenomenon 3: affective, emotional contagion, affective empathy). This may also happen after the observer projects himself into the target’s situation (phenomenon 4: cognitive, aesthetic empathy), or by imagining how the target thinks and feels (phenomenon 5, cognitive, perspective-taking), or by imagining how the observer would think and feel in the target’s place (phenomenon 6: cognitive, role-taking, “decentering”).

The other two phenomena are related to the second question, regarding the motivational or behavioural components of empathy. The observer responds sensitively through personal distress aroused by the target’s situation (phenomenon 7: outcomes, empathic distress) or feelings for the target (phenomenon 8: outcomes, empathic concern).

When being aware of the different phenomena, which are referred to as what empathy put out to answer in a real situation, it may help to understand how they are more representative of the cognitive, the affective, both parts in interaction, or the outcomes.

A model of empathy presented by Davis (1996; 2006) also includes several aspects that are significant for understanding the nature of empathy, constructed in an attempt to bring order to the theoretical approaches defining the concept. This model includes four aspects to consider when using the terminology observer and target: Antecedents (characteristics of the observer, target and situation, Processes (the mechanisms resulting in empathy, categorized as non-cognitive, simple cognitive or advanced cognitive), Intrapersonal outcomes (cognitive, affective and motivational outcomes within the observer, not manifested in behaviour towards the target, but arising from the exposure to the target), and Interpersonal outcomes (behaviour directed to and dependent on prior exposure to the target).

When considering the empathy concept and its relevance for a profession, it is most important to focus on the accuracy of people's assessments of targets' internal states, as this is the crucial initial stage according to Batson (2009). It involves several of the processes associated with the first question, regarding acquisition of knowledge about how the other is thinking and feeling. Batson (2009) also suggests that it is important to identify persons who have too much personal emotional arousal, as this interferes with the ability to help. In practical situations within a profession such as policing it would be associated with limited objectivity, i.e. limited capacity to understand the target's situation. This was the area of interest in Study III, investigating how the construct has been used in police contexts, and the way it has been conceptualized. The study showed that advances from SCN, focusing on several processes related to the first question along with emotion-regulating processes, are especially valuable for the police profession. According to this view there are several key aspects to address: the affective capacities to feel another person's experience as one's own experiences; the three cognitive capacities; self-other awareness, i.e. the ability to differentiate between one's own situation and the other's situation; emotional regulation, i.e. the ability to control one's emotions and perspective-taking, i.e. the ability to understand the perspective of others (Decety & Jackson, 2006).

The measurement of empathy

It is widely recognized that empathy varies between individuals and can be treated as either a continuous variable or a dichotomous risk factor (Jolliffe & Farrington, 2004). The choice (and hence the ideal kind of measuring instrument to use) depends on whether the objective is to measure people's levels of empathy or determine whether they have at least a threshold level of empathy. The most realistic approach, of course, is to treat it as not only continuous but also as a multidimensional construct, so ideally each dimension should be evaluated with the aim to assess the construct in a valid and reliable way.

Various instruments are available for measuring empathy, or parts of the construct. Some, such as the Questionnaire Measure of Emotional Empathy (QMEE), focus on the affective dimension (Mehrabian & Epstein, 1972). Others, such as the Hogan Empathy Scale (HES), focus on the cognitive dimension (Hogan, 1969). The most recent instruments, such as the Interpersonal reactivity index (IRI), address both the affective and cognitive dimensions (Davis, 1980). The motivational or behavioural dimension has also been evaluated (Eisenberg & Miller, 1987). Most instruments focus on the observer's self-reported level of empathy as the focus is on processes within the observer and intrapersonal outcomes. Hence, the observer is in the best position to know which processes are

active and on what level (in the absence of neurological technology such as functional Magnetic Resonance Imaging (fMRI) with far greater capacities than current instruments).

The conceptualisation of the empathy construct has had valuable advances due to research from the field of SCN (Decety & Jackson, 2006; Lamm, Batson, & Decety, 2007). Study III in this thesis provides argumentation for how the SCN perspective should be considered valuable for the police context and in Study IV, the Empathy Assessment Index (EAI; Gerdes, Lietz, & Segal, 2011), an instrument supporting several findings from SCN developed in the US, is translated and adapted to a Swedish context to promote further development.

5. Validity theory

Since this thesis deals with complex constructs and ways to measure them, validity is a key issue. Personal skills and qualities are latent variables, and in order to measure them accurately several aspects of validity included in the classical view of validity must be considered. However, when these latent variables are discussed in selection contexts, the modern view of validity, which incorporates consequences and value implications, is more appropriate. This section presents the background and definition of various aspects of validity and their relevance for the studies in this thesis. Practical considerations when using these ideas are then presented, according to guidelines regarding validity and validation in Standards for educational and psychological testing (AERA, APA, NCME, 1999).

The validity considerations based on the classical view and the practical considerations that were applied when using selected instruments in Studies I-IV are presented in the methodological section. The considerations more associated with the modern view are presented in the discussion section.

The classical view of validity

The concept of validity emerged early in the measurement literature and has traditionally been seen as a concept related to the measurement situation, and the test itself:

Two of the most important types of problems in measurement are those connected with the determination of what a test measures, and of how consistently it measures. The first should be called the problem of validity, the second, the problem of reliability. (Kelley, 1921, p.80).

Later conceptualizations of validity focused on the interpretation and use of the results obtained from a test, as argued by Cronbach and Meehl (1955). Four categories of validity were defined: predictive validity, concurrent

validity, content validity and construct validity. Categories more commonly divided into three categories, criterion, content, and construct validity.

Briefly, predictive validity concerns if a chosen criterion, measured by a test, can be obtained at a later occasion; if the chosen criterion is decided simultaneous with the test then it is defined as concurrent validity. When both predictive and concurrent validity are considered together, they are referred to as criterion validity. Content validity refers to the degree to which test items cover the region of interest, the subject area. Construct validity is relevant when the focus is on measuring a construct that cannot be easily separated or defined, but the test or instrument reflects meanings related to the construct. With testable hypotheses, the claims of the instrument can be accepted or rejected. Theoretical relationships between the focal construct and other constructs (the “nomological network”) are also crucial here: they must be presented clearly to allow others to reject or accept statements about procedures used to measure a proposed construct (Cronbach and Meehl, 1955). In the validity discourse, construct validity early got a central role as a unifying force, which to can be interpreted as the meaningfulness of how to interpret test scores and what can be done with them (Messick, 1995). Some validity theorists’ even defined validity as construct validity (Loevinger, 1957). However, subsequent theorists considered that solely investigating empirical results based on predictions of a theory regarding a theoretical construct has limitations and cannot be regarded as a general framework for validity (Kane, 2008).

The deficiencies identified with the limited conceptualization of validity opened up for other views. When limiting validity to the determination of what a test measures, problems outside the test but still related to the test were not included and these problems needed attention. Areas of interest are those that concern procedural problems, or how context might influence the results. For instance, results obtained using a test in one decade or country may differ from those obtained using the same test two decades later, or in another country. Differences in the measured construct between groups may also introduce bias (Newton, 2012).

The modern view of validity

Over time the view of validity and its implications has expanded. Messick (1989) presented a ground-breaking extension by raising the importance of considering the use of the test. In a progressive matrix, he related validity to measurement as an evidential basis through construct validity, but he also included the use and relevance of the scores for the applied purpose, by weighing the benefits of using a test in relation to the costs. He also introduced a consequential basis in terms of value implications or potential and actual social consequences, which is highly relevant to the instruments and criteria applied in law enforcement. The emphasis in this view focuses on the appropriateness of inferences and actions

based on the results. Messick (1989) also points out the importance of acknowledging differences between the concepts of validity and validation, where the latter represents a longer and continuing process. In the longer period of time, additional sources of evidence are considered, such as information about social consequences connected to the use of test results.

A few years later Kane (2001) attempted to combine the traditional view of validity with the broader view and increase the applicability of validity in real settings. He discussed the construct validity model, applicable to theoretical constructs, claiming that constructs could be evaluated by testing theory-based hypotheses against observations. The focus of searching for evidence here shifts from validation of the test to the interpretation of scores, as discussed by Cronbach (1971). Another matter considering this issue is the division between a weak program and a strong program related to the use of the theory describing the construct. The weak referred to as unreflected empiricism or more correctly exploratory empiricism where any correlation is used as evidence for validity, it is therefore recommended to explicitly state the theoretical ideas concerning a construct (Cronbach, 1989).

The validation framework offered by Kane (2001) is based on arguments that contain a number of inferences and assumptions. The strategy involves: stating the proposed interpretive argument clearly and explicitly; assembling all available evidence relevant to the inferences and assumptions in the interpretative argument; evaluating the most problematic arguments either empirically or logically; and repeating the process, starting by re-stating the interpretative argument. An interpretative argument that survives all challenges is provisionally accepted. Inferences are divided into five basic types: evaluation, generalization, extrapolation, explanation and decision-making (Kane, 2001). The argument-based approach to validation provides guidance on the research efforts needed or evidence of validity that needs to be strengthened, often focusing on the arguments that are most problematic or weak. Inferences made with the interpretative argument that prove not to hold should be revised or abandoned (Kane, 2001). The approach can also be seen as a chain, which is no stronger than its weakest link. Concerning consequences, Kane (2001) separates the interpretative argument in the descriptive part (which involves a network of inferences from the scores obtained with a test) and the prescriptive part (involving decisions based on the descriptive statements).

Practical considerations

The interpretations and recommendations presented in Standards for educational and psychological testing (AERA, APA, NCME, 1999) have guided the decisions when considering and applying instruments used in Studies I-IV. This reference book, used by many psychometricians, addresses many aspects of validity and types of evidence needed for validity. It also provides recommendations based on prevailing ideas concerning validity, which are

therefore closely linked to the theoretical descriptions presented above. According to the suggested standards (AERA, APA, NCME, 1999, p. 11-16), the following aspects should be considered in a validation process. The test content and the construct being measured should clearly influence the wording and choices of test items. The appropriateness and relevance of the instrument for the content domain are also important. In addition response processes should be considered, as they will influence the quality of the information (scores) obtained from the examinees. This information is valuable to enrich the definition of the construct under investigation.

It is also necessary to consider variations in interpretations by different sources (after excluding irrelevant factors), the internal structure of the instrument, and evidence based on theoretical assumptions made about the construct(s) under investigation. Problems will arise, for example, if a construct is measured in the belief that it is one-dimensional when in reality it is multi-dimensional. It is also important to acknowledge potential problems with differential functionality of items. Validity will be impaired if it is assumed that a construct investigated in one population can be generalized to another context without sufficient empirical evidence. Relationships of focal constructs with other variables are also important, as they will influence the quality of predictions based on them and their empirical performance. Two important concepts in this respect are convergent validity (the degree to which concepts or measures that should theoretically be related are actually correlated according to tests) and discriminant validity (the degree to which concepts or measures that should not theoretically be related are actually not correlated). The relationship between test results and the focal criterion is also crucial for predictive performance. Designs that are used to evaluate the evidence are predictive or concurrent, two different designs that can provide different information. The generalizability of validity regards how far validity evidence can be transferred to new situations without further investigation. Last, but not least, are the consequences of testing. This is a highlighted area in modern validity theory, and it is also important to distinguish between validity and social policy issues. It is therefore always important to consider the purpose of a test when considering the consequences of its application. If differences among groups can be traced to construct(s) under investigation and no other source, then validity when measuring the construct is intact; otherwise, there is a validity issue.

Reflections on validity

In summary, developments in validity theory have included a shift to including much more than the empirical evidence obtained from a specific test or set of results. It is now more a matter of providing a more complex joint view that considers the value of a measurement in a specific area. This view acknowledges that a measurement of a phenomenon is embedded in a

social context and that social consequences should be considered. In this view, a single study is not sufficient to determine the value of a specific measurement or instrument, and validity cannot be definitively and universally established. A more balanced view is to see validation as an argument-based activity (Kane, 2001) where specific weaknesses in an area can be targeted for future improvements. Cronbach (1988) presents the responsibility of validators as follows,

...the task of validation is not to uphold a test, practice, or theory. Ideally, validators will prepare as debaters do. Studying a topic from all angles, a debater grasps the arguments pro and con so well that he or she could speak for either side... (Cronbach, 1988, p. 3).

Ideas regarding validity are highly relevant to practical issues associated with the instruments used in the different studies of this thesis, in which tests based on the traditional view of validity were applied as discussed in the following Methods chapter. The unitary model of validity suggested by Messick (1980) is addressed in more detail in the Discussion chapter.

6. Methods

In this thesis, both quantitative and qualitative methods have been used in order to get valuable information from multiple sources, as recommended by the validation guidelines discussed above. Using mixed methods has been considered valuable in many fields. For instance, Creswell (2009) argues that it provides a better understanding of research problems by offering alternative perspectives of study topics and greater ability to capture complexities in focal research areas. This methodological approach is also recommended for attempts to apprehend the complex realities in social contexts where multiple viewpoints, positions, and standpoints are valuable (Johnson, Onwuegbuzie, & Turner, 2007). Furthermore, it can combine the benefits of qualitative and quantitative research, thereby enriching the types of data that can be acquired from participants regarding focal phenomena (Johnson, Onwuegbuzie, & Turner, 2007). Study I and Study II are examples of the different type of data that was provided were the interview material was richer and provided more on complexities while the data from the recruits was valuable for generalizability. This data made it possible to investigate the support for three competence areas developed from responses to items in an instrument. Here it should be mentioned that mixed methods do not necessarily have to be used in a single study, but can be applied in multiple studies to meet the aims of a research project (Creswell, 2009; Johnson, Onwuegbuzie, & Turner, 2007).

However, the most essential argument for choosing different methodological approaches for each study was to use methods that provided the most useful information for addressing the specific research questions posed. The initial step involved familiarization with the research field and previous studies, since the general aim was to understand what is valued in assessments of suitability in police contexts. Another of the aims was to identify and adapt an instrument for measuring empathy in the focal contexts, which requires sound grounding in theoretical understanding of the construct(s) and context(s), as discussed in validity literature (Messick, 1989), and below. A recommended approach when defining a construct theoretically, in order to conceptualize it, is to see how others have addressed the problem previously (Clark & Watson, 1995). Here it was valuable to frame a nomological net, as advocated by Cronbach and Meehl (1955), for the theoretical definition of the construct and its relations to other constructs.

Validity considerations

As previously described, the validity issue of most relevance when an external criterion is not easily attained or available is construct validity (Kane, 2008). In police research, it has generally been seen as difficult to find a valid performance criterion (Aamodt, 2004; Falkenberg, Gaines, & Cordner, 1991; Sanders, 2003), even if there are valuable studies where alternative criteria has been used (Annell, 2015; Annell, Lindfors, & Sverke, 2015).

The studies of this thesis all concern (to varying degrees) theoretical and conceptually defined constructs that are difficult to evaluate empirically in terms of their correlations to an external performance criterion. Their value has been primarily evaluated in relation to the theory describing what they entail. For this purpose, several practical suggestions concerning validity were followed (AERA, APA, NCME, 1999), including use of multiple sources of evidence, such as both decision-takers and recruits. This contributed to attempts to explore the content domain of suitability, empathy and all the items measuring subconstructs, and how well they fit theoretical models and theory-based structure. Other more complex questions were also addressed, notably whether the structure of a construct obtained from responses of one population could be generalized to another context. To better understand if a measure made sense, evidence of convergent validity was also at focus. These procedures required the use of appropriate methods for collecting data covering the content domain from multiple sources, such as literature searches and interviews with several formats. This was followed by statistical methods such as exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), to see if and how the theoretical assumptions of constructs

were reproduced in the data based on responses from test-takers representing the populations of interest.

Data and participants

A major objective when selecting material and participants for each of the studies was to ensure that they could provide as valuable information as possible to meet the specific research aims. In Study 1 selection experts from each of the two focal organizations, the Swedish police and Catalan Police, provided valuable perspectives on criteria often used when selecting future police officers. The numbers of participants were considered sufficient to obtain rich understanding of their views and the roles of key criteria, given their extensive knowledge of the subject and their proportion of the total population with this competence who participated. Study II was based on datasets from the RECPOL project, which are available for researchers and include information from several police organizations in European countries. These data were valuable since they enabled validation of an instrument measuring recruits' perspectives on what constitutes a successful/good police officer. The initial data contained information from 352 Swedish police recruits and 1280 Catalan police recruits, where the Catalan sample was later divided into samples of comparable sizes to the Swedish sample, which were the target sample. The number of participants in this study was considered to meet sample-size requirements for common factor methodology. The recruits were also considered to provide relevant information as they had just entered the organization. The material used in Study III included literature describing the police profile and documents that guide selection decisions in Sweden, together with published research concerning the value and role of empathy in policing and advances from the empathy field. In Study IV students from three different educational programmes were used when collecting data for the translation and adaptation of the instrument. The students met important criteria as they were of the same age as the target population and were enrolled on social science programs leading to professions that are assumed to require interpersonal empathy. Initially there were 340 participants, most (176) enrolled in the basic training programme for police officers, 92 in an economics program and 72 in teacher education.

Literature search

A literature review provided foundations for descriptions of the requirements in the police profession and suitability, in terms of competencies or qualities sought. As previously mentioned, previous reports of these (and related) issues are mainly based on information from studies that have focused on Anglo-Saxon police systems and traditions, published in English. However, several Swedish documents referring to requirements

for and the suitability of police recruits were found. These included government reports, which are based on recommendations of commissioned experts (often social scientists and/or practitioners) and hence generally reflect both views of experts and government policy (Furuhagen, 2015).

A similar approach was adopted when attempting to improve understanding of the empathy concept. The first step involved scanning literature with a wide scope, while the next step focused on literature and research that could provide useful insights for identifying a functional theoretical conceptualization. Much of the literature was in English, in the form of published research articles in journals, theses and textbooks, but there were also some valuable contributions in Swedish.

Focus groups

In Study I focus groups were used, which is a frequently used method in the social sciences when aiming to understand the meaning and the understanding of a specific area (Lunt & Livingstone, 1996). It was valuable for obtaining selection practitioners' perspectives regarding constructs, abilities and qualities that are often mentioned in connection to suitability in a police context. The approach has been used for both critical and mainstream research, and as a source of ideas for quantitative testing (Lunt & Livingstone, 1996). The "focus" label refers to the aim of the moderator, to keep the participants focused on a specific subject. It should be noted that several parameters can vary, and should be considered in the study design because they may influence the development of the discussions in the sessions. One of the central issues is how the participants are selected. In a quantitative study a survey sampling approach would be applied. However, in Study I the aim was to obtain information from experts in a specific topic. These participants were not randomly assigned and were seen more as consultants than as research objects. Another issue concerns the degree of structure imposed on the sessions, in terms of variables such as the material presented to the participants and the topic(s) covered. If the results are going to be used for comparative purposes, a relatively high level of structure is important. Structure is also relevant to consider when it comes to the procedures adopted during the sessions, if the participants are allowed to talk freely with limited directions or if there are controlled forms. Here, the critical approach to the method provides some valuable guidelines, where the interest of the researcher (often the moderator) is a relevant area that needs reflection (Ang, 2001). In Study I it concerned how the specific information from the participants is attained and what possibilities the participants are given to reflect on their own answers as a moderator can control and move a discussion in a preferred way. In the sessions with the selection practitioners, each participant was asked to fill out an instrument for ranking the relevance of statements regarding selection criteria privately,

then present the results to the other participants openly and discuss the rankings (and associated reasoning) of different constructs, abilities and qualities. This approach provided participants with the possibility to clarify what they meant or how they interpreted possibly vaguely defined constructs in a context where the role of social interaction was useful.

Common factor modelling

The analysis of the material from the RECPOL instrument (Study II), and the EAI instrument (Study IV) required methods based on the common factor model, since the focus was on theoretical (latent) constructs. The common factor has its roots in the assumption that observed measures (questions or indicators) in a questionnaire share a common cause, an underlying construct or latent variable. The objective is then to identify the latent variables that accounts for the covariance produced by a group of questions used in an instrument. A variance that is divided into the variance accounted for by the common factors and the unique variance divided into specific and random error variance (Preacher & MacCallum, 2003). When developing an instrument the usual procedure is to define the domain of a construct and develop items that measure the content of the domain, in desirable conditions these items share a common variance.

The approach used to analyse models with latent variables depends on whether the analysis is explorative or confirmative. Exploratory factor analysis (EFA), which is commonly used in the beginning of an evaluation of a psychometric instrument, investigates which questions/indicators relate to the focal underlying construct(s) (Schumacker & Lomax, 2012). An association between a question/indicator and the factor is sought. If there is no association, the item or question is not a useful indicator of the factor. This association is referred to as the loading, and can be described in statistical terms as the completely standardized estimate of the regression slope for the prediction of an item from a factor (Brown, 2015). The relation between the factors and the items provides valuable information when interpreting the main objective of the EFA, i.e. to evaluate how a set of multiple indicators (items) become interpretable factors. The exploratory part refers mainly to the fact that this is a data driven method where the information reveals the relations between the items and the factor in a descriptive manner, no restrictions are placed on these relations when starting out (Brown, 2015).

Several decisions must be taken in such analysis, including choosing a factor extraction method (Furr, 2011), and the number of factors to extract, which should be based on both statistical considerations, for which tests such as parallel analysis can be used (O'Connor, 2000), and theoretical interpretation, i.e. checking that the factors are coherent in the focal context. When having multidimensional constructs with several factors, a decision

has to be made of a rotation method, which is designed to find an interpretable factor solution. That is, where there is high loadings for a subset of items and low loadings for other subsets of items. An oblique rotation (promax) was used, assuming that the factors were correlated in both Study II and Study IV (Preacher & MacCallum, 2003). Prior research and theory about the constructs supported this correlation between the factors. Results of the EFA were then interpreted in terms of how items with high loadings were grouped under a single corresponding factor without any cross-loadings to other factors.

The process described here is theoretically quite simple, but in practice, several decisions are needed, which are not simple since the data material doesn't always behave as expected. There are some commonly agreed upon thresholds to consider when moving forward with the analyses. Here, it is equally important to include knowledge from the subject field as it is generally agreed that different areas of research may provide somewhat different results. In my opinion it is important to keep an eye on the research field of interest. When guidelines are presented it is often stated that it is difficult to develop a decision tree that describe all possible decisions, and that the guidelines should instead better be seen as roadmaps (Cabrera-Nguyen, 2010). Considering the fact that there were priori assumptions of how the items should be related to the factors, the explorative approach was also followed by a confirmative approach using a confirmative factor analysis (CFA).

Most of the terms discussed above have the same meanings and implications in CFA as in EFA, such as factor loadings, the relation between items and the factors, and unique variance, each item being related with a specific variance and random error variance. The difference lies in the confirmative (rather than explorative) element. Now the object is to investigate that the pattern of relationships that are hypothesized between the factors and the corresponding relations to the items are consistent with the empirical data (Diamatopoulos & Siguaw, 2000). Now there are some priori assumptions that are tested following the steps: model specification, identification, estimation, testing fit and respecification (Kelloway, 1998). Model specification concerns the operationalization of the model, in which each item is associated with a corresponding factor and the number of factors is specified in advance. For instance, in Study IV, the factors or latent variables were allowed to correlate with each other as it was assumed that the empathy construct consisted of several aspects. In confirmative factor analysis, theory plays a key role since all the modelled relations are guided by theory and prior research (Schumacker & Lomax, 2012).

The following steps are to specify a model and estimate the parameters using an appropriate estimation method. Commonly used estimation methods are Maximum Likelihood analysis (ML, based on several

assumptions such as multivariate normality) or Robust Maximum Likelihood analysis (RML, which provides more robust estimation when normality assumptions may have been violated). The assumption of a continuous scale can be questioned, but given the approximate normal distribution with scales having more than four categories and the widespread use of the same estimation methods in similar analyses, this was considered to be a reasonable approach (Byrne, 2001). The fit is then interpreted from the fit indices. The general object of these indices is to provide a measure (a numerical value) of the discrepancy between the computed covariance matrix as hypothesized according to theory in relation to the covariance matrix derived from the empirical data (Diamatopoulos & Sigauw, 2000). Here it is important to consider overall goodness of fit, local areas of strain and the interpretability of the model's parameters (Brown, 2015).

Methods applied in CFA can be further extended, refined or combined when developing or evaluating instruments. The use of EFA in a CFA context is one of these combinations that is valuable in applied research, according to Brown (2015). The technique was introduced by Jöreskog (1969) and provides more valuable information regarding hypothesized patterns of relations than EFA. By setting fixed anchor items for each factor and allowing other items associated with each factor load freely on the other factors, the significance of cross-loadings can be estimated, indicating the potential presence of salient error covariances (Brown, 2006).

Another extension applied was the use of a multiple group CFA model in Study II to investigate the cross-national applicability of a scale from the RECPOL instrument. This involved use of data from two different groups, a Catalan group and a Swedish group (of recruits) as reference. In this analysis, constraints were set on the unstandardized parameters of interest to see if the instrument behaves in a similar way in the groups studied. By using a chi-square difference test (Satorra-Bentler scaled difference chi-square test) statistical information is acquired about the plausibility of a constrained solution, where a significant reduction in fit indicates group differences. The practical steps include a stepwise procedure moving from the least restricted model to a more restricted model. From equal form to equal factor loadings and to equal intercepts when adopting the terminology of Brown (2015). A significant reduction in fit indicates differences between groups and if differences in the functionality of items is detected it is worth investigating the modification indices to identify the parameters that are not invariant and release equality constraints, in order to reach partial measurement invariance (Byrne, Shavelson, & Muthén, 1989).

The use of a Multiple Indicators Multiple Causes (MIMIC) model was also a valuable extension for estimating group differences on the created factors in Study II. Dichotomous covariates representing group membership were

used as predictors for the latent variables to investigate the group differences (Kline, 2010; Schumacker & Lomax, 2012).

Ethical considerations

The ethical considerations addressed in Studies I-IV were based on frequently discussed and applied requirements for researchers in the humanities and social sciences (Forskningsrådet, 1996). Central concerns include the responsibility of the researcher to present information about the aims of the study to prospective participants, and respect their decision if they decline to participate. Confidentiality is also an important general consideration, but no information was going to be acquired that could identify any specific participant in any of the studies. Furthermore, the information acquired in each study was only used for research purposes, as all of the participants were notified.

Some of the information acquired and presented in Study I was drawn from interviews with selection experts. The participants were informed about the aims of the study both before and during the interviews. The subject area was not considered to be particularly sensitive and the participants gave their consent for the interviews to be recorded. Study II was based on data collected in the RECPOL project purely for research purposes, from informed participants who could freely choose whether or not to participate. The information regarding recruits' views and attitudes towards competencies for the police profession was not considered to be sensitive. Study III was a review, involving no participants apart from the researchers, of theoretical descriptions of empathy and how the construct has been used in police contexts. Thus, this raised no ethical issues. Study IV involved the administration of a self-reported empathy instrument to participants who had already been admitted on educational programmes, thus participation was not considered to be a sensitive topic for the participants. Participants were informed both orally and in a cover letter that participation was voluntary before responding to the questions in the instrument.

Reflections concerning the methodological choices

The description of the steps and procedures applied in Studies I-IV clearly shows that the methodological approach involved numerous decisions, all of which were dependent on the researchers' experience, influences, and ideas. It is important to acknowledge the subjectivity associated with the steps taken, regardless of the methods used. However, the decisions have been thoroughly documented to allow the approach to be dissected, and if deemed necessary choose different options in future studies.

7. Summary of studies

Study I

Selection practitioners' views on recruitment criteria for the profile of police officers: A comparison between two police organizations.

The competencies and qualities required to become a successful police officer have been investigated in many studies, but not from selection practitioners' perspectives, and most have focused on American contexts.

Thus, the aim of Study I was to investigate selection practitioners' views of what constitutes suitability in the profile of a police officer. Further, to assess the stability and complexity of the criteria, views of practitioners active within two different police organizations (the Swedish National Police and Catalan Police in Spain) were addressed. To cover the practitioners' contexts and explore factors influencing these practices, official statements on criteria and documented selection procedures were also examined. The subjective views of the selection practitioners were investigated by identifying similarities and differences between them (and their contexts) based on responses to a set of statements about common criteria drawn from previous research (Aamodt, 2004; Burbeck & Furnham, 1984; Varela, Boccaccini, Scogin, Stump, & Caputo, 2004; Sanders, 2003).

To illuminate the contexts official documents guiding practitioners were analysed, and supplementary information was obtained from interviews in Catalonia, because the Catalan documents provided limited information about the reasoning behind recommended selection procedures. These interviews were structured and followed a guide previously used to investigate selection practices in the USA (Cochrane, Tett, & Vandecreek, 2003). The questions covered areas such as the procedures used to evaluate applicants and the value of the various tests applied in the decision-making process. To investigate the selection practitioners' views, focus group sessions were conducted in which five practitioners from each organization ranked 12 statements regarding qualities considered valuable to become a successful officer, by placing the most relevant at the top of a diamond with nine spaces, and less important statements at the bottom. The three least relevant were to be left out of the diamond since it only had nine spaces. The reasons for the rankings were then debated in group discussions.

The findings indicated that there were several similarities between the two contexts in the ideal profile of applicants and selection procedures. The practitioners' views supported the importance of psychological constructs including emotional stability, honesty, and understanding the perspective of others, in accordance with previous findings (Burbeck & Furnham, 1984; Hogue, Black, & Sigler, 1994; Perez, 2010; Sanders, 2003). There was also support for more vaguely defined constructs such as common sense,

supporting previous claims of its value in police contexts (Crank, 2004). However, there was more support for cognitive intelligence (relative to common sense) in Sweden than in Catalonia. In addition, the Catalan practitioners said that some applicants had biased views of police work oriented towards enforcement (influenced by how police work is usually portrayed on television, the good versus the bad), while the Swedish practitioners said that some applicants had biased views towards service, not realising that they may have to use force. The participants also noted several areas missing in the statements, including the relation of police to the citizens, the importance of values and respect for the citizens, and the value of communication skills. Overall, the selection practitioners' reasoning provided valuable illumination of the connections between psychological constructs and police profiles, together with strategies adopted when ranking different criteria. A major conclusion from this study was that it is important to have well-defined criteria, in order to interpret them and their implications in actual police work consistently and effectively.

Study II

Validating a measure of the 'successful/good' police officer using data from recruits in two different police systems

The objective of this study was to illuminate police recruits' attitudes about the successful/good officer, which can help avoid unwanted outcomes when training future police officers or making decisions about selection criteria.

A specific aim was to validate an instrument designed to assess attitudes to knowledge and skills relevant for success as a police officer, developed for comparisons of European police organizations. This is potentially highly problematic as valid comparisons of populations require cross-culturally valid (equivalent or stable) constructs. Thus, the study compared perspectives of recruits enrolled in two of the participating organizations: the Swedish National Police and the Catalan Police in Spain. The data, retrieved from the European Longitudinal Study of Recruitment, Education and Careers in the Police (RECPOL) project consisted of self-reported attitudes of 352 Swedish police recruits and 1280 Catalan police recruits, and demographic information. The tested measures consisted of self-reported answers on a 5-point Likert-scale to 21 statements about competence, knowledge, and skills relevant for the policing profession.

The analysis involved EFA of the factor structure in data obtained from Swedish recruits, EFA in a confirmative factor analysis (CFA) framework, and CFA (Brown, 2006; Jöreskog, 1969). The factor structure was then cross-validated with data obtained from Catalan recruits. Constraints were then applied to the parameters to test for invariance between the two groups, following a stepwise procedure: equal form, equal factor loadings, and equal intercepts. For between-group comparisons of latent means, invariance in

both factor loadings and indicator intercepts is required (Brown, 2006). Thus, non-invariant parameters were unconstrained to reach partial invariance (Byrne, Shavelson, & Muthen, 1989). All steps were required to identify cross-culturally valid items. A MIMIC model based on the samples of the two groups and demographic information from the recruits was then constructed to investigate the influence of covariates indicating group membership on the latent variables.

A three-factor (towards citizens, knowledge and leadership) model was supported. All three of these factors represent a set of related abilities that are valuable according to previous research. A comparison of the latent means indicated that there were statistically significant differences between the groups: the Swedish sample scoring significantly higher than the Catalan sample for the factors towards citizens and leadership, and the Catalan sample scoring significantly higher for the factor knowledge. The demographic constitution of the recruits explained some of these differences, according to the MIMIC model, but more could be explained if more variables were included. The finding that the factor structure is comparable will be valuable for future studies as the RECPOL project has a longitudinal design and includes police organizations of other countries.

Study III

Empathy from a police work perspective

Personal suitability is important in various professions that involve dealing with people, particularly people who are frequently in distressed or agitated states, such as police work. Empathy is considered to be valuable in such professions for several reasons, as outlined in the Introduction. Furthermore, previous research supported the use of empathy within different fields in policing, e.g. during interviews on offense disclosure (Holmberg & Christianson, 2002; Oxburgh & Ost, 2011; Oxburgh, Ost, & Cherryman, 2012) and victim support (Maddox, Lee and Barker, 2011).

Thus, the aim of Study III was to identify an appropriate conceptual model of empathy based on findings from relevant research and apply them in the police context.

To meet this aim databases such as Ebsco and Google Scholar were systematically searched using a combination of selected keywords to identify theories and advances that could be of genuine value for the policing profession, based on research and government documents describing competencies deemed to be requirement for success in the profession.

The findings illuminate the complexity of the construct and the value of integrating the latest findings from empathy research. Theories from social psychology (Davis, 2006) and social cognitive neuroscience (SCN) (Decety, 2011; Gerdes, 2011; Gerdes, Segal, & Liets, 2010; Singer et al., 2006) are considered to provide a unified description of the key processes involved in

interpersonal elements, i.e. affective responses, perspective-taking and regulatory components such as self-other awareness and regulation of emotions (Decety & Moriguchi, 2007; Gerdes et al., 2011; Lietz et al., 2011). Contextual elements, including situational aspects and external circumstances, were also considered to be relevant in police contexts (Barrett-Lennard, 1981; Davis, 2006). Besides the often-mentioned importance of perspective-taking (Gerdes et al., 2011; Rasoal, 2009; Roan et al., 2009; Segal, Wagaman, & Gerdes, 2012; Wang et al., 2003) other factors capturing a person's judgements are views on responsibility (e.g. society's responsibility to individuals and vice versa), and views of suspects, were also deemed to be relevant. Notably, previous research indicates that the perception of the target is influenced by how the observer perceives the target, and when the target is considered to act unfairly empathy is reduced, especially among men (Singer et al., 2006).

The components highlighted in this study should facilitate empirical tests of conceptual models of empathy, as they have been theoretically associated with specific policing contexts.

Study IV

Adaptation and development of the Empathy Assessment Index (EAI)

The aim of Study IV was to adapt (and subsequently develop) a self-reported multidimensional measure of interpersonal empathy, the Empathy Assessment Index (EAI; Gerdes, Lietz, & Segal, 2011), to the Swedish context. Much previous research indicates the importance of empathy in various situations encountered in police work (Lumb & Breazeale, 2002; Holmberg & Christianson, 2002; Maddox, Lee, & Barker, 2011; Oxburgh, Ost, & Cherryman, 2012), but the latest advances in other fields, focusing on the conceptualization and measurement of the construct, had not been integrated. Thus, this was a major specific aim.

To meet the aims the EAI instrument required appropriate adjustments for Swedish respondents and assessment of the adapted instrument's convergent validity was needed. Thus, the US version of the instrument was first adapted to the Swedish context following commonly agreed guidelines (Van de Vijer & Hambleton, 1996). Its convergent validity was then explored by investigating the relation between the central components and previous instruments used to measure similar components. A scale consisting of 25 items rated on a six-point Likert scale was administered to 340 university students enrolled on education (teaching) or economics programs, and 176 enrolled on the basic police training program. Factor structure was then investigated by sequential EFA, EFA within a CFA framework, and CFA, as previously outlined. The analyses were conducted using Lisrel 9.1 software (Jöreskog & Sörblom, 2013) and SPSS version 21 (IBM SPSS Statistics, Chicago, IL).

The results showed that a 4-factor 18-item model provided an acceptable fit to the data and should be suitable for future development according to both theoretical considerations and detailed information about its fit to the data. Internal consistency, in terms of coefficient alpha was also at acceptable levels considering that items had been deleted and also in comparison to the corresponding values in the US version of the instrument. The following values were reported for each subscale: affective response (AR) $\alpha = .72$, perspective taking (PT) $\alpha = .81$, emotion regulation (ER) $\alpha = .62$ and self-other awareness (SOA) $\alpha = .61$. The two factors measuring affective response (AR) and perspective taking (PT) showed moderate to high levels of correlation with the corresponding subscales from the Interpersonal Reactivity Index (IRI; Davis, 1980) providing evidence of convergent validity. The four factor structure, supported in the Swedish context was different from the five factor structure, supported in the US context. The main reason for this difference was that in the US context, there was support for a subscale named affective mentalizing (AM). In the Swedish sample and context AM was considered to be too similar to PT and difficult to distinguish due to possible translation issues or when using self-report instruments. Thus, items from the AM subscale were integrated with the PT subscale.

These findings should facilitate further studies with the instrument, and the instrument may facilitate candidate selection as it can provide information regarding several aspects of empathy either as a self-reported instrument or when conducting interviews.

8. Discussion

The overall aim of this thesis was to explore the personal skills and qualities that are relevant to suitability for the police profession, with a special focus on the concept of empathy, and investigate how it can be conceptualized and measured in the context of selection. The general aim has thus covered two more specific aims that are distinct but related: a) to explore what suitability involves in law enforcement professions, and b) to conceptualize and measure empathy in a police context.

The first aim was investigated by studying documents, research literature, and the views from selection practitioners and recruits. Personal skills and qualities considered valuable for policing were core concerns in all four studies, and both Studies I and II present empirical findings regarding these views. Selection practitioners' perspectives are documented in Study I and the recruits' perspectives on important aspects for being successful in the police profession are considered in Study II. Both studies broadened the Swedish context by contrasting the findings with comparable information

from a police organization in a different context (Catalonia). It was hypothesized that the organizations would provide interesting contrasts because of differences in their historical background and traditions, including association with the Anglo-Saxon and Continental police systems, respectively. The aims to conceptualize and measure empathy in police contexts (a construct supported both explicitly and implicitly in Studies I and II) were addressed in Studies III and IV, respectively.

The following sections present and discuss findings related to the research questions posed at the beginning of the thesis, their meaning, their importance, and their potential applications. Validity considerations are also discussed then some general conclusions are drawn. Finally, limitations of this research are addressed and suggestions for future research are made.

What personal skills and qualities are relevant to suitability from different perspectives?

The findings show that the studied police organizations and those influenced by the Anglo-Saxon tradition according to literature, clearly value personal skills and qualities that facilitate the establishment and maintenance of good relations with the citizens. These attributes are highlighted in the examined documents describing the desired qualities of future police officers. They include communication skills, interest in people, the ability to handle stressful situations and patience. The importance of these skills and qualities is also supported by selection practitioners. Study I showed that practitioners from both Swedish and Catalan contexts regarded emotional stability, honesty, understanding the role of the police, and understanding others' perspectives as important. Two of these qualities, emotional stability and understanding others' perspectives, can also be regarded as aspects of empathy as defined in SCN theory. A key argument for the importance of emotional stability is that police officers carry weapons, and thus must be stable. There was strong consensus regarding inclusion of these qualities both within each group and in both organizations. However, some other qualities were considered valuable in one of the contexts but not the other, or their relative importance differed (e.g. cognitive intelligence was valued more in Sweden in comparison to Catalonia, and loyalty to the organization was valued more in Catalonia than in Sweden). Further, psychological qualities were considered more valuable than physical fitness, in both contexts, but this could be slightly misleading because the applicants tended to have high levels of fitness anyway. Study II showed that there were also some differences in attitudes of new police recruits between the two contexts. Notably, their mean scores indicated that Swedish recruits regarded attitudes towards citizens and leadership as being more important to be a successful/good police officer than Catalan recruits. There were also demographic differences in these respects, as females and those with

experience of higher education recorded higher mean scores for the factors Towards citizens and Leadership than males and those with no experience of higher education, respectively. The factor Towards citizens was based on items related to empathy such as “ability to appreciate different points of view”. The findings regarding these groups’ higher scores for these factors are consistent with previous studies on links between demographic factors and views on relations with citizens (Paterson, 2011).

Further, the value of objectivity (in the sense of being in control of one’s emotions) and having the ability to understand others’ perspectives was confirmed in both Studies I and II. These processes or phenomena can be seen as individual constructs, like emotional stability and perspective-taking, but can also be seen as belonging to a multifaceted construct, i.e. empathy. Study I showed that both qualities were similarly ranked in both the Swedish and Catalan organizations. Here, it was considered valuable to use a construct that included these abilities in relation to other important theoretically associated aspects, to enable their more rigorous evaluation from a theoretical perspective in order to identify important criteria in selection contexts.

These findings provide indications from several perspectives that the relationship of police to the citizens is important and that empathy, which is regarded as an important quality for interpersonal relations, is valuable in this context (Decety, 2011). In addition, there are signs of differences between groups that may be significant in selection procedures (and practical police work) that warrant further research. These findings were also relevant for the second research question regarding the conceptualization of empathy in a relevant manner for the profession.

How can empathy be conceptualized and measured in a relevant manner for the police profession?

The complex issue of conceptualizing the multifaceted construct empathy was the focus of Study III. In a profession such as policing it is crucial to understand how the construct is manifested in actual work situations to grasp its value in sufficient depth. Thus, literature describing its use in policing (Johnson, Hughes, & Ireland, 2007; Maddox, Lee, & Barker, 2011) was viewed in the light of recent developments in empathy research, in order to find or construct a model that is theoretically aligned with police contexts. Several conclusions were drawn. Empathy is a construct with some parts categorized as cognitive (for instance “understanding the other’s perspective”) and others categorized as affective (for instance “feeling what another person is feeling”). In later studies it has been suggested that regulating processes should be included in the cognitive part of the construct (Decety & Jackson, 2006). These findings are valuable for a profession such as the police, since police officers must be able to cope with exposure to

highly emotional situations (Bakker & Heuven, 2006). Several theoretical models also mention the importance of the conditions of situations, for instance the “empathy cycle” presented by Barrett-Lennard, (1981) includes openness of the observer towards information provided by the target, and “antecedents” (characteristics of the observer, target and situation) are important in the model presented by Davis (1996). Thus, contextual elements are clearly relevant, as noted in related fields (Rasoal, 2009; Roan et al., 2009; Segal et al., 2012; Wang et al., 2003). Hence these elements were addressed in Study III. In addition, documents guiding the Swedish police selection process acknowledge the importance of having democratic values and understanding social changes, in accordance with theory and further highlighting the relevance of contextual components. The EAI instrument (Gerdes, Lietz, & Segal, 2011) was considered to be useful as it incorporates theoretical ideas from SCN and included the regulating processes in the cognitive part.

In Study IV, the US version of the EAI instrument (Gerdes, Lietz, & Segal, 2011) was translated and adapted to a Swedish context then its performance was evaluated in this context in preparation for further adaptation. The translation and adaptation were crucial steps, due to indications that several items were either too complicated or not suited for direct linguistic translation. The findings from Study IV supported the theoretical assumptions regarding interpersonal components drawn from Study III, which differed somewhat from the components included in the most recent version of the EAI instrument. More specifically, the component affective mentalizing (AM) had been added to the instrument but was considered to be too closely related to perspective-taking and therefore integrated with perspective-taking in the Swedish version of the EAI. The convergent validity of the adapted instrument was also explored using scales from the IRI instrument (Davis, 1980).

The findings indicated that the adapted instrument was suitable for the aims and intended context. They also indicated that further studies are required, but valuable steps had been taken towards having an instrument more conceptually valid including regulating components of empathy, thereby allowing further investigation of the validity of the theoretical assumptions.

Discussion concerning the overall aim

When considering the overall aims of the work reported in this thesis and placing the findings in a selection context, several other issues should also be discussed. Study I showed that several of the skills and qualities identified as contributing to suitability in previous research, official documents guiding selection of police applicants and the focus group sessions strongly aligned. It was also noted that some constructs could be interpreted differently and

thus required more thorough conceptualization. There was even ambiguity in the descriptions and interpretations of manifestations of important qualities in actual work situations. Notably, honesty was regarded as a crucial quality, but discussions of the selection practitioners clearly showed that absolute honesty is not required, indeed lying may be essential sometimes in the field. It was held to be more important for future police officers to be honest in a legal sense, i.e. keeping within the law. Thus, attempting to apply a simple criterion to qualities that are really nuanced and multifaceted may be problematic (for instance, an applicant who honestly states that he/she lies sometimes could be unfairly disadvantaged). It is therefore very important for each construct used by decision-making professionals to be thoroughly defined and consistently applied in order to ensure that all applicants are treated equally, regardless of whether they are assessed in interviews or questionnaires. The important message here is that further work is required when a construct has been identified and defined. It is particularly important to investigate the value of any proposed construct as manifested in real police work contexts in order to obtain valid information to improve decisions. This is directly related to validity threats, where the content of a construct depends on who judges it.

The findings are also considered consistent with reasoning applied in the Swedish police selection process. To understand the importance of suitability in a specific profession it is important to understand the objectives of the profession and how they are interpreted when embedded in the focal organization's societal context. As concluded in the selection section, several principles appear to guide the logic in the Swedish selection system, and the perceived roles of individuals within the Swedish police force. For instance, in accordance with the utilitarian principle, the role of the police is seen as acting for the good of the whole society, and there are signs that various decisions regarding the police have been intended to ensure that the police maintain good relations with the citizens (Furuhagen, 2015).

Use of a construct such as empathy in the criteria for selecting police officers may also have value in terms of face validity and organizational image (Cable & Judge, 1995), through the signals sent to applicants concerning what the profession is really about. It also raises questions about the complexities of the profession, and consequences of officers becoming emotionally aroused, or unable to understand others in emotionally charged situations.

Here it is also important to mention that the focal area covered of this research and thesis is only one of the important areas. Several findings in the studies show the value of other qualities for success in policing. This is hardly surprising, as police roles have traditionally included maintenance of order, law enforcement and service (Paoline, 2003, Rumbaut & Bittner,

1979), all of which require different qualities that need to be investigated in valid ways.

Validity considerations for measurements of constructs applied in police selection

This section discusses the findings regarding the wider or modern perspective of validity advocated by Messick (1989), which includes consideration of societal level consequences, which was beyond the scope of Studies I-IV. Messick (1989) uses a progressive matrix to describe facets of validity. One area in the matrix concerns the evidential basis in terms of how a test can be interpreted, suggesting that construct validity is of mayor importance, and then when combined with the relevance of the test use. The other areas concern the consequential basis and the test interpretation where value implications and the social consequences of test use are of relevance or also seen as construct validity with different features (Messick, 1995).

The value implications refer to effects of the choice of criteria to include in a selection profile on the selection system itself. Notably, when considering the measurement of a certain quality the interpretations and consequences of the measurement should be addressed, in addition to the results directly obtained from an instrument. For example, the description of qualities that are important in Swedish police selection has often been guided by qualities valued in a Swedish police officer, as exemplified by explicitly stated democratic ideas in official documents and direct directions about what the selection practitioners should prioritise (Furuhagen, 2015). In the value implications, there is correspondence to how the constructs that are considered valuable are perceived in the proposed context. In police culture, an example could be the theoretical versus the practical, where constructs such as *common sense* related to practice can have strong support without having documented empirical support for their importance. The social historical trends previously discussed about the visions seen in the selection process provide guidance about how the Swedish police organization values different constructs. From this perspective it is reasonable to include selection criteria that correspond to the higher goals of the organization. However, in the light of widely held ideas in society about the need for punishment and being harder on crime such criteria could be questioned, mostly because of the focus on the affective domain of empathy. Society would question the idea of dealing with criminals with empathic police officers, potentially threatening the organization's legitimacy. A simple solution could be to use and refer to the cognitive domain such as perspective-taking, which has been seen as valuable among US soldiers (Roan et al., 2009). Theoretically, though, it is important to consider both dimensions, as deficits in both cognitive and affective empathy have been associated with "dark triad traits" containing constructs such as psychopathy

(Jonason & Krause, 2013) and according to some studies such traits are even more strongly associated with deficits in the affective domain (Wai & Tiliopoulos, 2012).

There has been criticism of the social consequences element of the matrix presented by Messick (1989), including doubt that it should be considered part of validity (Borsboom, Mellenbergh, & van Heerden, 2004; Popham, 1997). However, when discussing significant constructs in police selection and their measurement the social consequences are clearly highly relevant, and when considering the social consequences in terms of how they may affect applicants to the police it is important to consider the key roles of the police in society, as previously discussed. Selection of police officers carries particularly high responsibilities because of the power and authority the applicants obtain as soon as they become police officers (Sanders, 2008). Police officers have the rights to carry a weapon and use violence if necessary. Thus, the power relation between test taker and test provider in the process of selecting police officers has a different meaning from a societal perspective. It is clearly essential to get the right people (and avoid getting the wrong people), thus there are strong arguments for placing greater reliance on tests or other methods, provided the tests and methods are consistent with theoretical assumptions.

Considering the police recruits' perspective or the police officers', it could be perceived as exaggerating the importance of individuals, as recruits adapt to the organization (Aylward, 1985). However, it might still be important to have knowledge of their views and qualities from the outset and investigate later changes. It is important to understand the interaction between the demands on individuals active in an organization and the pressure from the culture within the organization. This thesis discusses qualities that are suggested to be investigated in early phases when individuals enter the organization. But it is equally important to remember that the culture within each organization can have major influences to the new members of the organization sometimes referred to with the analogy that the focus should be in the barrel instead of the apple (Sanders, 2003). Another consideration is the fact that in Study I it was suggested by one of the participants that a construct such as emotional stability could change because of life events occurring and a concern based on this reasoning is if suitability would need to be investigated more regularly in professions such as the police considering the fact that these are professionals that are exposed to the dark side of humanity on regular basis.

Another valuable question to investigate in this respect is the role of social desirability, which groups are more prone to give socially desirable responses as it may be a serious social consequence and a threat to validity in this context considering that autonomy and police discretion are significant elements of police work (MacVean & Neyroud, 2012). The final reflexions

concern if it is even possible to be able to control for social desirability in a high stakes situation such as the police selection considering the interest for the profession in Sweden. One of the major concerns when measuring psychological constructs in this setting and when including scales of social desirability are the concerns if we move away from the actual domain of interest, interpersonal qualities. Instead, we find us answering questions of who can manipulate the tests or the interviews in the most realistic way. Therefore, I find it valuable to suggest approaches where the constructs of importance have to be investigated in different situations and from different perspectives. It should be remembered that the ultimate objective is to provide tools for the persons who have to decide which applicants have sufficient emotional stability to carry a weapon, have the potential to base decisions considering all the relevant facts without disavouring certain groups of people and who also have the potential to adapt to changes in society.

Conclusions

A main conclusion from the research presented in this thesis is that several personal skills and qualities reflected in multifaceted selection criteria contribute to suitability as a police officer. A multifaceted criterion such as empathy has support from several perspectives. It was found that it is important to investigate each criterion theoretically and test how the assumptions hold in real settings in order to develop a robust, valid profile in any selection system. When including certain selection criteria in a profile it is also important to reflect upon the characteristics of the job, demands on the selection system and the alignment of the criteria with these demands in order to achieve desired outcomes. This requires understanding of the organization's objectives and role in society. Further, when focusing on each criterion detailed knowledge of the history, conceptualization and interpretation of associated constructs and their value in relevant work environments is required. Such critical reflection on constructs' conceptualization is also important for refining the conceptualisation in order to minimise risks that interpretations of those who decide who to admit into the profession may vary, depending on their background and knowledge. This is an important validity concern that needs to be considered in all situations where suitability is addressed.

This thesis is a contribution to the field of policing and other professions that need to consider what should be included in suitability criteria. The findings are not limited to a Swedish context and some at least should be applicable internationally as several key validity issues have been addressed. The thesis not only focuses on the value of a specific construct but presents several of the initial steps to develop a strong program of

construct validation, as required for any criterion to be used in practice (Benson, 1998).

Moreover, each study should facilitate future research through the information provided and questions raised. Study I suggests an approach that proved successful for eliciting comparative information regarding focal themes or phenomena from practitioners who play active roles in organizations in the same field but different contexts. In addition, use of a diamond tool and developing statements in advance about the subject was found to be valuable for both introducing the subject and providing a structure that facilitated comparison. Study II highlights the importance of measurement invariance methodology, especially when making comparisons using translated and adapted instruments, which must provide equivalent (i.e. validly comparable) results when used to probe characteristics or views of groups in different contexts. This is essential because theoretical constructs may have different factorial structures or items may be perceived in different ways depending on their content and context. It is also a valuable approach when investigating items' functions across demographic groups, especially items in instruments used in selection systems, which must have high validity. Study III highlights the need for a robust theoretical rationale when promoting ideas that will influence measurement approaches, as assumptions that are valid in one context may require revision before application in a different context. Study IV is valuable as it provides an alternative to the most commonly used instrument, IRI (Davis, 1980), for measuring empathy. The Swedish version of the EAI presented and applied in Study IV may measure empathy more adequately in professional contexts, partly due to the inclusion of items designed to assess regulating mechanisms.

Limitations and future research

There are several limitations of this research that need to be addressed. Limitations are often associated with unavoidable restrictions in factors such as time, resources and access to an infinite population. Others may be due to the researcher failing to recognize the importance of influential variables that could potentially be controlled. These limitations are presented in this section where some are accompanied with planned or possible future studies.

The personal skills and qualities valued for suitability in law enforcement have been foci of all the studies presented in this thesis. Methods for assessing the constructs used to define and describe these skills and qualities require more attention in order to refine the relations between theory and empirical findings and the arguments concerning what to include in constructs such as empathy. For instance, some of the theoretically based suggestions presented in Study III regarding empathy were not incorporated

in Study IV. This was mainly due to the decision to focus on the interpersonal empathy constructs in Study IV, and leave consideration of situational and contextual factors for further, ongoing research.

A general question related to the measurement method of self-reported instruments in general and when measuring empathy exists, such as the dependency on the observers' metacognition or when developing conclusions based on incorrect information (Holm, 2001; Roan et al., 2009). Despite these weaknesses, it is the most common method for measuring empathy, as empathy involves internal processes of the observer (Bouton, 2014). An interesting approach would be to investigate how well self-reported empathy correlates with how other perceive this level, that is, both the observers own evaluation and the targets evaluation of the observer. This would be methodologically challenging, but some studies have found ways to include targets' views in a police context (Maddox, Lee, & Barker, 2011).

Other limitations are related to differences in conditions of the instruments used, notably in the cut-off values applied when using the two instruments applied in Studies II and IV. The instrument used in Study II had not been used in previous research, so the items had not been refined. Partly for this reason relatively relaxed cut-off values were used when evaluating loadings, for instance. Similarly, items that showed potential functionality were retained in the measurement model, based at least on theoretical grounds. In contrast, the instrument in Study IV had been applied in previous research and theoretical guidelines regarding expectations were available, so cut-offs for excluding items could be stricter.

Another limitation concerns the test-retest reliability of the instrument used in Study IV. It would have been valuable to re-test it with the same group of students, but practical problems precluded gathering the same large group again.

Other questions that need to be addressed concern how an instrument like the one used in Study IV would work in high-stake situations. The focus of such research shifts from the theoretical domain of the construct(s) measured towards issues related to social desirability, which raises validity threats when measuring constructs of personal qualities (Ziegler & Buehner, 2009). However, some research steps in this direction have been initiated. Notably, a 13-item version of the Marlowe-Crowne Social Desirability Scale developed by Reynolds (1982) has shown potential value in law enforcement (Greenberg & Weiss, 2012). Attempts have been made to develop a Swedish version, largely through adaptation rather than translation, and the items are considered to be conceptually valid in terms of social desirability in a Swedish context.

As a final comment, supporting the distinction between validity and validation (which is a continuous process), more studies are planned to further investigate subjectivity issues in self-reported measures of empathy

(Messick, 1989). The objective will be to investigate if test takers overestimate their accuracy by comparing self-reported empathy to results obtained using instruments with descriptions of real-life situations in a police context.

References

- Aamodt, M. G. (2004). *Research in law enforcement selection*. BocaRaton, FL: BrownWalker Press.
- American Educational Research Association [AERA]; American Psychological Association [APA]; & National Council on Measurement in Education [NMCE]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Ang, I. (2001). On the politics of empirical audience research. *Media and cultural studies: Keywords*, 177-197.
- Annell, S. (2012). Vilka sökande antas till och påbörjar polisutbildningen? [Which applicants are admitted to and begin police education?]. Karlstad: Rekryteringsmyndigheten.
- Annell, S. (2015). Hållbar polisrekrytering: Teoretiska, metodologiska och praktiska perspektiv på rekrytering och urval. [Sustainable police recruitment: Theoretical, methodological and practical perspectives on recruitment and selection]. Doctoral dissertation. Stockholm: Department of Psychology, Stockholm University.
- Annell, S., Lindfors, P., & Sverke, M. (2015). Police selection—implications during training and early career. *Policing: An International Journal of Police Strategies & Management*, 38(2), 221-238.
- Aylward, J. (1985). Psychological testing and police selection. *Journal of Police Science & Administration*.
- Bain, R. (1939). The policeman on the beat. *The Scientific Monthly*, 48, 450-458.
- Bakker, A. B., & Heuven, E. (2006). Emotional dissonance, burnout, and in-role performance among nurses and police officers. *International Journal of Stress Management*, 13(4), 423.
- Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology*, 28(2), 91.
- Batson, D. C. (2009). These things called empathy: eight related but distinct phenomena. In J. Decety & W. J. Ickes (Ed.). *The social neuroscience of empathy*. Cambridge, Mass: MIT Press.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17(1), 10-17.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality : a treatise in the sociology of knowledge*. Garden City, N.Y: Doubleday.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605-634.

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061.
- Boudreau, J. W., & Ramstad, P. M. (2003). Strategic industrial and organizational psychology and the role of utility analysis models. In W.C. Borman, D.R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology, Vol. 12.* (pp. 193-221). Hoboken, NJ: Wiley.
- Boudreau, J. W., Sturman, M. C., & Judge, T. A. (1994). *Utility analysis: What are the black boxes, and do they affect decisions?* Ithaca, New York: Cornell University, School of Hospitality Administration.
- Bouton, B. (2014). *Initial Development and Validation of the Teacher Empathy Scale.* Doctoral dissertation. Athens, GA.; University of Georgia.
- Brezis, E., & Crouzet, F. (2006). The role of higher education institutions: recruitment of elites and economic growth. *Institutions, Development, and Economic Growth*, 191-213.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guildford press.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research. (Second edition).* New York: Guildford press.
- Burbeck, E., & Furnham, A. (1984). Personality and police selection: Trait differences in successful and non-successful applicants to the metropolitan police. *Personality and Individual Differences*, *5*(3), 257-263.
- Burkhart, B. R. (1980). Conceptual issues in the development of police selection procedures. *Professional Psychology*, *11*(1), 121-129.
- Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, *1*(1), 55-86.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456.
- Cable, D. M., & Judge, T. A. (1995). The role of person-organization fit in organizational selection decisions. *CAHRS Working Paper Series*, 197.
- Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research*, *1*(2), 99-103.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309.

- Cochrane, R. E., Tett, R. P., & Vandecreek, L. (2003). Psychological Testing and the Selection of Police Officers A National Survey. *Criminal Justice and Behavior*, 30(5), 511-537.
- Crank, J. P. (2004). *Understanding police culture*. Cincinnati, OH: Anderson Pub.
- Creswell, J. W. (2009). *Research design : qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, Calif: Sage.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement theory and public policy* (pp. 147-171). Urbana: University of Illinois Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281-302.
- Dantzker, M. (2011). Psychological preemployment screening for police candidates: Seeking consistency if not standardization. *Professional Psychology: Research and Practice*, 42(3), 276.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85. Retrieved from http://www.uv.es/~friasnav/Davis_1980.pdf
- Davis, M. H. (1996). *Empathy : a social psychological approach*. Boulder, Colo: Westview Press.
- Davis, M. H. (2006). Empathy. In J. E. Stets & J. H. Turner (Eds.), *Handbook of the sociology of emotions* (pp. 443– 466). New York, NY: Springer US.
- Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion Review*, 3(1), 92-108.
- Decety, J., & Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Current directions in psychological science*, 15(2), 54-58.
- Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *BioPsychoSocial Medicine*, 1(1), 22.
- DeCoster-Martin, E., Weiss, W. U., Davis, R. D., & Rostow, C. D. (2004). Compulsive traits and police officer performance. *Journal of Police and Criminal Psychology*, 19(2), 64-71.
- Diamatopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL. A guide for the uninitiated*. London: Sage.
- Ds. (1996:11). Rekrytering och grundutbildning av poliser [Recruitment and initial training of police officers]. Stockholm: Ministry of justice.

- Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, 101(1), 91.
- Falkenberg, S., Gaines, L. K., & Cordner, G. (1991). An examination of the constructs underlying police performance appraisals. *Journal of Criminal Justice*, 19(4), 351-360.
- Forskningsrådet, H.-S. (1996). *Forskningsetiska principer i humanistisk-samhällsvetenskaplig forskning* [Ethical research principles in humanities and the social sciences research]. Stockholm: Vetenskapsrådet.
- Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. Los Angeles, Calif: SAGE.
- Furuhagen, B. (2015). The making of a social, representative and intellectual police force-Recruitment and education of police trainees in Sweden during the 20th century. *Nordisk politiforskning*, 2(01), 6-33.
- Gerdes, K. E. (2011). Empathy, Sympathy, and Pity: 21st-Century Definitions and Implications for Practice and Research. *Journal of Social Service Research*, 37(3), 230-241
- Gerdes, K. E., Lietz, C. A., & Segal, E. A. (2011). Measuring empathy in the 21st century: Development of an empathy index rooted in social cognitive neuroscience and social justice. *Social Work Research*, 35(2), 83-93.
- Gerdes, K. E., Segal, E. A., & Lietz, C. A. (2010). Conceptualising and Measuring Empathy. *British Journal of Social Work*, 40(7), 2326-2343.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of management review*, 18(4), 694-734.
- Greenberg, B., & Weiss, P. (2012). Validation of a Short Form of the Marlowe-Crowne for Use with Law Enforcement Personnel. *Journal of Police and Criminal Psychology*, 27(2), 123-128.
- Guffey, J. E., Larson, J. G., Zimmerman, L., & Shook, B. (2007). The development of a Thurstone Scale for identifying desirable police officer traits. *Journal of Police and Criminal Psychology*, 22(1), 1-9.
- Hibler, N. S., & Kurke, M. I. (1995). Ensuring personal reliability through selection and training. In M. I. Kurke & E. M. Scrivner (Ed.), *Police psychology into the 21st century* (pp. 57-91). Hillsdale, N.J: Lawrence Erlbaum Associates.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33(3), 307.
- Hogue, M. C., Black, T., & Sigler, R. T. (1994). Differential Use of Screening Techniques in the Recruitment of Police Officers, The. *Am. J. Police*, 13, 113.

- Holm, U. (2001). *Empati : att förstå andra människors känslor*. [Empathy: to understand other people's feelings]. Stockholm: Natur och kultur.
- Holmberg, U., & Christianson, S. Å. (2002). Murderers' and sexual offenders' experiences of police interviews and their inclination to admit or deny crimes. *Behavioral sciences & the law*, 20(1-2), 31-45.
- Jacoboni, M. (2008). Imitation, Empathy, and Mirror Neurons. *Annual Review of Psychology*, 60(1), 653-670.
- Jacobs, L. A. (2003). *Pursuing Equal Opportunities : The Theory and Practice of Egalitarian Justice*. West Nyack, NY, USA: Cambridge University Press.
- Jacquette, D. (2002). *Ontology*. Durham, GBR: Acumen.
- Johansson, M. (2004). *Empatisk förståelse : från inlevelse till osjälviskhet*. [Empathic understanding. From empathic understanding to unselfishness]. Doctoral dissertation. Lund: Department of Philosophy, Lund University.
- Johnson, H., Hughes, J. G., & Ireland, J. L. (2007). Attitudes towards sex offenders and the role of empathy, locus of control and training: A comparison between a probationer police and general public sample. *The Police Journal*, 80(1), 28-54.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), 112-133.
- Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and violent behavior*, 9(5), 441-476.
- Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences*, 55(5), 532-537.
- Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183-202.
- Jöreskog, K. G., & Sörbom, D. (2013). LISREL 9.1 for Windows [Computer software]. Skokie, IL: 620 Scientific Software International.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Kane, M. T. (2008). Terminology, emphasis, and utility in validation. *Educational Researcher*, 37(2), 76-82.
- Kelley, T. L. (1921). National Association of Directors of Educational Research. *The Journal of Educational Research*, 4(1), 75-80.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling : a researcher's guide*. Thousand Oaks, Calif: Sage.
- Kline, R. B. (2010). *Methodology in the Social Sciences: Principles and Practice of Structural Equation Modeling (3rd Edition)*. New York, NY, USA: Guilford Press.

- Kornai, J., Rothstein, B., & Rose-Ackerman, S. (2004). *Creating social trust in post-socialist transition*. New York: Palgrave Macmillan.
- Lamm, C., Batson, C. d., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Cognitive Neuroscience, Journal of*, 19(1), 42-58.
- Le, H., Oh, I.-S., Shaffer, J., & Schmidt, F. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from meta-analysis. *The Academy of Management Perspectives*, 21(3), 6-15.
- Lee, N., & Lings, I. (2008). *Doing Business Research: A Guide to Theory and Practice*. Los Angeles: SAGE.
- Lietz, C. A., Gerdes, K. E., Sun, F., Geiger, J. M., Wagaman, M. A., & Segal, E. A. (2011). The Empathy Assessment Index (EAI): A confirmatory factor analysis of a multidimensional model of empathy. *Journal of the Society for Social Work and Research*, 2(2).
- Lipsky, M. (1980). *Street-level bureaucracy : dilemmas of the individual in public services*. New York: Russell Sage Foundation.
- Loader, I., & Walker, N. (2001). Policing as a public good: Reconstituting the connections between policing and the state. *Theoretical Criminology*, 5(1), 9-35.
- Locke, J. (2000). *Two Treatises of Government*. Hamilton, Ontario, CAN: McMaster University Archive of the History of Economic Thought.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph supplement 9. *Psychological Reports*, 3(3), 635-694.
- Lough, J., & Von Treuer, K. (2013). A critical review of psychological instruments used in police officer selection. *Policing: An International Journal of Police Strategies & Management*, 36(4), 737-751.
- Lumb, R. C., & Breazeale, R. (2002). Police officer attitudes and community policing implementation: Developing strategies for durable organizational change. *Policing & Society*, 13(1), 91-106.
- Lunt, P., & Livingstone, S. (1996). Rethinking the focus group in media and communications research. *Journal of communication*, 46(2), 79-98.
- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55(2), 95.
- MacVean, A., & Neyroud, P. (2012). *Police Ethics and Values*. London: Learning Matters.
- Maddox, L., Lee, D., & Barker, C. (2011). Police empathy and victim PTSD as potential factors in rape case attrition. *Journal of Police and Criminal Psychology*, 26(2), 112-117.
- Mawby, R. I. (1999). *Policing across the world: issues for the twenty-first century*. London: Psychology Press.

- McGloin, J. M. (2003). Shifting paradigms: policing in Northern Ireland. *Policing: An International Journal of Police Strategies & Management*, 26(1), 118-143.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40(4), 525-543.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012.
- Messick, S. (1989) Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York: American Council on Education/Macmillan.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741.
- Mulgan, T. (2007). *Understanding Utilitarianism*. Durham, GBR: Acumen.
- Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research & Perspective*, 10(1-2), 1-29.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior research methods, instruments, & computers*, 32(3), 396-402.
- Oxburgh, G., & Ost, J. (2011). The use and efficacy of empathy in police interviews with suspects of sexual offences. *Journal of Investigative Psychology and Offender Profiling*, 8(2), 178-188.
- Oxburgh, G., Ost, J., & Cherryman, J. (2012). Police interviews with suspected child sex offenders: does use of empathy and question type influence the amount of investigation relevant information obtained? *Psychology, Crime & Law*, 18(3), 259-273.
- Paoline, E. A. (2003). Taking stock: Toward a richer understanding of police culture. *Journal of Criminal Justice*, 31(3), 199-214.
- Paterson, C. (2011). Adding value? A review of the international literature on the role of higher education in police training and education. *Police Practice and Research*, 12(4), 286-297.
- Perez, D. W. (2010). *Paradoxes of police work*. New York, NY: Cengage Learning.
- Popham, W. J. (1997). Consequential validity: Right Concern-Wrong Concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, 2(1), 13-43.
- Pugh, G. M. (1986). The good police officer: Qualities, roles, and concepts. *Journal of Police Science and Administration*, 14(1), 1-5.

- Rasoal, C. (2009). *Ethnocultural empathy: measurement, psychometric properties, and differences between students in health care education programmes*. Doctoral dissertation. Linköping: Department of Behavioural Sciences and Learning, Linköping University.
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of clinical psychology*, 38(1), 119-125.
- Rikspolisstyrelsen [RPS], 2012. Kompetensprofil för nyutbildad polis samt profil för antagning till polisutbildningen [Competence profile for newly examined police officer and selection to police training]. Stockholm: Rikspolisstyrelsen.
- Roan, L., Strong, B., Foss, P., Yager, M., Gehlbach, H., & Metcalf, K. A. (2009). *Social Perspective Taking* (No. ARIBSS-TR-1259). ECROSSCULTURE CORP BOULDER CO.
- Roberg, R., & Bonn, S. (2004). Higher education and policing: where are we now? *Policing: An International Journal of Police Strategies & Management*, 27(4), 469-486.
- Rumbaut, R. G., & Bittner, E. (1979). Changing conceptions of the police role: A sociological review. *Crime and Justice*, 239-288.
- Sanders, B. A. (2003). Maybe there's no such thing as a "good cop": Organizational challenges in selecting quality officers. *Policing: An International Journal of Police Strategies & Management*, 26(2), 313-328.
- Sanders, B. A. (2008). Using personality traits to predict police officer performance. *Policing: An International Journal of Police Strategies & Management*, 31(1), 129-147.
- Schumacker, R. E., & Lomax, R. G. (2012). *Beginner's Guide to Structural Equation Modeling (3rd Edition)*. Florence, KY, USA: Routledge.
- Segal, E. A., Wagaman, M. A., & Gerdes, K. E. (2012). Developing the social empathy index: An exploratory factor analysis. *Advances in Social Work*, 13(3), 541-560.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075), 466-469.
- SOU. (2007:39). *Framtidens polis : delbetänkande [Future police : preliminary report]*. Stockholm: Government commission report.
- SOU. (2008:39). *Framtidens polisutbildning : slutbetänkande [Future police training : final report]*. Stockholm: Government commission report.
- Stensöta, Helena. (2004). Den Empatiska Staten: Jämställdhetens Inverkan på Daghem och Polis 1950–2000. [The empathetic State. The impact of equality between men and women on kindergarten and police

- 1950-2000]. Doctoral dissertation. Gothenburg: Department of Political Science, University of Gothenburg.
- Urbina, S. (2014). *Essentials of Behavioral Science : Essentials of Psychological Testing (2nd Edition)*. Somerset, NJ, USA: Wiley.
- Wai, M., & Tiliopoulos, N. (2012). The affective and cognitive empathic nature of the dark triad of personality. *Personality and Individual Differences, 52*(7), 794-799.
- Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*(2), 89.
- Wang, Y.-W., Davidson, M. M., Yakushko, O. F., Savoy, H. B., Tan, J. A., & Bleier, J. K. (2003). The scale of ethnocultural empathy: Development, validation, and reliability. *Journal of Counseling Psychology, 50*(2), 221-234.
- Varela, J. G., Boccaccini, M. T., Scogin, F., Stump, J., & Caputo, A. (2004). Personality Testing in Law Enforcement Employment Settings A Metaanalytic Review. *Criminal Justice and Behavior, 31*(6), 649-675.
- White, M. D., & Escobar, G. (2008). Making good cops in the twenty-first century: Emerging issues for the effective recruitment, selection and training of police in the United States and abroad 1. *International Review of Law Computers & Technology, 22*(1-2), 119-134.
- Wolming, S. (1999). Ett rättvist urval? [Fair Selection?] *Pedagogisk forskning i Sverige, 4*(3), 245-258
- Ziegler, M., & Buehner, M. (2009). Modeling Socially Desirable Responding and Its Effects. *Educational and Psychological Measurement, 69*(4), 548-565.
- Zimdars, A. (2007). *Challenges to meritocracy? A study of the social mechanisms in student selection and attainment at the University of Oxford*. University of Oxford.

Appendix

Abbreviations

| | |
|--------|---|
| AM | Affective mentalizing, subscale of EAI instrument |
| AR | Affective response, subscale of EAI instrument |
| CFA | Confirmative factor analysis |
| CFI | Comparative fit index |
| CPI | California Psychological Inventory |
| Ds | Document from the Swedish Ministry of justice |
| EA | Empathic attitudes |
| EAI | Empathy assessment index, Swedish version and US version |
| EC | Empathic concern, subscale of IRI instrument |
| EFA | Exploratory factor analysis |
| ER | Emotion regulation, subscale of EAI instrument |
| fMRI | functional Magnetic Resonance Imaging |
| IRI | Interpersonal Reactivity Index, Swedish version and US version |
| KMO | Kaiser Meyer Olkin, measures the sampling adequacy to perform EFA |
| LISREL | Linear Structural Relations |
| MIMIC | Multiple Indicator Multiple Indicator Cause |
| ML | Maximum Likelihood |
| PT | Perspective taking, subscale of EAI instrument and IRI instrument |
| RECPOL | European Longitudinal Study of Recruitment, Education and Careers in the Police |
| RML | Robust Maximum Likelihood |
| RMSEA | Root Mean Square Error of Approximation |
| RPS | Rikspolisstyrelsen, Swedish National Police Board |
| SA | Self-awareness |
| SCN | Social cognitive neuroscience |
| SOA | Self-other awareness, subscale of EAI instrument |
| SOU | Statens offentliga utredningar, Official Reports of the Swedish Government |
| SPSS | Statistical package for the social sciences |
| SRMR | Standardized root mean residual |
| ToM | Theory of mind |

Statistical notation

| | |
|----------|---|
| α | Coefficient alpha (internal consistency, reliability) and probability level |
| CI | Confidence interval |
| df | Degrees of freedom |

| | |
|-------------|---|
| M | Mean |
| MAP | Minimum Average Partial test, gives guidance of the amount of factors to retain |
| NS | Not significant |
| p | Probability |
| r | Pearson's correlation coefficient |
| $SB \chi^2$ | Satorra-Bentler chi-square |
| SD | Standard deviation |
| SDCS | Satorra - Bentler scaled difference chi-square test |
| χ^2 | chi-square |