Stockholm University

Testifying in a non-native language as an eyewitness undermines credibility judgments but not reliability

Arman Raver

Doctoral Thesis in Psychology at Stockholm University, Sweden 2025

Testifying in a non-native language as an eyewitness undermines credibility judgments but not reliability

Arman Raver

Academic dissertation for the Degree of Doctor of Philosophy in Psychology at Stockholm University to be publicly defended on Friday 19 December 2025 at 10.00 in Lecture Hall 3, Albanovägen 22.

Abstract

Eyewitness testimony plays a critical role in legal proceedings. Yet, its usefulness hinges on both the reliability of the information from the witness, and how that information is perceived. When eyewitnesses testify in a non-native language, language barriers may shape not only what is communicated but also how their testimony is interpreted and judged. This thesis investigates whether testifying in a non-native versus native language affects the reliability and perceived credibility of honestly reported eyewitness testimonies. Across three empirical studies, I examined how language influences memory accuracy, suggestibility, the confidence-accuracy relation, observer-rated credibility, investigative questioning practices, and non-verbal behavior. **Study I** comprised two experiments. In **Study Ia** (N = 121), mock eyewitnesses provided testimony in either a non-native (English; n = 61) or their native (Swedish; n = 60) language after viewing a mock-crime video. Testimonies accuracy and suggestibility were measured, and the confidence-accuracy relation computed. Witnesses also rated their own credibility and cognitive effort when giving the testimony. No differences emerged between witness groups in accuracy, suggestibility, cognitive effort, or self-rated credibility. However, in line with the confidence-accuracy relation, witnesses were more confident overall in correct than incorrect details. Importantly, non-native speakers reported significantly lower confidence than native speakers, despite being equally accurate. In **Study Ib** (N = 202), mock interrogators were randomly assigned to assess either a non-native or native video-recorded testimony for credibility and to formulate follow-up interrogation questions. The interrogators judged non-native speakers as less credible than native speakers. **Study II** examined the interrogation questions (N = 806) formulated in Study Ib with respect to how leading, doubt-casting, and open-ended they were. No differences emerged in the types of questions posed to non-native versus native speakers. However, when interrogators perceived a witness as more credible, they posed fewer questions casting doubt, but this effect was only observed for native-speaking witnesses. **Study III** examined whether the presence (vs. absence) of a set of predefined non-verbal cues predicted the accuracy of 680 statements, witnesses' self-reported confidence, and observer-rated credibility. None of the cues reliably predicted these outcomes, challenging assumptions about the diagnostic value of non-verbal cues in legal contexts. Together, this thesis provides converging evidence that language barriers can introduce bias in how eyewitness testimony is received and acted upon, not due to differences in testimony reliability, but due to how the testimonies are perceived in terms of credibility. The findings underscore the need for legal systems to better account for challenges placed on non-native speakers and to adopt practices that promote more fair credibility assessments.

Keywords: *eyewitness testimony, non-native speakers, memory accuracy, suggestibility, confidence-accuracy, investigative interviews, non-verbal cues, perceived credibility.*

Stockholm 2025
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-249101

ISBN 978-91-8107-416-1
ISBN 978-91-8107-417-8

Department of Psychology

Stockholm University, 106 91 Stockholm

TESTIFYING IN A NON-NATIVE LANGUAGE AS AN EYEWITNESS UNDERMINES CREDIBILITY JUDGMENTS BUT NOT RELIABILITY

Arman Raver

Stockholm
University

Testifying in a non-native language as an eyewitness undermines credibility judgments but not reliability

Arman Raver

©Arman Raver, Stockholm University 2025

ISBN print 978-91-8107-416-1
ISBN PDF 978-91-8107-417-8

Cover by Cihangir

Printed in Sweden by Universitetsservice US-AB, Stockholm 2025

To my brother, *Bover*

Abstract

Eyewitness testimony plays a critical role in legal proceedings. Yet, its usefulness hinges on both the reliability of the information from the witness, and how that information is perceived. When eyewitnesses testify in a non-native language, language barriers may shape not only what is communicated but also how their testimony is interpreted and judged. This thesis investigates whether testifying in a non-native versus native language affects the reliability and perceived credibility of honestly reported eyewitness testimonies. Across three empirical studies, I examined how language influences memory accuracy, suggestibility, the confidence-accuracy relation, observer-rated credibility, investigative questioning practices, and non-verbal behavior.

Study I comprised two experiments. In **Study Ia** ($N = 121$), mock eyewitnesses provided testimony in either a non-native (English; $n = 61$) or their native (Swedish; $n = 60$) language after viewing a mock-crime video. Testimonies accuracy and suggestibility were measured, and the confidence-accuracy relation computed. Witnesses also rated their own credibility and cognitive effort when giving the testimony. No differences emerged between witness groups in accuracy, suggestibility, cognitive effort, or self-rated credibility. However, in line with the confidence-accuracy relation, witnesses were more confident overall in correct than incorrect details. Importantly, non-native speakers reported significantly lower confidence than native speakers, despite being equally accurate. In **Study Ib** ($N = 202$), mock interrogators were randomly assigned to assess either a non-native or native video-recorded testimony for credibility and to formulate follow-up interrogation questions. The interrogators judged non-native speakers as less credible than native speakers.

Study II examined the interrogation questions ($N = 806$) formulated in **Study Ib** with respect to how leading, doubt-casting, and open-ended they were. No differences emerged in the types of questions posed to non-native versus native speakers. However, when interrogators perceived a witness as more credible, they posed fewer questions casting doubt, but this effect was only observed for native-speaking witnesses.

Study III examined whether the presence (vs. absence) of a set of predefined non-verbal cues predicted the accuracy of 680 statements, witnesses' self-reported confidence, and observer-rated credibility. None of the cues reliably predicted these outcomes, challenging assumptions about the diagnostic value of non-verbal cues in legal contexts.

Together, this thesis provides converging evidence that language barriers can introduce bias in how eyewitness testimony is received and acted upon, not due to differences in testimony reliability, but due to how the testimonies are perceived in terms of credibility. The findings underscore the need for legal systems to better account for challenges placed on non-native speakers and to adopt practices that promote more fair credibility assessments.

Keywords: eyewitness testimony, non-native speakers, memory accuracy, suggestibility, confidence-accuracy, investigative interviews, non-verbal cues, perceived credibility.

Sammanfattning

Vittnens minnen av en brottshändelse spelar en avgörande roll i rättsliga förfaranden. Vittnesmålens användbarhet beror inte bara på tillförlitligheten i det som återges, utan också på hur denna information uppfattas och bedöms. När ett vittne ger sitt vittnesmål på ett språk som inte är hens modersmål kan språkbarriärer påverka både vad som kommuniceras och hur vittnesmålet tolkas och värderas. Denna avhandling undersöker huruvida vittnesmål som ges på ett annat språk än modersmålet påverkar tillförlitligheten och den upplevda trovärdigheten i ärligt återgivna vittnesutsagor. Tre empiriska studier undersökte hur språk påverkar korrektheten i vittnesutsagan, påverkbarhet, sambandet mellan säkerhet och korrekthet, observatörers trovärdighetsbedömningar, förhörsfrågors utformning samt icke-verbalt beteende.

Studie I bestod av två experiment. I **Studie Ia** ($N = 121$) lämnade fingerade vittnen sina vittnesmål antingen på ett annat språk än modersmålet (engelska; $n = 61$) eller på sitt modersmål (svenska; $n = 60$) efter att ha sett en film med ett iscensatt brott. Vittnesmålen bedömdes med avseende på korrekthet, påverkbarhet och sambandet mellan säkerhet och korrekthet. Vittnena rapporterade även sin upplevda kognitiva ansträngning och skattade sin egen trovärdighet. Resultatet visade inga signifikanta skillnader mellan språkgrupperna vad gäller korrekthet, påverkbarhet, kognitiv ansträngning eller självskattad trovärdighet. I linje med tidigare forskning kring ett vittnesmåls korrekthet och vittnets säkerhet var vittnen generellt mer säkra på korrekta än på felaktiga minnen. Resultatet visade dock att vittnen som gav sitt vittnesmål på ett språk som inte var deras modersmål var mindre säkra än de som vittnade på sitt modersmål, trots att deras utsagor var lika korrekta. I **Studie Ib** ($N = 202$) fick deltagare i rollen som fingerade förhörsledare dels bedöma trovärdigheten i ett videoinspelat vittnesmål från ett vittne som antingen talade sitt modersmål eller ett annat språk, samt formulera uppföljande förhörsfrågor. Resultaten visade att vittnen som inte vittnade på sitt modersmål bedömdes som mindre trovärdiga än de som talade sitt modersmål.

I **Studie II** analyserades de förhörsfrågor ($N = 806$) som formulerats **i Studie Ib** med avseende på hur ledande de var, om de uttryckte tvivel, och om de var öppna eller inte. Resultatet visade inga skillnader i frågetyper mellan vittnen som talade ett icke-modersmål och de som talade sitt modersmål. Däremot

ställdes färre frågor som uttryckte tvivel till vittnen som förhörsledarna uppfattade som mer trovärdiga, detta samband var dock begränsat till vittnen som vittnade på sitt modersmål.

Studie III undersökte huruvida förekomsten (jämfört med frånvaron) av tolv fördefinierade icke-verbala beteenden var kopplade till korrekthet i enskilda vittnesutsagor (N = 680), vittnens självskattade säkerhet samt observatörers bedömningar av trovärdighet. Inget av de undersökta icke-verbala beteendena visade något samband med korrekthet, säkerhet eller bedömd trovärdighet. Resultatet indikerar att icke-verbala beteenden inte är tillförlitliga markörer för bedömning av korrekthet av vittnens utsagor i rättsliga sammanhang.

Sammanfattningsvis visar denna avhandling samstämmiga belägg för att språkbarriärer kan påverka hur vittnesmål uppfattas och används, inte på grund av skillnader i tillförlitlighet, utan på grund av hur vittnesmålen upplevs i termer av trovärdighet. Resultaten understryker behovet av att rättsväsendet bättre beaktar de särskilda utmaningar som vittnen möter när de vittnar på ett språk som inte är deras modersmål, samt att utveckla mer rättssäkra rutiner för trovärdighetsbedömningar.

List of studies

This thesis is based on the following published studies, henceforth referred to by their assigned roman numerals.

I. **Raver**, **A**., Lindholm, T., Gustafsson, P. U., & Alm, C. (2023). Memory accuracy, suggestibility and credibility in investigative interviews with native and non-native eyewitnesses.
Frontiers in Psychology, *14*, 1240822.
https://doi.org/10.3389/fpsyg.2023.1240822

II. **Raver**, **A**., Lindholm, T., Liljestrand Hassoun, S., & Alm, C. (2024). Interrogation questions to native and non-native eyewitnesses: The role of witness credibility.
Legal and Criminological Psychology.
https://doi.org/10.1111/lcrp.12301

III. **Raver**, **A**., Lindholm, T., & Alm, C. (2025). Non-verbal cues in eyewitness testimonies do not predict accuracy or credibility assessments.
Scientific Reports, *15*(1), 5265.
https://doi.org/10.1038/s41598-025-89825-0

Author contributions

Author contributions are categorized according to the Contributor Roles Taxonomy (CRediT).

I. *Arman Raver*: writing—original draft, writing—review and editing, conceptualization, methodology, formal analysis, investigation, data curation, and visualization.
Torun Lindholm: writing—review and editing, conceptualization, methodology, resources, visualization, supervision, project administration, and funding acquisition.
Philip U. Gustafsson: writing—review and editing, resources.
Charlotte Alm: writing—review and editing, conceptualization, methodology, resources, visualization, supervision, project administration, and funding acquisition.

II. *Arman Raver*: writing—original draft; writing—review and editing; conceptualization; methodology; formal analysis; investigation; project administration; data curation; visualization.
Torun Lindholm: writing—review and editing; conceptualization; methodology; resources; visualization; supervision; project administration; funding acquisition; data curation.
Sofie Liljestrand Hassoun: writing—review and editing; methodology; investigation; project administration; data curation; visualization.
Charlotte Alm: writing—review and editing; conceptualization; methodology; resources; visualization; supervision; project administration; funding acquisition; data curation.

III. *Arman Raver*: writing—original draft, writing—review and editing, conceptualization, methodology, formal analysis, investigation, data curation, and visualization.
Torun Lindholm: writing—review and editing, conceptualization, methodology, resources, supervision, project administration, and funding acquisition.
Charlotte Alm: writing—review and editing, conceptualization, methodology, resources, supervision, project administration, and funding acquisition.

Contents

Abstract……i
Sammanfattning……ii
List of studies……iii
Author contributions……iv
Contents……v
Introduction……1
The role of eyewitness testimony in legal investigtions……1
The memory process……1
Cognitive capacity and non-native language use: Theoretical basis…… 2
Language and judgments of eyewitness testmony……3
Testifying in a non-native language……4
Challenges in practice……4
Empirical gaps in the literature……5
Research questions and aims……6
Operationalizations……7
Core components……7
Memory accuracy……7
Accuracy of non-native vs. native eyewitnesses……8
Confidence-accuracy relation……9
Confidence-accuracy relation for non-native vs. native eyewitnesses……9
Suggestibility……10
Suggestibility of non-native vs. native eyewitnesses……11
Credibility assesment……11
Credibility of non-native vs. native eyewitnesses……12
The role of non-verbal cues in judgments of eyewitness testimony……13
Non-verbal cues of non-native vs. native eyewitnesses……13
How core components relate to eyewitness testimony involving language barriers……14
Methodological overview……16
Study samples……16
Pre-registered sample size justification……17
Materials, Measures, and Coding procedures……18
Crime stimulus……18
Investigative interview procedure……18
Suggestibility……19
Video-recorded eyewitness testimonies……19
Credibility ratings……20

Eyewitness statements, memory accuracy and the confidence-accuray relation……..……..……..20
Interrogation questions……..21
Non-verbal cues……..……22
Cognitive effort……23
Experimental designs and procedures……...23
Ethical considerations……...26
Statistical analyses……26
A note on interpretation 1……..……...27
Study summaries……..28
Study I: Memory accuracy, suggestibility and credibility in investigative interviews with native and non-native eyewitnesses……..…….……..28
Study Ia……28
Methods……..……...28
Results and discussion……..……28
Study Ib……31
Methods……..……...31
Results and discussion……..……31
General conclusion……..……..…32
Study II: Interrogation questions to native and non-native eyewitnesses: The role of witness credibility……..……...33
Methods……..……..33
Results and discussion……..……34
Conclusion……..……...36
Study III: Non-verbal cues in eyewitness testimonies do not predict accuracy or credibility assessments……..……..…..……36
Methods……..……..37
Results and discusion……..……..……37
Conclusion……..……...38
General discussion……..39
Synthesis of findings across studies……..40
A note on interpretation 2……..……...42
Cross-linguistic and cultural biases in legal systems……43
Methodological considerations……43
Samples, validity and generalizability……..……..43
Statistical considerations……..……..46
Conclusion……47
Theoretical implications……..……47
Practical implications……48
Directions for future research……...49
Concluding remarks……..49
References……..……51
Acknowledgements……..63

Introduction

The role of eyewitness testimony in legal investigations

Eyewitness testimony remains one of the most influential forms of evidence in legal proceedings. Legal professionals (e.g., police officers, interviewers, judges, jurors, and other fact-finders) often place substantial weight on witness statements, particularly when the witness is perceived as detailed, confident, and emotionally compelling (Wells et al., 2006; Wixted et al., 2018). However, research has consistently shown that eyewitnesses are not always reliable, even when they report events with sincere intent to assist the investigation (Garrett, 2011; Meissner, 2021). Here, I refer to such accounts as *honestly reported eyewitness testimonies* and focus specifically on their legal implications. In fact, eyewitness error is one of the leading causes of wrongful convictions across the world. According to the Innocence Project (2023; a non-profit legal organization dedicated to exonerating individuals who have been wrongly convicted of crimes), mistaken eyewitness identification has played a role in approximately 70% of convictions overturned through DNA evidence in the United States. Also, in the United States, 55% of exonerees were convicted based on eyewitness memory errors, while in Europe, 34 such cases have been documented (The National Registry of Exonerations, 2025; European Registry of Exonerations, 2024). Such errors typically arise not from witnesses' intentional deception, but from sub-optimal practices when interviewing sincere witnesses and from the natural fallibility of human memory (Albright, 2017; Memon et al., 2010; Wade et al., 2018).

The memory process

The memory process unfolds across three stages: encoding, storage, and retrieval (Melton, 1963). *Encoding* refers to the initial process of perceiving and attending to an event. This stage is sensitive to external (e.g., lighting, distance, and duration) and internal factors (e.g., attention, stress, and emotional state), which can limit and affect what information is registered in the first place (Albright, 2017; Loftus, 2005). *Storage* involves maintaining the encoded information over time. During this phase, memories are not fixed but are subject to decay and distortion from both internal (e.g., imagination) and

external influences (e.g., post-event misinformation; Loftus, 2005; Schacter, 2012). *Retrieval* is the process of recalling stored information, such as during investigative interviews or courtroom testimony. Importantly, retrieval is not like replaying a video of our past experiences. Rather, retrieval is an active and reconstructive process: people reconstruct past experiences from incomplete fragments shaped by associations, prior knowledge and beliefs, and the present context. As part of this metacognitive process, people often rely on a subjective sense of confidence to judge the accuracy of their memories, a judgment thought to be based on retrieval-related cues such as retrieval ease, fluency or coherence rather than direct access to memory accuracy (Koriat, 2012). In legal contexts, this brings into focus the concept of testimony *reliability*, operationalized in this thesis as the extent to which information reported by a witness is accurate (i.e., corresponds to what actually happened) and resistant to suggestion (i.e., remains stable when exposed to external influence).

The reliability of reported memories is never guaranteed. Because memory is shaped across the stages of encoding, storage, and retrieval, eyewitness accounts are susceptible to distortion, omission, and influence (Dodson & Schacter, 2015; Wells & Loftus, 2003). Notably, research on suggestibility has shown that even sincere eyewitnesses can report inaccurate information when questioned in a misleading or suggestive manner (Loftus, 2005; Gudjonsson, 2003). Such inaccuracies do not reflect deception but rather the malleability of human memory under social (e.g., suggestive questions, interpersonal pressure) and cognitive (e.g., confidence, cognitive effort) pressures due to different *investigative interview practices*. This becomes particularly important when witnesses testify in a non-native language. As credibility judgments often rely on how the statement is delivered, rather than its content, non-native speakers may be especially vulnerable to biased assessments, yet this remains an underexplored issue in the literature.

Cognitive capacity and non-native language use: Theoretical basis

According to theoretical models of working memory and attentional control, retrieving and articulating memories is a cognitively demanding task. These models posit that memory retrieval requires cognitive resources to search for, evaluate, and verbalize relevant information (Baddeley, 2012; Van Acker et al., 2018). From this perspective, working memory has a limited capacity and must coordinate multiple operations simultaneously, such as maintaining relevant details while suppressing distractions or irrelevant cues (Cowan, 2010; Cowan et al., 2024; Engle, 2002). These demands can increase during investigative interviews, where witnesses must retrieve, monitor, and

articulate memories under pressure. When this is done in a non-native language, the cognitive load can increase even further (Chen & Chang, 2009). Producing speech in a non-native language taxes cognitive capacity for word selection, grammatical construction, and suppression of interference of ones' native language (Abutalebi, 2008; Abutalebi & Green, 2007; Branzi et al., 2016; Green, 1998; Green & Abutalebi, 2013). This added load may reduce the capacity to monitor and structure memory output effectively (Segalowitz, 2010). While the underlying memory trace may remain intact, the act of reporting it becomes more effortful. Importantly, these cognitive costs may affect not only the content of the testimony but also its delivery – potentially leading to less fluent, less confident, or more hesitant responses, which in turn may shape how fact-finders perceive the witness. Thus, based on theories of cognitive capacity it can be expected that second-language use may not necessarily impair memory reliability, but instead might increase cognitive demands at the point of retrieval and verbalization, creating conditions under which both the witness's subjective experience and observers' assessments may become biased.

Language and judgments of eyewitness testimony

Due to rising global migration (United Nations, 2020), legal professionals increasingly encounter eyewitnesses who testify in a language other than their native one. While memory errors generally are well-documented (Wixted et al., 2018), psychological research has largely overlooked how testifying in a non-native language influences what witnesses are able to communicate, how accurately they do so, their vulnerability to suggestion, and how their testimony is perceived. These concerns relate not only to the actual reliability of a testimony (i.e., memory accuracy and suggestibility) but also to the perceived *credibility* of the witness. Credibility refers to how trustworthy or believable a witness's statement is perceived to be, and it plays a decisive role in legal outcomes (Boccaccini, 2002). Crucially, credibility judgments are shaped not only by the factual content of a statement but also by how it is delivered and how the witness is perceived. This creates potential for bias, particularly when witnesses testify in a non-native language.

Despite formal safeguards provided by legal systems, such as the right to an interpreter (Directive 2010/64/EU, 2010) and the principle of equal treatment under the law (European Convention on Human Rights, 1950), legal professionals' credibility assessments are often influenced by extra-legal factors. These include preconceptions about how (dis)honest witnesses are expected to behave (Denault et al., 2020, 2024; Kaufmann et al., 2003; Rogers et al., 2014; Wessel et al., 2006), and how they should express emotion (Bandes, 2014). Furthermore, language-related characteristics, such as foreign accents

or non-fluent speech, can also bias credibility assessments. Research shows that statements from ethnic outgroup (vs. ingroup) members and accented (vs. non-accented) speakers are often judged as less credible, even when no actual differences in accuracy are present (Lindholm, 2005, 2008; Lindholm & Yourstone, 2013; Frumkin, 2000, 2007, 2020; Rakic et al., 2011).

Testifying in a non-native language

Producing speech in a non-native language requires greater cognitive effort (Abutalebi, 2008). Moreover, previous research suggests that when the language of retrieval matches that of encoding, recall tends to be more accurate – an effect explained by language-dependent memory and encoding specificity, which hold that memory is enhanced when the retrieval context mirrors the original learning conditions (Marian & Kaushanskaya, 2004, 2007; Schroeder & Marian, 2014; Tulving & Thomson, 1973). Cultural background may also shape how eyewitnesses express their memories and how their accounts are evaluated. For instance, research on cross-cultural eyewitness testimony has shown that individuals from some cultures tend to report more specific event-related details, while individuals from others are more likely to include contextually rich or interpretive content, such as background information or reflections on the situation (Anakwah et al., 2020; de Bruïne et al., 2025). Such cross-cultural differences in communication style may influence how testimony is interpreted and assessed in legal settings (Hope et al., 2022). Other language-related barriers may arise from level of non-native language proficiency, accent, and cross-linguistic differences in how words and emotions are understood or expressed, all of which can increase the risk of misunderstanding in high-stakes legal contexts (Fausey & Boroditsky, 2011; Itzhak et al., 2017). Despite these findings, few studies have in fact examined how non-native language use play out in high-stakes investigative interviews.

Challenges in practice

While interpreters are commonly used when non-native speaking eyewitnesses testify, literal or culturally insensitive translations can distort meaning and affect testimony reliability (Dhami et al., 2017; Goodman-Delahunty & Howes, 2019; Hope et al., 2025; Lai & Mulayim, 2014; Vredeveldt et al., 2023; Walsh et al., 2020). To support early memory retrieval and to reduce potential interviewer influence, tools such as the Self-Administered Interview have been developed (SAI is a standardized written protocol designed to help witnesses provide detailed statements without the need of an interviewer; Gabbert et al., 2009; Hope et al., 2011). Meta-analytic evidence suggests that the SAI increases the number of correctly recalled details compared to free recall, although it also results in more incorrect details and a small drop in overall

accuracy (Horry et al., 2020). While potentially useful in resource-limited contexts (e.g., interpreter availability) or when language barriers are present, the SAI poses practical constraints for non-native speakers. For instance, Ernberg and Mac Giolla (2023) showed that the SAI took substantially longer to complete than face-to-face interviews, increased cognitive demands, and required translation, factors that can delay the investigative process and elevate witness stress. Participants also reported higher levels of nervousness in the SAI condition, potentially due to increased cognitive demands.

Empirical gaps in the literature

To date, three main unresolved gaps characterize the existing literature. First, while some studies have addressed non-native language use in legal contexts such as asylum cases and interpreter-mediated interviews (Allison et al., 2022; Walsh et al., 2020), very few have focused specifically on honestly reported eyewitness testimony. Second, prior research on how testifying in a non-native language affects memory reliability, operationalized as memory accuracy and resistance to suggestion, has produced mixed and inconclusive findings (see section Operationalizations). Third, no existing studies have systematically examined how non-native language use influences credibility judgments or the nature of questioning in investigative interviews involving honest eyewitnesses. Taken together, there is a need for greater empirical clarity, specifically, a better understanding of when, how, and to what extent testifying in a non-native language influences honestly reported eyewitness testimonies and how they are evaluated. This thesis directly addresses these issues by systematically examining how testifying in a non-native versus native language affects memory accuracy, susceptibility to suggestion, perceived credibility, and the nature of investigative questioning.

To clarify, this thesis does not examine cross-cultural variation or compare specific languages; rather, it focuses on the impact of testifying in a non-native versus native language. Importantly, this thesis does not assume that non-native speakers are less capable of remembering accurately. Rather, it investigates whether the reporting and perception of testimonies are adversely affected by language barriers, with downstream consequences for how eyewitness statements are evaluated in legal contexts.

Research questions and aims

Given that eyewitness testimonies often serve as crucial evidence in criminal investigations, the overarching aim of this thesis is to examine how testifying in a non-native language may impact the reliability and credibility assessments of such testimony. Specifically, it investigates whether language barriers influence what eyewitnesses report, how accurately they do so, and how their statements are perceived and assessed. The thesis further explores potential sources of bias in these assessments, including extra-legal factors such as non-verbal behavior, that may unfairly disadvantage non-native speakers. In doing so, it contributes to the understanding of how language barriers influence memory performance, credibility judgments, and the fairness of investigative practices.

Study I consisted of two separate experiments that examined testimony reliability and credibility, respectively. **Study Ia** examined whether memory accuracy, suggestibility and self-rated cognitive effort differed between non-native and native eyewitnesses. This study also investigated the confidence-accuracy relation and whether this relation differed between non-native and native eyewitnesses. **Study Ib** examined whether independent observers' credibility assessments of non-native and native eyewitnesses differed.

Study II examined the role of perceived credibility in investigative interview questioning and specifically if non-native and native speaking eyewitnesses are questioned differently. I investigated whether non-native eyewitnesses received more leading questions, questions that cast doubt on the witness, and fewer open-ended questions compared to native eyewitnesses. Additionally, I investigated whether the perceived credibility was associated with the degree to which questions were leading, casting doubt, or open-ended.

Study III investigated the predictive value of non-verbal cues in honestly reported eyewitness testimonies. Specifically, I examined whether a set of non-verbal cues was associated with the accuracy of eyewitness statements, witnesses' self-reported confidence, and observers' ratings of witness credibility. Additionally, I explored whether these associations differed between non-native and native eyewitnesses.

Operationalizations

Core components

Scientific progress depends on conceptual clarity and exactitude. Defining key terms sharpens the focus of inquiry, allowing researchers to identify and examine the phenomena most relevant to their aims. To address my research questions, I begin by outlining key psychological components relevant to eyewitness reliability (i.e., memory accuracy and suggestibility) and perceived credibility, as well as investigative interviewing and extra-legal factors. The following section reviews memory accuracy, the confidence-accuracy relation, suggestibility, credibility judgments, and non-verbal cues, with a focus on how each may be affected when testimony is delivered in a non-native language. For each component, an overview is provided followed by empirical evidence on differences (if any) between non-native and native speakers.

Memory accuracy

In the context of eyewitness testimony, memory accuracy refers to the degree to which reported details correspond to what actually occurred during a target event. However, human memory is not a literal recording of reality. Instead, remembering is an active, reconstructive process shaped by how the event was perceived and interpreted (Schacter, 2012).

In laboratory settings, accuracy can be objectively verified by comparing a witness's report to the known facts of a mock-crime event. In real-world legal contexts, however, such verification is rarely possible for eyewitness testimonies, especially in the absence of corroborating evidence such as CCTV footage or cross-validation from other sources (Albright, 2017; Wells et al., 2006). Courts must therefore often rely on indirect indicators of witness accuracy, such as perceived credibility, confidence, consistency, and plausibility.

While eyewitness testimonies can be a reliable and valuable source of information, it is of importance to understand *when and how* memory can fail. To this end, research in the eyewitness memory paradigm has focused on identifying situational and individual factors that shape memory performance. These include, for example, acute stress (Deffenbacher et al., 2004), age and developmental-related effects (Ceci et al., 2007; Otgaar et al., 2018), alcohol intoxication (Jores et al., 2019), exposure durations to the witnessed event (Memon et al., 2003), the presence of a weapon during the event (Fawcett et

al., 2013), and investigative interview practices (Fisher & Schreiber, 2017). These findings on factors related to eyewitness accuracy is further supported by a recent survey showing broad consensus among eyewitness memory researchers that even honest witnesses are vulnerable to well-established memory distortions, including the effects of attitudes and expectations, long retention intervals, misinformation, suggestive questioning, divided attention, false memory formation, and cross-race bias (Seale-Carlisle et al., 2024). If fact-finders assume that memory operates like a video-recorder, they risk misjuding its reliability and drawing flawed conclusions. However, despite these advances in eyewitness memory research, relatively little is known about how testifying in a non-native language may affect memory accuracy and the perceived reliability of eyewitness accounts. This gap is particularly striking given the growing linguistic diversity in legal contexts worldwide (United Nations, 2020).

Accuracy of non-native vs. native eyewitnesses. While eyewitness memory has been extensively studied, relatively few studies have examined how testifying in a non-native language affects eyewitness accuracy. Prior to the work presented in this thesis, only a handful of published studies had addressed this issue. One previous study, Alm et al. (2019), had Swedish participants watch a mock-crime film and then provide a detailed free verbal recall as if they had witnessed the event in real life, either in English (non-native) or Swedish (native). Despite reporting higher levels of cognitive effort, non-native speakers were just as accurate as native speakers. Hu and Naka (2022) reported that bilingual eyewitnesses' accuracy varied depending on the type of detail recalled: object (e.g., weapon descriptions) and action (e.g., stabbing) details were reported more accurately in the native language, whereas location details (e.g, parking lot) were better recalled in the non-native language. Moreover, Ernberg and Mac Giolla (2022) compared accuracy across three interview formats: native language with interpreter, non-native language without interpreter, and a SAI translated to witnesses' native language. Witnesses interviewed in either the with- or without-an-interpreter condition showed similar accuracy levels, while the SAI condition exhibited slightly lower accuracy. However, the differences observed between groups were only minor. In sum, the direction and size of any language-related effects on testimony accuracy from these studies remain unclear. These results call for further scrutiny, and the present thesis contributes to this effort by examining whether testifying in a non-native language influences the accuracy of honestly reported eyewitness testimony.

Confidence-accuracy relation

The confidence-accuracy relation refers to the degree to which a witness's self-reported confidence in a specific reported statement reflects the actual accuracy of that memory. This relation has long been a topic of theoretical and practical interest, as confidence is frequently treated as a proxy of memory reliability by legal professionals (Fitzgerald et al., 2021; Yates, 2017). However, research has repeatedly demonstrated that confidence is not a flawless predictor of accuracy. Generally, under pristine conditions (such as when recall is immediate and uncontaminated by external influence; Penrod & Cutler, 1995), witnesses tend to be more confident in correct than incorrect detail information, suggesting some degree of calibration between peoples subjective confidence in their memory, and the objective veracity of that memory (Juslin et al., 1996; Mickes et al., 2017, 2025; Sporer et al., 1995; Wixted & Wells, 2017). While this difference is statistically reliable in the long-run, the overlap between confidence ratings for correct and incorrect memories, both between- and within-subjects, remains, limiting its precision at the individual level. That is, even well-calibrated witnesses may report high confidence in inaccurate statements, or report low confidence in correct ones. Moreover, the confidence-accuracy relation is highly malleable by situational and procedural influences (i.e., non-pristine conditions). Factors such as delays between the event and recall, feedback from investigators, stress, and repeated questioning have all been shown to disrupt the confidence-accuracy relation (Brewer & Wells, 2006; Mickes et al., 2025; Wixted & Wells, 2017; Wixted et al., 2015). Because such non-pristine conditions remain common in real-world investigative settings, despite best-practice recommendations, confidence ratings must be interpreted with caution when pristine conditions are not met. This becomes especially relevant when examining witnesses who testify in a non-native language.

Confidence-accuracy relation for non-native vs. native eyewitnesses. Testifying in a non-native language typically places high cognitive demands on witnesses (Abutalebi, 2008), which can interfere with the cues they normally rely on, such as how easily a memory comes to mind, to assess the reliability of their memories (Lindholm et al., 2018; Gustafsson et al., 2019). According to Koriat and Goldsmith (1996), confidence reflects a metacognitive-monitoring process in which individuals evaluate the fluency of retrieval, and use this fluency to decide whether a memory is accurate enough to report. When retrieval feels effortful, the rememberers' confidence tends to be lower, regardless of accuracy. Under increased cognitive load, such as when testifying in a non-native language, witnesses may experience reduced retrieval fluency. his may lead witnesses to omit details they struggle to retrieve, not as a function of confidence, but due to the effort involved in accessing less accessible information. While the details they do report may still be accurate and easier to retrieve, the overall effortful retrieval experience might nonetheless lower

their expressed confidence. Whether this pattern disrupts the typical confidence-accuracy relation remains an open empirical question, but the theoretical rationale for expecting such effects among non-native speakers is well-founded.

Suggestibility

Suggestibility refers to how external influences, such as suggestive questions or interpersonal pressure, can distort an eyewitness's account, even when the witness is honest and motivated to be accurate (Gudjonsson, 2003; Loftus, 2005). Eyewitnesses have repeatedly been shown to be susceptible to misleading or suggestive information, particularly in studies on the misinformation effect (i.e., memory distortion following exposure to misleading post-event information; Edmar et al., 2025; Loftus & Palmer, 1974) and the impact of question framing, where subtle wording shifts (e.g., "smashed into" vs. "hit") change what is remembered or reported (Lindsay & Johnson, 1989; Loftus & Palmer, 1974; Loftus, 2005). Gudjonsson (2003) defines suggestibility as the tendency to accept and incorporate misleading information into one's initial memory account, typically as a result of uncertainty, weak memory trace, or confusion. Importantly, it distinguishes suggestibility from compliance, where witnesses may report inaccurate information not because they believe it, but to align with perceived expectations or to reduce social discomfort (Gudjonsson & Clark, 1986). These dual influences (i.e., suggestibility and compliance) make witnesses especially prone to suggestion when they experience doubt, low confidence, or heightened anxiety (Gudjonsson & Clark, 1986).

The structure and phrasing of questions in investigative interviews play a decisive part in increasing the reliability of witness testimony. Open-ended questions that invite free narrative recall are considered the gold standard and yield more accurate and uncontaminated responses (Brewer et al., 2018; Fisher & Geiselman, 1992; Geiselman & Fisher, 1989; Meissner et al., 2021). In contrast, leading questions often result in memory distortion by narrowing the scope of response, suggesting answers, or implying that the initial report was inadequate, for example, "*You saw the knife in his hand, didn't you?*" (Hope & Gabbert, 2019). Even subtle forms of suggestive questioning, such as questions that cast doubt, for example, "*Are you sure that's what you saw?*" can undermine witness' confidence, and influence what witnesses report and how they report it (Chan et al., 2017; Kebbell et al., 2010). Thus, when questions guide or constrain the witness too strongly, the interviewer's influence on the testimony can become so substantial that it raises concerns about authorship of the provided testimony: Does the account reflect the witness's initial memory or the interviewer's suggestions?

Despite strong empirical support for best-practices in investigative interviews, problematic questioning techniques remain common in the real-world legal procedures (Cederborg et al., 2013; Fisher, 1995; Hyman Gregory et al., 2023; Schreiber Compo et al., 2012).

Suggestibility of non-native vs. native eyewitnesses. As noted, individuals testifying in a non-native language are increasingly common in legal contexts worldwide (United Nations, 2020). However, little research has examined how language barriers affect susceptibility to suggestions. Testifying is cognitively demanding for any witness, but non-native speakers face additional challenges since retrieving and articulating memories in a language other than the primary one requires greater cognitive effort through interference by the native language (Abutalebi, 2008). This interference may reduce the cognitive and regulatory resources necessary to resist misleading questions, making non-native speakers more vulnerable to suggestions (Alm et al., 2019; Otgaar et al., 2012). Combined with other cognitive factors such as lower confidence, heightened anxiety, and uncertainty about how they are perceived – common under linguistic strain – these conditions may increase susceptibility to suggestibility and compliance (Gudjonsson & Clark, 1986). Thus, non-native eyewitnesses may be particularly prone to yielding to leading or misleading suggestions because of the cognitive and social pressures (e.g., deference to authority, fear of negative evaluation, or a desire to avoid conflict) embedded in the act of testifying in a non-native language.

On top of this, non-native speakers often face communication challenges due to differences in how their native language structures information. For example, some languages allow events to be described without specifying the agent (e.g., "*The vase broke*" instead of "*She broke the vase*"), which can lead investigators to increase the use of directed, closed or otherwise suggestive questions (Boroditsky et al., 2002; Filipović, 2011, 2020). Furtheremore, non-native speakers often belong to ethnic outgroups relative to interviewers. Such factors can trigger increased skepticism from legal professionals (Scassa, 1994), potentially reducing perceived credibility. Prior research shows that when interviewers doubt a witness's credibility, they tend to shift from open-ended, information-gathering approaches to more confrontational questioning (van Veldhuizen et al., 2017; de Bruïne et al., 2023). Such questioning approaches not only risks distorting the content of testimony but can also erode the witness's confidence over time. Repeated questioning has been shown to influence confidence independently of actual accuracy (Shaw III, 1996; Shaw III & McClure, 1996), thereby disrupting the confidence-accuracy relation. As a result, witnesses may express unwarranted confidence in inaccurate details, potentially misleading both themselves and fact-finders, and causing such detail information to appear more accurate than it truly is.

Credibility assessement

Credibility refers to how trustworthy and believable an eyewitness appears to legal professionals. It is not determined by objective accuracy, but by whether the observer perceives the witness as a convincing source of information. Credibility assessments are often shaped by a combination of epis-

temic cues (e.g., a testimony's internal consistency, level of detail, and witness' self-reported confidence; Evans & Fisher, 2011; Fisher & Cutler, 1995; Smeets et al., 2004) and extra-legal factors (e.g., the witness's physical appearance, non-verbal behavior, or perceived traits such as warmth and competence; Boccaccini, 2002; Denault et al., 2020, 2024; Hansen et al., 2017; Neal et al., 2012). However, eyewitnesses who provide detailed accounts are typically seen as more credible, even though the number of details reported is only weakly related to actual testimony accuracy (Wells & Leippe, 1981; Bell & Loftus, 1988). Moreover, observers often equate confident or emotionally expressive behavior with credibility, even when such testimony can be objectively inaccurate, while uncertainty or restraint tends to be interpreted as less credible, even when such testimony can be objectively accurate (Kaufmann et al., 2003; Rogers et al., 2014; Wessel et al., 2006).

Credibility of non-native vs. native eyewitnesses. Although no previous studies have directly examined how language barriers influence observer-rated credibility assessments of eyewitness, related research offers strong reason for concern. Experimental studies show that ethnic minority witnesses are often perceived as less credible than majority group members, even when their testimonies are equally accurate (Lindholm, 2005; 2008a; 2008b). This disparity appears rooted in how the testimony is delivered and perceived rather than in its actual content (Boccaccini, 2002; Kaufmann et al., 2003). Research has demonstrated that a foreign accent or poor fluency in speech can trigger negative judgments of credibility, regardless of accuracy (Hansen et al., 2017; Rakic et al., 2011). For example, studies by Frumkin (2000, 2007, 2020) has shown that witnesses with foreign accents were judged as less credible, less accurate, and more dishonest than native-accented speakers, even when the actual statements were identical. Such biases seem to arise from implicit norms and stereotyped expectations about how a credible witness should speak or behave.

Moreover, non-native speakers may pause more, search for words, or exhibit less fluent speech, which are natural byproducts of speaking in a non-native language. Yet observers may interpret such para-verbal cues (e.g., non-word fillers, hedges, pauses, hesitations, speech rate) as signs of uncertainty and poor memory (cf. Gustafsson et al., 2019, 2021, 2022; Lindholm et al., 2018). Consequently, non-native speaking witnesses may be unfairly penalized, not because of what they say, but because of how they say it.

The role of non-verbal cues in judgments of eyewitness testimony

In legal contexts, legal professionals often rely on non-verbal behaviors, such as facial expressions, body posture, and eye contact, when assessing eyewitness credibility. Yet, most research on non-verbal cues has focused on whether such cues differ between liars and truth-tellers, hence for detecting deception. Meta-analyses show that such cues are only weakly and inconsistently related to deception, with effect sizes typically small or negligible (DePaulo et al., 2003; Hartwig & Bond, 2011; Vrij & Granhag, 2012, 2014). Luke (2019) goes further, arguing that much of the existing deception-cue literature may be compromised by publication bias, low statistical power, and inflated effect sizes, casting doubt on the evidential value of the field as a whole.

However, while deception may not be discerned by non-verbal cues, it remains unknown how non-verbal behavior relate to honestly reported eyewitness testimonies. In these cases, witnesses are not trying to deceive but to recall events as accurately as possible. Yet, signs of nervousness, such as fidgeting or avoiding eye contact, may still be misinterpreted as signs of dishonesty or uncertainty (Bothwell & Jalil, 1992). These behaviors can stem from the stress of the interview, rather than from their account being false or inaccurate. This mismatch between the witness's behavior and the truthful content of their statement, known as content–behavior incongruence, can lead observers to judge otherwise accurate testimony as untrustworthy (Kaufmann et al., 2003; Wessel et al., 2006; Rogers et al., 2015)

While non-verbal behaviors do affect impressions of people in general (Burgoon et al., 2011; Esteve-Gibert & Guellaï, 2018; Moore et al., 2014; Patterson et al., 2023), their diagnostic value for memory accuracy is limited. This is especially problematic in legal contexts where stakes are high. For instance, Chalmers et al. (2022) found that mock jurors relied on gaze, posture, and signs of nervousness when evaluating witness credibility. Similarly, Denault et al. (2024) showed that judges often view gaze aversion, body movements, gestures, and facial expressions as indicative of a witness's credibility. As one judge remarked, "*some witnesses reveal their true personalities in the courtroom despite their best denials*" (Denault, 2022; p. 14), reflecting the belief that non-verbal behavior expose a witness's true intentions or credibility, even in the absence of empirical support.

Taken together, these findings suggest that credibility judgments may be unduly shaped by impressions of behaviors that bear little to no relation to the accuracy of the testimony.

Non-verbal cues of non-native vs. native eyewitnesses. The increased cognitive effort when testifying in a non-native language may manifest in non-verbal behaviors such as fidgeting, gaze aversion, or animated facial expressions, which often reflect cognitive strain rather than (un)certainty. However,

such behaviors can influence how confident a speaker appears to others, aligning with research on observers' inferences about internal states like Feeling of Knowing (i.e., a speaker's internal sense of certainty about a statement; Swerts & Krahmer, 2005). Non-native speakers may be especially vulnerable to these biases: Lindholm (2008) found that while credibility ratings increased with testimony accuracy for ethnic ingroup witnesses, high-accuracy ethnic outgroup witnesses, who continued to display behavioral cues associated with uncertainty, were rated no more credible than their low-accuracy outgroup counterparts. Yet little is known about whether non-verbal behavior in honestly reported testimonies systematically differs across language groups. Given the risk of misinterpretation, the role of non-verbal cues in credibility assessments warrants continued scrutiny within legal and investigative contexts.

How core components relate to eyewitness testimony involving language barriers

This thesis examines several core psychological components central to eyewitness testimony, namely: *memory accuracy*, *confidence-accuracy relation*, *suggestibility*, *credibility assessement*, and *non-verbal cues*. While these components are often studied independently, they collectively influence the content and evaluation of eyewitness accounts, especially in high-stakes legal contexts involving non-native speakers. The three empirical studies presented in this thesis investigate how each component is affected by language barriers, with the goal of promoting fair evaluations and recognizing the unique challenges faced by non-native witnesses in legal proceedings.

To contextualize the studies, it is important to clarify how these core components outlined above are not isolated phenomena. At the center of evaluating any eyewitness account lies the question of accuracy: to what extent do the reported details reflect what actually happened? This is a critical concern in legal contexts. In real-world cases, the ground truth is rarely verifiable, making it difficult to assess whether a testimony is accurate. Laboratory settings, by contrast, allow for direct comparisons between reported details and known events, enabling systematic investigation of eyewitness memory performance. Therefore in **Study Ia**, I aimed to investigate the question of whether *memory accuracy* differed depending on whether witnesses testified in their non-native versus native language. Since memory accuracy is not directly observable, legal professionals also often use a witness' self-reported confidence as a proxy. While this *confidence-accuracy* relation holds moderately under pristine conditions, confidence can be influenced by internal cues (e.g., how easily a memory comes to mind) and external factors (e.g., social feedback or question framing), which complicates its diagnostic value. In **Study Ia**, I investigated whether the confidence-accuracy relation differed for

non-native versus native speakers. In **Study Ia**, I also investigated witnesses' susceptibility to suggestions. *Suggestibility* further complicates testimony reliability, as suggestive questions may alter witnesses' recollections, thereby affecting both accuracy and self-reported confidence. These distortions can, in turn, undermine the witness's *perceived credibility*, the subjective assessment of whether a witness is a trustworthy source of information. These judgments often rely on superficial or intuitive cues such as the consistency of the account, emotional expressiveness, non-verbal behaviour, or how confident the witness seems. To capture witnesses' subjective experience of testifying and the cognitive effort involved, their self-rated credibility and perceived cognitive effort were also assessed. In **Study Ib**, I investigated whether non-native and native speakers were perceived differently in terms of credibility by independent observers.

In investigative interviews, the interrogation questions posed by interviewers can shape both what is remembered and how it is reported. Suggestive questions, such as leading questions or those that cast doubt, can affect the reliability of testimony and alter witness' confidence. In **Study II**, I investigated whether non-native (vs. native) speakers were more likely to receive such questions by an interviewer, and whether their perceived credibility influenced how they were questioned. Specifically, I investigated whether the degree to which interrogation questions were leading, casting doubt, and open-ended differed between language groups – to examine the investigative interview questioning practices employed in settings where language barriers exist.

Finally, credibility is often influenced by extra-legal factors such as misconceptions about *non-verbal cues*. In **Study III**, I explored this issue by investigating whether a witness' specific non-verbal cues predicted memory accuracy, self-reported confidence, and perceived credibility, and whether these associations differed between non-native and native speaking eyewitnesses.

Methodological overview

All studies were preregistered on the Open Science Framework (https://osf.io/ne4uk/), with materials, data, and statistical analysis code available in their respective project repositories. Ethical approvals were obtained from the Swedish Ethical Review Authority (File number 2020-00624).

Study samples

Across all three studies, I examined the role of language in eyewitness testimony by comparing non-native (English) and native (Swedish) speakers. English was selected as the non-native language for practical reasons, including time, resource constraints, and feasibility (see Methodological Considerations for limitations). Participants were recruited largely through Accindi (an online research platform), and also through in-person recruitment on Stockholm University's campus, university bulletin boards, and the Social Psychology Laboratory's Facebook group.

In **Study Ia**, participants (N = 121) viewed a mock-crime film from the perspective of eyewitnesses and were randomized to provide their testimonies in either their non-native (n = 61) or native (n = 60) language. Eyewitnesses assigned to the non-native language group were informed after viewing the mock-crime video that the subsequent interview would be conducted in English. In **Study Ib**, independent observers (N = 202) were randomly allocated to watch a film with either a non-native or native eyewitness testimony and formulated follow-up interrogation questions. They also assessed the testimony in terms of credibility. **Study II** extended this investigation by having a new group of independent raters (N = 207) evaluating a random subset of these interrogation questions with regard the degree to which the questions were leading, casting doubt, or was open-ended without watching the eyewitness testimony. Finally, in **Study III**, the presence (vs. absence) of non-verbal cues in the filmed eyewitness testimonies were coded to examine whether such cues predicted statement accuracy, self-reported confidence, and credibility ratings, and whether these associations differed between non-native and native speakers. For clarity, participants in **Study Ia** are henceforth referred to as eyewitnesses, those in **Study Ib** as interrogators, and those in **Study II** as raters. **Study III** used research assistants as coders, who are henceforth referred to as coders.

To ensure valid comparisons between language conditions, I carefully controlled for language proficiency and comprehension levels across studies. In **Study Ia**, eyewitnesses rated their proficiency in English after providing their testimony using the Common European Framework of Reference for Languages (CEFR) Global Scale (2019), as Basic, Independent, or Proficient. They then rated their comfort in speaking English on a 7-point Likert scale (1 = not at all comfortable, to 7 = very comfortable; $M = 5.30$, $SD = 0.91$, $Mdn = 6$). Non-native speakers who rated themselves to be both Proficient in the English language and having a maximum comfort level (7/7) were excluded to maintain the intended language manipulation.

In **Study Ib**, interrogators rated their own English comprehension proficiency on a 7-point Likert scale (1 = not at all comfortable, to 7 = very comfortable; $M = 6.09$, $SD = 0.97$, $Mdn = 6$) to ensure they were proficient enough to accurately evaluate the credibility of the testimonies. Interrogators who rated their comprehension below 3 were excluded from analysis to maintain a minimum proficiency threshold.

For **Study II** and **Study III**, language proficiency was not assessed: in **Study II**, all interrogation questions had been formulated in Swedish, and in **Study III**, coders evaluated the presence (vs. absence) of non-verbal cues rather than verbal content.

Across all studies, data on participant demographics (age, gender, and education level) were collected. To examine whether these variables influenced the outcomes, hierarchical multiple regression analyses were conducted in each study. As none of the covariates had statistically significant effects (all $ps > .05$), they are not reported further.

Pre-registered sample size justifications. For sample size justifications, I conducted a priori power analyses (to determine the required sample size for a desired effect size and power) and sensitivity analyses (to assess the smallest effect size detectable with the available sample size). The choice of method was based on practical constraints including time, financial resources, and feasibility.

For **Study Ia**, an a priori power analysis was conducted using G*Power (Erdfelder et al., 1996) to determine the minimum required sample size. An effect size of $d = 0.51$ was selected as a meaningful threshold, given the limited prior research on effects of language on eyewitness accuracy. With 60 eyewitnesses per language condition ($N = 120$ total), this design provided 80% power to detect an effect size of $d = 0.51$ at $\alpha = 0.05$ (two-tailed test). For **Study Ib**, a sensitivity analysis was conducted using Judd et al.'s (2017) power analysis tool for statistical models with random intercepts for eyewitnesses and interrogators. Given time and resource constraints, 202 interrogators were recruited, each randomly assigned to assess the credibility of a one of 36 eyewitness testimony. This assignment was done with replacement, meaning that

each of the 36 video testimonies was shown to multiple interrogators. On average, each testimony was viewed by 5 to 6 interrogators. This sample provided 80% power to detect an effect size of $d = 0.63$ at $\alpha = 0.05$ (two-tailed test).

For **Study II**, a sensitivity analysis was conducted using the *simr* package in R (Green et al., 2016), that allow for power calculations based on simulated data. Recruitment continued until at least 200 raters with usable data were obtained, providing 95.7% power to detect an effect size of $d = 0.30$.

No sample size justification was computed for **Study III**, as the complete dataset was used to code the presence (vs. absence) of non-verbal cues.

Materials, measures, and coding procedures

Crime stimulus

The stimulus material consisted of a mock-crime film, in line with standard procedures in the eyewitness research paradigm, to approximate a real-world violent event while adhering to ethical research standards. The 36-second film, filmed from a witness perspective, depicted a confrontation at a parking lot (Gustafsson et al., 2021). The sequence began with a man (the victim) engaging in conversation with another man (the first perpetrator) next to a stationary car. Moments later, a second perpetrator exited the car and suddenly attacked the victim with a knife while the first perpetrator restrained him from behind. The scene ended with both perpetrators getting into the vehicle, which had an additional occupant (the third perpetrator) in the driver's seat throughout. The knife was briefly visible during the stabbing, which lasted around three seconds.

The film was shot using a handheld camera with amateur actors. While the level of violence was minimal compared to real-world crimes and what is commonly depicted in films, television, and news media, the design of the footage provided a controlled yet ecologically valid setting for studying eyewitness testimony. This controlled setting allows for precise verification of reported details and allowing comparability across participants and studies. To minimize mock witnesses' potential emotional distress, participants were informed about the study's nature in advance.

Investigative interview procedure

Following the mock-crime film, eyewitnesses took part in a filmed structured three-phase investigative interview with an interviewer to assess testimony accuracy and susceptibility to suggestion (presented below). The first two phases followed the classic Cognitive Interview (CI) framework, as outlined by Fisher and Geiselman (1992). First, in the Free recall phase, eyewitnesses were asked to provide a comprehensive and detailed account of the

event as if they had witnessed it in real life. This was followed by the Cued recall phase, where witnesses answered seven open-ended questions designed to elicit additional details about the crime scene and individuals involved (e.g., "*You mentioned a person who was assaulted. Can you describe what that person looked like*?"). As eyewitnesses responded, the interviewer documented as many details reported by the witness as possible, such as descriptions of clothing, vehicles, and physical appearances, using a numbered sheet in order to test the confidence-accuracy relation (presented below). Finally, in the third phase, to assess susceptibility to suggestions, eyewitnesses completed the Gudjonsson Suggestibility Scales (GSS; Gudjonsson, 1997).

All three interview phases, Free recall, Cued recall, and the GSS, were conducted in Zoom and video-recorded to enable analysis and to be used in the forthcoming studies.

Suggestibility

Eyewitnesses susceptibility to suggestions was assessed using a modified version of the GSS (Gudjonsson, 1997), adapted to the mock-crime film. Here, eyewitnesses responded to 20 questions, which included both non-leading (n = 5; e.g., " *Did the victim speak with anyone before the assault?*") and leading questions (n = 15; e.g., "*Was there blood on the ground?*"). A subset (n = 5) of the leading questions contained false-alternatives (e.g., "*Was the victim stabbed two or three times?*", where the correct response was once). After completing the first round of questioning (Yield 1), regardless of actual performance, eyewitnesses were given mock negative feedback stating that some of their answers were incorrect. This follows the GSS procedure, where suggestibility is operationalized as the extent to which participants accept and act upon misleading feedback, even when their initial answers were in fact correct. This feedback is designed to induce interpersonal pressure and uncertainty, two key components underlying interrogative suggestibility (Gudjonsson, 1997). Participants were then asked to respond to the same set of questions a second time (Yield 2). The Shift score was calculated based on the number of answers changed between the first and second round, regardless of direction (i.e., from correct to incorrect or vice versa), and reflects the extent to which participants modify their responses after being told that some of their answers were incorrect. The sum of Yield 1 and Shift scores constitute the Total Suggestibility score, providing a measure of each eyewitnesses' susceptibility to misleading information and social pressure in the context of questioning.

Video-recorded eyewitness testimonies

From the full set of 121 video-recorded interviews originally collected as part of **Study Ia**, a selected subsample of 36 testimonies was used as stimulus

material in **Studies Ib** and **III**. The selection was based on a screening process designed to control for extraneous variability and ensure stimulus quality. Specifically testimonies with (1) poor audio or video quality, (2) signs of poor internet connection affecting transmission or recording, (3) contaminated data (e.g., eyewitnesses referring to the film's technical elements, such as "*the camera panned away"),* and (4) environmental disturbances (e.g., excessive peripheral movement or distracting background noise) were exlcuded.

From the remaining eligible recordings, the selected testimonies were matched across language and gender, resulting in four equally sized groups: nine native-speaking women, nine native-speaking men, nine non-native-speaking women, and nine non-native-speaking men. Matching criteria included (a) the length of the testimony, (b) levels of background noise or visual disturbances, and (c) overall performance in terms of memory accuracy (correct vs. incorrect) and self-rated cognitive effort. Furthermore, the interviewer's voice was replaced with a standardized audio introduction for the investigative interview using the DaVinci Resolve editing software, ensuring consistency across all recordings.

In **Study Ib**, only the free recall phase was used for interrogators to assess witness credibility and to formulate interrogation questions, to avoid bias from interviewer interactions during the cued recall. In **Study III**, both the free and cued recall phases were used for coding the presence (vs. absence) of non-verbal cues, in order to maximize the number of statements used for analysis.

Credibility ratings

Following Lindholm's (2008) framework, perceived credibility was assessed across multiple dimensions, including the witness's trustworthiness, ability to convey their memory, and the testimony's usefulness for a criminal investigation. Specifically, eight items were rated on a 7-point Likert scale (1 = not at all, 7 = very much; for item g, 1 = very poor, 7 = very good), asking to what extent: (a) the witness's memory was accurate, (b) the witness was able to convey their memory, (c) the witness was confident in their statement, (d) the witness was motivated to remember the event, (e) the witness was truthful, (f) the testimony was useful for the investigation, (g) how the witness experienced the situation of being questioned about the event, and (h) how nervous the witness appeared during the interview.

In **Study Ia**, eyewitnesses rated their own credibility using the eight items. One item ("*How nervous were you during the interview?*") showed low factor loading and was excluded from the credibility index (Cronbach's $\alpha = .78$).

In **Study Ib**, interrogators rated all eight items after watching only the free recall phase of a video-recorded interview (Cronbach's $\alpha = .83$).

Separate analyses of individual items across **Studies I–III** yielded similar results and are available in the respective OSF repositories linked with each

study. To maintain parsimony, only the composite credibility index is reported in this thesis.

Eyewitness statements, memory accuracy and the confidence-accuracy relation

Eyewitness testimonies were first transcribed verbatim and segmented into individual statements describing specific details observed during the mock-crime film. Only statements that could be clearly verified against the film were included in the final dataset.

Memory accuracy was scored using a predefined coding scheme based on the mock-crime film (Gustafsson et al., 2021), with accuracy defined as the absolute count of correctly and incorrectly reported details, respectively. Each verifiable detail was scored as either correct (1 correct point; matched the film) or incorrect (1 incorrect point; deviated from the film). For instance, the statement "*the one with the beige jacket grabbed the victim*" was scored as 1 correct point, while "*he had a black jacket*" was scored as 1 incorrect point if the correct color was brown. More complex statements like "*the victim's jacket was not black, not blue nor green*" (correct answer = red) were scored by adding each correct and incorrect point, in this case this sentence equalled three correct points. Details repeated within or across free and cued recall were only scored once.

Self-reported confidence was assessed after the investigative interview, based on the details the interviewer had elicited during the cued recall phase. As eyewitnesses reported their testimony, the interviewer recorded as many details as possible in real time on a numbered sheet (e.g., "*red jacket*", "*silver car*", and "*victim shorter than perpetrator*"). Because this was done during the live interview, not all recalled details were captured. Following the interview, eyewitnesses were presented with each of the recorded details read aloud one at a time by the interviewer (e.g., "*You mentioned that the victim had a red jacket, how confident are you?*"). For each detail, eyewitnesses provided a confidence rating on a scale from 0% to 100%, in 20% increments (i.e., 0%, 20%, 40%, 60%, 80%, or 100%). Each detail was then scored as 1 (correct) or 0 (incorrect), based on the predefined coding scheme described above. These binary scores were used to model the confidence–accuracy relation.

Interrogation questions

The interrogation questions were collected in **Study Ib**, where 202 participants were assigned the role of interrogators. After watching one of the video-recorded free recall testimony, each participant formulated 3–5 questions in Swedish with the (bogus) aim at gathering additional information to help solve the crime. To ensure quality and analytical consistency, the interrogation questions were screened prior to analysis based on a set of predefined inclusion

criteria. Specifically, questions that (1) deviated from the task instructions, (2) addressed the authenticity or technical aspects of the mock-crime video (e.g., camera angles), (3) were ambiguous or poorly formulated, or (4) were unrelated to the crime or investigation were excluded. Minor adjustments (e.g., punctuation) were allowed, but questions requiring major edits or rewording were excluded. Questions related to emotions, thoughts, and viewing conditions (e.g., lighting, distance, duration) were retained if clearly connected to the event or investigative purpose. Questions such as "*How do you feel?*" were excluded. Only interrogation questions aligned with investigative relevance and requiring no substantial revision were included in the final dataset. Out of 871 total questions, a total of 806 usable interrogation questions were retained (n_{native} = 397; $n_{non\text{-}native}$ = 409), meaning that approximately 7.5% of questions were excluded.

In **Study II**, each interrogation question was assessed on three dimensions: as leading, casting doubt, and as open-ended. The assessment of how leading each question is was based on categories used in Cederborg et al. (2013), that is, either option-posing (providing the witness with a choice between specific alternatives; for example: "*Was he wearing a black or a brown jacket?*") or suggestive prompts (implying a preferred answer; for example: "*He had a knife in his hand, right?*"). Next, in line with Cairns-Lee et al. (2022), I defined questions casting doubt as those where an interrogator explicitly or implicitly questions the veracity of something the witness has stated (for example: "*Are you sure that's what you saw?*" or "*That doesn't quite add up, could you be mistaken?*"). I defined open-ended questions as those allowing the witness to freely recount events from memory without constraining their answer to options or suggestions (for example: "*Can you describe what happened next?*"; Cairns-Lee et al., 2022). These examples were included in the rater instructions to illustrate interrogation questions that are leading, casting doubt and open-ended, respectively.

Each question was rated on a 7-point Likert scale, where 1 indicated that the characteristic was "not at all" present and 7 indicated it was present "to a very high degree".

Non-verbal cues

For the coding of non-verbal cues in **Study III**, a total of 680 objectively verifiable eyewitness statements were extracted from the 36 video-recorded mock eyewitness testimonies, selected and described in previous sections. Each statement represented a discrete unit of a verbally recalled detail related to the witnessed mock crime event. Of these, 479 statements were classified as correct and 201 statements as incorrect, based on comparison with the ground truth coding scheme outlined by Gustafsson et al. (2022). Non-verbal behaviour was coded at the statement level for the presence (1) or absence (0)

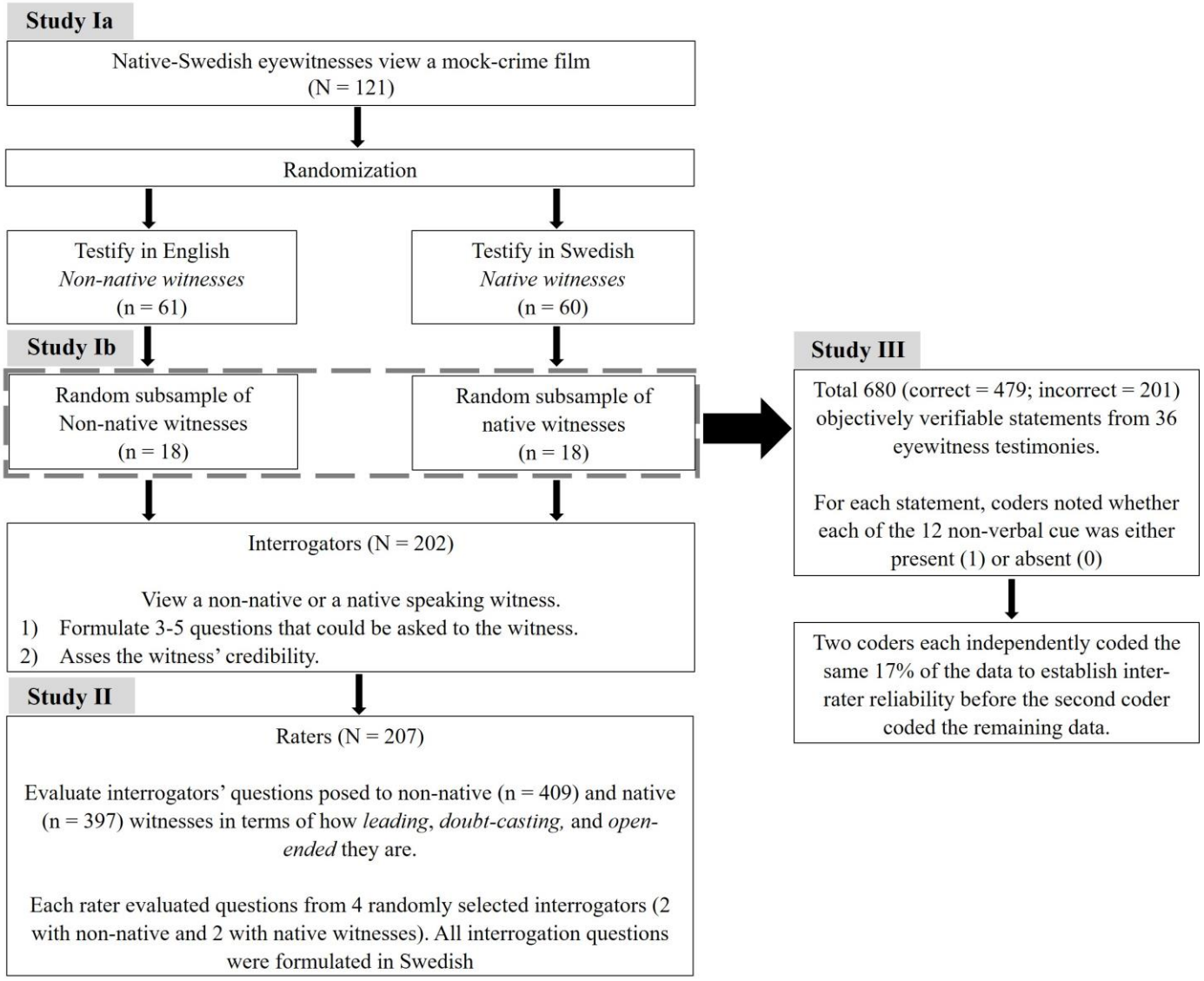
of twelve specific cues: *fidgeting*, *illustrators* (hand movements accompanying speech), *facial shielding*, *shrugs*, *head nods*, *head shakes*, *other body movement* (e.g., posture shifts, trunk movement), *smile*, *eye shifts*, *eyes closed*, *brow movement*, and *other facial expressiveness* (e.g., grimacing, eyelid tightening). The final set of cues was derived through a structured four-step selection process. Drawing on definitions from DePaulo et al. (2003) and considering technical constraints of the video data (e.g., camera angle, visibility, and quality), an initial list of 44 behaviours was reduced to 17 empirically and theoretically relevant cues. These were then piloted on a small sample of eyewitness testimonies (n = 5), and refined based on coder feedback. Then, similar behaviour categories (e.g., head movements and postural shifts were combined into *other body movement*, and high and low gaze into *eye shifts*) were merged, resulting in the final twelve cues. Coding was performed using both the video recordings and corresponding transcripts to ensure alignment between verbal content and non-verbal behaviour. Coders were blind to statement accuracy, self-reported confidence, and credibility ratings. Two trained research assistants independently coded the same 17% of the data, achieving percentage agreement ranging from 83.3% to 100% across cues. Discrepancies were resolved through discussion before one of the research assistants proceeded to code the remaining dataset.

Cognitive effort

Eyewitnesses' self-reported cognitive effort was measured in **Study Ia** using four items adapted from Englert et al. (2015) and Alm et al. (2019). Participants rated each item on a 7-point Likert scale (1 = not at all; 7 = very much). Specifically the items asked: (a) how difficult they found the task, (b) how effortful the task felt, (c) how mentally exhausted they felt at that moment, and (d) how much effort they had to exert to ensure their story appeared accurate. Item (d) was excluded due to low factor loading. A composite cognitive effort index was computed by averaging the three retained items (Cronbach's α = .61).

Experimental designs and procedures

The studies (**I-III**) in this thesis were interconnected, with the data collected from the non-native versus native eyewitnesses re-used across the studies. **Study Ia** investigated testimony reliability, while **Study Ib** assessed the perceived credibility of the testimonies. **Study II** investigated how these eyewitnesses were questioned, and **Study III** investigated the non-verbal cues exhibited by the two groups of eyewitnesses. A summary of the study flow is provided in Figure 1.


Study Ia
Native-Swedish eyewitnesses view a mock-crime film
(N = 121)
Randomization
Testify in English
Non-native witnesses
(n = 61)
Testify in Swedish
Native witnesses
(n = 60)
Study Ib
Random subsample of
Non-native witnesses
(n = 18)
Random subsample of
native witnesses
(n = 18)
Interrogators (N = 202)
View a non-native or a native speaking witness.
1) Formulate 3-5 questions that could be asked to the witness.
2) Asses the witness' credibility.
Study II
Raters (N = 207)
Evaluate interrogators' questions posed to non-native (n = 409) and native (n = 397) witnesses in terms of how leading, doubt-casting, and open-ended they are.
Each rater evaluated questions from 4 randomly selected interrogators (2 with non-native and 2 with native witnesses). All interrogation questions were formulated in Swedish
Study III
Total 680 (correct = 479; incorrect = 201) objectively verifiable statements from 36 eyewitness testimonies.
For each statement, coders noted whether each of the 12 non-verbal cue was either present (1) or absent (0)
Two coders each independently coded the same 17% of the data to establish inter-rater reliability before the second coder coded the remaining data.

In **Study 1a,** I investigated whether memory accuracy, suggestibility and the confidence-accuracy relation differed depending on whether witnesses testified in their non-native versus native language. As can be seen in Figure 1, in **Study Ia**, eyewitnesses (N = 121) watched the mock-crime film. After viewing the film, they were randomly assigned to provide their testimony in either a non-native (English, n = 61) or their native (Swedish, n = 60) language. These video-recorded eyewitness testimonies formed the dataset used in the subsequent studies. The testimonies consisted of (1) Free recall, (2) Cued recall, and (3) the GSS. During the Cued recall phase, I collected the self-reported confidence ratings of reported details to investigate the confidence-accuracy relation. Next, eyewitnesses provided ratings of their own credibility and cognitive effort during recall.

In **Study Ib**, I investigated the perceived credibility of non-native versus native speaking eyewitnesses. In this study, a new sample of participants (N = 202) was recruited and randomly assigned as interrogators to view one of 36 pre-selected eyewitness testimonies (free recall only) from **Study Ia**. First, these interrogators formulated questions to be asked to the eyewitness, with the instruction to gather additional information to help solve the crime. After completing this task, they assessed the credibility of the eyewitness who provided the testimony. **Study Ib** examined whether non-native and native eyewitnesses were judged differently in terms of credibility. The interrogation questions from **Study Ib** were used in **Study II**. The credibility ratings obtained from **Study Ib** were later incorporated into **Study II** to assess their influence on the interrogation questions framing, and to **Study III** to explore their relation with non-verbal cues.

In **Study II**, the aim was to examine how non-native versus native speaking eyewitnesses were questioned. The interrogation questions analysed in this study (N = 409 for non-native and N = 397 for native eyewitnesses) were collected during the **Study Ib**. To assess the nature of these questions, a new sample of 207 independent raters each evaluated a random subset of questions (n = 10–16) in terms of how leading, doubt-casting, or open-ended they were. Furthermore, **Study II** examined whether the ascribed credibility (from **Study Ib**) was associated with the nature of the questions posed to the non-native versus native eyewitnesses.

In **Study III**, I investigated the non-verbal cues exhibited by the eyewitnesses. The study used same 36 video-recorded eyewitness testimonies as in **Study Ib** and **Study II** (from **Study Ia)**, including both the Free recall and Cued recall phase. Then, for each statement provided, two trained coders assessed whether each of the twelve predefined non-verbal cues was present or absent. The coding process was conducted independently, with inter-rater reliability established before coding the full dataset. Thus, this study explored whether specific non-verbal cues were associated with testimony accuracy and self-reported confidence (**Study Ia**) and credibility assessements (**Study Ib**),

and whether this differed between non-native and native speaking eyewitnesses.

Ethical considerations

This project was approved by the Swedish Ethical Review Authority (File number: 2020-00624).

Study I involved the collection of video-recorded eyewitness testimonies. Participants gave written informed consent to (a) be interviewed in either their native or non-native language about a mock-crime video, (b) be video- and audio-recorded, and (c) allow their recordings to be used in subsequent research studies within the project. They were informed about the study's aims, the fictional nature of the mock-crime film, and their right to withdraw at any time. The mock-crime film depicted a brief, non-explicit instance of physical violence, consistent with standard practices in eyewitness research. Participants were informed of the content in advance to minimize potential distress and were advised not to participate if they were sensitive to such material. All recordings and data were stored securely on Stockholm University's cloud service (with servers located at SU) and handled in compliance with data protection regulations. The data did not include sensitive personal information.

Studies II and **III** re-used the approved video materials from **Study I**, in line with the ethical approval. These subsequent studies involved either assessing or coding the previously recorded testimonies.

All procedures adhered to GDPR and national data protection legislation.

Statistical analyses

Across all studies, linear, generalized, or ordinal mixed-effects models were employed to account for the nested structure of the data. These models allowed us to handle repeated measures within participants (e.g., multiple statements per eyewitness, multiple ratings per interrogator or rater), while quantifying group-level differences and individual variability. All analyses were conducted in R (R Core Team, 2022; RStudio Team, 202) using packages *lme4* (Bates et al., 2015), *ordinal* (Christensen, 2015), and *lmerTest* (Kuznetsova et al., 2017), with data wrangling and visualization via *tidyverse* (Wickham et al., 2019).

Random intercepts were included to account for baseline variation across witnesses, interrogators, and raters. For example, I assumed each eyewitness has a different baseline in expressing any given non-verbal cue. This means that the approach allowed for estimations not only whether non-native versus native speakers differed overall, but also whether these effects persisted when accounting for individual variability.

In **Study I**, linear mixed-effects models assessed differences in confidence-accuracy, suggestibility, and perceived credibility between non-native and native speakers. For testimony accuracy and self-rated cognitive effort, I conducted *t*-tests with Holm correction to compare the experimental conditions. In this summary of the thesis, additional analyses were conducted to assess the confidence-accuracy relation by both Confidence-accuracy Calibration (CAC) and Receiver Operating Characteristic (ROC) analyses. CAC curves illustrate the average accuracy at different confidence levels and are used to assess calibration in terms of how well subjective confidence aligns with actual memory accuracy. ROC curves evaluate the diagnostic value, that is, the ability of confidence to discriminate between correct and incorrect responses across all confidence thresholds. The Area Under the Curve (AUC) summarizes this diagnostic value such that values near 1.0 reflect strong discrimination, 0.5 indicates chance-level, and values below 0.5 suggest a negative relation between confidence and accuracy.

In **Study II**, I used ordinal mixed-effects models to evaluate how interrogation questions were rated along dimensions of being leading, casting doubt, or being open-ended, with fixed effects for language, perceived credibility, and their interaction. Also, to further investigate the language and credibility (high vs. low; categorized as ±1 standard deviation from the mean) interaction, I conducted tests of simple effects utilizing the error terms from the mixed-effects model using the *emmeans* package (Lenth et al., 2018).

In **Study III**, generalized linear models tested whether the presence (vs. absence) of each specific non-verbal cues predicted accuracy, confidence, or perceived credibility, and whether these associations varied by language. As this involved multiple comparisons, a Bonferroni-corrected alpha threshold ($p < .004$) was applied to minimize Type I error.

I based the model comparisons on Akaike weights (Burnham & Anderson, 2002; Wagenmakers & Farrell, 2004), which quantify the relative fit of competing models while penalizing complexity.

A note on interpretation 1. In line with the null hypothesis significance testing framework, I report conventional statistical outputs (e.g., effect sizes, confidence intervals, and *p*-values) but caution against interpreting these metrics in isolation (Amrhein et al., 2019; Gigerenzer, 2004). While useful for statistical inference, such artefacts do not reflect practical importance (Anvari & Lakens, 2021; Lakens, 2022). In legal psychology, this is particularly problematic as interpretation of findings must consider theoretical coherence, methodological rigor, contextual relevance, and the real-world implications of findings, while also weighing cost–benefit trade-offs. Even statistically significant and precisely estimated effects can be trivial in practice (Funder & Özer, 2019; Cohen, 1988; Panzarella et al., 2021). Accordingly, I have aimed to interpret my findings not through dichotomized thresholds, but through careful consideration on what the effects may imply in practice.

Study summaries

Study I: Memory accuracy, suggestibility and credibility in investigative interviews with native and non-native eyewitnesses

This study's aim was two-fold: **Study Ia** investigated whether testifying in a non-native language affects eyewitnesses' memory accuracy, suggestibility, and self-reported confidence, compared to testifying in one's native language. I expected that non-native speakers would (H1a) report fewer correct details and (H1b) report more incorrect details across free and cued recall, (H2) be more susceptible to suggestion, (H3) report lower confidence in their memories, (H4) report higher perceived cognitive effort, and (H5) rate themselves as less credible than native speakers. **Study Ib** investigated whether non-native speaking eyewitnesses were judged as less credible than native speaking ones by interrogators. Here, I expected that (H6) non-native speakers would be rated as less credible.

Study Ia.

Methods

A total of 121 eyewitnesses viewed a mock-crime video and were interviewed in Zoom either in a non-native (English; $n = 61$) or their native language (Swedish; $n = 60$). The non-native group were informed after watching the mock-crime video that the interview would be conducted in English. The interview included three phases: free recall, cued recall, and the modified GSS. During the cued recall phase, witnesses reported details about the crime and later rated their confidence in these details (0–100%). After the interview, participants also rated their own perceived cognitive effort during the recall, followed by their self-rated credibility as an eyewitness.

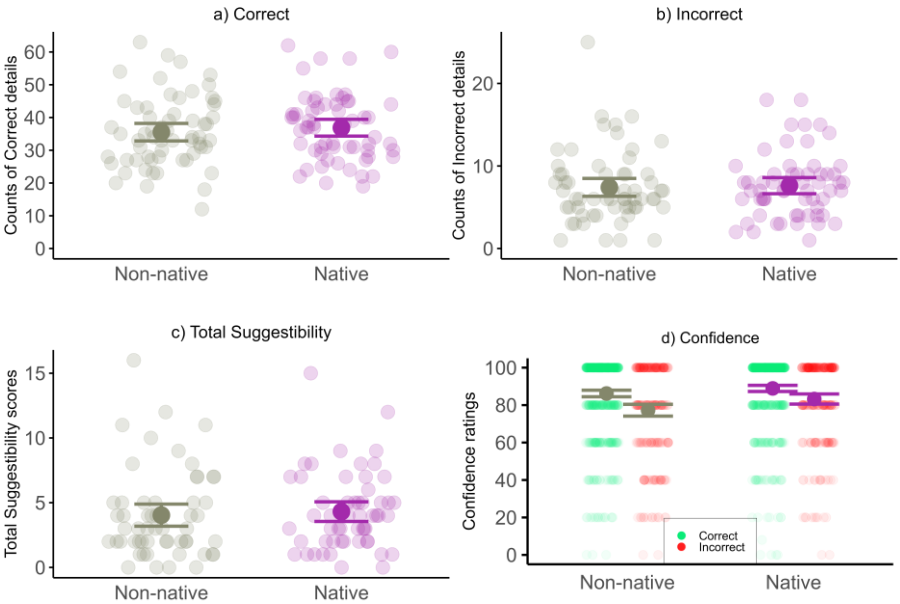
Results and discussion

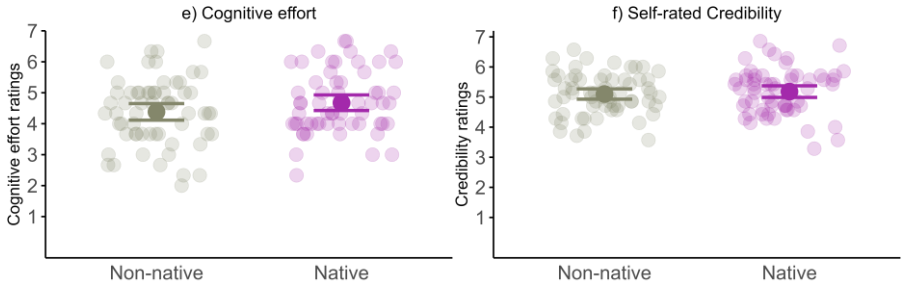
See Figures 2a-f for Means, 95% confidence intervals and distribution of data points across all reported outcome variables.

Non-native and native-speaking eyewitnesses did not differ in memory accuracy. Non-native speakers reported on average $M = 35.52$ ($SD = 10.66$) correct and $M = 7.41$ ($SD = 4.31$) incorrect details, while native speakers reported $M = 36.88$ ($SD = 10.12$) correct and $M = 7.62$ ($SD = 3.86$) incorrect details. No statistically significant differences were found for correct details ($d = 0.13$, $p = .474$) or incorrect details ($d = 0.05$, $p = .782$).

Similarly, no statistically significant differences were found between non-native versus native speaking witnesses for suggestibility. For Yield 1, non-native speakers scored on average $M = 2.35$ ($SD = 1.79$) and native speakers $M = 2.27$ ($SD = 1.52$), $OR = 0.98$, 95% CI [0.73, 1.32]); for Yield 2, non-native speakers scored $M = 2.60$ ($SD = 2.31$) and native speakers $M = 3.12$ ($SD = 2.59$), $OR = 0.75$, 95% CI [0.52, 1.09]); and for Shift, non-native speakers scored $M = 1.70$ ($SD = 2.02$) and native speakers $M = 2.04$ ($SD = 2.26$), $OR = 0.72$, 95% CI [0.45, 1.14]). For Total Suggestibility, no difference between language groups: non-native speakers scored $M = 4.04$ ($SD = 3.40$) and native speakers $M = 4.31$ ($SD = 2.99$), $b = -0.03$, 95% CI [-0.08, 0.03]).

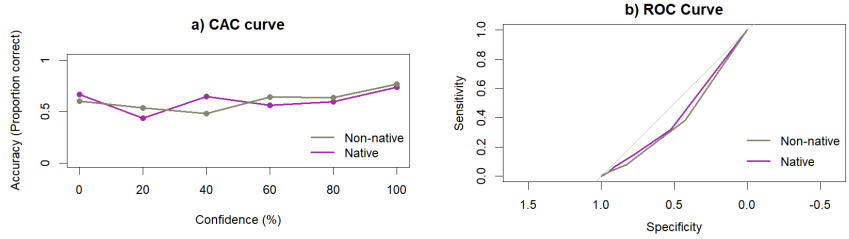
Figures 2a-f. Means, 95% Confidence Intervals and distribution of data points for Non-native and Native speakers.





Self-reported confidence was higher for correct statements (M = 87.55, SD = 20.65) than for incorrect statements (M = 80.35, SD = 24.06; d = -0.33), a difference that reached statistical significance (b = 6.63, 95% CI [3.59, 9.66], p < .001). However, regardless of testimony accuracy, non-native speakers reported significantly lower confidence than native speakers (b = -4.76, 95% CI [-9.39, -0.14], p = .045), with native-speaking witnesses reporting M = 87.10 (SD = 20.86) and non-native-speaking witnesses M = 83.49 (SD = 23.01; d = 0.16). While both non-native and native speakers reported higher confidence in correct than incorrect details overall, the CAC curves (Figure 3a) suggest modest calibration, with only small differences between the language groups across confidence levels. The receiver operating characteristic (ROC) analysis further showed that confidence was a poor discriminator of accuracy on a trial-by-trial basis in both language groups ($AUC_{non\text{-}native}$= 0.39; AUC_{native} = 0.42), with no statistically significant difference between them, t(1640.2) = 0.93, p = .35 (see Figures 3a-b).

Figures 3a-b. CAC and ROC curves, respectively, demonstrating the predictive value of confidence for memory accuracy across language conditions.



Note. [a] Confidence-accuracy Characteristic (CAC) curves showing the proportion of correct responses at each confidence level, separated by language condition. [b] Receiver Operating Characteristic (ROC) curves showing the ability of confidence ratings to discriminate between correct and incorrect responses. Sensitivity (true positive rate) represents the proportion of correctly identified positives. Specificity (true negative rate) represents the proportion of correctly identified negatives. The ROC curve illustrates the trade-off between sensitivity and (1 – specificity) across different decision thresholds.

There were no statistically significant differences between language groups in self-reported cognitive effort ($d = 0.29$, $p = .118$) or self-rated credibility ($d = 0.11$, $p = .542$).

To explore whether self-rated comfort speaking in the non-native language (1–7 Likert-scale) predicted accuracy, suggestibility, self-rated credibility, or cognitive effort within the non-native language group, separate regression models was conducted for each outcome. No significant main effects emerged for accuracy, suggestibility, credibility, or cognitive effort (all ps > .05). However, one significant effect was observed: lower comfort speaking non-native language predicted a weaker confidence–accuracy relation, $b = –3.39$, 95% CI [-6.30, -0.47], $p = .027$.

Unexpectedly, speaking a non-native language did not increase experienced cognitive effort, impair memory accuracy, or make witnesses more susceptible to suggestion. However, non-native language did lower witnesses' self-reported confidence in their testimonies, hence there was a mismatch between how accurate their memories were and how confident they felt. While confidence was moderately predictive of accuracy overall, additional ROC and CAC analyses revealed below-chance discrimination for both groups, indicating metacognitive miscalibration such that eyewitnesses were often overconfident in incorrect but underconfident in correct statements, regardless of language. Furthermore, the follow-up analyses showed that eyewitnesses who felt less comfortable speaking a non-native language were just as accurate as those who felt more comfortable, but significantly less confident. These findings indicate that the metacognitive miscalibration observed between non-native and native speakers was even more pronounced within the non-native language group.

Study Ib

Methods

Interrogators ($N = 202$) were each randomly assigned to view one of 36 pre-selected eyewitness testimonies from **Study Ia** (free recall only) and rated the witness' credibility using 8 items (1 = not at all; 7 = very much). Each interrogator also formulated 3–5 follow-up interrogation questions aimed at gathering additional information to help solve the crime (basis for **Study II**).

Results and discussion

See Figure 4 for Means, 95% confidence intervals and distribution of data points across all reported outcome variables.

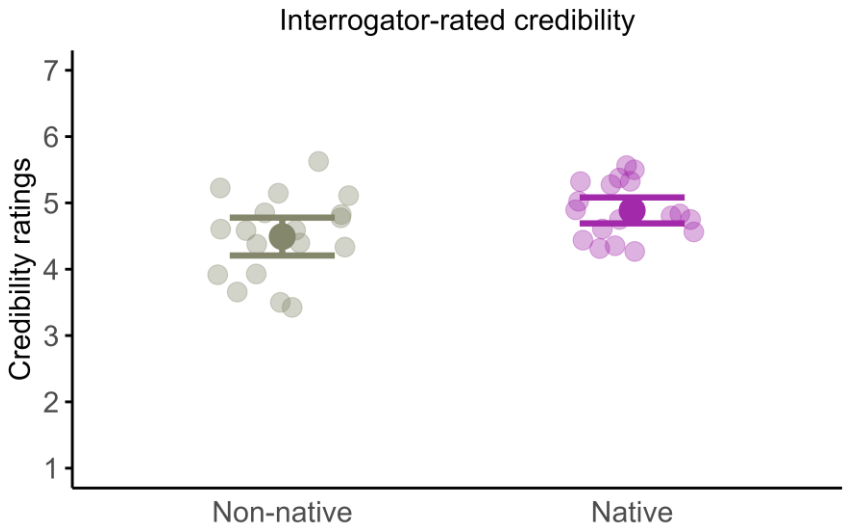
A statistically significant effect of language on interrogator-rated credibility was found. Witnesses who testified in a non-native language ($M = 4.49$, $SD = 1.04$) were judged as less credible than those who testified in their native language ($M = 4.89$, $SD = 0.80$), ($b = -0.36$, 95% [-0.71, -0.01] $p = .047$). The

observed effect size (Cohen's $d = 0.37$) corresponds to a Probability of Superiority 60.3%, meaning there is a 60.3% chance that a randomly selected individual from the native condition was rated as more credible than a randomly selected individual from the non-native condition (see Grissom & Kim, 2005; Lakens, 2013). Moreover, the random-intercept model yielded an intraclass correlation of 0.18, indicating that 18% of the variance in credibility ratings was attributable to differences between individual eyewitnesses, regardless of language condition.

To examine whether eyewitnesses' subjective experience predicted observer-rated credibility, two separate linear regression models were conducted with language condition, self-rated credibility and confidence, respectively, and their interaction as predictors of observer-rated credibility. Neither analysis yielded statistically significant effects (all ps > .05), suggesting no significant association between eyewitnesses' subjective experience and how credible they were judged by independent observers.

Taken together, these findings suggest that speaking in a non-native (vs. native) language affects credibility assessment negatively, despite comparable levels of testimony accuracy in **Study Ia**.

Figure 4. Means, 95% confidence intervals and distribution of data points for Non-native and Native speakers.



General conclusion

In **Study Ia**, I investigated whether testifying in a non-native (vs. native) language affected eyewitnesses' memory accuracy, suggestibility, self-reported confidence, self-reported cognitive effort, and self-rated percieved

credibility. I found no support for H1a or H1b, as non-native and native speakers did not differ in the number of correct or incorrect details recalled. H2 was also not supported, with no significant differences across any suggestibility components (Yield 1, Yield 2, Shift, or Total Suggestibility). However, H3 was supported since non-native speakers reported significantly lower confidence in their memories than native speakers. Finally, I found no support for H4 or H5, as there were no statsistically significant differences between the language groups in self-reported cognitive effort or self-rated credibility.

In **Study Ib**, I tested whether non-native (vs. native) speaking eyewitnesses were perceived as less credible by independent observers. Here, H6 was supported since non-native speakers were rated as less credible than native speakers, despite the content of their testimonies being comparable in terms of accuracy.

Overall, **Study I** demonstrated that non-native and native speakers provided equally reliable eyewitness testimonies. However, their spoken language systematically influenced both self-reported confidence in their testimonies and interrogator-rated credibility which are the two cues often used as proxies for reliability in legal settings. These findings provide empirical evidence of a systematic bias in that non-native speakers were perceived as less credible and reported lower confidence, despite similar levels of accuracy. This underscores how language-related biases can distort credibility judgments and compromise the fairness of investigative processes.

Study II: Interrogation questions to native and non-native eyewitnesses: The role of witness credibility

This study investigated whether testifying in a non-native versus native language affected the degree to which interrogation questions were leading, casting doubt, and open-ended. It also examined whether perceived credibility, as judged by the interrogators, was associated with how these questions were framed. I expected that non-native speaking eyewitnesses would receive (H1) more leading, (H2) more doubt-casting, and (H3) less open-ended questions than native speakers. I further expected that higher perceived credibility would be associated with (H4) lower degree of leading questions, (H5) lower degree of questions casting doubt, and (H6) higher degree of open-ended questions.

Methods

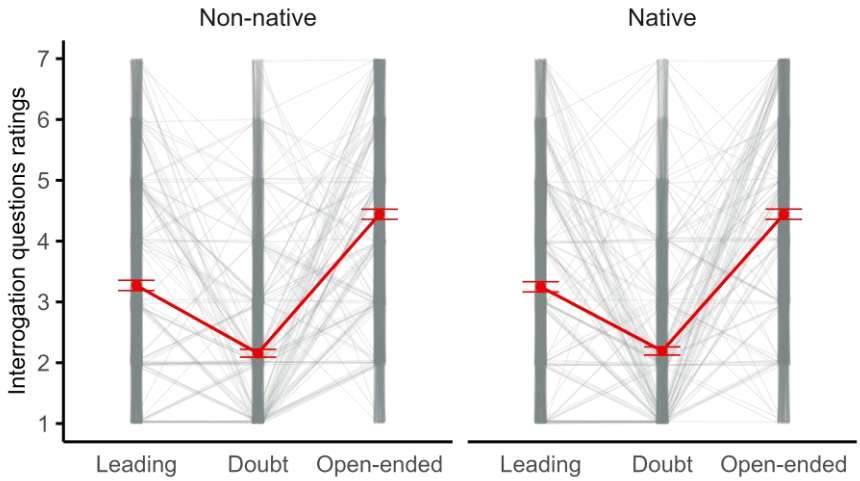
A sample of interrogators (N = 202; from **Study Ib**) were randomly assigned to view a video-recorded free recall testimony from **Study Ia**, given either by a non-native or native speaker. After viewing the testimony, each interrogator formulated 3–5 follow-up questions in Swedish intended to help

solving the crime, yielding a total of 806 interrogation questions. These questions were subsequently rated by a new group of 207 independent raters on three dimensions: (1) how leading they were, (2) the degree to which they casted doubt about something the witness had said, and (3) how open-ended they were. Before these ratings, raters completed a brief familiarization round with six interrogation questions, receiving immediate feedback (correct vs. incorrect) on their identification of each of the three types of questions. Then, ratings during the main test were made on 7-point Likert scales (1 = not at all, 7 = to a very high degree). Ratings were aggregated and used to examine whether question type varied depending on the language (native vs. non-native) of the eyewitness and their perceived credibility, as judged in **Study Ib**.

Results and discussion

I opted to use ordinal mixed-effects models to test the degree to which questions were framed as leading, casting doubt, or open-ended varied by language and perceived credibility. Contrary to expectations, language did not affect question type: as Figure 5 shows, non-native and native witnesses were equally likely to receive all three question types. For leading questions, non-native speakers received on average $M = 3.27$ ($SD = 2.13$) and native speakers $M = 3.25$ ($SD = 2.13$), ($OR = 1.09$, 95% CI [0.86, 1.40], $p = .45$). For questions casting doubt, non-native speakers received $M = 2.16$ ($SD = 1.63$) and native speakers $M = 2.19$ ($SD = 1.70$), ($OR = 0.92$, 95% CI [0.69, 1.22], $p = .56$). For open-ended questions, non-native speakers received $M = 4.44$ ($SD = 2.09$) and native speakers $M = 4.44$ ($SD = 2.12$), ($OR = 0.97$, 95% CI [0.76, 1.23], $p = .79$). While the same interrogation questions ($N_{total} = 806$; $n_{\text{Non-native}} = 409$; $n_{\text{Native}} = 397$) often overlapped across dimensions, simple regressions confirmed expected relations: leading questions were less likely to be perceived as open-ended ($r = -.61$) and more likely to cast doubt ($r = .33$); questions casting doubt were less likely to be perceived as open-ended ($r = -.31$), all ps $< .001$.

Figure 5. Mean ratings of question type by language condition.



Note. The red lines indicate group means (CI 95%) for each question type (Leading, Doubt-casting, Open-ended), by language condition (Non-native vs. Native). Grey lines represent individual interrogation question ratings, showing within-language-condition variability across question types.

Lower (vs. higher) ascribed credibility was associated with significantly more questions casting doubt (*OR* = 0.66, 95% CI [0.54, 0.80], $p < .001$; for native speakers: *M* = 2.07, *SD* = 1.68 at high vs. *M* = 3.02, *SD* = 2.13 at low credibility; for non-native speakers: *M* = 2.43, *SD* = 1.78 at high vs. *M* = 1.97, *SD* = 1.57 at low credibility). However, the interaction with language indicated that the effect was limited to native speakers (*OR* = 1.73, 95% CI [1.32, 2.28], $p < .001$). No effect of credibility emerged for leading questions (*OR* = 0.89, 95% CI [0.75, 1.06], $p = .19$; for native speakers: *M* = 3.31, *SD* = 2.24 at high vs. *M* = 3.61, *SD* = 2.06 at low credibility; for non-native speakers: *M* = 3.31, *SD* = 2.14 at high vs. *M* = 3.15, *SD* = 2.11 at low credibility). Nor was there a credibility effect for open-ended questions (*OR* = 1.19, 95% CI [0.99, 1.44], $p = .055$; for native speakers: *M* = 4.56, *SD* = 2.18 at high vs. *M* = 3.92, *SD* = 2.04 at low credibility; for non-native speakers: *M* = 4.39, *SD* = 2.01 at high vs. *M* = 4.43, *SD* = 2.21 at low credibility). Although interactions with language were observed for both leading and open-ended questions, follow-up tests of simple effects showed no statistical significant differences within or between language groups. This may reflect limited statistical power to detect effects within each language group.

Overall, inter-rater agreement was poor ($M_{ICC} = 0.39$), prompting follow-up analyses to ensure robustness of the findings. Specifically, I repeated all main analyses using a filtered dataset including only raters who had performed

above 60% accuracy in the familiarization phase and using only the interrogation questions that reached acceptable inter-rater reliability (ICC $\geq$ 0.70). These follow-up analyses yielded results consistent with the main findings and can be found in the OSF repository linked to the study.

Conclusion

In **Study II**, I tested whether language (non-native vs. native) and perceived credibility influenced the interrogation questions posed across three dimensions: leading, casting doubt, and open-ended. I found no support for H1–H3, as there were no statistically significant differences between non-native and native speakers in the types of questions they received. That is, language alone did not affect the degree to which questions posed to witnesses were leading, casting doubt, or open-ended. I also found no support for H4 or H6, as perceived credibility did not affect the degree to which questions were leading or open-ended. However, H5 was supported, as lower perceived credibility was associated with a higher degree of questions casting doubt, although this effect was only observed for native speaking eyewitnesses.

Although the effect was small, it may indicate that interrogators adjusted their questioning based on perceived credibility but only when the witness spoke their native language. Given that credibility judgments often reflect extra-legal factors, and that questions casting doubt are discouraged in investigative interviews, this pattern may reflect unwarranted treatment of those native speakers who were perceived as less credible, despite no indication of lower memory accuracy.

Study III: Non-verbal cues in eyewitness testimonies do not predict accuracy or credibility assessments

This study investigated whether non-verbal cues are associated with (Research Question 1) the accuracy of eyewitness statements, (RQ2) witnesses' self-reported confidence, and (RQ3) interrogators' ratings of witness credibility. I also examined (RQ4) whether these associations differed between non-native and native speaking eyewitnesses. Here, I focused on 12 non-verbal cues based on the meta-analytic review by DePaulo et al (2003). Table 1 presents the list of the cues, a definition of each one, and the inter-rater percentage agreement.

Table 1. Non-verbal cues, their definition and inter-rater percentage agreement.

Cue	Definition	% Agreement
Fidgeting	Directly observable object fidgeting, self-fidgeting, or facial fidgeting	99.17
Illustrators	Hand movements that accompany and illustrate speech	100
Facial Shielding	The speaker shielding their face	100
Shrugs	Up and down movement of the shoulders, or the palms of the hand are open and the hands are moving up and down	100
Head nods	Affirmative head nods, that is vertical head movements	99.17
Head shakes	Negating head shakes, that is side-to-side head movements	100
Other body movement	Movements of the head (including all movements of head except head shakes and head nods), arms, postural adjustments, leanings, trunk movements, or repositioning of the body	86.78
Smile	The speaker is smiling (even silently)	99.17
Eye shifts	Eye movements/shifts in gaze. For example, low gaze (looking downwards), high gaze (looking upwards), or side gaze (looking to the left or the right)	91.74
Eyes closed	Eyes are closed	98.35
Brow movement	Eyebrows are raised or lowered	95.87
Other facial expressiveness	The speaker's face is animated, expressive, or deviating from a neutral expression (e.g., grimacing, eyelids tightening)	98.35

Note. Gwet's AC1 ranged from 0.59 to 1.00, with a mean of 0.89 (*SD* = 0.11). Only *Other Body Movement* was below Gwet's AC1 = 0.80, with an AC1 of 0.59.

Methods

The dataset consisted of 680 objectively verifiable statements (479 correct, 201 incorrect) drawn from the 36 video-recorded eyewitness testimonies collected in **Study Ib**. Each statement originated from the free or cued recall phases and had been transcribed and verified for accuracy against the coding scheme based on the mock-crime video (**Study Ia**). Confidence was self-reported by witnesses on cued recall statements only (0–100%; **Study Ia**) and

credibility was rated by the interrogators on free recall only (from 1 = not at all to 7 = very much; **Study Ib**).

Each statement was coded for the presence (1) or absence (0) of each of the twelve cues, using both transcript and video. Two trained research assistants performed the coding. The first coder coded 17% of the data, and then a second coder independently coded the same 17% for inter-rater reliability. Discrepancies were resolved through discussion, after which the second coder completed the coding of the remaining dataset. Coders were blind to statement accuracy, confidence, and credibility ratings.

Results and discussion

I used generalized mixed-effects models to test whether the presence of non-verbal cues was predicted by statement accuracy, self-reported confidence, and interrogator-rated credibility, and whether any associations differed between non-native and native speakers.

Across all models, the presence (vs. absence) of non-verbal cues was not reliably predicted by statement accuracy, self-reported confidence, or interrogator-rated credibility. Specifically, no cue distinguished between correct and incorrect statements with ORs ranging from 0.32 to 4.02 ($M = 1.24$; $SD = 0.75$; $Mdn = 1.03$), and p-values ranging from .009 to .998 ($M = .585$; $SD = 0.3$; $Mdn = .674$). Similarly, no cue predicted perceived credibility with ORs ranging from 0.07 to 6.15 ($M = 1.26$; $SD = 1.22$; $Mdn = 0.96$), and p-values ranging from .010 to .920 ($M = .451$; $SD = 0.32$; $Mdn = .459$). However, for self-reported confidence, one exception emerged: *Other facial expressiveness* was statistically associated with lower confidence, though the effect was negligible ($OR = 0.97$, 95% CI [0.96, 0.98], $p < 0.001$). Otherwise, no cue was statistically associated with confidence with ORs ranging from 0.96 to 1.02 ($M = 0.99$; $SD = 0.02$; $Mdn = 0.99$), and the p-values ranging from <.001 to .731 ($M = 0.356$; $SD = 0.236$; $Mdn = 0.319$).

Although non-native speakers reported lower confidence (**Study Ia**) and were generally perceived as less credible (**Study Ib**), these differences were not explained by systematic variation in their exhibited non-verbal behavior. Specifically, no cue predicted perceived credibility, and aside from a negligible effect for *Other facial expressiveness*, non-verbal cues did not predict self-reported confidence. Notably, *Other facial expressiveness* was not associated with statement accuracy.

Conclusion

Contrary to popular belief and persistent misconceptions in real-world legal contexts, non-verbal cues do not reliably predict accuracy, self-reported confidence or credibility in honestly reported eyewitness testimonies – regardless of language. This challenges their evidentiary value and suggests caution in using such cues for high-stakes legal decision-making.

General discussion

The aim of this thesis was to investigate whether non-native and native speaking eyewitnesses differ in terms of testimony *reliability* – operationalized as memory accuracy and suggestibility – and perceived *credibility*, and to examine the *investigative interview practices* and *extra-legal factors* in this case non-verbal cues that may underlie such potential differences. While eyewitnesses are expected to report what they have seen as accurately and truthfully as possible, the way their accounts are received and evaluated may depend on more than just the content of their testimony. This is a critical question, as language barriers are increasingly common in legal contexts worldwide, and may systematically shape how eyewitness testimony is received and judged.

To address these questions, the thesis empirically examined:

i. whether non-native speakers differed from native speakers in memory accuracy, suggestibility, perceived cognitive effort, self-reported credibility, and the confidence-accuracy relation (**Study Ia**; testimony reliability);
ii. whether non-native speakers were perceived as less credible by independent observers (**Study Ib**; testimony credibility);
iii. whether non-native speakers received more suggestive interrogation questions, and if the perceived credibility influenced this practice (**Study II**; investigative interview practices); and
iv. whether non-verbal cues predicted accuracy, confidence, or credibility, and whether these associations differed depending on the spoken language (**Study III**; extra-legal factors).

By systematically examining both actual and perceived differences between non-native and native-speaking witnesses, this thesis aimed to clarify whether observed disparities in legal settings reflect genuine differences in testimony quality or are instead driven by biased credibility judgments. This has direct implications for equal treatment in legal proceedings. The following sections synthesize the findings across studies, relate them to broader psychological and legal frameworks, and reflect on the methodological, statistical, and practical implications of the results. See Table 2 for an overview of all

outcome variables showing the (non)significant differences between non-native and native eyewitnesses.

Synthesis of findings across studies

Based on the present findings, there was no evidence that testimony reliability – assessed through memory accuracy and susceptibility to suggestion – differed between non-native and native speakers. Both groups recalled the witnesssed mock-crime event with similar accuracy and were equally resistant to suggestive questions, aligning with previous findings showing that testifying in a non-native language does not necessarily impair testimony accuracy (Alm et al., 2019; Ernberg & Mac Giolla, 2023; Hu & Naka, 2022). This undermines any straightforward claim that non-native witnesses are less reliable due to language barriers. My initial assumption was that testifying in a non-native language would impose greater cognitive effort, potentially impairing testimony accuracy. However, non-native and native speakers reported similar levels of self-reported cognitive effort. This may help explain why the differences in suggestibility between language groups were absent, in contrast to previous research showing that elevated cognitive load increases suceptibility to suggestions (Otgaar et al., 2012).

Importantly however, non-native speakers were less confident in their testimonies and were perceived as less credible by interrogators. These findings, though not reflective of actual accuracy, have clear forensic relevance. Confidence is often used by fact-finders as a proxy for reliability (Mickes et al., 2025; Wixted & Wells, 2017; Sporer et al., 1995). Yet, when non-native speakers report lower confidence, it may reflect the subjective experience of effortful retrieval under increased language demands rather than impaired memory accuracy (Gustafsson et al., 2019; Lindholm et al., 2018; Koriat & Goldsmith, 1996). This pattern was more pronounced within the non-native group: those less comfortable speaking English were even less confident, yet equally accurate. However, I did not find any differences in self-reported cognitive effort between non-native and native speakers.

While confidence ratings were moderately aligned with accuracy at group level, the CAC curves revealed only modest calibration across confidence levels, and ROC analysis showed that confidence was a poor discriminator of accuracy on a trial-by-trial basis in both language groups. This suggests that, although confidence can reflect accuracy on average, its diagnostic value in evaluating individual statements remains limited – particularly in high-stakes forensic contexts.

Similarly, credibility assessements shape legal decisions, including how eyewitness testimonies are evaluated and whether they are pursued further (Boccaccini, 2002). The pattern observed here is therefore problematic: non-

native speakers were seen as less credible, despite being equally reliable in terms of content. This echoes broader findings that extra-legal factors such as language proficiency, speech-fluency, and group membership influence perceptions of witness credibility, even in the absence of actual differences in terms of reliability (Frumkin, 2020; Lindholm, 2005, 2008; Rakic et al., 2011). Moreover, the results revealed that 18% of the variance in interrogator-rated credibility was attributable to idiosyncratic differences between eyewitnesses. This suggests that credibility judgments are shaped not only by manipulated variables such as language, but also by other person-specific characteristics (e.g., demeanor, speaking style, accent, or emotional display), highlighting the subjective nature of credibility assessments in legal contexts.

Table 2. Overview of statistically significant differences between non-native and native speaking eyewitnesses across studies.

Investigated variable	**Expected difference**[1]	**Observed difference**
Study 1a – testimony reliability		
Memory accuracy	Yes	No
Confidence-accuracy	Yes	**Yes**
Suggestibility	Yes	No
Cognitive effort	Yes	No
Study 1b – testimony credibility		
Perceived credibility	Yes	**Yes**
Study II – interview practices[2]		
Leading	Yes	No
Doubt	Yes	No
Open-ended	Yes	No
Study III – extra-legal factors		
Non-verbal cues	-[3]	No

Note. [1] The column "Expected difference" refers to whether the a priori hypotheses expected differences between language groups. [2] Credibility effects emerged only for questions casting doubt. [3] I had no directional expectations on the effects of language on non-verbal cues.

Interestingly, despite the observed statistically significant differences in self-reported confidence and credibility, the type of interrogation questions posed to non-native and native speakers was similar overall. All eyewitnesses received a comparable proportion of questions that were leading, casting

doubt, and open-ended. However, a noteworthy pattern emerged when examining the link between perceived credibility and questions casting doubt. For native speakers, interrogators adjusted their doubt as expected: lower ascribed credibility was associated with more questions casting doubt, and vice versa. For non-native speakers however, this link was absent. Interrogators posed questions reflecting the same level of doubt regardless of the witness's perceived credibility. This raises concerns given that even subtle shifts in question phrasing or structure can contaminate memory reports, even when witnesses are honest and motivated to be accurate (Fisher & Geiselman, 1992; Gudjonsson, 2003; Hope & Gabbert, 2019; Loftus, 2005).

Witness' suggestibility is shaped by both cognitive factors (e.g., weak memory traces, source confusion) and social factors (e.g., deference to authority, desire to avoid conflict), making witnesses especially vulnerable when they experience low confidence or heightened anxiety (Gudjonsson & Clark, 1986). Previous research suggests that non-native speakers may be especially vulnerable to suggestibility due to increased cognitive demands when retrieving and articulating memories in a non-native language (Abutalebi, 2008; Alm et al., 2019; Otgaar et al., 2012). Although the present results did not show increased suggestibility or self-reported cognitive effort among non-native speakers, this remains a theoretically and practically plausible concern. Nonetheless, the absence of a credibility–doubt link for non-native speakers indicates that skepticism was applied uniformly, irrespective of perceived credibility. This aligns with related research showing that witnesses from ethnic outgroups are often judged as less credible and are questioned more suggestively (de Bruïne et al., 2023; Lindholm, 2005, 2008; Scassa, 1994; van Veldhuizen et al., 2017). Nonetheless, the overall levels of suggestive questions were low in the present thesis.

Lastly, non-verbal cues such as fidgeting, facial expressiveness, or gaze aversion, did not predict statement accuracy, self-repored confidence, or perceived credibility. This directly challenges the common belief, especially among legal professionals, that such behaviors are diagnostically meaningful (Chalmers et al., 2022; Denault et al., 2020, 2024). In the data, non-verbal behavior was not informative with regard to memory accuracy. This raises concerns about relying on such cues in real-world legal contexts, particularly when evaluating linguistically or culturally diverse witnesses whose expressiveness may differ from those of interrogators.

A note on interpretation 2. As discussed, statistical metrics should not be interpreted in isolation (Amrhein et al., 2019; Funder & Özer, 2019; Gigerenzer, 2004). In legal psychology, small effects can carry real-world weight, while large effects may prove negligible. For instance, in **Study Ib**, non-native speakers were rated as less credible, an effect that may have meaningful real-world consequences even if small – given the weight credibility assessments carries in legal decision-making. Therefore, the aim is not always to estimate

how frequently or how strongly an effect occurs, but to establish that it *can* occur at all. As such, even when the data may appear noisy or overlapping upon ocular inspection, the consequences can still be meaningful in high-stake real-world cases.

Cross-linguistic and cultural biases in legal systems

This thesis taps into broader concerns regarding systemic biases in legal contexts, and how linguistic and cultural factors may influence judicial processes. Legal systems, though ideally objective, often rely on subjective judgments of witness reliability and credibility, which are susceptible to distortion by cross-linguistic and cultural dynamics (Anakwah et al., 2020, 2024; de Bruïne et al., 2025; Frumkin, 2000, 2007, 2020; Hope et al., 2022, 2023; Lindholm, 2005, 2008a, 2008b; Rakic et al., 2011; Scassa, 1994). By examining honestly reported eyewitness testimonies, this thesis explicitly addressed how language barriers influence memory accuracy, the confidence-accuracy relation (Mickes et al., 2025; Wixted & Wells, 2017; Juslin et al., 1996), suggestibility (Loftus, 2005; Gudjonsson & Clark, 1986), credibility judgments (Boccaccini, 2002; Lindholm, 2008), and non-verbal behavior (Chalmers et al., 2022; Denault et al., 2020, 2024), all of which are central to legal decision-making. Again, this thesis did not assume that non-native speakers are inherently less capable of remembering accurately, but rather that the reporting, communication, and perception of their memories may be negatively affected by language barriers.

From the empirical findings in this thesis, one can extrapolate a troubling reality: non-native speakers' risk being penalized in credibility assessments, despite comparable testimony accuracy and resistance to suggestion as their native-speaking counterparts. This bias presents a critical challenge for fairness in increasingly diverse legal landscapes. Addressing these issues is important for ensuring that justice transcends language barriers and cultural divides, upholding principles of fairness and objectivity across judicial systems.

Methodological considerations

Samples, validity and generalizability

The main methodological concern in this thesis relates to the operationalization of non-native versus native language. The language manipulation was implemented through Swedish-speaking participants providing testimony in either their native language (Swedish) or in English. While this approach allowed for experimental control and feasibility within the scope of the current

project, it also presents inherent limitations. Participants in the non-native condition were generally proficient in English, reflecting Swedes' high average language proficiency. Although proficiency was measured and statistically accounted for, the language challenges faced by these participants likely underestimate those experienced by real-world non-native witnesses, such as newly arrived migrants or individuals with limited language proficiency undergoing high-stakes investigative interviews (Itzhak et al., 2017). Consequently, the observed effects may be conservative estimates of the disparities likely to arise in the real world. Nonetheless, I argue that foundational experimental work of the present kind is critical for building a robust empirical base from which more ecologically valid studies can evolve.

A potential limitation concerns the interrogator sample, specifically their English language comprehension and lack of expertise in interviewing crime witnesses. However, a follow-up analysis including their self-rated English language comprehension as a predictor showed no statistically significant effects, suggesting that neither credibility ratings nor question formulation were meaningfully influenced by the interrogators' own language comprehension. Furthermore, there is no clear evidence that trained professionals would formulate fundamentally different types of interrogation questions. In fact, despite wide-spread awareness of best-practice techniques, such as the Cognitive Interview, legal professionals still often use closed, multiple-choice, and other-wise suggestive questions that risk compromising reliability (Brewer et al., 2018; Cederborg et al., 2013; Fisher, 1995; Fisher & Geiselman, 1992; Geiselman & Fisher, 1989; Hyman Gregory et al., 2023; Meissner et al., 2021; Schreiber Compo et al., 2012).

A related concern is the asynchronous format: interrogators generated follow-up questions in writing, after viewing the recorded testimony, rather than during a live (synchronous) exchange. While this may reduce certain aspects of interactional spontaneity, the task remained justified given that participants were explicitly instructed to formulate questions as if preparing for a follow-up interview.

In terms of construct validity, one of the central outcome variables in this thesis, namely credibility, was operationalized using a composite rating scale encompassing eight dimensions: perceived memory accuracy, clarity of communication, confidence, motivation, truthfulness, investigative utility, emotional expression, and nervousness. While credibility is typically defined in legal contexts as the extent to which a person is seen as a trustworthy source of information, this broader operationalization reflects the complex and multifaceted nature of real-world credibility assessments. However, it also blends distinct evaluative targets: some items (e.g., nervousness, emotional expression, or perceived usefulness) may be only weakly related to testimonial reliability and more reflective of interpersonal impression formation. Thus, although the scale demonstrated satisfactory internal consistency, the construct it captures is conceptually heterogeneous and may conflate credibility with

extra-legal factors that influence judgments in legal contexts. However, I also conducted separate analyses for each individual item across all outcomes in the three studies, but these did not yield any different results.

A similar concern arises in the assessment of cognitive effort. The scale used in **Study Ia** was adapted from sport psychology research, where cognitive strain is tied to task performance in competitive or high-pressure environments. While it offered a structured approach to gauging self-reported cognitive effort, its generalizibility to investigative interview settings is unclear.

The measurement of suggestibility also raises questions about ecological validity. The GSS, while widely used in research, consists of 15 direct and in some cases blatantly leading questions. While this ensures comparability, it risks drawing participants' attention to the experimental manipulation itself. In real-world investigative interviews, suggestive questions can be more subtle, context-dependent, and interwoven with rapport and pressure. Hence, the GSS may not fully capture how susceptible individuals are in real-world investigative interviews. Moreover, although suggestibility and compliance are conceptually distinct, where compliance refers to reporting inaccurate information not because it is believed to be true but to meet perceived expectations or avoid conflict (Gudjonsson & Clark, 1986), the Gudjonsson Compliance Scale was not included in this thesis. Given that the thesis focuses on honestly reported testimony and does not assume that non-native speakers are less capable of remembering accurately, including a measure of compliance could have provided a valuable complement by clarifying the extent to which any inaccurate reporting reflects social pressure rather than memory distortion. Notably, however, Alm et al. (2019) included a compliance scale and found no significant association with memory accuracy, suggesting that such effects may be limited or context-dependent.

A broader epistemological limitation lies in the nomothetic approach underpinning this thesis, as well as in much quantitative psychological research. While group-level analyses help identify general patterns and inform policy, applying them uncritically to individuals risks ecological fallacy (i.e., group-based findings may not accurately reflect individual-level processes or outcomes; Fisher et al., 2018; Molenaar, 2004). The legal system adjudicates individual cases, not aggregates. Even robust group-level findings may fail individual-level inferences. A native speaking witness who is confident, accurate, and credible may still be disbelieved, and a non-native speaker may be unduly trusted. While idiographic approaches, such as single-N studies or qualitative designs, can help capture individual variability, this nomothetic-idiographic dilemma remains a core limitation when applying legal psychological research to the real-world.

Statistical considerations

Given the nested data structure (e.g., repeated statements per witness and multiple ratings per question) I opted to use mixed-effects models to account for individual variability and ensure more precise estimates. Although statistical inference often relies on binary decision-making, we should not be prisoners of dichotomized thinking. All statistical results, like the ones presented here, are subject to uncertainty and is conditional on specific data and modeling choices. One example warranting caution is the observed pattern linking facial expressiveness, perceived confidence, and presumed accuracy. Although no formal path analysis was conducted, each link in this chain – (1) more facial expressiveness predicting lower confidence, and then (2) lower confidence predicting lower accuracy – was statistically significant in isolation. This raises a potential interpretative pitfall: a witness who displays more facial expressiveness might be perceived as less confident, which in turn may lead observers to infer lower accuracy and, ultimately, reduced credibility. In this way, facial expressiveness may be equated with lower credibility through a chain of assumptions (e.g., facial expressiveness → low confidence → low accuracy → low credibility). However, such a conclusion would be both practically and theoretically fragile, which underscores why statistical artefacts alone should not dictate legal decision-making. Consequently, these findings should be interpreted as probabilistic tendencies, not deterministic rules, particularly when extrapolated to real-world legal contexts that are case-specific by nature.

Conclusion

What, then, can be concluded from this thesis? I examined testimony reliability, perceived credibility, investigative interviewing practices, and extra-legal factors in relation to whether eyewitnesses testified in their native or non-native language. Did the studies answer these key questions? Does this thesis reveal new insights or merely confirm existing concerns? Moreover, can the findings be applied universally across languages and cultural groups? Finally, can it now state with empirical support that non-native speakers face systematic discrimination not because of what they say, but because of *how* they say it?

Theoretical implications

An important strength of this thesis is that it isolates the effect of testifying in a non-native language while controlling for several potential confounds. Participants in the non-native group were generally proficient speakers, not predominantly from minority backgrounds, and did not differ from native speakers in self-reported cognitive effort or in displayed non-verbal behaviour. As such, the key findings, namely, lower confidence and reduced perceived credibility in non-native speakers, despite comparable accuracy and susceptibility to suggestions, can be attributed more directly to the act of testifying in a non-native language.

Building on this, the findings of this thesis offer a theoretically grounded account of how second-language use affects eyewitness testimony in high-stakes legal settings. Drawing on models on cognitive load (Baddeley, 2012; Cowan, 2010; Engle, 2002; Van Acker et al., 2018) and non-native language use (Abutalebi, 2008 Abutalebi & Green, 2007; Branzi et al., 2016; Chen & Chang, 2009; Green, 1998; Green & Abutalebi, 2013; Segalowitz, 2010), the results demonstrate that testifying in a non-native language lowers witnesses' confidence despite intact memory accuracy. This dissociation supports theoretical claims that confidence reflects retrieval effort rather than actual memory (Koriat & Goldsmith, 1996). This is further substantiated by the follow-up analyses within the non-native condition, showing that witnesses who felt less as opposed to more comfortable speaking in the non-native language were also less confident, yet equally accurate. This supports the interpretation

that reduced confidence may reflect greater retrieval effort rather than weaker memory, consistent with theoretical accounts of metacognitive monitoring. Moreover, the effect of non-native language use on observer-rated credibility underscores how delivery, rather than content, can bias fact-finders' evaluations. The separation between what is remembered and how it is expressed is crucial in legal contexts, as credibility is often inferred from extra-legal factors (Boccaccini, 2002; Denault et al., 2020, 2024; Lindholm & Yourstone, 2013; Frumkin, 2000, 2007, 2020; Rakic et al., 2011).

Practical implications

My conclusion is that the findings in this thesis do have practical implications. Imagine a courtroom or police interview where a non-native speaker is doing their best to recount a witnessed crime. Their speech is proficient, their grammar near-perfect, they display only few signs of nervousness or fidgeting. Yet their testimony is still met with added skepticism. This skepticism does not arise from what they say, but rather from how their testimony is perceived. As a result, the honest non-native speaking witness's account, however accurate, may be dismissed or undervalued. This scenario is plausible and illustrates a risk that may exist in many legal systems: that language barriers, even when not accompanied by low proficiency or visible distress, can influence credibility judgments and contribute to an unfair judgment on non-native speakers. A burden that grows even larger when coupled with visible cues of ethnic outgroup membership. This highlights the importance of evaluating testimony based on its content rather than on extra-legal factors.

To prevent such miscarriages of justice, legal systems must adopt evidence-based practices that weigh the cost of dismissing a strong testimony due to delivery-related biases (viz. language barriers) against the benefit of fair and accurate fact-finding. Training and education programs should equip legal professionals with the tools to evaluate testimony content less subjectively and to recognize how language and cultural differences can bias their judgments. While awareness alone will not eliminate these biases, it is a necessary first step toward safeguarding fairness. As Pezdek and Reisberg (2022) emphasize, empirical misconceptions about memory, perception, and judgment persist partly because scientific findings often fail to reach legal practitioners. To address this, efforts are needed to translate psychological science into actionable guidelines through expert testimony, judicial education, and policy reform. Kassin (2022) similarly advocates that legal actors be trained not only to recognize cognitive myths, such as relying on demeanor to infer truthfulness, but also to replace these myths with evidence-based practices. Without change and active engagement from researchers and institutions, even well-inten-

tioned reforms may fall short. In high-stakes contexts where credibility assessments carry a heavy legal weight, overlooking such disparities is not just an oversight, it is a risk to justice itself.

Directions for future research

Future research should begin with direct replications of the current findings to test their robustness. This includes both (a) exact replications using the same design with larger sample sizes, and, more importantly, (b) by involving other language groups to assess the generalizability across linguistic and cultural contexts. This should be prioritized to determine whether the observed effects reflect a genuine signal or merely statistical noise (Gigerenzer, 2004; Loken & Gelman, 2017; Smaldingo & McElreath, 2016).

Second, future research should extend beyond experimental paradigms to examine real-world investigative interviews with non-native speakers. One approach would be to analyse video-recorded police interviews or transcripts, using structured coding schemes to assess question types (e.g., proportion of suggestive vs non-suggestive questions), response quality (e.g., level of detail, clarity, consistency), and extra-legal factors (e.g., non-verbal behavior, ethnic ingroup vs. outgroup membership). These could then be compared across non-native and native speakers, while controlling for case type and severity. Such ecologically grounded designs would provide important understanding into how language barriers shape not only testimony but downstream legal consequences.

Finally, a critical avenue for future work is to investigate the interpreter-assisted interviews. These are increasingly common when language barriers are present, yet little is known about how interpretation influences questioning strategies, witness responses, and credibility judgments. Research is needed to clarify what communicative dynamics emerge in interpreter-assisted interviews, particularly whether the question posed is the same as the one being answered, and what may be lost, altered, or implicitly added in translation. Such distortions may subtly shift the trajectory of an interview and, consequently, the outcome of a case.

Concluding remarks

The findings across the three empirical studies in this thesis do, in fact, point to a troubling pattern: non-native speakers were just as reliable as native speakers in terms of memory accuracy and susceptibility to suggestions, yet

they were less confident in their testimonies and were perceived as less credible by independent observers. Specifically, **Study Ia** showed that non-native and native speakers were equal in testimony accuracy and suggestibility; **Study Ib** demonstrated that credibility assessments favored native speakers; **Study II** demonstrated that interrogators adjusted their level of doubt based on credibility for native, but not for non-native speakers; and **Study III** showed that non-verbal cues did not predict statement accuracy, confidence, or credibility judgments, challenging assumptions about their diagnostic value in legal settings.

Together, these results demonstrate that non-native speakers can face additional hurdles in legal settings, not due to poorer memory recall, but due to how their testimony is perceived and responded to. Although the differences observed were not large, even small biases may carry serious consequences in high-stakes contexts. This underscores the need for legal procedures that account for the communicative demands placed on non-native speaking eyewitnesses and promote fairer credibility assessments. While non-native and native speakers may be equally reliable, these findings provide empirical evidence that language barriers can still lead to differential treatment, ultimately compromising fairness in judicial proceedings.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*, 466–478. https://doi.org/10.1016/j.actpsy.2008.03.014

Abutalebi, J., & Green, D. W. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242–275. doi:10.1016/j.jneuroling.2006.10.003

Albright, T. D. (2017). Why eyewitnesses fail. *Proceedings of the National Academy of Sciences*, *114*(30), 7758-7764. https://doi.org/10.1073/pnas.1706891114

Allison, M., Gerwing, J., & Gadaire, C. B. (2022). When the eyewitness to a crime is an English language learner: Identifying and resolving troubles in understanding in interviews. *Journal of Investigative Psychology and Offender Profiling*, *19*(3), 185-203. https://doi.org/10.1002/jip.1588

Alm, C., Helmy Rehnberg, N., & Lindholm, T. (2019). Language and eyewitness suggestibility. *Journal of Investigative Psychology and Offender Profiling*, *16*(3), 201–212. https://doi.org/10.1002/jip.1529

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305-307. https://doi.org/10.1038/d41586-019-00857-9

Anakwah, N., Horselenberg, R., Hope, L., Amankwah-Poku, M., & Van Koppen, P. J. (2020). Cross-cultural differences in eyewitness memory reports. *Applied Cognitive Psychology*, *34*(2), 504-515. https://doi.org/10.1002/acp.3637

Anvari F., & Lakens D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology*, 96, Article 104159. https://doi.org/10.1016/j .jesp.2021.104159

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual review of psychology*, *63*(1), 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, B. E., & Loftus, E. F. (1988). Degree of Detail of Eyewitness Testimony and Mock Juror Judgments 1. *Journal of Applied Social Psychology*, *18*(14), 1171-1192. https://doi.org/10.1111/j.1559-1816.1988.tb01200.x

Boccaccini, M. T. (2002). What do we really know about witness preparation?. *Behavioral sciences & the law*, *20*(1-2), 161-189.

Boroditsky, L., Ham, W., & Ramscar, M. (2002). What is universal in event perception? Comparing English and Indonesian speakers. In *Proceedings of the 24th annual meeting of the cognitive science society*. Fairfax, VA

Bothwell, R. K., & Jalil, M. (1992). The credibility of nervous witnesses. *Journal of Social Behavior & Personality*, *7*(4), 581–586.

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex, 26*(6), 2367-2380. doi:10.1093/cercor/bhv052

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and human behavior*, *42*(5), 458. https://doi.org/10.1037/lhb0000294

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*(1), 11. https://doi.org/10.1037/1076-898X.12.1.11

Burgoon, J. K., Guerrero, L. K., & Manusov, V. (2011). Nonverbal signals. In The SAGE Handbook of Interpersonal Communication. SAGE Publications: Thousand Oaks, CA, USA, 2011; pp. 239–282

Burnham, K. P. & Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd Edition, Springer-Verlag, New York. http://dx.doi.org/10.1007/b97636

Cairns-Lee, H., Lawley, J., & Tosey, P. (2022). Enhancing researcher reflexivity about the influence of leading questions in interviews. *The Journal of Applied Behavioral Science*, *58*(1), 164-188. https://doi.org/10.1177/00218863211037446

Ceci, S. J., Kulkofsky, S., Klemfuss, J. Z., Sweeney, C. D., & Bruck, M. (2007). Unwarranted assumptions about children's testimonial accuracy. *Annual Review of Clinical Psychology*, *3*(1), 311-328. https://doi.org/ 10.1146/annurev.clinpsy.3.022806.091354

Cederborg, A.-C., Alm, C., Lima da Silva Nises, D., & Lamb, M. E. (2013). Investigative interviewing of alleged children – An evaluation of a new training programme for police officers in Sweden. *Police Practice and Research: An International Journal*, *14*, 242-254. https://doi.org/10.1080/15614263.2012.712292

Chalmers, J., Leverick, F., & Munro, V. E. (2022). Handle with care: Jury deliberation and demeanour-based assessments of witness credibility. The International Journal of Evidence & Proof, 26(4), 381-406. https://doi.org/10.1177/13657127221120955

Chan, J. C., Manley, K. D., & Lang, K. (2017). Retrieval-enhanced suggestibility: A retrospective and a new investigation. *Journal of applied research in memory and cognition*, *6*(3), 213-229. https://doi.org/10.1016/j.jarmac.2017.07.003

Chen, I. J., & C. C. Chang. (2009). Cognitive load theory: An empirical study of anxiety and task performance in language learning. *Electronic Journal of Research in Educational Psychology 7*(2), 729–746.

Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal. *R-package version*, *28*, 406.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

Common European Framework of Reference for Languages. (2019). *Global Scale*. https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Council of Europe. (1950). *European Convention for the Protection of Human Rights and Fundamental Freedoms*, as amended by Protocols Nos. 11 and 14. Retrieved from https://www.echr.coe.int/

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current directions in psychological science*, *19*(1), 51-57. https://doi.org/10.1177/09637214093592

Cowan, N., Bao, C., Bishop-Chrzanowski, B. M., Costa, A. N., Greene, N. R., Guitard, D., ... & Ünal, Z. E. (2024). The relation between attention and memory. *Annual Review of Psychology*, *75*(1), 183-214. https://doi.org/10.1146/annurev-psych-040723-012736

de Bruïne, G., Vredeveldt, A., & van Koppen, P. J. (2023). Culture and credibility: the assessment of asylum seekers' statements. *Psychology, Crime & Law*, 1-23. https://doi.org/10.1080/1068316X.2023.2279328

de Bruïne, G., Vredeveldt, A., & van Koppen, P. J. (2025). The way we remember and report: an experiment testing cultural differences in eyewitness memory. *Memory*, 1-10. https://doi.org/10.1080/09658211.2025.2505213

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and human behavior*, *28*(6), 687-706. https://doi.org/10.1007/s10979-004-0565-x

Denault, V., Plusquellec, P., Jupe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., ... & Koppen, P. J. V. (2020). The analysis of nonverbal communication: The dangers of pseudoscience in security and justice contexts. *Anuario de Psicología Jurídica*, *30*(1), 1-12. doi:10.5093/apj2019a9

Denault, V., Leclerc, C., & Talwar, V. (2024). The use of nonverbal communication when assessing witness credibility: a view from the bench. Psychiatry, Psychology and Law, 31(1), 97-120. https://doi.org/10.1080/13218719.2023.2175068

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. Psychological bulletin, 129(1), 74. DOI: 10.1037/0033-2909.129.1.74

Dhami, M. K., Goodman-Delahunty, J., & Desai, S. (2017). Development of an information sheet providing rapport advice for interpreters in police interviews. *Police practice and research*, *18*(3), 291-305. http://dx.doi.org/10.1080/15614263.2017.1291580

Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings, 2010 O.J. (L 280) 1–7.

Dodson, C. S., & Schacter, D. L. (2015). Memory distortion. In *Handbook of cognitive neuropsychology* (pp. 445-463). Psychology Press.

Edmar, B. A., Enquist, N. S. A., & Luke, T. J. (2025, April 26). What Fifty Years of Research on the Misinformation Effect Can Tell Us: A Meta-Analytic Review. https://doi.org/10.31234/osf.io/m68kf_v1

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19-23. https://doi.org/10.1111/1467-8721.00160

Englert, C., Persaud, B. N., Oudejans, R. R. D., & Bertrams, A. (2015). The influence of ego depletions on sprint start performance in athletes without track and field experience. *Frontiers in Psychology*, *6*, 1207. https://doi.org/10.3389/fpsyg.2015.01207

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers*, *28*(1), 1-11. https://doi.org/10.3758/BF03203630

Ernberg, E., & Mac Giolla, E. (2022). Interviewing witnesses in a second language: A comparison of interpreter-assisted, unaided, and self-administered interviews. *Legal and Criminological Psychology*. https://doi.org/10.1111/lcrp.12231

Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. Frontiers in Psychology, 9, 338. https://doi.org/10.3389/fpsyg.2018.00338

EUREX. (2025). The European Registry of Exonerations. Online: https://www.registryo fexonerations.eu/

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology*, *25*(3), 501-508. https://doi.org/10.1002/acp.1722

Fausey, C., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin and Review*, *18*(1), 150-157. doi:10.3758/s13423-010- 0021-5

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: A meta-analytic review of the 'weapon focus' literature. *Psychology, Crime & Law*, *19*(1), 35-66. https://doi.org/10.1080/1068316X.2011.59932

Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, *15*(4), 466-485. doi:10.1177/1367006911403062

Filipović, L. (2020). Bilingual memory advantage: Bilinguals use a common linguistic pattern as an aid to recall memory. *International Journal of Bilingualism*, *24*(3), 542-555. doi:10.1177/1367006918814381

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106-E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy, and Law*, *1*(4), 732.

Fisher, R. P., & Cutler, B. L. (1995). The relation between consistency and accuracy of eyewitness testimony. *Psychology, law, and criminal justice: International developments in research and practice*, 21-28.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C Thomas Publisher.

Fisher, R. P., & Schreiber, N. (2017). Interview protocols to improve eyewitness memory. In *The Handbook of Eyewitness Psychology: Volume I* (pp. 53-80). Psychology Press.

Fitzgerald, R. J., Rubínová, E., & Juncu, S. (2021). Eyewitness identification around the world. *In Methods, measures, and theories in eyewitness identification tasks* (pp. 294-322). Routledge.

Frumkin, L. A. (2000). *The effect of eyewitnesses' accent, nationality, and authority on the perceived favorability of their testimony*. University of Maryland, Baltimore County.

Frumkin, L. (2007). Influences of accent and ethnic background on perceptions of eyewitness testimony. *Psychology, Crime & Law*, *13*(3), 317–331. https://doi.org/10.1080/10683160600822246

Frumkin, L. (2020). The impact of British accents on perceptions of eyewitness statements. *Journal of Language and Discrimination*, *4*(1), 119-138. https://doi.org/10.1558/jld.39368

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in methods and practices in psychological science*, *2*(2), 156-168. https://doi.org/10.1177/2515245919847202

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and human behavior*, *33*, 298-307. DOI: 10.1007/s10979-008-9146-8

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Harvard University Press. https://doi.org/10.4159/harvard.9780674060982.c10

Geiselman, R. E., & Fisher, R. P. (1989). The cognitive interview technique for victims and witnesses of crime. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 191–215). Springer Publishing Company.

Gigerenzer, G. (2004). Mindless statistics. *The journal of socio-economics*, *33*(5), 587-606. https://doi.org/10.1016/j.socec.2004.09.033

Goodman-Delahunty, J., Martschuk, N., Hale, S.B., Brandon, S.E. (2020). Interpreted Police Interviews: A Review of Contemporary Research. In: Miller, M.K., Bornstein, B.H. (eds) Advances in Psychology and Law. *Advances in Psychology and Law*, vol 5. Springer, Cham. https://doi.org/10.1007/978-3-030-54678-6_4

Green, D. J. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67-81. https://doi.org/10.1017/S1366728998000133

Green, D.W., & Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. doi:10.1080/20445911.2013.796377

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Lawrence Erlbaum Associates Publishers.

Gudjonsson, G. H. (1997). *The Gudjonsson Suggestibility Scales*. Hove: Psychology Press.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. John Wiley & Sons.

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour*, *1*(2), 83–104.

Gustafsson, P. U., Lindholm, T., & Jönsson, F. U. (2019). Predicting accuracy in eyewitness testimonies with memory retrieval effort and confidence. *Frontiers in psychology*, *10*, 703. https://doi.org/10.3389/fpsyg.2019.00703

Gustafsson, P. U., Lindholm, T., & Jönsson, F. U. (2021). Judging the accuracy of eyewitness testimonies using retrieval effort cues. *Applied Cognitive Psychology*, *35*(5), 1224-1235. https://doi.org/10.1002/acp.3854

Gustafsson, P. U., Lindholm, T., & Jönsson, F. U. (2022). Eyewitness accuracy and retrieval effort: Effects of time and repetition. *Plos one*, *17*(9), e0273455

Hansen, K., Rakić, T., & Steffens, M. C. (2017). Competent and warm?. *Experimental Psychology*. *64*(1). https://doi.org/10.1027/1618-3169/a000348

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological bulletin*, *137*(4), 643. https://doi.org/10.1037/a0023589

Hope, L., & Gabbert, F. (2019). Interviewing witnesses and victims. In N. Brewer, & A. B. Douglass (Eds.), *Psychological science and the law (pp. 130–156). The Guildford Press.*

Hope, L., Gabbert, F., & Fisher, R. P. (2011). From laboratory to the street: Capturing witness memory using the Self-Administered Interview. *Legal and Criminological Psychology*, *16*(2), 211–226. https://doi.org/10.1111/j.2044-8333.2011.02015.x

Hope, L., Anakwah, N., Antfolk, J., Brubacher, S. P., Flowe, H., Gabbert, F., ... & Anonymous. (2022). Urgent issues and prospects at the intersection of culture, memory, and witness interviews: Exploring the challenges for research and practice. *Legal and criminological psychology*, *27*(1), 1-31. https://doi.org/10.1111/lcrp.12202

Hope, L., Kontogianni, F., Rechdan, J., Tavitian-Emladjian, L., Soubra, N. A., Abu Marak Brome, D. M., Gibson, V., & Anakwah, N. (2023). Exploring cultural differences in eyewitness accounts using a self-administered reporting technique. *Psychology, Crime & Law*, 1-19. https://doi.org/10.1080/1068316X.2023.2279330

Hope, L., Hawkins de Namor, N., Gabbert, F., Ferenczi, N., & Oxburgh, G. (2025). From an investigator's perspective: challenges and opportunities in building and maintaining rapport in cross-cultural investigative interviewing contexts. *Journal of Criminal Psychology*. https://doi.org/10.1108/JCP-11-2024-0123

Horry, R., Hughes, C., Sharma, A., Gabbert, F., & Hope, L. (2021). A meta-analytic review of the Self-Administered Interview©: Quantity and accuracy of details reported on initial and subsequent retrieval attempts. *Applied Cognitive Psychology*, *35*(2), 428-444. https://doi.org/10.1002/acp.3753

Hu, Z., & Naka, M. (2022). Eyewitness testimony in native and second languages. *Psychology, Crime & Law*, 1-17. https://doi.org/10.1080/1068316X.2022.2030332

Hyman Gregory, A., Wolfs, A., & Schreiber Compo, N. (2023). Witness/victim interviewing: a survey of real-world investigators' training and practices. *Psychology, Crime & Law*, *29*(9), 957-981. https://doi.org/10.1080/1068316X.2022.2043312

Itzhak, I., Vingron, N., Baum, S. R., & Titone, D. (2017). Bilingualism in the real world: How proficiency, emotion, and personality in a second language impact communication in clinical and legal settings. *Translational Issues in Psychological Science*, *3*(1), 48–65. https://doi.org/10.1037/tps0000103

Jores, T., Colloff, M. F., Kloft, L., Smailes, H., & Flowe, H. D. (2019). A meta-analysis of the effects of acute alcohol intoxication on witness recall. *Applied cognitive psychology*, *33*(3), 334-343. https://doi.org/10.1002/acp.3533

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual review of psychology*, *68*(1), 601-625. https://doi.org/10.1146/annurev-psych-122414-033702

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1304-1316. https://doi.org/10.1037/0278-7393.22.5.1304

Kassin, S. M. (2022). It's time to bury three justice-corrupting myths once and for all. *Journal of Applied Research in Memory and Cognition*, *11*(2), 161-165.

Kaufmann, G., Drevland, G. C., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *17*(1), 21-34. https://doi.org/10.1002/acp.842

Kebbell, M. R., Evans, L., & Johnson, S. D. (2010). The influence of lawyers' questions on witness accuracy, confidence, and reaction times and on mock jurors' interpretation of witness accuracy. *Journal of Investigative Psychology and Offender Profiling*, *7*(3), 262-272. https://doi.org/10.1002/jip.125

Koriat, A. (2012). "The subjective confidence in one's knowledge and judgements: some metatheoretical considerations," in *Foundations of Metacognition*, eds M. J. Beran, J. Brandl, J. Perner, and J. Proust (Oxford: Oxford University Press), 213–233. doi: 10.1093/acprof:oso/9780199646739.003.0014

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological review*, *103*(3), 490. https://doi.org/10.1037/0033-295X.103.3.490

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D. (2022). Improving Your Statistical Inferences. Retrieved from https://lakens.github.io/statistical_inferences/. https://doi.org/10.5281/zenodo.6409077

Lai, M., & Mulayim, S. (2014). Interpreter linguistic intervention in the strategies employed by police in investigative interviews. *Police Practice and Research*, *15*(4), 307-321. https://doi.org/10.1080/15614263.2013.809929

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means. *AKA least-squares means*, *1*(7).

Lindholm, T. (2005). Group-based biases and validity in eyewitness credibility judgments: Examining effects of witness ethnicity and presentation modality. *Journal of Applied Social Psychology*, *35*, 1474–1501. https://doi.org/10.1111/j.1559-1816.2005.tb02180.x

Lindholm, T. (2008). Validity in judgments of high- and low-accurate witnesses of own and other ethnic groups. *Legal and Criminological Psychology*, *13*, 107–121. https://doi.org/10.1348/135532506X152949

Lindholm, T. (2008). Who can judge the accuracy of eyewitness statements? A comparison of professionals and lay-persons. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*(9), 1301-1314. https://doi.org/10.1002/acp.1439

Lindholm, T., & Yourstone Cederwall, J. (2013). Ethnicity and gender biases in the courtroom. In P. A. Granhag (Ed.), *Forensic psychology in context: Nordic and international approaches* (pp. 228–246). London, UK: Routledge Taylor & Francis Group.

Lindholm, T., Jönsson, F. U., & Liuzza, M. T. (2018). Retrieval effort cues predict eyewitness accuracy. *Journal of Experimental Psychology: Applied*, *24*(4), 534. https://doi.org/10.1037/xap0000175

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*(3), 349-358.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & memory*, *12*(4), 361-366. DOI: 10.1101/lm.94705

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589. https://doi.org/10.1016/S0022-5371(74)80011-3

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584-585. DOI: 10.1126/science.aal3618

Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. Perspectives on Psychological Science, 14(4), 646-671. https://doi.org/10.1177/1745691619838258

Marian, V., & Kaushanskaya, M. (2004). Self-construal and emotion in bicultural bilinguals. *Journal of Memory and Language, 51*(2), 190–201. doi:10.1016/j.jml.2004.04.003

Marian, V., & Kaushanskaya, M. (2007). Language context guides memory content. *Psychonomic Bulletin & Review*, 14, 925–933. doi:10.3875/BF03194123

Meissner, C. A. (2021). "What works?" Systematic reviews and meta-analyses of the investigative interviewing research literature. *Applied Cognitive Psychology*, *35*(2), 322-328. https://doi.org/10.1002/acp.3808

Meissner, C. A., Kleinman, S. M., Mindthoff, A., Phillips, E. P., and Rothweiler, J. N. (2021). "Investigative interviewing: a review of the literature and a model of science-based practice," in *Oxford Handbook of Psychology and Law*, eds D. DeMatteo and K. Scherr's (Oxford: Oxford University Press).

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 1–21. https://doi.org/10.1016/S0022-5371(63)80063-8

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British journal of psychology*, *94*(3), 339-354. https://doi.org/10.1348/000712603767876262

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. Psychology, public policy, and law, 16(4), 340. https://doi.org/10.1037/a0020518

Mickes, L., Clark, S. E., & Gronlund, S. D. (2017). Distilling the confidence-accuracy message: A comment on Wixted and Wells (2017). *Psychological Science in the Public Interest*, *18*(1), 6-9. https://doi.org/10.1177/1529100617699240

Mickes, L., Wilson, B. M., & Wixted, J. T. (2025). The cognitive science of eyewitness memory. *Trends in Cognitive Sciences*.

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Moore, N.-J., Hickson, M. & Stacks, D. W. *Nonverbal Communication: Studies and Applications* (Oxford University Press, 2014).

Neal, T. M., Guadagno, R. E., Eno, C. A., & Brodsky, S. L. (2012). Warmth and competence on the witness stand: Implications for the credibility of male and female expert witnesses. *Journal of the American Academy of Psychiatry and the Law Online*, *40*(4), 488-497.

Otgaar, H., Alberts, H., & Cuppens, L. (2012). How cognitive resources alter our perception of the past: Ego depletion enhances the susceptibility to suggestion. *Applied Cognitive Psychology*, *26*, 159–163. https://doi.org/10.1002/acp.1810

Otgaar, H., Howe, M. L., Merckelbach, H., & Muris, P. (2018). Who is the better eyewitness? Sometimes adults but at other times children. *Current Directions in Psychological Science*, *27*(5), 378–385. https://doi.org/10.1177/0963721418770998

Panzarella E., Beribisky N., Cribbie R. A. (2021). Denouncing the use of field-specific effect size distributions to inform magnitude. *PeerJ*, 9, Article e11383. https://doi.org/10.7717/peerj.11383

Patterson, M. L., Fridlund, A. J., & Crivelli, C. (2023). Four misconceptions about nonverbal communication. Perspectives on Psychological Science, 18(6), 1388-1411. https://doi.org/10.1177/17456916221148142

Pezdek, K., & Reisberg, D. (2022). Psychological myths about evidence in the legal system: How should researchers respond?. *Journal of Applied Research in Memory and Cognition*, *11*(2), 143. https://doi.org/10.1037/mac0000037

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, *1*(4), 817.

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of personality and social psychology*, *100*(1), 16.

R Core Team (2022). R: A language and environment for statistical computing, ver. 4.2.2. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rogers, H., Fox, S., & Herlihy, J. (2015). The importance of looking credible: The impact of the behavioural sequelae of post-traumatic stress disorder on the credibility of asylum seekers. Psychology, Crime & Law, 21(2), 139-155. https://doi.org/10.1080/1068316X.2014.951643

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Scassa, T. (1994). Language standards, ethnicity and discrimination. *Canadian Ethnic Studies*, *26*(3).

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*(8), 603. https://doi.org/10.1037/a0029869

Schreiber Compo, N., Hyman Gregory, A., & Fisher, R. (2012). Interviewing behaviors in police investigators: A field study of a current US sample. *Psychology, Crime & Law*, *18*(4), 359-375. https://doi.org/10.1080/1068316X.2010.494604

Schroeder, S.R., & Marian, V. (2014). Bilingual episodic memory: How speaking two language influences remembering. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of bilingual memory* (pp. 111–132). Springer Science + Business Media. doi:10.1007/978-1-4614-9218-4_6

Seale-Carlisle, T. M., Quigley-McBride, A., Teitcher, J. E., Crozier, W. E., Dodson, C. S., & Garrett, B. L. (2024). New insights on expert opinion about eyewitness memory research. *Perspectives on Psychological Science*, 17456916241234837. https://doi.org/10.1177/17456916241234837

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Shaw III, J. S. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied*, *2*(2), 126.

Shaw III, J. S., & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, *20*(6), 629-653.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society open science*, *3*(9), 160384. https://doi.org/10.1098/rsos.160384

Smeets, T., Candel, I., & Merckelbach, H. (2004). Accuracy, completeness, and consistency of emotional memories. *The American journal of psychology*, 595-609. https://doi.org/10.2307/4148994

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315. https://doi.org/10.1037/0033-2909.118.3.315

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. Journal of Memory and Language, 53(1), 81-94. https://doi.org/10.1016/j.jml.2005.02.003

The National Registry of Exonerations. (2025). *Exonerations in the United States Map*. Exonerations in the United States map.

The Innocence Project. Innocence Project. [Internet]. 2025 [cited 2025 Jun 10]. Available from: https://www.innocenceproject.org/

United Nations, Department of Economic and Social Affairs, Population Division. (2020). *International Migration Report 2020: Highlights (ST/ESA/SER.A/404)*. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd_2020_international_migration_highlights.pdf

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352-373. doi:10.1037/h0020071.

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, technology & work*, *20*(3), 351-365. https://doi.org/10.1007/s10111-018-0481-3

van Veldhuizen, T. S., Horselenberg, R., Landström, S., Granhag, P. A., & van Koppen, P. J. (2017). Interviewing asylum seekers: A vignette study on the questions asked to assess credibility of claims about origin and persecution. *Journal of Investigative Psychology and Offender Profiling*, *14*(1), 3-22. https://doi.org/10.1002/jip.1472

Vredeveldt, A., Given-Wilson, Z., & Memon, A. (2023). Culture, trauma, and memory in investigative interviews. *Psychology, Crime & Law*, 1-21. https://doi.org/10.1080/1068316X.2023.2209262

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2), 110-117. https://doi.org/10.1016/j.jarmac.2012.02.004

Vrij, A., & Granhag, P. A. (2014). Eliciting information and detecting lies in intelligence interviewing: An overview of recent research. *Applied Cognitive Psychology*, *28*(6), 936-944. https://doi.org/10.1002/acp.3071

Wade, K. A., Nash, R. A., & Lindsay, D. S. (2018). Reasons to doubt the reliability of eyewitness memory: Commentary on Wixted, Mickes, and Fisher (2018). *Perspectives on Psychological Science*, *13*(3), 339-342. https://doi.org/10.1177/1745691618758261

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*(1), 192-196. https://doi.org/10.3758/bf03206482

Walsh, D., Oxburgh, G. E., & Amurun, T. (2020). Interpreter-assisted interviews: Examining investigators' and interpreters' views on their practice. *Journal of Police and Criminal Psychology*, *35*, 318-327. https://doi.org/10.1007/s11896-020-09366-2

Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications? Using memory for peripheral detail can be misleading. *Journal of Applied Psychology*, *66*(6), 682. https://doi.org/10.1037/0021-9010.66.6.682

Wells, G. L., & Loftus, E. F. (2003). Eyewitness memory for people and events. *Handbook of psychology: Forensic psychology*, *11*, 149-160.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45–75. https://doi.org/10.1111/j.1529-1006.2006.00027.x

Wessel, E., Drevland, G. C., Eilertsen, D. E., & Magnussen, S. (2006). Credibility of the emotional witness: a study of ratings by court judges. Law and human behavior, 30(2), 221. doi:10.1007/s10979-006-9024-1

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686.

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger III, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, *70*(6), 515.

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, *13*(3), 324-335. https://doi.org/10.1177/1745691617734878

Wixted, J. T., Wells, G. L., Loftus, E. F., & Garrett, B. L. (2021). Test a witness's memory of a suspect only once. *Psychological Science in the Public Interest*, *22*(1_suppl), 1S-18S.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10-65. https://doi.org/10.1177/1529100616686966

Yates S. Q. (2017, January 6). *Memorandum for heads of department law enforcement components, all department prosecutors*. Washington, DC: Office of the Deputy Attorney General, U.S. Department of Justice. Retrieved from https://www.justice.gov/archives/opa/press-release/file/923201/download

Acknowledgements

I owe my deepest gratitude to *Charlotte Alm*. Throughout this journey, you met ideas with calm, clarity, and good sense. You have made space for difficult questions and trusted me to shape my own path. I have learned so much from your example. Thank you for your time and patience, for your generosity, and for a kind leadership that sets the tone for good science and good collaboration. I am grateful for your trust and belief in me.

Torun Lindholm Öjmyr, from my BA and MA theses onward (we go way back!), you've been the person who keeps things honest and moving. Your guidance, paired with real perseverance, taught me to cut through noise. You challenged me when it mattered and backed me when it counted. I've learned as much from your steadiness as from your advice. Thank you.

My sincere thanks to *Mats Nilsson*, for discussions on all things matters. Your precision and range reshaped my scientific thinking. I'm deeply grateful.

Philip "Puggen" Gustafsson, thank you for taking me under your wing. You welcomed me in and steered me through, salute for being the best partner in crime I could have wished for.

Freja Isohanni, thanks for showing up with wit and a gentle light that made the days kinder.

Sofie Liljestrand Hassoun, beyond your meticulous assistance, thank you for the fun and the friendship.

Göran Kecklund, thank you for the unfailingly positive and impactful presence you bring to me and to the department (Forza Kecklund-bullen!).

My sincere thanks to *Christian Portin, Valentina Barrios Pizarro, Daniela Brandel, Henrik Dunér, Jenny Bourelius, Erica Grahn, Mats Gröndal, Beatriz Helgotsson, Anastasia Helperin, Henric Bergqvist, Tommy Olin,* and *Stefan Gelov* the silent engines and backbone of the department. I am grateful for everything you have done for me.

I am grateful to my colleagues in CLIP (Research unit for Criminal, Legal and Investigative Psychology), University of Gothenburg, *Emelie Ernberg, Erik Mac Giolla, Sara Landström, Leif Strömwall, David Neequaye, Karl Ask, Mikaela Magnusson*, *Malin Joleby, Pär-Anders Granhag,* and *Minhwan Jang*, for launching my academic path. In our line of work, you set the bar: integrity, thoughtful critique and, above all, the collaborative spirit that science depends on.

Timothy J. Luke, you're the MJ #23 of my academic world: precision, clutch, and standards that push me higher. Thank you for the inspiration.

To my Sumi-hitos, *Elli van Berlekom* and *Astrid Hallman*, you brought the jokes, the sparkle, and the perfect chaos – highs, lows, and all the glorious in-betweens. Thank you.

To *Sofia "Braccs" Bracco* and *Clara Plückelmann*, or should I say Kobe–Shaq, or even Totti–Cassano... different games and styles, but same magic. The duo I wouldn't trade for anything in life.

To *Martin Oscarsson* and *Tim Lachmann*: Martin, first a comrade, and then a colleague; Tim, an ally from day one. Both, brothers in arms. We ride together; we die together – PhDs for life (then the beat goes...).

Anna Sofia Tanimoto, we started this thing together... but didn't finish together (I see you). Thank you for all the long and fruitful discussions every morning on the 2.

Mårten Hammarlund and *Joel Gruneau Brulin*, genuinely kind and quietly brilliant. Each in your own way. Thanks for the decency, the brains, and your steady presence – I salute you.

And now to the old ones, the new ones, the good ones and the... There have been many comings and goings, each with their own beautiful presence that indeed made these years feel like a joy. In no particular order (?), my thanks to *Lisa Axelsson, Hanna Björlin Avdic, Nhi Nguyen, Caroline Kummelstedt, Billy Gerdenfeldt, Ida Nilsen, Niklas "Dusty Blue" Leitner, Melanie Wiehe, Cecilia Salzmann, Andreas Sarling, Erik Forsberg, Joanna Lindström, William Fredborg, Rebecka Fältén, Philip Ström, Franziska Müller, Ingela Clausén Gull, Thomas Jack Samuels, Elahe Tavakoli Mahramnejad, Yannick Klein, Miriam Napadow, Erik Bojerud, Zaur Afandiyev, Niklas Carlson, Annika Andersson, Lillian Döllinger, Marta Zakrzewska, Ivo Todorov, Nichel Bo Gonzalez,* and *Sebastian Cancino Montecinos*.

To my senior (in academic terms) colleagues *Armita Golkar, Jacob Håkansson, Tina Sundelin, Maria Larsson, Laura Ferrer-Wreder, Marie Gustafsson-Sendén, Sabina Cehajic-Clancy, Håkan Fischer, Ann-Christin Cederborg, Petri Laukka, Stefan Wiens, Ronald Van den Berg, Magnus Sverke, Petra Lindfors, Fredrik Jönsson, Pehr Granqvist*, *Matilda Frick* and *Johan Lundin Kleberg* thank you for your generosity and guidance.

To *Tomas "Zigao", "Tompaõ", "thinks he knows NBA...", "but he doesn't" Lindmark*, the score is now 134-98 – you are welcome.

To my dearest friend *Christopher Frånberg*, for every time you called and I didn't pick up – I appreciate you.

Shout-out to my brothers *Adam, Bobo, CK, Il Bergano, Faffa, Robbin, Jeppe, Stea, Kåre, John, DK, Charlie, Dogge, Erik, Carlos,* and *Mantilla*, for backing me, always.

Gabriel Olausson, thank you for all things IDA in life – if u know u know.

To my 540-angels, *Lina Nyström* (aka Gary Oldman) and *Elsa Fanni* (aka Nicole Kidman), you two are my favourite people. Thank you for being my pel, for your unwavering love, and for keeping the fire burning throughout.

To my family. The grandmas & the grandpas, the aunties *Tulay* & *Nesrin,* and the uncs *Erol* & *Cihangir,* to my cousins *Alex* and *Kaan*, thank you for the warmth and the sense of home that travels with me.

My mother *Olcay* and father *Cengiz*, for your unconditional love. Thank you.

To *Bover*, you are my fallback and my force. Your support has carried me more than you know.

Ez sıma zaf hesxon