Psychology of Language and Communication 2023, Vol. 27, No. 1

sciendo

DOI: 10.58734/plc-2023-0012

Keith Wylie[1] and Jacqueline R. Evans[2]
[1] Psychology Department, Emporia State University, United States
[2] Psychology Department, Florida International University, United States

Investigating the utility of cognitive interview mnemonics among non-native English speakers

Best practice eyewitness interviewing mnemonics have not been tested with linguistically diverse samples. Cognitively complex mnemonics may overload non-native speakers' cognitive resources, which are already engaged in speaking a non-native language. Social facilitation mnemonics may help non-native speakers, who might be hesitant to report details. The current study tested the reverse order mnemonic (cognitively complex) and a set of introductory instructions (social facilitation) compared to control interviews among native and non-native English speakers. Native speakers provided more details than non-native speakers, particularly in the control interview condition. Both the reverse order and control conditions elicited newly generated details because of repeated questioning. Accuracy rates were comparable across language and interview conditions. Future research should develop an interviewing protocol that is sensitive to the challenges faced by non-native speakers.

Key words: language, non-native speakers, investigative interviewing, cognitive interview, eyewitness memory.

Address for correspondence: Keith Wylie
Psychology Department, Emporia State University, Emporia, United States
Email: kwylie@emporia.edu

This is an open access article licensed under the CC BY NC ND 4.0 License.

Over 20% of people in the United States speak a language other than English in their homes, with over 25 million of them speaking English "less than very well" (Zeigler & Camarota, 2019). Compared to native and bilingual English speakers, non-native English speakers often live in areas with relatively high crime rates (Gostjev & Nielsen, 2017). Thus, they may witness or be victims of crimes, potentially requiring them to provide statements to police investigators. However, police are unlikely to seek interpretation when interviewing witnesses and victims who do not speak English well (Shaffer & Evans, 2018; Wakefield et al., 2015). Indeed, The New York Times has covered many such instances over the past several years (e.g., Rivera & Zraick, 2010; Southall, 2017; Turkewitz, 2014). As such, if an interview happens at all, it typically happens in English—the witness or victim's second language. Although there has been much research on best practice witness interviewing guidelines, little research has focused specifically on interviewing witnesses in their second languages. Importantly, speaking a second language introduces additional cognitive and social demands that may not be present in first-language interactions. These additional cognitive and social demands of may adversely affect the amount and reliability of details that witnesses provide.

In the current study, we compared the number and accuracy of details elicited in interviews with native and non-native English-speaking samples. Further, we manipulated the interview strategies employed in order to determine their utility with non-native English speakers. Specifically, from the literature on best-practice witness interviews (particularly the cognitive interview, CI, Fisher & Geiselman, 1992), we identified two strategies to test. First, a cognitive strategy (the reverse order instruction) and second, a social strategy (the "transfer of control" instruction).

Interviewing Witnesses and Victims

The literature and best practice guidelines on witness and victim interviewing are largely dominated by one well-researched approach: the CI (Fisher et al., 2014; Wells et al., 2020). The CI (Fisher & Geiselman, 1992) comprises several investigative interviewing techniques and mnemonics that are designed to improve recall and has been shown to yield substantially more correct details compared to a standard comparison interview in meta-analyses ($d \geq 0.80$; Köhnken et al., 1999; Memon et al., 2010).Two broad foci of the CI are cognitive psychological principles that increase memory output and social psychological principles that facilitate communication.

The cognitive aspect of the CI can include encouraging multiple and varied retrieval attempts. The reverse order mnemonic is one specific way to implement varied retrieval included in the original CI (Fisher & Geiselman, 1992). The reverse order technique requires witnesses to recount events backward, starting with the end of the event. Asking witnesses to recount their memories in reverse

order helps them to search their memory "frame by frame" instead of relying on a script or schematic reconstruction (Geiselman et al., 1986). This can help witnesses uncover incidental or nonschematic details, which can be useful to investigators (Geiselman & Callot, 1990).

The social prong of the CI can include various instructions given to witnesses during the interview. For example, the transfer of control instruction emphasizes that the witness, not the interviewer, knows what happened and should do most of the talking during the interview (Fisher & Geiselman, 1992). Similarly, an instruction to "report everything" helps witnesses understand that they should report any detail they think of, even if it seems trivial. The "report everything" instruction is not meant to pressure or overwhelm the witness. Rather, its purpose is to emphasize that the witness should report anything that comes to mind during the interview and not censor themselves or make assumptions about what the interviewer is interested in (Paulo et al., 2013). Indeed, police officers trained in the CI acknowledge the report-all instruction as one of the most frequently used and most useful CI mnemonics (Kebbell et al., 1999).

Whereas the efficacy of the CI as a whole is well established (e.g., Memon et al., 2010), research on the effectiveness of specific mnemonics, like the reverse order instruction or the transfer of control instruction, is limited, and the findings are mixed. One study demonstrated the relative benefit of CI mnemonics—including change order and change perspective—compared to the control condition (Boon & Noon, 1994). However, in another study (Milne & Bull, 2002), CI mnemonics generally elicited fewer details when used in isolation compared to when used together. When tested together as part of a cohesive interview strategy, change order and change perspective elicited more details than a comparison, standard interview, but about the same amount of details as an interview using motivated free recall instead (Davis et al., 2005); mental reinstatement of context and a report-all instruction were most effective. It is important to determine which CI mnemonics are most effective because police investigators often use truncated or altered versions of the CI in real investigations (Dando et al., 2008). As such, a secondary goal of the current study was to add to the limited research on the utility of CI interview techniques in isolation.

Can the Cognitive Interview Benefit Non-Native Speakers?

Whether the CI's beneficial effect is present for non-native English speakers has never been tested. Witnesses who are non-native speakers face specific cognitive and linguistic challenges that may impede their responses to typical CI mnemonics.

First, speaking a second language is cognitively demanding (Ardila, 2003). Novice speakers often translate messages from one language to another using word association (e.g., "biblioteca" = library, "library" = place with books). As speakers gain expertise in a language, they are able to directly associate words

in their second language to concepts rather than words ("biblioteca" = a place with books), circumventing the additional cognitive step (Kroll & Curley, 1988). Memory output and/or accuracy may decrease if witnesses' attention is focused on complex tasks (e.g., reporting in reverse order; Vrij et al., 2012) rather than message formulation or metacognitive monitoring. For example, the trade-off hypothesis suggests that speakers' attention can be divided among the complexity, accuracy, and fluency of speech being produced (Skehan, 2009). As the demand for complexity increases—for example, in the case of asking witnesses to report in reverse order—attention available to monitor accuracy decreases[1]. This trade-off, compounded by the attentional demands of producing speech in a second language, may introduce inaccuracies in non-native speakers' reports. This is important to consider because misinformation presented in one language can affect memory traces encoded in another language (Shaw et al., 1997).

Second, monitoring the social situation also requires attentional resources. Although a speaker might be able to produce speech in their non-native language, they may feel uncomfortable doing so. Proficiency of the conversational partner, physical setting, goals, desire to communicate with a specific person, and self-confidence (in the moment) are situational factors that can motivate or demotivate a non-native speaker to communicate (Hatch, 1992; Macintyre et al., 1998). For example, when the police interviewer is a native speaker, non-native speaking witnesses may be passive and avoidant. As an additional example, a non-native speaker may be very confident communicating in familiar context, like an academic setting (e.g., a classroom) but less confident communicating in novel context, like a legal setting (e.g., a police station). Non-native speakers also report that they expect stigmatization and problems communicating because of their accents (Gluszek & Dovidio, 2010). As Gluszek and Dovidio suggest, these expectations may lead to a self-fulfilling prophecy in which non-native speakers expect to be stigmatized and misunderstood and therefore choose not to report as many details as they could (Russo et al., 2017). Thus, CI mnemonics like the report-all and transfer of control instructions may be empowering to non-native speakers. However, those mnemonics still need to be tested.

A limited body of research has examined investigative interviewing of non-native English speakers. Some of this literature focuses on interviewing witnesses, but most focuses on interrogation (i.e., interviewing suspects or sources of human intelligence). Nonetheless, as this research often includes experimental conditions in which the interviewee is innocent of wrongdoing (like a witness) and reports relevant outcome measures (e.g., number of details provided), it can inform the present study, which focused on witness interviewing. Some of these studies specifically examined the role of interpreters in suspect

1 Skehan (2009) defined accuracy in terms of vocabulary and grammatical choices. Our definition of accurate details in the current study—those that reflect verifiable details from the critical event—is not identical, but might be conceptualized as a speaker's ability choose correct vocabulary.

interviews. Those that include an interpreter-absent condition can also inform the current study. For example, compared to participants interrogated in their native language, participants interrogated in their non-native language provided fewer details overall (Ewens, Vrij, Leal, Mann, Jo, & Fisher, 2016). Similarly, model statements—detailed, yet unrelated, narratives provided by the interviewer to demonstrate the level of detail expected from the interviewee—increased the number of details given by native speakers but not non-native speakers (Ewens, Vrij, Leal, Mann, Jo, Shaboltas, et al., 2016).

As previously noted, some CI mnemonics—like the reverse order technique—may be cognitively demanding for interviewees. The reverse order technique requires witnesses to retrieve memories and then mentally manipulate the order of those reported memories (i.e., report them backward rather than chronologically). Because the witness must not only retrieve the memories of the event but also manipulate the order in which those memories are reported, such a technique requires more cognitive resources than simple free recall instructions. The effects of imposing a load upon cognitive resources (cognitive load) has been studied in other investigative interviewing contexts. For example, in the deception detection literature, cognitively overloaded liars are easier to identify than liars who can devote their full attention to the deception task (Vrij et al., 2012). Importantly, lack of detail is the most predictive cue to deceit (Vrij, 2019). This effect suggests the reverse order technique interferes with the ability to lie convincingly, perhaps by reducing the number of details the interviewee provides. Lying and speaking a non-native language are both cognitively demanding tasks for an interviewee. In the same way that cognitively overloading liars reduces their ability to deceive, cognitively overloading non-native speakers may reduce their ability to report details. As such, implementing the reverse order technique may overload non-native speakers and interfere with their ability to provide detailed reports.

To our knowledge, only Ewens, Vrij, Mann, and Leal (2016) have tested the reverse order technique specifically with non-native English speakers. Their participants were instructed to either lie or tell the truth, either in English or in their native languages with the assistance of an interpreter. Liars were instructed to provide false information about a video that they watched whereas truth-tellers were instructed to provide as much truthful detail about the video as possible. Additionally, all participants were instructed that they should convince the interviewer that they were telling the truth, adding an additional cognitive task on top of lying or recalling their memory. The video included dialogue (dubbed in the participant's native language) about a technical piece of spy equipment and plans to plant the device in a specific location. All participants were interviewed both in chronological order then in reverse order.

Although Ewens, Vrij, Mann, and Leal's (2016) design does not exactly map onto our research question, the truth-telling condition of their study is most similar to our current research question. Ewens, Vrij, Mann, and Leal (2016) did

not find statistically significant differences in the number of details reported—neither chronologically nor in reverse order—between language conditions (i.e., comparing reports in the participants' first language to reports in English. They also did not find statistically significant differences in the number of commission errors reported across language conditions. That is, participants made as many errors in their native language as they did in English.

Importantly, Ewens, Vrij, Mann, and Leal (2016) noted that they were not testing the effectiveness of the reverse order mnemonic per se. Rather, they were testing it as a tool to aid in lie detection. We aimed to directly test and compare the effectiveness of the reverse order technique among non-native speakers. Further, because Ewens, Vrij, Mann, and Leal's (2016) study primarily focused on lie detection, the video stimulus was quite different from what might normally be studied in eyewitness contexts. Specifically, their participants had to recall verbal conversational information from the video. However, witnesses and victims also often report on person characteristics (describing the perpetrator) or actions (what the perpetrator did).

In contrast to the potential cognitive overload that could result from using some CI components with non-native speakers, there are components of the CI that may be especially helpful with non-native speakers. Witnesses are frequently hesitant to communicate with interviewers for many reasons (e.g., describing personal, stressful events). Non-native speakers may feel the additional burden of not being able to communicate effectively on top of the typical discomfort. Interviewers can manipulate the social context to help non-native speakers overcome barriers to reporting. Specifically, instructing witnesses to report everything they can remember without concerning themselves with predicting what the interviewer is likely to find important and transferring control to the witness (e.g., by telling witnesses that they are the expert in the interview) may help witnesses feel more comfortable and empowered during the unfamiliar context of a police interview. Such instructions may help to increase witnesses' self-confidence in the moment (by taking on the role of the expert). Furthermore, if witnesses know that seemingly unimportant details are welcome and that they need not be concerned with taking too much time, they may feel more secure speaking a non-native language. As such, introductory instructions recommended by the CI may be especially beneficial for non-native speakers.

The Current Study

The current study tested CI mnemonics with native and non-native English speakers. The primary aim was to test whether these CI techniques are effective interviewing techniques for non-native speaking witnesses, while the secondary aim was to test the utility of the mnemonics in isolation. The techniques were (a) the reverse order instruction and (b) an introductory set of instructions including both a "report-all and transfer of control instruction (henceforth "CI

instructions").

Hypotheses

1. We expected both native speakers and non-native speakers to provide more details in response to the CI instructions compared to the control instructions (interview main effect).

2. We expected native speakers, but not non-native speakers, to provide more details to the reverse order mnemonic than a "try again" prompt (language × interview interaction).

3. We expected the accuracy rates of non-native speakers' reports to be lower than native speakers' (language main effect).

Method

Design

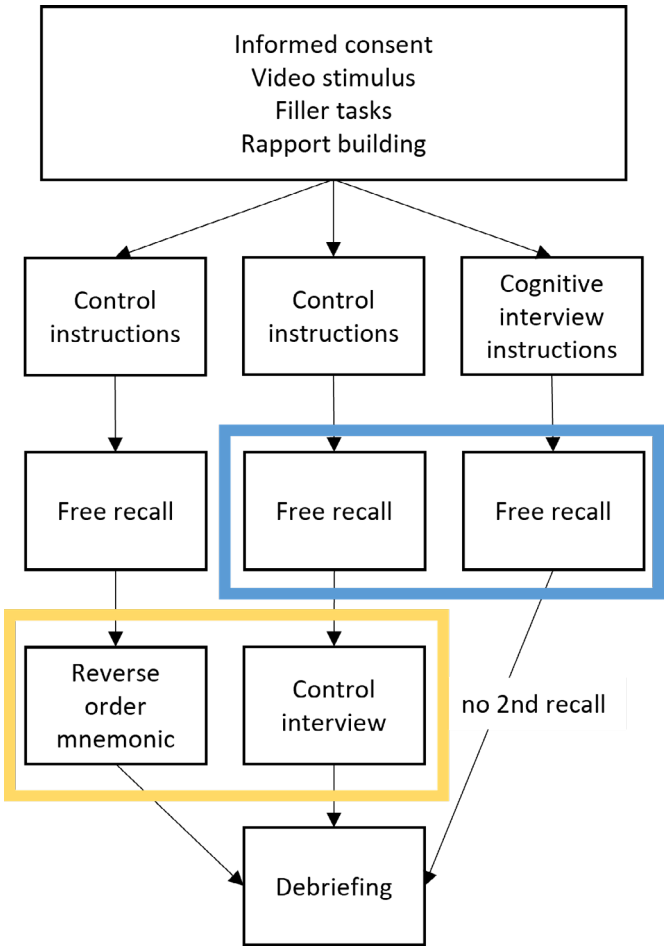
The study conformed to a 3 × 2 (interview technique [control interview, CI instructions, and reverse order] × English proficiency [native, non-native]) between-participants design, with participants randomly assigned to the interview technique condition via block randomization (see Figure 1).

Participants

Participants (N = 111) over the age of 18 (M = 23, SD = 6; 65% female, 35% male) were recruited. Of the sample, 63 (56.8%) were native speakers, and 48 (43.2%) were non-native speakers of English. Participants were recruited from the psychology department participant pool (via Sona Systems) of a large U.S. university, from the on-campus English Language Institute (ELI) via class announcements and flyers, and from the community via flyers. Participants recruited from Sona Systems were compensated with research credit in their psychology course. Participants recruited from the ELI and from the community after January 2019 were compensated with a $10 Starbucks gift card. Participants recruited from the ELI before January 2019 volunteered to participate for no compensation.

We conducted an a priori power analysis based on our preplanned 2 × 2 analyses of variance (ANOVA). The a priori power analysis indicated that 128 participants would yield 80% power to detect medium-sized effects in a 2 × 2 design. Thus, we planned to collect 32 participants per cell for a total sample size of 192 across all six cells. However, given the constraints faced when collecting data, we collected 111 total participants (as further discussed in the Limitations section). Thus, we conducted a sensitivity analysis for the final sample size in each set of analyses (Lakens, 2022). The sensitivity analysis revealed that the final sample of 74 participants for the reverse order analyses yielded 80% power to detect large effects of $\eta^2 > .10$ (Cohen's $d > 0.66$) and 60% power to detect medium effects of $\eta^2 > .06$ (Cohen's $d > 0.51$). The final sample of 70 participants in the CI

Figure 1. *Experimental design with critical comparisons highlighted. Blue represents the conditions we compared to test Hypothesis 1, effects of cognitive interview instructions. Yellow represents the condition we compared to test Hypothesis 2, effects of reverse order mnemonic.*



instructions analysis yielded 80% power to detect large effects of $\eta^2 > .10$ (Cohen's $d > 0.68$) and 60% power detect medium effects of $\eta^2 > .07$ (Cohen's $d > 0.54$).

Language Condition Classification

Participants from the psychology department participant pool were assumed to be native speakers unless they rated their own English proficiency as less than "very well" on the language history questionnaire (described below). That is, on a scale from 1 (*very poor*) to 7 (*native-like*), participants who reported that their English speaking ability was less than 6 (*very well*) (i.e., 5 [*well*] or

Table 1. *Participant Characteristics*

	Native (n = 54)	Non-native (n = 47)	t	d
English ability *M* (*SD*)				
Listening ability	6.82 (0.43)	5.45 (1.38)	6.91	1.38
Speaking ability	6.72 (0.53)	5.11 (1.56)	7.14	1.43
Reading ability	6.82 (0.52)	5.64 (1.21)	6.52	1.30
Writing ability	6.56 (0.88)	5.21 (1.38)	5.89	1.18
Age *M* (*SD*)	22.63 (5.81)	24.58 (7.09)		
Gender *n* (%)				
Female	37 (58.73%)	33 (68.75%)		
Male	25 (39.68%)	13 (27.08%)		
Not reported	1 (1.59%)	2 (4.17%)		
Education, n (%)				
High school	7 (11.11%)	6 (12.50%)		
College (BA/BS)	53 (84.13%)	30 (62.50%)		
Graduate (MA/MS)	1 (1.59%)	7 (14.58%)		
Doctoral	0 (0.00%)	2 (4.17%)		
Not reported	2 (3.18%)	3 (6.25%)		

Note. All t test degrees of freedom = 99, $p < .001$. d = Cohen's d. Degrees of freedom is less than total sample size because of incomplete LHQ data. English Ability scales range from 1 (*very poor*) to 7 (*native-like)*, with 5 indicating "*well*."

below) were categorized as non-native English speakers. Community member participants were categorized in the same way. Any student or community member participants with incomplete language data (e.g., questions regarding English language ability were skipped or not filled out for English) were grouped with native speakers. All participants recruited from the ELI were considered non-native English speakers, regardless of their self-reported proficiency.

Among participants classified as non-native speakers, the most commonly reported native languages were Spanish (n = 19, 40.4%), Russian (n = 5, 10.6%), and Portuguese (n = 3, 6.4%). Participant demographics across English proficiency classification can be found in Table 1. Participants provided self-ratings of their English speaking, reading, writing, and listening. Independent samples t-tests indicated that participants categorized as native speakers reported statistically significantly higher (ps < .001) ratings for all measures than participants categorized as non-native speakers (see Table 1).

Materials

Language History Questionnaire

Each participant completed a language history questionnaire (LHQ; Li et al., 2019). The LHQ included questions about the respondent's native language, context of first and second language use (e.g., at work, in the classroom, in casual

conversations), and cultural identity. The LHQ included demographic questions including age, sex, education level, country of origin, and country of residence. It also included self-report questions assessing context and self-rated proficiency of all languages spoken by the participant. For example, one question asked the age at which participants starting using each of their languages at home, at school, and at work. Another question asked participants to rate their ability to listen, speak, read, and write each of their languages. Other questions asked participants to estimate the number of hours per day they spend consuming media in different languages (e.g., television, radio, reading/writing for school or work) and the number of hours per day they spend speaking with family members, friends, classmates, and coworkers in different languages.

Video Stimulus

The to-be-remembered event was a video of a mock theft created by the first author. The video was 2 minutes, 10 seconds long. The video was shot so that the participant saw the action from a first-person perspective. In the video, the viewer travels through a university building and enters the psychology department of the building. In the psychology department, the viewer approaches an office space in which two people are stealing items. A male and female perpetrator take several items, including a laptop, some books, and a backpack from the office space. They leave the office space and exit the psychology hallway. The video included no spoken language and minimal language content (e.g., the word "Psychology" appears on the hallway, there is a sign on the office door). Figure 2 presents two frames from the video.

Informed Consent

Informed consent was provided in English, Spanish, or Portuguese, depending on participant preference. For participants who were native speakers of other

Figure 2. *Two frames from the mock crime video showing the "theft." Panel A shows the scene of the event. Panel B shows a closeup shot of the perpetrators and includes the sign on the office door.*



languages, the research assistant offered to reschedule the session to obtain appropriately translated consent documents, but all such participants indicated a preference for English.

Procedure

Other than the consent process, the entire study occurred in English for all participants. Participants were told that they would watch a video, and they were asked to pay attention to the video because the research assistant would ask questions about it later. Participants then viewed the mock theft video. Following the video, participants completed the LHQ in approximately 20 minutes. After completing the LHQ, the research assistant began the interview according to the randomly assigned interview condition. All interviews were conducted in English and were audio recorded in their entirety. All interviews began with a short rapport-building phase. To establish rapport, the research assistant asked questions to the participant (e.g., "Are you graduating this semester?" "Did you do the GRE yet?") and responded briefly to draw a connection with the participant (e.g., "Oh, I didn't know that programs in Europe don't require the GRE!"). Each experimental condition then diverged according to the assigned protocol. All interviewers were trained not to interrupt at any point.

Control Interview Condition

The control interview began with a set of control instructions. The participant was informed that they just saw a video of a crime being committed and that the research assistant would ask some questions about the crime. Then, participants were asked a free recall question ("Please tell me everything you can remember about what happened"). After the participant finished recalling, the research assistant asked the participant to engage in a second free recall ("Please tell me, again, everything that you remember about what happened"). Participants were asked to orally recount the instruction to confirm their understanding. Research assistants were trained to re-explain the instruction if participants incorrectly described the instruction or indicated misunderstanding. At the end of the second narrative, the research assistant confirmed that the participant had finished recalling by asking "Is there anything else you can tell me about what happened?" and then concluded the interview.

Reverse Order Interview Condition

The reverse order interview was identical to the control condition with one exception: For the second recall attempt, the research assistant asked the participant to recall the crime in reverse order, starting with the end of the crime and working backward to the beginning. ("Okay, now I want you tell me

everything that happened but in reverse order. [pause] So, start with the end of the crime video, and go backward to the beginning. For example, if you were telling me about everything you did this morning before you left your house, you would start with walking out your door and end with you waking up in bed. Ok, just to make sure you understand, can you tell me what you're supposed to do?").

Cognitive Instrument Instructions Interview Condition

The CI instruction interview began with the control instructions, which were followed the CI instructions. These CI instructions transferred control to the participant ("This is not like a TV interview. I'm not going to ask a lot of questions. Instead, I want you to do most of the talking. You saw the crime, so you know what happened. I don't know what happened, so I need you to tell me. I'm just going to take notes while you talk.") and encouraged the participant to report everything ("Every detail that you can remember is important. I want you to tell me everything you can possibly remember, even if it seems small"). Then, participants were asked the same free recall question as in the other conditions ("Please tell me everything you can remember about what happened"). After the participant finished recalling, the research assistant confirmed that the participant was finished recalling ("Is there anything else you can tell me about what happened?"), and then concluded the interview. There was no second recall phase in the CI instructions condition.

Following the interview, participants rated the perceived difficulty of the interview, which served as a manipulation check. Specifically, participants rated how hard it was to complete the interview in their assigned condition (i.e., "How hard was it to report your memory in reverse order?" or "How hard was it to report your memory a second time?" or "how hard was it to report everything you could remember?" on a scale where 1 = *not hard at all* and 10 = *very hard*). The research assistant subsequently debriefed and dismissed the participant.

Transcription and Scoring

The recording of each interview was transcribed, and transcripts were coded for two constructs: units of information and detail accuracy. Units of information were operationalized as the smallest statements made by participants that contained verifiable information and included nouns, adjectives, active verbs (i.e., not "was" or "had"), adverbs, and prepositions. One research assistant was trained on the unitization protocol and then unitized all the transcripts. To illustrate the coding scheme, the statement "she was holding it [a bag] by the side as if it was heavy" contains four units: (a) she, (b) was holding it, (c) by the side, (d) as if it was heavy.

The unitized transcripts were subsequently scored for uniqueness and accuracy. A master list of critical details was compiled from the video by two independent research assistants. The master list and the video itself were provided

to the scorers. Scorers first determined whether the reported detail was new or repeated. Repeated details were not scored further. Unique (nonrepeated) details were scored as accurate if they appeared on the master list or were verified by checking the video, or scored as inaccurate if they contradicted the details in the master list and/or contradicted a detail in the video. Details were scored as subjective if they reflected an opinion or assumption made by the participant (e.g., "they seemed freaked out"). Details were considered "not scoreable" if they did not the fit into the previous categories. That is, non-scoreable details referred to the video but could not be verified for accuracy and were not considered subjective. Using the example statement provided above ("she /was holding it [a bag] /by the side, /as if it was heavy"), Detail A is accurate (there was a female thief), Details B and C are inaccurate (she was not holding the bag, and what she was holding [books] was not by its side), and Detail D ("heavy") is subjective/an assumption.

Inter-Rater Reliability

Two independent research assistants were trained to score the unitized transcripts. Both scorers scored 10% of the transcripts to assess inter-rater reliability. Intraclass correlation coefficients (ICCs) for consistency between the two scorers were computed in SPSS using two-way mixed models. Reliability analyses revealed acceptable levels of consistency between scorers across all scores. In general, ICCs above .90 indicate excellent reliability, ICCs between .75 and .89 indicate good reliability, and ICCs between .50 and .74 indicate moderate reliability (Cicchetti, 1994; Koo & Li, 2016). In the current study, excellent reliability was achieved for accurate (ICC = .97) and inaccurate (ICC = .90) details. Reliability for subjective details was good (ICC = .75). After achieving acceptable reliability, each coder independently scored half of the remaining transcripts.

Results

Manipulation Checks

Participants in the reverse order condition reported that the interview was harder (M = 5.97, SD = 2.51) than participants in the control condition (M = 3.49, SD = 2.10), $t(68) = 4.54$, $p < .001$, Cohen's $d = 1.08$. The large, statistically significant difference in difficulty ratings between these interview groups suggests that the reverse order technique is indeed more cognitively challenging than a second recall attempt.

To test whether the interview manipulations were effective, we ran a 3 × 2 (interview condition [control, CI instructions, and reverse order] × language [native, non-native speaker]) ANOVA on the manipulation check question. There was a large, statistically significant two-way interaction, $F(2, 100) = 8.58$, $p <$

.001, partial $\eta^2 = 0.15$. For native speakers, tests of simple main effects revealed that the reverse order interview was more difficult than the control interview, $t(100) = 4.68$, $p < .001$. However, there was no statistically significant difference between the control and CI instructions interviews, $t(100) = 0.55$, $p = .581$. Non-native speakers did find the CI instructions interview more difficult than the control interview, $t(100) = 3.13$, $p < .001$. For non-native speakers, there was no statistically significant difference between the control and reverse order interviews, $t(100) = 1.06$, $p = .292$. This manipulation check suggests that the reverse order interview was cognitively demanding for native speakers, as expected but did not introduce additional cognitive demands among non-native speakers. Instead, non-native speakers may have experience increased cognitive demand from the CI instructions compared to control instructions.

Further, an independent-samples *t* test was conducted on participants' responses to the question, "How comfortable were you reporting your memory in English?" (1 = *very uncomfortable* and 10 = *very comfortable*). Welch's *t* was used to account for heterogeneity of variances between the two groups. Results indicated that non-native English speakers reported feeling statistically significantly less comfortable ($M = 6.58$, $SD = 2.46$) than native English speakers ($M = 9.18$, $SD = 2.01$), $t(78.06) = 5.73$, $p < .001$, Cohen's $d = 1.15$. This large, statistically significant difference provides evidence that participants did differ in terms of social experience between the two language conditions, as expected.

Unique Details

To test Hypotheses 1 and 2, analyses in the current section were performed with number of unique, verifiable details as the dependent variable. Specifically, we included accurate and inaccurate units reported by participants; subjective, unscoreable, and repeated details were excluded. Means and standard deviations are reported in text for statistically significant comparisons only. Complete descriptive statistics are reported in Tables 2 and 3.

Importantly, to test Hypotheses 1 and 2, we conducted two separate analyses. The difference in number of recall attempts across conditions makes it inappropriate to conduct a single 3 × 2 analysis. Instead, we conducted two 2 × 2 analyses for each hypothesis (see Figure 2). The 2 × 2 analyses described in the following sections isolate the specific recall attempts where we hypothesized differences for each interview technique.

Control versus Cognitive Interview Instructions Condition

The critical comparison to test hypothesis 1, regarding the impact of CI instructions, was the number of unique details reported to the first open-ended question asking participants to recall everything they could remember in the CI instructions condition versus the control condition. A 2 × 2 (language [native,

non-native speaker] × interview condition [control, CI instructions]) ANOVA revealed a large, statistically significant main effect of language, $F(1, 73) = 10.90$, $p = .001$, partial $\eta^2 = .13$. Specifically, native speakers ($M = 32.03$, $SE = 2.03$) provided statistically significantly more unique details than non-native speakers ($M = 21.79$, $SE = 2.35$) in response to the initial open-ended question. There was neither a statistically significant main effect of interview condition, $F(1, 73) = 0.71$, $p = .402$, partial $\eta^2 = .01$, nor a statistically significant interaction, $F(1, 73) = 0.39$, $p = .537$, partial $\eta^2 = .01$ (see Figure 3).

Control versus Reverse Order Condition

The critical comparison to test Hypothesis 2 (regarding the impact of reverse order recall instructions) was the number of unique details reported in the second recall attempt asking participants to try again (i.e., control condition) versus the number reported when asked to recall again in reverse order. A 2 × 2 (interview condition [control, reverse order] × language [native, non-native speaker]) ANOVA revealed medium, statistically significant main effects of both interview condition, $F(1, 69) = 5.19$, $p = .026$, partial $\eta^2 = .07$, and language proficiency, $F(1, 69) = 4.38$, $p = .040$, partial $\eta^2 = .06$. Participants in the control condition

Figure 3. *Total number of accurate and inaccurate details reported to the first interview question between the control and instructions conditions. Error bars represent 95% confidence intervals.*

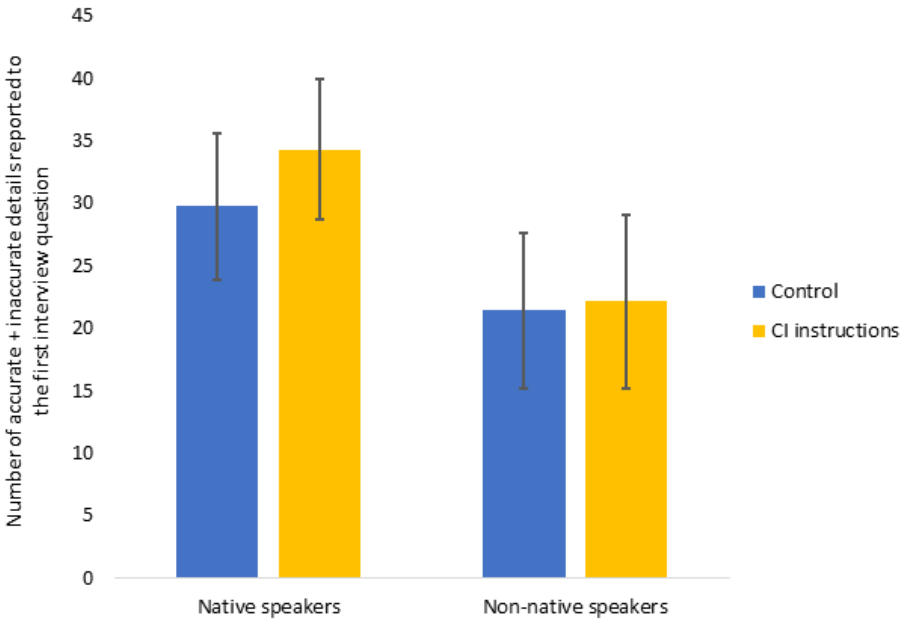
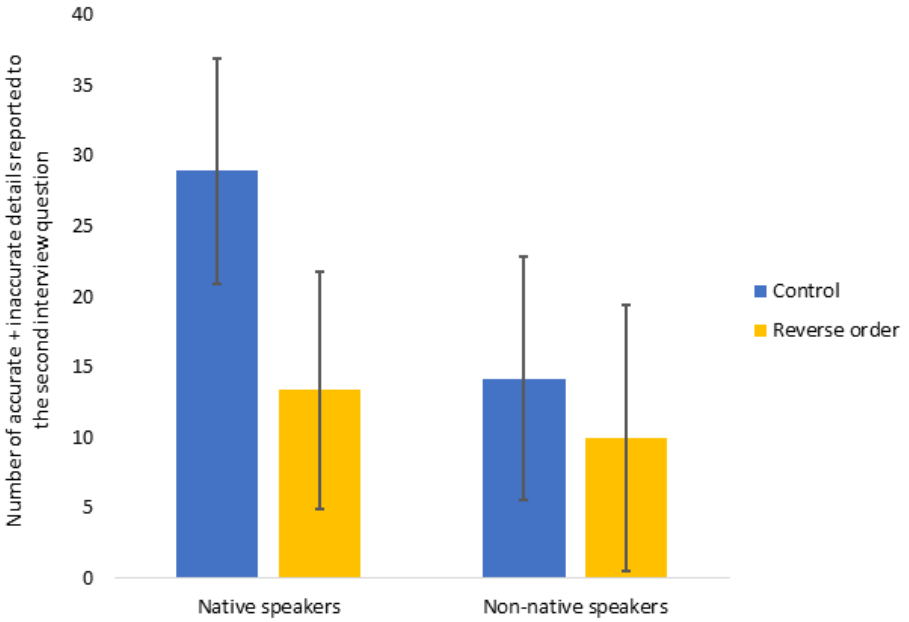


Figure 4. *Total number of accurate and inaccurate details reported to the second interview question between the control and reverse order conditions. Error bars represent 95% confidence intervals.*



(M = 21.54, SE = 2.96) provided statistically significantly more unique details than participants in the reverse order condition (M = 11.65, SE = 3.18) during their second recall attempt. Moreover, native speakers (M = 21.14, SE = 2.91) provided statistically significantly more unique details than non-native speakers (M = 12.05, SE = 3.22). However, inconsistent with our hypothesis, there was not a statistically significant interaction of interview condition and language, $F(1, 69) = 1.70$, $p = .197$, partial $\eta^2 = .02$ (see Figure 4).

New Information in Response to Repeated Recall Prompts

It is important to establish that asking follow-up questions during the interview did elicit new, unique details. Thus, we looked at the number of unique details provided during the first recall attempt and compared this to cumulative unique details provided after the second recall attempt (i.e., total unique details provided across both recall attempts). This allowed us to see if there was an increase in unique details after the second attempt, and if this varied as a function of our independent variables. As such, we ran a 2 × 2 × 2 (recall attempt [first recall only, first + second recall] × language [native, non-native] × interview [control, reverse order]) repeated-measures ANOVA, with cumulative unique details as

the outcome, recall attempt as the repeated-measures factor, and language and interview as the between-subjects factors.

The main effect of interview did not meet statistical significance, $F(1, 69) = 3.24$, $p = .076$, partial $\eta^2 = .05$. However, there was a large, statistically significant main effect of recall attempt, $F(1, 69) = 65.14$, $p < .001$, partial $\eta^2 = .49$, as well as a medium, statistically significant main effect of language proficiency, $F(1, 69) = 6.13$, $p = .016$, partial $\eta^2 = .08$ (see Table 4). These main effects were qualified by two interactions.

First, there was a medium, statistically significant interaction between language and recall attempt, $F(1, 69) = 4.62$, $p = .035$, partial $\eta^2 = .06$. Specifically, while there was a statistically significant increase in cumulative unique details after the second recall attempt for both language proficiency groups ($ps < .001$ after Bonferroni correction). The increase was greater for the native speakers (mean difference = 21.99) than for non-native speakers (mean difference = 12.74; see Table 2 and Figure 5).

Further, there was a medium, statistically significant interaction between interview condition and recall attempt, $F(1, 69) = 6.17$, $p = .015$, partial $\eta^2 = .08$. Specifically, while there was a statistically significant increase in cumulative unique details after the second recall attempt in both interview conditions ($ps \leq .002$ after Bonferroni correction), the increase in unique details was greater for

Figure 5. *Total number of unique, non-repeated details reported during the first recall attempt and cumulative number of unique, non-repeated details reported across both recall attempts. The increase in previously reported details was statistically significant for both language groups but was statistically significantly larger for native speakers than non-native speakers. Error bars represent 95% confidence intervals around the means.*

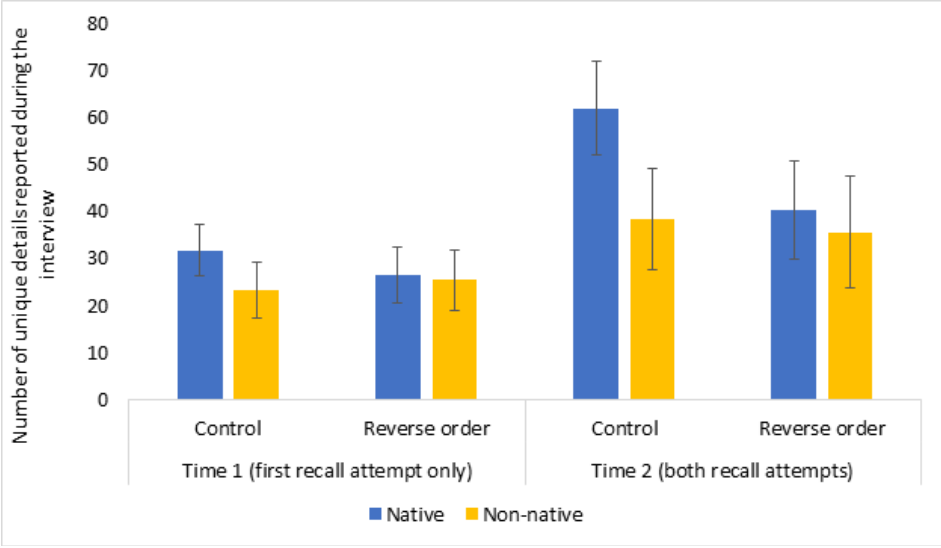


Table 2. *Descriptive Statistics for Analyses of New Information*

Cumulative unique details	Language	Interview	*M*	*SD*	95% CI	
					Lower	Upper
Time 1	Native English	Control	31.81	12.56	26.26	37.36
		Reverse order	26.53	11.68	20.69	32.36
		Across interviews	29.30	13.14	25.14	33.20
	Non-native	Control	23.22	13.14	17.23	29.22
		Reverse order	25.40	13.82	18.83	31.97
		Across interviews	24.21	13.29	19.86	28.76
	Across languages	Control	27.85	13.38	23.43	31.60
		Reverse order	26.03	12.48	21.57	30.36
Time 2	Native English	Control	61.95	30.77	51.92	71.99
		Reverse order	40.36	17.88	29.82	50.92
		Across interviews	51.70	27.42	43.88	58.44
	Non-native	Control	38.50	22.38	27.66	49.34
		Reverse order	35.60	15.72	23.73	47.47
		Across interviews	37.18	19.40	29.01	45.09
	Across languages	Control	51.13	29.37	42.84	57.61
		Reverse order	38.26	16.88	30.04	45.93

Note. Time 2 means represent total number of unique details provided during the interview cumulatively across Time 1 and Time 2.

the control condition (mean difference = 22.71) than for reverse order condition (mean difference = 12.02).

The interaction between language and interview was not statistically significant, $F(1, 69) = 2.91$, $p = .092$, partial $\eta^2 = .04$). Further, the three-way interaction was not statistically significant, $F(1, 69) = 1.70$, $p = .197$, partial $\eta^2 = .02$.

Information Quality

To test Hypothesis 3 (regarding the impact of language proficiency on accuracy rates), we assessed information quality using the previously described coding categories: accurate, inaccurate, and subjective. Specifically, to assess the accuracy rate and inaccuracy rate, we divided the number of accurate and inaccurate respectively by the number of unique details.

Control versus CI Instructions Condition

The critical outcome when testing the impact of the CI instructions condition was the accuracy rate for details reported to the first open-ended free recall question. A 2 × 2 (interview condition [control, CI instructions] × language [native, non-native speaker] ANOVA revealed neither statistically significant main effects of interview condition, $F(1, 73) = 0.44$, $p = .508$, partial $\eta^2 = .01$, nor language, $F(1, 73) = 0.14$, $p = .714$, partial $\eta^2 < .01$, and no statistically significant interaction, $F(1, 73) = 1.19$, $p = .278$, partial $\eta^2 = .02$ (see Table 3). Overall, accuracy was quite high for all groups.

Table 3. *Estimated Marginal Means for Proportion of Accurate Details in Response to the First Interview Question*

Interview	Language	*Mean*	*SE*	95% Confidence Interval	
				Lower	Upper
Control	Native	0.96	0.03	0.91	1.01
	Non-native	0.92	0.03	0.86	0.97
Instructions	Native	0.95	0.03	0.90	1.00
	Non-native	0.97	0.03	0.90	1.03

Control versus Reverse Order Condition

The critical outcome when testing the impact of the reverse order instruction was the accuracy rate of details reported in the second recall attempt. A 2 × 2 (interview condition [control, reverse order] × language [native, non-native speaker]) ANOVA revealed no statistically significant main effects of interview condition, $F(1, 64)^2 = 0.08$, $p = .773$, partial $\eta^2 < .01$, nor language, $F(1, 64) = 0.07$, $p = .792$, partial $\eta^2 < .01$, and no statistically significant interaction, $F(1, 64) = 0.45$, $p = .505$, partial $\eta^2 = .01$. Participants were equally accurate across language proficiency and interview conditions, and were generally quite accurate (see Table 4).

Discussion

We tested the utility of two forensic investigative interview mnemonics with an underrepresented population. Non-native speakers face cognitively taxing tasks when providing statements in their non-native language, and some CI mnemonics are cognitively demanding. Thus, non-native speakers may not have the cognitive bandwidth to engage in challenging CI mnemonics while also formulating statements in their second language.

Table 3. *Estimated Marginal Means for Proportion of Accurate Details in Response to the First Interview Question*

Interview	Language	*Mean*	*SE*	95% Confidence Interval	
				Lower	Upper
Control	Native	0.88	0.06	0.76	1.00
	Non-native	0.85	0.06	0.73	0.98
Reverse order	Native	0.82	0.06	0.69	0.95
	Non-native	0.88	0.07	0.74	1.02

2 Note that degrees of freedom differ from other analyses because three participants reported no accurate or inaccurate details at this stage (e.g., they reported only subjective details); thus, their accuracy rates were not able to be included in this analysis.

However, some CI components—like the introductory instructions encouraging witnesses to be active participants—were expected to be particularly helpful to non-native speakers. Although these mnemonics have been tested among participants speaking a native language, the current study is among the first to test best practices investigative interviewing strategies with witnesses speaking a non-native language.

Effects of Language

Overall, native speakers provided more information than non-native speakers. The accuracy of the information they provided did not differ between language groups (providing no support for Hypothesis 3). Importantly, our results provide no reason to distrust responses given by non-native speakers, as is often the case (Hanzlíková & Skarnitzl, 2017; Lev-Ari & Keysar, 2010). In legal settings, observers rate statements given by non-native speakers as low in credibility (Frumkin, 2007), and they frequently show a lie bias when judging statements by non-native speakers (e.g., Evans & Michael, 2014; Leach et al., 2017). One explanation for this effect is that accented speech is more effortful for listeners to process. Such an explanation suggests that exposure to specific accents should increase the listener's processing fluency, thereby increasing the listener's trust in the speaker (Boduch-Grabka & Lev-Ari, 2021).

That said, recent research suggests that processing fluency is not the only predictor of trust in accented speech—social judgments about the speaker may also influence trustworthiness. In one recent study, accent was not predictive of trustworthiness ratings but rather predicted social status judgments (Foucart et al., 2020). Perceived group membership—whether the speaker is from the listener's ingroup or not—and speaker confidence are also predictive of trustworthiness (Jiang et al., 2020). Even regional native accents are rated differently in terms of social status and trustworthiness, for example, with "lower class" British accents rated as less trustworthy than British Received Pronunciation (Frumkin & Stone, 2020; Frumkin & Thompson, 2020). Such social judgments are important to consider because non-native speakers may be seen as less credible than native speakers even when processing fluency is not an issue. For instance, in our sample, non-native speakers were reasonably proficient in English: They rated their proficiency in speaking English, on average, 5 out of 7, indicating they spoke English "well."

The results of our study may not generalize to witnesses who are very low proficiency, like recent immigrants or international travelers. However, the sample likely reflects the segment of non-native speakers most likely to be interviewed by police in English: those proficient enough to communicate in English at a basic level (e.g., Shaffer & Evans, 2018).

Utility of Individual CI Mnemonics

A secondary goal of the current study related to the utility of individual CI mnemonics absent the rest of the CI. Results indicated that a simple second recall prompt elicited more unique details than the reverse order mnemonic in both native and non-native speakers. Furthermore, this control procedure elicited more new, previously unreported details compared to the reverse order mnemonic. However, note that both the reverse order and repeated free recall mnemonics did elicit new, previously unreported details. Previous findings on individual CI mnemonics are mixed, and the current study adds to the mixture. In contrast to previous findings, we found a benefit of the control second-recall attempt relative to the reverse order second-recall attempt in native English speakers. Differences among previous studies might be attributed to variations in how the reverse (or change) order mnemonic is implemented. For example, Boon and Noon (1994) asked participants to provide a written statement, whereas other studies (including the current study) collected verbal reports. Writing and speaking rely on different cognitive procedures. For example, writing allows participants to reflect and revise before submission, whereas speaking does not.

Including introductory transfer of control instructions did not result in statistically significant differences in number of details reported, and this did not depend on whether the interviewee was a native or non-native English speaker. Furthermore, neither of the CI mnemonics produced significant differences in accuracy of reports between language conditions. The lack of differences suggests that—consistent with previous research on the CI (Memon et al., 2010)—CI mnemonics produce reports with similar accuracy rates as comparison interviews. Overall, we did not find support for the utility of the reverse order technique or transfer of control instructions when used isolation, regardless of language proficiency. That said, there was clear evidence of the benefit of asking for a statement multiple times, as a second report generated more unique details for both native and non-native speakers.

Limitations

Sample Characteristics

Native and non-native speakers were recruited from slightly different pools of people, possibly complicating comparisons between the two groups. Native speakers were recruited from the departmental participant pool, whereas non-native speakers were recruited from English programs, the community, and the participant pool. Given the international population at the university, several participants from the department participant pool identified as non-native English speakers. Thus, we grouped those participants with the non-native speakers recruited from other sources. Although this confounded language proficiency with

recruitment source, all participants were recruited from the same local area—for example, members of the same metropolitan community. As a result, non-native English speakers reported a larger range of English proficiency (*SD* of speaking ability = 1.46) than native English speakers (*SD* of speaking ability = 0.50).

Many (at least 35.6% [n = 23]) native English speakers were bilingual in English and another language, having indicated speaking two languages since in early childhood (e.g., from birth). Only two non-native speakers reported being bilingual (excluding English, i.e., they spoke their native language and another, non-English language). As some participants completed the LHQ incorrectly or incompletely, the estimate of 35.6% being fully bilingual is likely conservative. When speakers have two relatively complete sets of linguistic rules (i.e., a balanced bilingual, who speaks two languages equally well), they may construct speech differently than speakers with only one set of linguistic rules (Filipović, 2011).

Furthermore, educational attainment differed between the language groups. Whereas most native English speakers reported being enrolled in college (i.e., an undergraduate program), nearly 20% of non-native speaking participants reported being enrolled in or having completed graduate school. The variation in education could add noise to the analyses.

Finally, a sensitivity power analysis determined that we had appropriate power (80%) to detect large effects ($\eta^2 > .10$, Cohen's $d > 0.66$) and 60% power to detect medium effects ($\eta^2 > .06$, Cohen's $d > 0.51$). The statistically significant effects of interview type in the current study were medium ($\eta^2 = .05–.08$, Cohen's $d = .46–.59$). These effect sizes are smaller than others in the CI literature (Memon et al., 2010). However, there is a preponderance of research demonstrating the effectiveness of the CI. Informally, we believe there is a relatively strong prior probability that the CI is effective (Wegener et al., 2022). Thus, we are reasonably confident that the effects observed in the current study represent true effects of the CI. The current results are generally consistent with the larger narrative of CI research.

We found effects of neither interview type nor language on accuracy of reports. These null results may reflect either no true effects or type II errors. Previous research suggests that, compared to standard interviews, the CI may elicit small increases in inaccurate details, but those increases are outweighed by the large increases in accurate details (Memon et al., 2010). Thus, we would not expect the CI to have a main effect on accuracy rates. Indeed, we did not find main effects of the CI on accuracy in the current study. Although we expected language to affect accuracy rates, we found no such effect. We cannot conclude that this lack of effects is not a type II error, and we recommend that future research further address the question. That said, despite our limited power, we believe that our research is an important first step for future studies in this field.

Non-native speakers are a challenging population to recruit for several reasons, including limited enrollment in ELI, limited willingness to travel to the university, and inability to move the study off-campus without sacrificing experimental control (e.g., no distraction-free location to test participants).

Further, we collected data in 2018 and 2019, prior to the COVID-19 pandemic. Collecting additional data would have involved adapting the study (e.g., by using teleconferencing software or conducting the interviews in masks), which would create confounds and introduce new theoretical concerns (e.g., how masks affect communication, which is beyond the scope of the current study). However, there are very few studies including this underrepresented population in the interviewing literature. Thus, the current study represents an important early step in understanding the experience of non-native speakers, even if large samples are difficult to recruit.

Methodological Limitations

The current study employed a relatively strong control interview protocol. Specifically, research assistants were trained to ask a broad, open-ended free recall question and to not interrupt participants during their recall. Although the free recall question is absent other memory-enhancing mnemonics, free recall is perhaps the most commonly recommended "best practice" investigative interviewing technique. Furthermore, it is not necessarily consistent with real-world witness interviewing techniques. Investigators often interrupt their witnesses and use complex, multi-part questions (Schreiber Compo et al., 2012). Thus, the use of a high-quality control condition may have made it particularly difficult to find significant differences across interview conditions. As such, the results in the current study may underestimate the true effects of these interviewing mnemonics in ecologically valid scenarios. However, the use of such a control group was important for ensuring that the differences across conditions were well controlled, maximizing internal validity. Further, had low quality interview procedures been used in the control group, the design would not have been a strong test of the hypotheses.

Future Directions

The current study is a first step toward recruiting diverse populations in investigative interviewing research. Future research should specifically seek to recruit participants from ethnically, culturally, and linguistically diverse samples. Increasing generalizability on these dimensions is critical for making policy recommendations in a society that is rapidly diversifying.

A key takeaway from the current study is that non-native speakers reported fewer details than native speakers. Thus, future research should identify ways to increase the number of details reported by non-native speakers. For example, interviews scaffolded with cued recall questions—like "Tell me more about [previously mentioned detail]"—have been useful for minimally verbal adults (Bearman et al., 2019). Cued recall questions may help non-native speakers overcome limitations in vocabulary by providing the cue word within the question.

Further, it may help mitigate feelings of stress that non-native speakers might feel when responded to an open-ended free recall question in a second language. However, investigators should be cautious of introducing new information in cued recall questions: recent research indicates that witnesses speaking a second language might be particularly vulnerable to misleading suggestions from interviewers (Alm et al., 2019).

Future research on eyewitness memory among non-native speakers should also test other interviewing techniques, from the CI and otherwise, that facilitate communication. Sketching (Dando et al., 2008) or constructing a timeline (Hope et al., 2013) could be effective ways to decrease cognitive load, encourage nonverbal communication, and improve interview performance for non-native speaking witnesses. Social factors, like increasing confidence in a witness's language ability or decreasing state-level anxiety, may be critical components of building rapport with non-native speakers. Future research could identify the key predictors of detailed reporting among non-native speakers and design an interviewing strategy to maximize reporting.

In sum, it is important to recognize the challenges that non-native speakers face during investigative interviews as well as the challenges that interviewers face when a language barrier exists between the interviewer and witness/victim. The next step is to work toward developing an interview strategy that is sensitive to these challenges, so that non-native speakers can have equitable access, representation, and justice within the legal system.

Acknowledgments

We would like to acknowledge Melissa Baralt, Ronald Fisher, Deborah Goldfarb, and Nadja Schreiber Compo for support throughout this research. We would also like to acknowledge Alexander Perry and Jonathan Moreira for exceptional efforts in data collection.

Conflict of Interest Disclosure

The Authors have no conflicts of interest to declare.

Funding

This work was funded by the Psychology Department at Florida International University's graduate student seed fund and an American Psychology–Law Society Grant-in-Aid.

Research Ethics Statement

This study received institutional approval: Florida International University Institutional Review Board #IRB-18-0217-CR01.

References

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

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders, 36*(3), 233–240. https://doi.org/10.1016/S0021-9924(03)00022-4

Bearman, M., Brubacher, S. P., Timms, L., & Powell, M. (2019). Trial of three investigative interview techniques with minimally verbal adults reporting about occurrences of a staged repeated event. *Psychology, Public Policy, and Law, 25*(4), 239–252. https://doi.org/10.1037/law0000206

Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive Science, 45*(11). https://doi.org/10.1111/cogs.13064

Boon, J., & Noon, E. (1994). Changing perspectives in cognitive interviewing. *Psychology, Crime & Law, 1*(1), 59–69. https://doi.org/10.1080/10683169408411936

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Dando, C., Wilcock, R., & Milne, R. (2008). The cognitive interview: Inexperienced police officers' perceptions of their witness/victim interviewing practices. *Legal and Criminological Psychology, 13*(1), 59–70. https://doi.org/10.1348/135532506X162498

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*(1), 75–93. https://doi.org/10.1002/acp.1048

Evans, J. R., & Michael, S. W. (2014). Detecting deception in non-native English speakers. *Applied Cognitive Psychology, 28*(2), 226–237. https://doi.org/10.1002/acp.2990

Ewens, S., Vrij, A., Leal, S., Mann, S., Jo, E., & Fisher, R. P. (2016). The effect of interpreters on eliciting information, cues to deceit and rapport. *Legal and Criminological Psychology, 21*(2), 286–304. https://doi.org/10.1111/lcrp.12067

Ewens, S., Vrij, A., Leal, S., Mann, S., Jo, E., Shaboltas, A., Ivanova, M., Granskaya, J., & Houston, K. (2016). Using the model statement to elicit information and cues to deceit from native speakers, non-native speakers and those talking through an interpreter. *Applied Cognitive Psychology, 30*(6),

854–862. https://doi.org/10.1002/acp.3270

Ewens, S., Vrij, A., Mann, S., & Leal, S. (2016). Using the reverse order technique with non-native speakers or through an interpreter: Reverse order technique using foreign speakers or with an interpreter. *Applied Cognitive Psychology, 30*(2), 242–249. https://doi.org/10.1002/acp.3196

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. https://doi.org/10.1177/1367006911403062

Fisher, R. P., Compo, N. S., Rivard, J., & Hirn, D. (2014). Interviewing Witnesses. In T. Perfect & D. S. Lindsay, *The SAGE handbook of applied memory* (pp. 559–578). SAGE Publications Ltd. https://doi.org/10.4135/9781446294703.n31

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

Foucart, A., Costa, A., Morís-Fernández, L., & Hartsuiker, R. J. (2020). Foreignness or processing fluency? On understanding the negative bias toward foreign-accented speakers. *Language Learning, 7*0(4), 974–1016. https://doi.org/10.1111/lang.12413

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., & Stone, A. (2020). Not all eyewitnesses are equal: Accent status, race and age interact to influence evaluations of testimony. *Journal of Ethnicity in Criminal Justice, 18*(2), 123–145. https://doi.org/10.1080/15377938.2020.1727806

Frumkin, L., & Thompson, A. (2020). 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

Geiselman, R. E., & Callot, R. (1990). Reverse versus forward recall of script-based texts. *Applied Cognitive Psychology, 4*(2), 141–144. https://doi.org/10.1002/acp.2350040206

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology, 99*(3), 385. https://doi.org/10.2307/1422492

Gluszek, A., & Dovidio, J. F. (2010). Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology, 29*(2), 224–234. https://doi.org/10.1177/0261927X09359590

Gostjev, F. A., & Nielsen, A. L. (2017). Speaking the same language? English language fluency and violent crime at the neighborhood level. *The Sociological Quarterly, 58*(1), 111–139. https://doi.org/10.1080/00380253.2016.1246901

Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research in Language, 15*(3), 285–298. https://doi.org/10.1515/rela-2017-0016

Hatch, E. M. (1992). *Discourse and language education*. Cambridge University Press.

Hope, L., Mullis, R., & Gabbert, F. (2013). Who? What? When? Using a timeline technique to facilitate recall of a complex event. *Journal of Applied Research in Memory and Cognition, 2*(1), 20–24. https://doi.org/10.1016/j.jarmac.2013.01.002

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology, 73*(1), 55–79. https://doi.org/10.1177/1747021819865833

Kebbell, M. R., Milne, R., & Wagstaff, G. F. (1999). The cognitive interview: A survey of its forensic effectiveness. *Psychology, Crime & Law, 5*(1–2), 101–115. https://doi.org/10.1080/10683169908414996

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime & Law, 5*(1–2), 3–27. https://doi.org/10.1080/10683169908414991

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (Vol. 2, pp. 389-395). John Wiley & Sons.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), 33267. https://doi.org/10.1525/collabra.33267

Leach, A.-M., Snellings, R. L., & Gazaille, M. (2017). Observers' language proficiencies and the detection of non-native speakers' deception: Observer proficiency and deception detection. *Applied Cognitive Psychology, 31*(2), 247–257. https://doi.org/10.1002/acp.3322

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 1–7. https://doi.org/10.1017/S1366728918001153

Macintyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

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), 340–372. https://doi.org/10.1037/a0020518

Milne, R., & Bull, R. (2002). Back to basics: A componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology, 16*(7), 743–753. https://doi.org/10.1002/acp.825

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2013). The enhanced cognitive interview: Towards a better use and understanding of this procedure. *International Journal of Police Science & Management, 15*(3), 190–199. https://doi.org/10.1350/ijps.2013.15.3.311

Rivera, R., & Zraick, K. (2010, November 18). Audit says police fall short in providing interpreters. *New York Times*. https://www.nytimes.com/2010/11/19/nyregion/19translate.html

Russo, M., Islam, G., & Koyuncu, B. (2017). Non-native accents and stigma: How self-fulfilling prophesies can affect career outcomes. *Human Resource Management Review, 27*(3), 507–520. https://doi.org/10.1016/j.hrmr.2016.12.001

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

Shaffer, S. A., & Evans, J. R. (2018). Interpreters in law enforcement contexts: Practices and experiences according to investigators. *Applied Cognitive Psychology, 32*(2), 150–162. https://doi.org/10.1002/acp.3388

Shaw, J. S., Garcia, L. A., & Robles, B. E. (1997). Cross-language postevent misinformation effects in Spanish–English bilingual witnesses. *Journal of Applied Psychology, 82*(6), 889–899. https://doi.org/10.1037/0021-9010.82.6.889

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Southall, A. (2017, May 25). Police must tackle language barrier in domestic abuse cases. *The New York Times*. https://www.nytimes.com/2017/05/25/nyregion/police-tackle-language-barrier-in-domestic-abuse-cases.html

Turkewitz, J. (2014, May 11). Language barrier continues to thwart victims of crimes. *The New York Times*. https://www.nytimes.com/2014/05/12/nyregion/language-barrier-continues-to-thwart-victims-of-crimes.html

Vrij, A. (2019). Deception and truth detection when analyzing nonverbal and verbal cues. *Applied Cognitive Psychology, 33*(2), 160–167. https://doi.org/10.1002/acp.3457

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law, 18*(6), 579–594. https://doi.org/10.1080/1068316X.2010.515987

Wakefield, S. J., Kebbell, M. R., Moston, S., & Westera, N. (2015). Perceptions

and profiles of interviews with interpreters: A police survey. *Australian & New Zealand Journal of Criminology, 48*(1), 53–72. https://doi.org/10.1177/0004865814524583

Wegener, D. T., Fabrigar, L. R., Pek, J., & Hoisington-Shaw, K. (2022). Evaluating research in personality and social psychology: Considerations of statistical power and concerns about false findings. *Personality and Social Psychology Bulletin, 48*(7), 1105–1117. https://doi.org/10.1177/01461672211030811

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior, 44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Zeigler, K., & Camarota, S. A. (2019, October 29). 67.3 million in the United States spoke a foreign language at home in 2018. *Center for Immigration Studies*. https://cis.org/Report/673-Million-United-States-Spoke-Foreign-Language-Home-2018