Received: 6 May 2019 | Revised: 14 October 2019 | Accepted: 23 October 2019

DOI: 10.1002/acp.3615

RESEARCH ARTICLE

WILEY

Why don't you believe me? Detecting deception in messages written by nonnative and native speakers

Sarah Volz[1] | Marc-André Reinhard[1] | Patrick Müller[2]

[1] Department of Psychology, University of Kassel, Kassel, Germany

[2] Faculty of Civil Engineering, Building Physics, and Business, University of Applied Sciences Stuttgart, Stuttgart, Germany

Correspondence
Sarah Volz, Department of Psychology, University of Kassel, Holländische Straße 36-38, 34127 Kassel, Germany.
Email: sarah.volz@uni-kassel.de

Summary

Detecting lies is crucial in numerous contexts, including situations in which individuals do not interact in their native language. Previous research suggests that individuals are perceived as less credible when they communicate in a nonnative compared with native language. The current study was the first to test this effect in truthful and fabricated messages written by native and nonnative English speakers. One hundred native English speakers judged the veracity of these messages, and overall, they proved less likely to believe and to correctly classify nonnative speakers' messages; differences in verbal cues between native and nonnative speakers' messages partly explained the differences in the judgments. Given the increased use of nonnative languages in a globalized world, the discrimination against nonnative speakers in veracity judgments is problematic. Further research should more thoroughly investigate the role of verbal cues in written and spoken nonnative language to enable the development of effective interventions.

KEYWORDS

credibility, deception, language proficiency, lie detection, nonnative speakers, verbal cues

1 | INTRODUCTION

The ability to detect lies is essential in daily interactions and in high-stakes scenarios (see, e.g., Ekman, 2009; Vrij, 2008). For instance, professionals in the forensic context (e.g., police officers) must detect suspects' lies to solve criminal cases, whereas recruiters try to spot applicants' lies to ensure that the candidate truly possesses the claimed traits and experiences. With the world becoming more interconnected on economic, cultural, political, and social levels (see, e.g., Jackson, 2008), spoken and written communication increasingly takes place in languages other than native languages. Therefore, detecting lies of nonnative speakers became increasingly relevant (e.g., at border controls or when nonnative speakers apply for jobs) and has thus also emerged as a topic of interest for science. Researchers investigated biases and accuracy rates when the veracity of spoken messages by nonnative and native speakers was judged (e.g., Cheng & Broadhurst, 2005; Da Silva & Leach, 2013; Evans & Michael, 2014). Given that international communication also takes place in written form (e.g., social media, international email communication, and international job applications), our study closes a gap in research by examining for the first time veracity judgments about messages written by nonnative speakers.

To date, empirical evidence suggests that spoken messages by nonnative speakers are perceived as less credible compared with those by native speakers, implying higher mistrust towards nonnative speakers. In some studies, the lower credibility manifested in a lie bias towards nonnative speakers (e.g., Castillo, Tyson, & Mallard, 2014; Da Silva & Leach, 2013; Leach, Snellings, & Gazaille, 2017); in others, there was a truth bias for native speakers, which was not present for nonnative speakers (e.g., Elliott & Leach, 2016; Leach & Da Silva, 2013). Still, in other studies, lower credibility manifested in a smaller truth bias for

nonnative than for native speakers (e.g., Akehurst, Arnhold, Figueiredo, Turtle, & Leach, 2018; Evans & Michael, 2014). Despite the strong empirical evidence for a lower credibility of nonnative speakers, what underlying mechanisms elicit this perception of low credibility remain unclear.

Studies testing the influence of senders' language proficiency on judges' ability to accurately discern between lies and truths yielded mixed results. In some studies, judges were less accurate in their veracity judgments when evaluating nonnative senders' than when evaluating native senders' messages (e.g., Akehurst et al., 2018; Da Silva & Leach, 2013; Leach & Da Silva, 2013). Other studies found the opposite; that is, accuracy rates were higher for evaluations of nonnative compared with native speakers' messages (Evans, Michael, Meissner, & Brandon, 2013; Evans, Pimentel, Pena, & Michael, 2017). Beyond that, certain studies reported accuracy rates that did not significantly differ between native and nonnative senders' messages (e.g., Castillo et al., 2014; Cheng & Broadhurst, 2005).

Various explanations for the effects of nonnative language use on judges' accuracy rates and biases have been put forward in the literature. For instance, Leach et al. (2017) tried to identify the source of the lower perceived credibility by comparing native and nonnative speakers' judgments about messages by native and nonnative senders. As their judgments did not differ, Leach and her colleagues concluded that neither familiarity with nonnative speech (see improving effect of familiarity on lie detection accuracy, Reinhard, Sporer, & Scharmach, 2013; Reinhard, Sporer, Scharmach, & Marksteiner, 2011) nor an ingroup bias (i.e., judges trusting their own ingroup more) could explain previous findings. They suggested that processing fluency (e.g., through nonnative accents, Lev-Ari & Keysar, 2010) could explain the lower credibility of nonnative senders because fluency equally affects native and nonnative judges and is connected to low perceived credibility (Reber & Schwarz, 1999; Unkelbach, 2007).

Effects of nonnative language use on senders' display of emotions are also often used as explanation (e.g., Elliott & Leach, 2016; Evans et al., 2017; Evans & Michael, 2014) because laypersons commonly believe lies can be detected based on senders' display of emotions (e.g., The Global Deception Research Team, 2006). Nonnative language use is thought to affect senders in two different ways: On the one hand, nonnative language use has an emotionally distancing effect (Bond & Lai, 1986; Dewaele, 2008; Harris, Ayçíçeğí, & Gleason, 2003; Keysar, Hayakawa, & An, 2012), and liars feel their lies less strongly (Caldwell-Harris & Ayçiçeği-Dinn, 2009). On the other hand, lying in a nonnative language was found to increase senders' stress levels (Caldwell-Harris & Ayçiçeği-Dinn, 2009). Thus, nonnative language use could affect veracity judgments when judges rely on senders' emotional display to detect deception.

Another prominent explanation is based on the idea that lying and speaking in a nonnative language increase cognitive load (e.g., Ardila, 2003; Vrij, 2015; Vrij, Fisher, Mann, & Leal, 2008; Vrij, Granhag, Mann, & Leal, 2011). When senders complete a cognitively taxing task in addition to lying, they often display more signs of cognitive load (see cognitive load approach to detect deception, e.g., Vrij, 2015; Vrij et al., 2008), signs that are thought to influence veracity judgments. This explanation will be discussed in more detail later.

Additional factors that might play a role are based on senders' race, accent, and the activation of stereotypes (e.g., see Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012; Gluszek & Dovidio, 2010; Ruby & Brigham, 1996; Vrij & Winkel, 1992), as well as on culture-specific social norms (e.g., Castillo & Mallard, 2011). Judges might infer deception if senders do not behave according to the norms. Potential mismatches of senders' and judges' social norms can result in expectancy violations and in a bias towards nonnative speakers who were usually not raised in the same culture as the judges.

So far, most studies could merely speculate which factors are responsible for the gaps between judges' biases and accuracy rates for native and nonnative speakers. Most studies used videos as stimulus material for the veracity judgments. Thus, nonverbal and visual (e.g., race), verbal (e.g., signs of cognitive load), and paraverbal (e.g., accent) cues were simultaneously available, allowing no final conclusions regarding underlying theoretical processes (see, e.g., Sporer & Schwandt, 2006, for a definition of the different kinds of cues).

A study by Akehurst et al. (2018) attempted to discern between different explanations by studying the effects of isolated components of nonnative language speech. They modified videotaped messages into "visual and audio," "audio only," and "visual only," and presented these modified messages along with interview transcripts to judges. Judges exhibited a truth bias across all modifications for native but not for nonnative senders, suggesting that no single speech component was responsible for the lower credibility of nonnative speakers. For accuracy, paraverbal indicators were most influential. Only when paraverbal cues were included (audio only, visual, and audio) were judges less accurate for nonnative than native senders. When paraverbal indicators were not available (visual only and transcripts), accuracy rates for native and nonnative senders' messages did not differ.

We followed an approach similar to Akehurst and colleagues but investigated for the first time written communication by nonnative and native speakers (senders typed their messages). By cutting off audio and visual cues, we set the focus on verbal cues, depriving the judges of stereotypical information (e.g., race and accent) about the sender. Verbal cues (e.g., plausibility, detailedness, and consistency) are related to beliefs about deception (e.g., Reinhard, Burghardt, Sporer, & Bursch, 2002; Strömwall & Granhag, 2003; The Global Deception Research Team, 2006; Ulatowska, 2017; Zuckerman, Koestner, & Driver, 1981) as well as to actual deception (see DePaulo et al., 2003, for a meta-analysis). Thus, we assumed that if the verbal cues of truth and deception differ between native and nonnative speakers' messages, they should at least account in part for differences in judges' bias and accuracy rates regarding native and nonnative senders' written communication.

Note that the cognitive processes involved in producing and processing written language differ from those involved in producing and processing spoken language. For instance, when senders write (or type) a message, they can take planning time and edit their message, whereas planning and editing are limited when senders deliver their message verbally. Thus, the findings of our study cannot be generalized to spoken language without further investigation. Nevertheless, this study provides an important step into researching written

nonnative language in the deception detection context as well as into more systematically investigating why nonnative speakers are perceived differently in veracity judgments.

1.1 | Language proficiency and limited ability to express oneself

Using a nonnative language when delivering a message can impact the quality and quantity of cues that are typically taken as indicators of a message's veracity. For instance, individuals tend to include fewer idea units (i.e., the smallest information unit in a narrative) when recalling a memory about an event in a language they did not use when they experienced the event (Javier, Barroso, & Muñoz, 1993; Marian & Neisser, 2000). This can occur, for example, when a job interview was conducted in the applicant's nonnative language and then the applicant must recall job experiences he or she had instead in a native language. Fewer idea units in the recalls likely reduce the vividness and detailedness of truthful messages. This might lower one's credibility, as vividness and detailedness are commonly seen as indicators of a message's truthfulness (Hansen & Wänke, 2010; Reinhard et al., 2002; Strömwall & Granhag, 2003).

Nonnative speakers' vocabulary limitations (e.g., Karlsen, Lyster, & Lervåg, 2017; Simos, Sideridis, Mouzaki, Chatzidaki, & Tzevelekou, 2014; Szabo, 2016; Wolter, 2001) can also contribute to the lack of idea units, in addition to implausibility and incoherence of messages. Nonnative speakers sometimes lack the words needed to express themselves; as a result, they might use incorrect words or leave out information they cannot communicate in their nonnative language. Incomprehensible or missing information might elicit impressions of implausibility and incoherence, signs that are commonly taken as indicators for deception (Akehurst, Köhnken, Vrij, & Bull, 1996; Reinhard et al., 2002; Ulatowska, 2017).

Even if one has memorized the word needed, words often do not come to mind immediately. These so-called tip-of-the-tongue states are relatively frequently experienced by nonnative speakers (Gollan & Acenas, 2004). When individuals choose an alternative word that comes to their mind faster but might be less suitable, the sentence can sound odd or the alternative word can even obscure the meaning. When nonnative speakers are under time constraints and take the time to try to remember the word, their messages might be shorter than those of native speakers. Such differences in message length can affect veracity judgments, for instance, as message length was found to be a valid cue to deception (Hauch, Blandón-Gitlin, Masip, & Sporer, 2015).

1.2 | Efficiency of text production and cognitive load

Differences in message length might not only stem from tip-of-the-tongue states; in fact, other factors further influence the amount of text native and nonnative liars and truth-tellers produce. For instance, nonnative speakers were found to take more time to make lexical decisions (Ransdell & Fischler, 1987). Thus, the increased thinking time likely prevents them from producing as much text as native speakers would within the same time. In general, nonnative speakers seem to produce fewer words in their writing than do native speakers and have a tendency to revise their texts more (see Silva, 1993), again reducing the amount of text produced in a limited time frame.

Message length likely also differs between lying and truth-telling senders in written (typed) communication. Liars, in an effort to appear credible, are thought to be more deliberate and controlled in their communication than are truth-tellers (e.g., DePaulo et al., 2003; Zuckerman, DePaulo, & Rosenthal, 1981). We assume that liars' higher levels of monitoring and deliberateness lead to more thinking time and text editing. Supporting this assumption, Derrick, Meservy, Jenkins, Burgoon, and Nunamaker (2013) found that liars used the backspace key more often than truth-tellers when typing messages. Potentially resulting from this increased editing, deceptive typed messages were shorter than truthful typed messages in the study by Derrick et al., as well as in various other studies (see Hauch et al., 2015, for a meta-analysis).

In addition, lying and using a nonnative language likely impact message length due to the high demands on cognitive resources. Individuals usually choose from several alternatives the task they perceive as least cognitively demanding. High cognitive effort is considered aversive; thus, individuals usually try to avoid it (e.g., Dunn, Lutes, & Risko, 2016; Kool, McGuire, Rosen, & Botvinick, 2010). Liars and nonnative speakers in particular should engage in effort-reducing strategies because their cognitive resources are more taxed than those of native truth-tellers (see, e.g., Ardila, 2003; Vrij et al., 2011; Vrij, Fisher, et al., 2008). Senders could, for example, reduce their effort by writing less as well as by working more slowly, both resulting in shorter messages under time constraints.

Further, high cognitive demands do affect not only the quantity but also the quality of messages. Liars must make up their lies, ensure the consistency within their stories, and invent details to make their story sound plausible and credible (Buller & Burgoon, 1996; Zuckerman, DePaulo et al., 1981). Meta-analytic findings revealed that liars do not fully succeed in this invention process as indicated, for example, by lies containing fewer details than truths (DePaulo et al., 2003). Hauch et al. (2015) found that lies were less elaborate (lower word variety) and less complex (fewer exclusive words such as except or without) than were truths, likely resulting from the high cognitive demands of lying.

Like the high cognitive demands of lying, the demands of speaking in a nonnative language (Ardila, 2003) potentially also influence the quality of a message along with factors such as vocabulary limitations (see Silva, 1993, for an overview of differences between native and nonnative speakers' writing). Messages by nonnative speakers were found to feature more signs of cognitive load than those by native speakers and were also perceived as less credible (Evans et al., 2013; Evans & Michael, 2014). As signs of cognitive load are commonly assumed to be indicators of deception (e.g., low detailedness, Ulatowska, 2017), concentrated occurrence of such signs in messages by nonnative senders seems to affect credibility.

1.3 | The present research

The first aim of our study was to research the influence of senders' nonnative language use on veracity judgments about written

communication. Therefore, instead of videotaping senders, we asked them to type their messages in order to test whether previous findings could be replicated using written messages. The second aim was to examine how differences in verbal cues between truthful and deceptive messages by native and nonnative speakers would relate to (a) whether judges believe a message and (b) whether judges evaluate it correctly. To our knowledge, all previous studies compared the characteristics of messages and the judgments made about those messages on a group level, lacking a direct link between message characteristics and judgments. This direct link should, however, advance our understanding of why nonnative senders are perceived differently compared with native senders.

On the basis of the theoretical reasoning that messages by native speakers are more in line with laypersons' idea of a truthful message (see, e.g., Akehurst et al., 1996; Reinhard et al., 2002; The Global Deception Research Team, 2006; Ulatowska, 2017), we hypothesized that messages written by native speakers would be more likely to be rated as true than those by nonnative speakers.

If judges base their evaluations on verbal cues, native speakers' messages should overall be more likely to be evaluated correctly compared with nonnative speakers' messages (H2). We expected a congruency between what laypersons believe truthful versus deceptive messages look like and what truthful versus deceptive messages by native senders actually look like (e.g., high vs. low plausibility, high vs. low detailedness, and long vs. short messages). In contrast, for nonnative speakers, we expected an incongruency between what laypersons believe truthful versus deceptive messages look like and what truthful versus deceptive messages by nonnative senders actually look like. Thus, we predicted a significant interaction between message veracity and senders' language proficiency (H3): When evaluating native speakers, truthful messages should be more likely to be correctly judged than deceptive messages; this effect should not be present for nonnative speakers' messages (H4).

Note that the logic regarding the (in)congruency of laypersons' beliefs and actual characteristics of messages works only if valid cues are coded in the studies. We chose to code detailedness and plausibility as verbal cues because, on the one hand, they are valid cues to detect deception in native speakers (DePaulo et al., 2003), and on the other hand, they are commonly taken as indicators of deception (e.g., Akehurst et al., 1996; Reinhard et al., 2002; Ulatowska, 2017). Thus, variations in detailedness and plausibility likely affect both accuracy and bias. In addition, message length has been selected because it might serve as a strong visual cue regarding typed messages, and message length has proved a valid cue to discern truthful and deceptive typed messages (Hauch et al., 2015).

On the basis of higher cognitive load of nonnative speakers and liars, increased thinking time, and editing behavior, we expected native truth-tellers' messages to be longer than the messages in the other three conditions (H5). Because native truth-tellers do not have the (cognitively) demanding tasks of lying or speaking in a nonnative language, their messages should also be more detailed (H6) and more plausible (H7) than messages by nonnative speakers and lies by native speakers.

We hypothesized that the verbal cues would mediate the effect of message status (i.e., nonnative vs. native sender and deceptive vs. truthful message) on the probability that a message is judged as true (H8) as well as on the probability that it is evaluated correctly (H9).

2 | METHOD

2.1 | Participants and design

One hundred native English speakers (68 males, 30 females, one other, and one not specified) with the mean age of 35.97 years (*SD* = 11.14) participated as judges in the online study. Judges were recruited via the online crowdsourcing platform Prolific (www.prolific.ac) and received £1.25 for their participation. Prolific is a pool of registered individuals, mostly from the United Kingdom, other European countries, and the United States, who participate in online studies for remuneration. Prescreening filters can be employed to recruit a particular sample. For our online study, filters ensured judges self-identified as native English speakers. Judges were from the United Kingdom (85 judges), the United States (13), Australia (one), Ireland (one), Turkey (one), and one unspecified country (one).

The study employed a 2 (Sender Language Proficiency: native vs. nonnative speaker) × 2 (Message Veracity: lie vs. truth) within-subjects design. Thus, each judge rated deceptive and truthful messages by native and nonnative senders.

2.2 | Stimulus material

An online study was employed to create the stimulus material with a 2 (English Language Proficiency: native vs. nonnative speaker) × 2 (Message Veracity: lie vs. truth) between-subjects design. Two hundred fifty-one participants who self-identified as native or nonnative English speakers (hereafter referred to as senders) were randomly assigned to either lie or tell the truth about a previous job. Senders were asked to list their native language(s) and were categorized as native speaker if English was listed. The 79 female and 172 male senders were on average 33.96 years old (*SD* = 9.18). They were recruited for a study on letters of motivation using Prolific and Amazon Mechanical Turk (MTurk) and received a reimbursement of £1.20 on Prolific or $1.20 on MTurk. Prescreening ensured senders had work experience.

Informed consent was collected before senders listed general demographic data, their native language(s), and the language(s) they speak at home. Nonnative English speakers filled in questions based on the Language History Questionnaire (Li, Zhang, Tsai, & Puls, 2014). For instance, they indicated at which age they started learning English in terms of listening, speaking, reading, and writing. They stated how many years they studied English in a classroom setting and how many years they had been actively using English. Further, they rated their language abilities in terms of listening, speaking, reading, and writing on 7-point scales ranging from 1 (*very poor*) to 7 (*native-like*).

To be able to verify the truthfulness of their messages, senders had to indicate job titles throughout their lives and business areas of those jobs. Next, senders were randomly assigned to either select a previously listed job (truth condition) or to invent a job they had not reported before (lie condition). Senders wrote a text in which they had to answer questions regarding the selected job: (a) "For what company did you work?" (b) "How long did you work there and which area did you work in?" (c) "What tasks did you have?" (d) "What was the most valuable thing you learned during the time on the job?" (e) "What did you like about your job?" and (f) "What did you not like about your job?" Text production was limited to 5 min 30 s after which the text was automatically submitted. Senders could not move on to the next page before the time was over.

Senders' *motivation to appear convincing* ("I was motivated to write a convincing text."), *perceived message credibility* ("When someone reads my text from the previous page, they would believe that I really held the described job."), *general perceived credibility in relation to language proficiency* ("Because of my language abilities, I feel that people often do not believe me when I communicate in English in everyday life."), and *cognitive load* were measured on 7-point scales ranging from 1 (*certainly not*) to 7 (*certainly*). Cognitive load was assessed with the three items adapted from Vrij, Mann, Leal, and Fisher (2010); (a) "Writing the text required a lot of thinking (cognitive effort)."; (b) "Writing the text was mentally difficult."; and (c) "While writing the text I had to concentrate a lot." (Cronbach's α = .87). Finally, senders were debriefed and asked whether their texts could be used in future research. Senders were thanked and directed back to MTurk or Prolific for remuneration.

Prior to analyzing the messages, we excluded messages of senders who did not consent to the further use of their messages (n = 10). Messages of senders who were not able to type their message due to a temporary bug in the questionnaire software were also excluded, as were very short messages (n = 39) that would not allow well-grounded veracity judgments. At least three questions had to be answered including at least two of the content producing questions (c) to (f). Messages of senders who misunderstood the task (e.g., writing about general working experiences without reference to a specific job) were also removed (n = 36). Some senders, despite being in the lie condition, wrote about a position they had listed as a job previously held. Even if these senders lied by reporting fabricated tasks within the selected job, the truthfulness of their messages remains unclear. Thus, these messages (n = 46) were excluded.

Out of the remaining messages (nonnative lie 23, nonnative truth 30, native lie 33, and native truth 34), we randomly selected 80 messages to create five sets, each incorporating 16 messages (four of each condition). Senders were featured only once. The size of the sets was based on a pilot study and previous research (see meta-analysis, Bond & DePaulo, 2006). Five sets were created, as the nonnative lie condition contained too few messages for more sets. Information in the messages indicating a sender's identity (such as names of companies) was replaced to ensure anonymity.

The senders who wrote the selected messages (53 females and 27 males) were on average 33.93 years old (SD = 8.84). Thirty nine of the native speakers currently lived in the United States, and one resided in Spain. The 40 nonnative senders were from India (17), Italy (six), Greece (four), Netherlands and Portugal (3 each), Germany (two), and Finland, Hungary, Israel, Spain, and Sweden (one each).

Nonnative senders on average learned written English at the age of 9.38 (SD = 3.76, range: 4–25 years). They rated their English writing skills on average as 5.65 (SD = 1.03, range: 3–7), thus as ranging between good and very good on a 7-point scale from 1 (*very poor*) to 7 (*native-like*). They stated even better English reading skills (M = 6.08, SD = 1.05, range: 3–7); thus, we assumed that senders were able to understand the instructions in the study.

One trained rater, Rater A, coded the detailedness and subsequently the plausibility of all selected messages on 7-point semantic differential-type scales (see Burgoon, Buller, Ebesu, White, & Rockwell, 1996; Stiff & Miller, 1986) ranging from 1 (*little detailed*) to 7 (*very detailed*), and from 1 (*implausible*) to 7 (*plausible*). A second trained independent rater, Rater B, coded the detailedness of a random sample of 20 messages. The two raters discussed their ratings for these 20 messages, potential explanations for disagreements, and solutions to resolve disagreements. Subsequently, Rater A revisited the other 60 ratings that were already done, and Rater B rated the detailedness of the remaining 60 messages. The detailedness ratings of Rater A and Rater B were averaged per message and taken for the subsequent analyses. The same procedure was followed for plausibility. The ICC2 was .843 for detailedness and .766 for plausibility, indicating a high agreement between the raters. Message length was operationalized as the number of words of a message.

2.3 | Procedure

Judges gave informed consent and learned that they had to make veracity judgments about 16 written accounts of job experiences. They were informed that some accounts were true accounts of someone's job experiences whereas other accounts contained fabricated jobs and experiences. Judges were randomly assigned to one of the five sets of stimulus material, and the 16 messages were presented in random order. Each message contained the respective job description in the heading and was accompanied by the veracity judgment question as well as the confidence measure. Judges were asked whether or not they thought that the author of the message had indeed held the stated job and had had the described experiences. The judges' confidence was measured by asking how confident they were that their judgment was correct (11-point scale ranging from 1 [*0% pure guess*] to 11 [*100% absolutely sure*] in steps of 10%).

In the end, judges stated general demographic data, were thanked and debriefed, and were directed back to Prolific for remuneration.

3 | RESULTS

We used multilevel models to analyze the data because, unlike general regression models, they do not assume independence, so they can be estimated even though judgments depend on the particular judge and

message. In addition, multilevel models can prevent Type I error inflation and allow generalizations of the results across both judge and sender samples (Judd, Westfall, & Kenny, 2012; Watkins & Martire, 2015).

To run the analyses on the judgment level, we created a dataset with one row of data for each veracity judgment. The 1,600 rows contained information on the type (0 = rated as lie, 1 = rated as truth) and correctness (0 = incorrect, 1 = correct) of each judgment. Because these variables were binary, we calculated logistic multilevel models using the glmer function from the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). Predicted likelihoods (back-transformed from the log odds), including 95% confidence intervals, for the experimental conditions are reported to facilitate the interpretation of the results.

3.1 | Likelihood of messages being rated as true

We estimated a logistic multilevel model with senders' language proficiency as fixed effect (effect coded: −0.5 = native, 0.5 = nonnative) and random intercepts for judges and senders to predict the likelihood of a message being rated as true. Proficiency was a significant predictor, $\beta = -0.85$, $SE = 0.24$, $z = -3.49$, $p < .001$. Thus, senders' language proficiency predicted the probability of messages being rated as true after controlling for the variance associated with the judges and senders. As hypothesized (H1), messages by native senders (M = 74.38%, CI [67.24%, 80.42%]) were more likely to be rated as true compared with those by nonnative senders (M = 55.40%, CI [46.98%, 63.51%]). Pseudo-R^2 for the fixed effect (calculated using the procedure suggested by Nakagawa & Schielzeth, 2013) was .04.[1]

3.2 | Likelihood of messages being judged correctly

A logistic multilevel model was estimated to predict the likelihood of a message being judged correctly from the fixed effects of message veracity (effect coded: −0.5 = lie, 0.5 = truth), senders' language proficiency (effect coded: −0.5 = native, 0.5 = nonnative), and their interaction. Random intercepts for judges and senders were added.

As hypothesized (H2), proficiency, $\beta = -0.50$, $SE = 0.23$, $z = -2.14$, $p = .032$, significantly predicted the likelihood that a message was judged correctly. Native speakers' messages (M = 59.48%, CI [51.29%, 67.15%]) were overall more likely to be evaluated correctly compared with those by nonnative speakers (M = 47.13%, CI [39.21%, 55.18%]). Veracity was also significant, $\beta = 1.28$, $SE = 0.23$, $z = 5.47$, $p < .001$, with a higher likelihood of truths (M = 68.48%, CI [60.92%, 75.18%]) compared with lies (M = 37.59%, CI [30.32%, 45.44%]) to be judged correctly.

In line with Hypothesis 3, the interaction was significant, $\beta = -1.71$, $SE = 0.47$, $z = -3.65$, $p < .001$. As predicted (H4), pairwise comparisons revealed that native speakers' truths (M = 80.96%, CI [72.59%, 87.34%]) were more likely to be evaluated correctly than their lies (M = 33.46%, CI [24.18%, 44.35%]), $z = 6.35$, $p < .001$. For nonnative speakers, truths (M = 52.31%, CI [41.12%, 63.53%]) and lies (M = 41.76%, CI [31.37%, 53.10%]) did not differ in the likelihood of being evaluated correctly (see Figure 1), $z = 1.32$, $p = .187$. Pseudo-R^2 for the fixed effects was .14.[2]

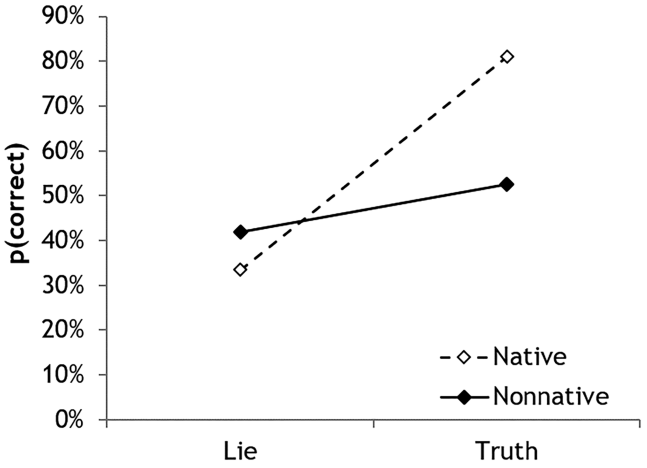

FIGURE 1 Estimated likelihood that truthful and deceptive messages by native and nonnative senders were judged correctly after controlling for the variance associated with the judges and senders

3.3 | Mediation analysis truth-judgment model

For the mediation analyses, we applied a multilevel approach as suggested by Krull and MacKinnon (2001) to avoid data aggregation while directly linking a message's characteristics to the judgments made about it. Following the recommendation of Hayes (2018) for correlated mediators, we simultaneously added the verbal cues as parallel mediators instead of calculating one model per verbal cue. The conceptual mediation model for the likelihood that a message was rated as true (truth-judgment model), and its statistical model are depicted in Figure 2. Within the terminology of Krull and MacKinnon (2001), the model constitutes a 2–2–1 mediation model (predictors and mediators are level 2 sender variables and the outcome is a level 1 judgment variable).

Following the procedure suggested by Hayes (2018), the mediation analysis was calculated in two steps. In Step 1, we predicted the mediators (verbal cues) from the independent variables (proficiency and veracity). A standard (single-level) linear regression model was calculated for each verbal cue because the predictors and the outcome variables were measured on the sender level. In Step 2, we estimated a

[1]Male and female judges showed similar tendencies to rate messages by native and nonnative speakers as true. When judges' gender (effect coded: −0.5 = male, 0.5 = female) was added in the model as fixed effect, gender, $\beta = -0.05$, $SE = 0.14$, $z = -0.37$, $p = .714$, and the interaction with senders' language proficiency, $\beta = -0.39$, $SE = 0.25$, $z = -1.58$, $p = .114$, were nonsignificant.

[2]Male and female judges were similarly likely to evaluate messages by native and nonnative speakers correctly. Entering judges' gender in the model yielded a nonsignificant effect of gender, $\beta = -0.15$, $SE = 0.14$, $z = -1.04$, $p = .300$. The interactions with veracity, $\beta = -0.12$, $SE = 0.25$, $z = -0.48$, $p = .630$, with senders' language proficiency, $\beta = 0.03$, $SE = 0.25$, $z = 0.13$, $p = .895$, and the three-way interaction, $\beta = -0.76$, $SE = 0.51$, $z = -1.50$, $p = .134$, were also not significant.

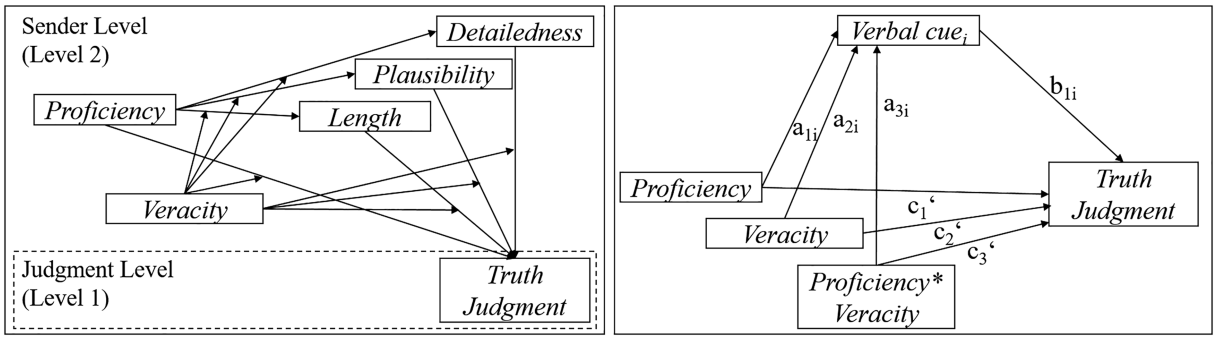

FIGURE 2 Conceptual and statistical diagrams of the mediation model to estimate the effect of senders' language proficiency and message veracity through verbal cues (parallel mediators) on the likelihood that a message was rated as true (truth-judgment model)

multilevel model to predict the likelihood that a message was rated as true on the basis of a sender's language proficiency and message veracity controlling for the mediators (verbal cues).

> *Step* 1. We tested whether message veracity (effect coded: −0.5 = lie, 0.5 = truth) and senders' language proficiency (effect coded: −0.5 = native, 0.5 = nonnative) impacted the verbal cues of the messages (see also Table 1 for descriptive statistics). Thus, we wanted to establish whether the verbal cues were diagnostic for detecting deception and whether they could mediate the effect of language proficiency and message veracity on the lie detection outcomes.

Table 2 presents the results of the linear regression models for plausibility, length, and detailedness of the messages. Language proficiency significantly predicted message length and detailedness, both $|t|$s > 3.87, both ps < .001, but not plausibility, $t = -1.71$, $p = .092$. Veracity did not significantly predict any of the three cues, all ts < 1.56, all ps > .125. The interaction between proficiency and veracity was significant for plausibility and detailedness, both $|t|$s > 2.32, ps < .023, but not for length, $t = -1.78$, $p = .079$.

Despite the nonsignificant interaction for length, we followed up on it because the strong main effect of language proficiency likely decreased the power to detect the hypothesized ordinal interaction (see, e.g., Bobko, 1986; Strube & Bobko, 1989). As predicted (H5), pairwise comparisons revealed that truthful messages by native speakers were significantly longer than the messages in the other three conditions, all $t(76)$s > 2.14, all ps < .036. In addition, native senders' lies were longer than nonnative senders' truths, $t(76) = 2.05$, $p = .044$, whereas all other comparisons were not significant. In line with Hypotheses 6 and 7, native senders' truths were more detailed and more plausible than the messages in the other three conditions, all$|t(76)|$s > 2.30, ps < .025. All other pairwise comparisons of detailedness and plausibility were not significant, indicating that the messages in the other three conditions were similarly detailed and plausible.

> *Step* 2. To predict the likelihood that a message was rated as true, we estimated a multilevel model from the fixed effects of senders' language proficiency, veracity, and their interaction, with and without verbal cues. Random intercepts for judges and senders were included.

Without the verbal cues (see Model 1 in Table 3), senders' language proficiency, $\beta = -0.86$, $SE = 0.23$, $z = -3.64$, $p < .001$, and the interaction were significant, $\beta = -1.00$, $SE = 0.47$, $z = -2.13$, $p = .033$, whereas veracity was not, $\beta = 0.27$, $SE = 0.23$, $z = 1.15$, $p = .252$. When the verbal cues were included (see Model 2 in Table 3), language proficiency and the interaction were no longer significant, both $|z|$s < 1.79, all ps > .073, whereas plausibility, $\beta = 0.50$, $SE = 0.11$, $z = 4.49$, $p < .001$, and length, $\beta = 0.30$, $SE = 0.15$, $z = 2.08$, $p = .037$, were significant. The positive coefficients indicated that messages were more likely to be rated as true the longer or the more plausible they were. Detailedness was not significant, $\beta = 0.11$, $SE = 0.16$, $z = 0.67$, $p = .500$, suggesting that detailedness was no mediator of the effect of language proficiency and veracity on the likelihood that a message was rated as true when controlling for plausibility and length. Thus, the results support Hypothesis 8 for plausibility and message length as mediators, but not for detailedness.

TABLE 1 Means (standard deviations) of plausibility, message length, and detailedness for the four message conditions

	Native senders		Nonnative senders	
Measure	Lie	Truth	Lie	Truth
Plausibility	4.73 (1.43)	5.98 (1.06)	4.93 (1.43)	4.68 (1.77)
Message length	143.50 (69.80)	176.60 (46.28)	117.70 (34.17)	111.80 (37.40)
Detailedness	4.47 (1.51)	5.70 (1.51)	4.03 (1.29)	3.68 (1.61)

Note. Plausibility and detailedness were measured on a 7-point scale (1 to 7). Higher values represent higher degrees of the variable. $N = 80$ messages.

TABLE 2 Prediction of message length, plausibility, and detailedness (all standardized) based on message veracity, senders' language proficiency, and their interaction

		Plausibility		Message length		Detailedness	
Path (see Figures 2 and 3)		***β* (*SE*)**	**t**	***β* (*SE*)**	**t**	***β* (*SE*)**	**t**
	(Intercept)	0.00 (0.11)	0	0.00 (0.10)	0	0.00 (0.10)	0
a_1	Proficiency	−0.36 (0.21)	−1.71	−0.83** (0.20)	−4.14	−0.78** (0.20)	−3.88
a_2	Veracity	0.33 (0.21)	1.55	0.25 (0.20)	1.24	0.27 (0.20)	1.37
a_3	Proficiency × Veracity	−0.99* (0.43)	−2.32	−0.72 (0.40)	−1.78	−0.99* (0.40)	−2.47
		R^2 = .12 F(3, 76) = 3.57 p = .018		R^2 = .22 F(3, 76) = 7.29 p < .001		R^2 = .23 F(3, 76) = 7.67 p < .001	

Note. Message veracity (−0.5 = lie, 0.5 = truth) and senders' language proficiency (−0.5 = native, 0.5 = nonnative) effect coded. Plausibility, detailedness, and length standardized. N = 80 messages.

*p < .05. **p < .001.

TABLE 3 Results of the multilevel model to predict the likelihood that a message was rated as true on the basis of a message's veracity and its sender's language proficiency with and without the inclusion of the verbal cues (mediators)

	Model 1			Model 2	
Predictor	***β* (*SE*)**	***z***	**Path (see Figure 2)**	***β* (*SE*)**	***z***
(Intercept)	0.64** (0.12)	5.31		0.64** (0.09)	7.51
Proficiency	−0.86** (0.23)	−3.64	c_1'	−0.31 (0.17)	−1.78
Veracity	0.27 (0.23)	1.15	c_2'	−0.03 (0.16)	−0.18
Proficiency × Veracity	−1.00* (0.47)	−2.13	c_3'	−0.11 (0.33)	−0.32
Plausibility	—		b_{1p}	0.50** (0.11)	4.49
Length	—		b_{1l}	0.30* (0.15)	2.08
Detailedness	—		b_{1d}	0.11 (0.16)	0.67

Note. Message veracity (−0.5 = lie, 0.5 = truth) and senders' language proficiency (−0.5 = native, 0.5 = nonnative) effect-coded. Plausibility, detailedness, and length standardized. N = 1,600 judgments.

*p < .05. **p < .001.

3.4 | Mediation analysis correctness model

Similar as for the likelihood that a message was rated as true, we calculated a two-step mediation analysis to probe whether verbal cues mediated the effect of senders' language proficiency and message veracity on the likelihood that the respective message was judged correctly (correctness model). Step 1 was identical to Step 1 in the truth-judgment model. Step 2 in the correctness model differed from Step 2 in the truth-judgment model regarding the inclusion of the verbal cues. Given that judges take verbal cues (e.g., detailedness) as indicators of truth (e.g., Ulatowska, 2017), a more detailed message should be more likely to be rated as true. Thus, verbal cues were included as normal predictors in the truth-judgment model. In the correctness model, however, veracity was included as moderator of the verbal cues because detailed messages are only more likely to be judged correctly when they are truthful. Thus, the interaction term of verbal cues and veracity was necessary to predict the likelihood of a message being judged correctly from verbal cues. Figure 3 specifies the conceptual and statistical diagram of the correctness model.

Step 2. We estimated a multilevel model with the fixed effects of message veracity, senders' language proficiency, their interaction, and the interactions of each verbal cue with veracity. Random intercepts for judges and senders were added (see Model 2 in Table 4). Model 1 in Table 4 shows the same model without the inclusion of the mediators and equals the model reported in the analysis for Hypothesis 2 to 4.

Despite the inclusion of the mediators, veracity remained a significant predictor, β = 1.26, SE = 0.16, z = 7.84, p < .001, indicating that truths were more likely to be judged correctly than lies. The main effect of language proficiency, β = −0.06, SE = 0.18, z = −0.34, p = .730, and the interaction between veracity and proficiency was no longer significant, β = −0.57, SE = 0.36, z = −1.59, p = .111. Instead, the interactions between veracity and plausibility, β = 1.07, SE = 0.22, z = 4.76, p < .001, as well as between veracity and length, β = 0.75, SE = 0.31, z = 2.44, p = .015, were significant predictors. Similar to the truth-judgment model, the interaction term with detailedness did

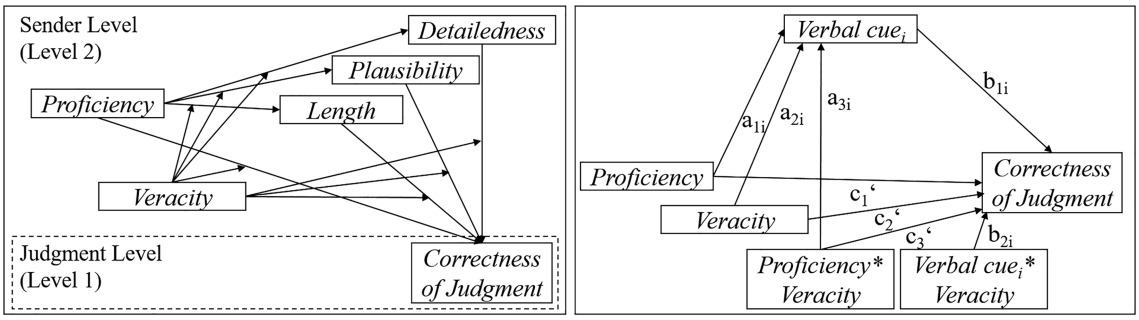

FIGURE 3 Conceptual and statistical diagrams of the mediation model to estimate the effect of senders' language proficiency and message veracity through verbal cues (parallel mediators) on the likelihood that a message was judged correctly (correctness model)

TABLE 4 Results of the multilevel model to predict the likelihood that a message was judged correctly on the basis of a message's veracity (ver) and its sender's language proficiency (prof) with and without the inclusion of the verbal cues (mediators)

	Model 1			Model 2	
Predictor	***β* (*SE*)**	**z**	**Path (see Figure 3)**	***β* (*SE*)**	**z**
(Intercept)	0.13 (0.12)	1.12		−0.01 (0.08)	−0.14
Proficiency (Prof)	−0.50* (0.23)	−2.14	c_1'	−0.06 (0.18)	−0.35
Veracity (Ver)	1.28*** (0.23)	5.47	c_2'	1.26*** (0.16)	7.84
Prof × Ver	−1.71*** (0.47)	−3.65	c_3'	−0.57 (0.36)	−1.59
Plausibility	—		b_{1p}	0.16 (0.11)	1.43
Length	—		b_{1l}	0.16 (0.16)	0.28
Detailedness	—		b_{1d}	−0.26 (0.16)	−1.56
Plausibility × Ver	—		b_{2p}	1.07*** (0.22)	4.76
Length × Ver	—		b_{2l}	0.76** (0.31)	2.44
Detailedness × Ver	—		b_{2d}	0.07 (0.33)	0.23

Note. Veracity (−0.5 = lie, 0.5 = truth) and senders' language proficiency (−0.5 = native, 0.5 = nonnative) effect coded. Plausibility, detailedness, and length standardized. n = 1,600 judgments.

*p < .05. **p < .01. ***p < .001.

not reach the level of significance, β = 0.07, SE = 0.33, z = 0.23, p = .820. Thus, the results support Hypothesis 9 for plausibility and message length as mediators, but not for detailedness. The model is plotted in Figure 4 for plausibility and length.

3.5 | Additional analyses

Using the same procedure as for the verbal cues, we tested further potential explanations for differences in the veracity judgments made about native and nonnative speakers' messages. We ran the same multilevel mediation models as for the verbal cues for senders' motivation to appear convincing, their cognitive load, their perceptions of their messages' credibility, and their general credibility in relation to language proficiency. We found differences between lying and truth-telling native and nonnative senders on some of these variables, but none mediated the relationship between senders' language proficiency, message veracity, and the likelihood that a message was evaluated correctly or rated as true.

We also investigated the relation between senders' language proficiency, message veracity, and senders' self-reported cognitive load, as well as the relation of cognitive load and verbal cues. The reported cognitive load of native-speaking truth-tellers was lower than the load of senders in the other three conditions, whereas the levels of cognitive load were similar in these three conditions. However, the correlations between senders' self-reported cognitive load and verbal cues were not significant, all $|r|$s < .13, all ps > .252, indicating that there was no direct relation between senders' perception of their own cognitive load and the verbal cues of their messages.

To facilitate comparisons with previous studies, Table 5 provides the signal detection measures d' (discrimination ability) and C (bias), calculated following the procedure by Stanislaw and Todorov (1999), as well as the raw overall accuracy scores for native and nonnative senders' messages. Accuracy, discrimination ability, and the tendency to rate messages as true were higher for native than for nonnative senders' messages, all $t(99)$s > 4.10, all ps < .001. Judges were able to discriminate between lies and truths by native speakers as indicated

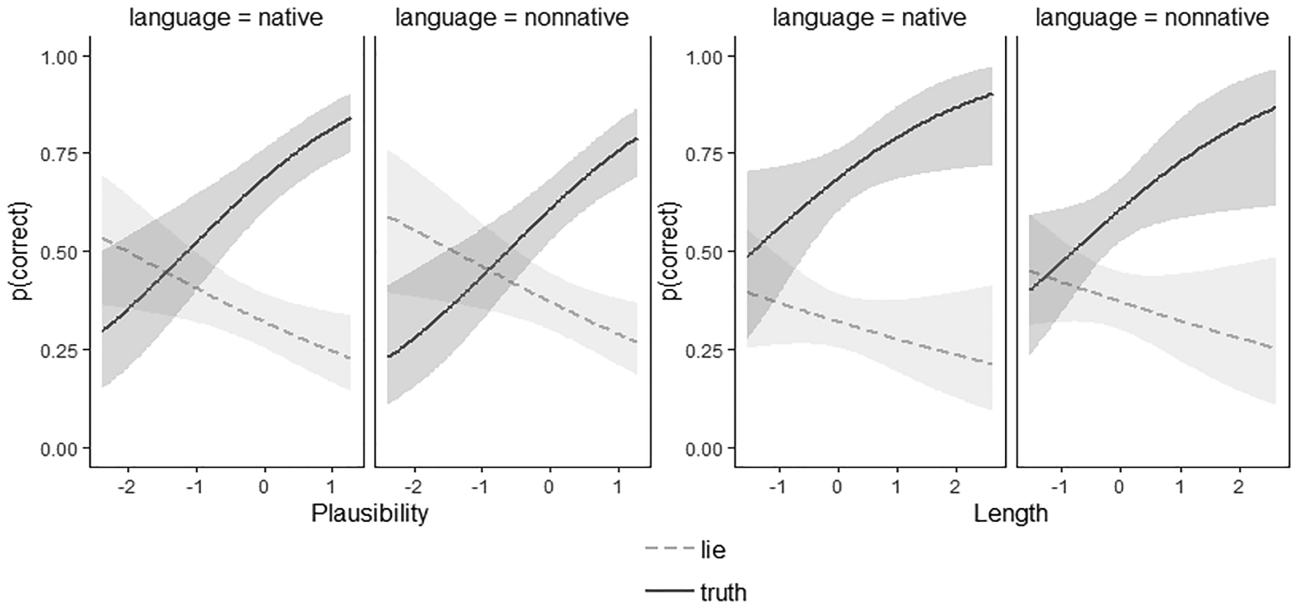

FIGURE 4 The likelihood that a message was judged correctly as a function of its veracity, its sender's language proficiency, and the respective verbal cues plotted for the mediators of plausibility and length (fixed effects with 95% confidence intervals)

TABLE 5 Comparisons of mean overall accuracy rates, discrimination ability, and response bias for messages by native and nonnative senders

Predictor	Native *M* (*SD*)	Nonnative *M* (*SD*)	*t*(99)
Overall accuracy	57.00% (16.70%)	47.75% (17.80%)	4.10***
Discrimination (d')	0.34 (0.82)	−0.12 (0.88)	4.13***
Response bias (*C*)	0.50 (0.43)	0.13 (0.46)	5.66***

***$p < .001$.

by d' for native speakers' messages being significantly higher than 0 (i.e., no sensitivity), $t(99) = 4.12$, $p < .001$, which was not the case for messages by nonnative speakers, $t(99) = -1.33$, $p = .187$.[3]

4 | DISCUSSION

This work contributes to research on the perception of nonnative speakers in veracity judgments by focusing on written instead of the previously investigated spoken communication. Using written messages, thereby eliminating paraverbal and nonverbal cues, we investigated differences in judges' evaluations of native and nonnative speakers' truthful and deceptive messages. We further tested whether verbal cues of those messages could explain why native and nonnative speakers were evaluated differently.

We found similar results regarding nonnative speakers' credibility as studies employing videotaped messages (e.g., Akehurst et al., 2018; Cheng & Broadhurst, 2005; Da Silva & Leach, 2013; Evans et al., 2017; Evans & Michael, 2014; Leach & Da Silva, 2013). As hypothesized, nonnative speakers were perceived as less credible than were native speakers. To investigate why their messages were evaluated differently, we estimated a multilevel mediation model with the verbal cues as parallel mediators. Plausibility and length mediated the effect of senders' language proficiency and message veracity on the likelihood that a message was rated as true while controlling for the respective message's detailedness.

Our findings align with those of Akehurst et al. (2018) who found that nonnative compared with native speakers' messages were less often believed to be true when the veracity judgments were based only on verbal cues of spoken language (transcripts of videotaped messages). Thus, nonnative speakers seem to face similar disadvantages regarding the credibility of their spoken and written communication because verbal cues in both cases seem to elicit lower ratings of nonnative speakers' credibility.

As outlined above, one rationale behind using written messages was to reduce the impact of stereotypes (e.g., regarding nonnative speakers' race or accented speech; see Fuertes et al., 2012; Gluszek & Dovidio, 2010; Ruby & Brigham, 1996; Vrij & Winkel, 1992). One might argue that the messages, despite the exclusion of paraverbal and nonverbal cues, activated stereotypes that led to judges more often mistrusting nonnative speakers. We cannot fully dispel these doubts, but the mediating role of verbal cues suggests that verbal cues, rather than stereotypes, are responsible for the lower credibility of nonnative speakers. Still, further research is needed to investigate whether verbal cues have a direct effect on judgments or whether stereotypes have a further (mediating) role regarding the effect of verbal cues on judgments. Regardless of the potential impact of stereotypes, the results of our study and the research of Akehurst et al. (2018) suggest that communication without visual information does not counteract nonnative speakers' disadvantages in deception detection.

[3]These additional analyses as well as the dataset of the study are available upon request from the corresponding author.

Judging communication more often as truthful than as deceitful constitutes a functional strategy outside the research environment, as most social interactions are not deceitful (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996); a truth bias therefore increases the overall likelihood of making a correct judgment (see also Street, 2015, for a discussion on biases as adaptive strategies in lie detection). Judges' likelihood to rate a message as true was considerably larger than 50% for messages by native speakers (74.38%) but only slightly larger for messages by nonnative speakers (55.40%). If the implicit strategy of a truth bias does not come into effect for nonnative speakers, or does so only to a limited extent, those speakers are, presumably, more often wrongfully suspected of lying than are native speakers (more Type I errors). However, research has yet to clarify whether judges mistrust nonnative speakers only when prompted to judge a message's veracity or whether senders' language proficiency also impacts the initiation of veracity judgments. If either scenario was the case, nonnative speakers would not only be judged more harshly but also more frequently become the object of veracity judgments than would native speakers.

Similar to studies on spoken language (e.g., Da Silva & Leach, 2013; Elliott & Leach, 2016; Leach et al., 2017; Leach & Da Silva, 2013), we found that messages by native speakers were overall more likely to be evaluated correctly compared with those by nonnative speakers. Native speakers' truths were more likely to be evaluated correctly than their lies, whereas there was no difference between lies and truths by nonnative speakers.

As hypothesized, verbal cues (plausibility and length) mediated the effect of senders' language proficiency and message veracity on the likelihood of a message being evaluated correctly. Low plausibility and detailedness, as well as a message being short, were valid indicators of deception in native, but not in nonnative speakers' messages. More importantly, native speakers' messages matched laypersons' beliefs about what truthful versus deceptive communication looks like (see, e.g., Akehurst et al., 1996; Reinhard et al., 2002; Ulatowska, 2017), whereas nonnative speakers' messages did not. Thus, discerning between truth and deception based on verbal cues was impaired in nonnative speakers' messages.

4.1 | Future research

Qualitative and quantitative differences between native and nonnative speakers' truthful and deceptive messages seem to impact veracity judgments. We have stated several potential reasons for why the messages differ regarding verbal cues (e.g., because of different levels of cognitive load or language difficulties). Further research is needed to clarify which factors are responsible for the differences in verbal cues.

Importantly, the findings of our study cannot be generalized to spoken language without further investigations. Further research is needed to clarify whether the influential components of nonnative language use differ between written and spoken language as our findings do not align with those of Akehurst et al. (2018). They found no difference in accuracy when judges evaluated transcripts of videotaped messages (i.e., verbal cues of spoken language). Instead, paraverbal cues were identified as the responsible component for judges' lower accuracy when evaluating videotaped messages by nonnative speakers. Even though paraverbal cues were not available in our study, we found a difference in accuracy, indicating that verbal cues of written and spoken language might affect veracity judgments differently. Given the practical importance of the topic, further research is needed on both written and spoken communication by nonnative speakers in the face of lie detection.

Following the recommendations of Watkins and Martire (2015) and of Judd et al. (2012), we modeled each judgment in relation to the respective message's characteristics by using a multilevel approach. Thus, we did not average data over judges or senders and therefore avoided information loss. In line with research suggesting that lie detection accuracy is determined by senders' ability to lie rather than by judges' ability to detect lies (e.g., Hartwig & Bond, 2011), between-sender variance was higher than between-judge variance in our study. The simulations by Judd et al. suggest that Type I error rates are especially high when variance between stimulus material is large (i.e., between senders) and stimulus material is not treated as random factor. We therefore suggest the consideration of multilevel analyses in future lie detection research.

4.2 | Limitations

The self-categorization of senders into native and nonnative speakers on the one hand allowed comparisons to earlier studies using this binary distinction (e.g., Akehurst et al., 2018; Cheng & Broadhurst, 2005; Da Silva & Leach, 2013; Evans & Michael, 2014). On the other hand, a granular definition of language proficiency (see Elliott & Leach, 2016; Evans et al., 2017) could have offered more nuanced insights into how different language proficiency levels affect veracity judgments in written messages.

Native-speaking senders were mostly from the United States, whereas judges were mostly native British English speakers. The differences between American and British English might have impacted our findings and should, in future research, be investigated in the face of lie detection.

In the stimulus material questionnaire, we set a fixed time for the message production task (i.e., senders could work neither longer nor shorter on this task). This restriction was included to standardize the text production so that the messages would not vary too much in length. One might argue that this measure to ensure internal validity limits the ecological validity of the results, as we do not know if nonnative speakers just need more time to produce similar texts as native speakers. However, people often face serious time restrictions in everyday life when writing texts of which the credibility might be assessed. For instance, in business contexts, individuals usually must complete tasks within a certain time frame; thus, they cannot take unlimited time to formulate e-mails. Further, written assignments in assessment centers or real-time chat communication can put individuals under time pressure while trying to appear as credible as possible.

We excluded messages based on a minimum number of questions answered, rather than on a minimum number of words. We chose this approach to avoid excluding senders who chose to answer questions briefly while still ensuring a certain amount of content as a basis for the veracity judgments. Message production scenarios without any restrictions, as well as other ways of standardizing the messages (e.g., setting length requirements), should be investigated in future studies.

Senders' self-reported cognitive load was not related to the verbal cues of their messages. However, it is unclear whether such indirect subjective measures correspond to actual cognitive load (see, e.g., Bruenken, Plass, & Leutner, 2003). Therefore, more objective measures in future studies could provide insight into whether and how senders' cognitive load affects how senders are perceived and evaluated. In addition, message production tasks with no prescribed time limit should be investigated because our time limit might have imposed a mild time pressure, thus making the task even more difficult, especially for nonnative speakers.

Both data collections for this research were carried out online using crowdsourcing platforms. Even though data quality of such studies is often questioned, research shows that it is not lower than in laboratory studies (e.g., Casler, Bickel, & Hackett, 2013; Clifford & Jerit, 2014; Hauser & Schwarz, 2016; Necka, Cacioppo, Norman, & Cacioppo, 2016). We did not force judges to spend a minimum amount of time on a particular page, which could have increased data quality but in turn may have provided an anchor regarding how much time judges should take to form a judgment. Veracity judgments could thus have been affected. Large effects across all sets in the judgment study indicate that the findings are not likely a random result from judges not paying attention or rushing through the online questionnaire. Yet it must be mentioned that various messages were excluded because senders did not answer enough questions or misunderstood the task.

5 | CONCLUSION

The present research advances our understanding of how senders' language proficiency affects veracity judgments in written language through verbal cues. Considering the increasing prevalence and relevance of nonnative written communication in a globalized world, our findings make an important contribution to the research on deception detection in nonnative speakers. Despite the restricted range of available information in written messages, senders' use of a nonnative language seems to impact both judges' bias and ability to correctly classify nonnative speakers' messages. As this was the first study employing written messages, additional research is required, in particular to determine why verbal cues differ between native and nonnative speakers' messages and whether these effects also occur in spoken language. A more thorough understanding of the underlying processes likely allows the development of effective interventions to prevent discrimination against nonnative speakers when the veracity of their messages is assessed.

CONFLICT OF INTEREST

The research was conducted in the absence of any commercial or financial relationships that could be interpreted as a potential conflict of interest.

ORCID

Sarah Volz https://orcid.org/0000-0002-5958-5002

REFERENCES

Akehurst, L., Arnhold, A., Figueiredo, I., Turtle, S., & Leach, A.-M. (2018). Investigating deception in second language speakers: Interviewee and assessor perspectives. *Legal and Criminological Psychology*, *10*(1), 1–22. https://doi.org/10.1111/lcrp.12127

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*(6), 461–471. https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413>3.0.CO;2-2

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

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

Bobko, P. (1986). A solution to some dilemmas when testing hypotheses about ordinal interactions. *The Journal of Applied Psychology*, *71*(2), 323–326. https://doi.org/10.1037/0021-9010.71.2.323

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Bond, M. H., & Lai, T.-M. (1986). Embarrassment and code-switching into a second language. *Journal of Social Psychology*, *126*(2), 179–186. Retrieved from https://www.tandfonline.com/toc/vsoc20/current

Bruenken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53–61. https://doi.org/10.1207/S15326985EP3801_7

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242. https://doi.org/10.1111/j.1468-2885.1996.tb00127.x

Burgoon, J. K., Buller, D. B., Ebesu, A. S., White, C. H., & Rockwell, P. A. (1996). Testing interpersonal deception theory: Effects of suspicion on communication behaviors and perceptions. *Communication Theory*, *6*(3), 243–267.

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, *71*(3), 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Castillo, P. A., & Mallard, D. (2011). Preventing cross-cultural bias in deception judgments: The role of expectancies about nonverbal behavior. *Journal of Cross-Cultural Psychology*, *43*(6), 967–978. https://doi.org/10.1177/0022022111415672

Castillo, P. A., Tyson, G., & Mallard, D. (2014). An investigation of accuracy and bias in cross-cultural lie detection. *Applied Psychology in Criminal Justice*, *10*(1), 66–82. Retrieved from http://www.apcj.org/

>Cheng, K. H. W., & Broadhurst, R. (2005). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology and Law*, *12*(1), 107–118. https://doi.org/10.1375/pplt.2005.12.1.107

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*(2), 120–131. https://doi.org/10.1017/xps.2014.5

Da Silva, C. S., & Leach, A.-M. (2013). Detecting deception in second-language speakers. *Legal and Criminological Psychology*, *18*(1), 115–127. https://doi.org/10.1111/j.2044-8333.2011.02030.x

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., & Nunamaker, J. F. (2013). Detecting deceptive chat-based communication using typing behavior and message cues. *ACM Transactions on Management Information Systems*, *4*(2), 1–21. https://doi.org/10.1145/2499962.2499967

Dewaele, J.-M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics*, *40*(10), 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002

Dunn, T. L., Lutes, D. J. C., & Risko, E. F. (2016). Metacognitive evaluation in the avoidance of demand. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1372–1387. https://doi.org/10.1037/xhp0000236

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. New York: WW Norton & Company.

Elliott, E., & Leach, A.-M. (2016). You must be lying because I don't understand you: Language proficiency and lie detection. *Journal of Experimental Psychology. Applied*, *22*(4), 488–499. https://doi.org/10.1037/xap0000102

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

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition*, *2*(1), 33–41. https://doi.org/10.1016/j.jarmac.2013.02.002

Evans, J. R., Pimentel, P. S., Pena, M. M., & Michael, S. W. (2017). The ability to detect false statements as a function of the type of statement and the language proficiency of the statement provider. *Psychology, Public Policy, and Law*, *23*(3), 290–300. https://doi.org/10.1037/law0000127

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120–133. https://doi.org/10.1002/ejsp.862

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review: An official journal of the Society for Personality and Social Psychology, Inc*, *14*(2), 214–237. https://doi.org/10.1177/1088868309359288

Gollan, T. H., & Acenas, L.-A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and Tagalog–English bilinguals. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(1), 246–269. https://doi.org/10.1037/0278-7393.30.1.246

Hansen, J., & Wänke, M. (2010). Truth from language and truth from fit: The impact of linguistic concreteness and level of construal on subjective truth. *Personality and Social Psychology Bulletin*, *36*(11), 1576–1588. https://doi.org/10.1177/0146167210386238

Harris, C. L., Ayçíçeğí, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, *24*(4), 561–579. https://doi.org/10.1017/S0142716403000286

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659. https://doi.org/10.1037/a0023589

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, *19*(4), 307–342. https://doi.org/10.1177/1088868314556539

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. In *Methodology in the social sciences* (Second ed.). New York, London: The Guilford Press. Retrieved from https://ebookcentral.proquest.com/lib/potsdamuni/detail.action?docID=5109647

Jackson, J. (2008). Globalization, internationalization, and short-term stays abroad. *International Journal of Intercultural Relations*, *32*(4), 349–358. https://doi.org/10.1016/j.ijintrel.2008.04.004

Javier, R. A., Barroso, F., & Muñoz, M. A. (1993). Autobiographical memory in bilinguals. *Journal of Psycholinguistic Research*, *22*(3), 319–338. https://doi.org/10.1007/BF01068015

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Karlsen, J., Lyster, S.-A. H., & Lervåg, A. (2017). Vocabulary development in Norwegian L1 and L2 learners in the kindergarten-school transition. *Journal of Child Language*, *44*(2), 402–426. https://doi.org/10.1017/S0305000916000106

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, *23*(6), 661–668. https://doi.org/10.1177/0956797611432178

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249–277. https://doi.org/10.1207/S15327906MBR3602_06

Leach, A.-M., & Da Silva, C. S. (2013). Language proficiency and police officers' lie detection performance. *Journal of Police and Criminal Psychology*, *28*(1), 48–53. https://doi.org/10.1007/s11896-012-9109-3

Leach, A.-M., Snellings, R. L., & Gazaille, M. (2017). Observers' language proficiencies and the detection of non-native speakers' deception. *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., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(03), 673–680. https://doi.org/10.1017/S1366728913000606

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, *129*(3), 361–368. https://doi.org/10.1037/0096-3445.129.3.361

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PloS ONE*, *11*(6), 1–19. https://doi.org/10.1371/journal.pone.0157732

Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language*, *26*(4), 392–405. https://doi.org/10.1016/0749-596X(87)90098-2

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338–342. https://doi.org/10.1006/ccog.1999.0386

Reinhard, M.-A., Burghardt, K., Sporer, S. L., & Bursch, S. E. (2002). Alltagsvorstellungen über inhaltliche Kennzeichen von Lügen. *Zeitschrift für Sozialpsychologie*, *33*(3), 169–180. https://doi.org/10.1024//0044-3514.33.3.169

Reinhard, M.-A., Sporer, S. L., & Scharmach, M. (2013). Perceived familiarity with a judgmental situation improves lie detection ability. *Swiss Journal of Psychology*, *72*(1), 43–52. https://doi.org/10.1024/1421-0185/a000098

Reinhard, M.-A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011). Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology*, *101*(3), 467–484. https://doi.org/10.1037/a0023726

Ruby, C. L., & Brigham, J. C. (1996). A criminal schema: The role of chronicity, race, and socioeconomic status in law enforcement officials' perceptions of others. *Journal of Applied Social Psychology*, *26*(2), 95–112. https://doi.org/10.1111/j.1559-1816.1996.tb01840.x

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, *27*(4), 657. https://doi.org/10.2307/3587400

Simos, P. G., Sideridis, G. D., Mouzaki, A., Chatzidaki, A., & Tzevelekou, M. (2014). Vocabulary growth in second language among immigrant school-aged children in Greece. *Applied Psycholinguistics*, *35*(03), 621–647. https://doi.org/10.1017/S0142716412000525

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: a meta-analytic synthesis. *Applied Cognitive Psychology*, *20*(4), 421–446. https://doi.org/10.1002/acp.1190

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stiff, J. B., & Miller, G. R. (1986). "Come to think of it …": Interrogative probes, deceptive communication, and deception detection. *Human Communication Research*, *12*(3), 339–357. https://doi.org/10.1111/j.1468-2958.1986.tb00081.x

Street, C. N. H. (2015). ALIED: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition*, *4*(4), 335–343. https://doi.org/10.1016/j.jarmac.2015.06.002

Strömwall, L., & Granhag, P. A. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology, Crime & Law*, *9*(1), 19–36. https://doi.org/10.1080/10683160308138

Strube, M. J., & Bobko, P. (1989). Testing hypotheses about ordinal interactions: Simulations and further comments. *The Journal of Applied Psychology*, *74*(2), 247–252. https://doi.org/10.1037/0021-9010.74.2.247

Szabo, C. (2016). Exploring the mental lexicon of the multilingual: Vocabulary size, cognate recognition and lexical access in the L1, L2 and L3. *Eurasian Journal of Applied Linguistics*, *2*(2), 1–25. https://doi.org/10.32601/ejal.461007

The Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74. https://doi.org/10.1177/0022022105282295

Ulatowska, J. (2017). Teachers' beliefs about cues to deception and the ability to detect deceit. *Educational Psychology*, *37*(3), 251–260. https://doi.org/10.1080/01443410.2016.1231297

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Vrij, A. (2008). Detecting lies and deceit: Pitfalls and opportunities. In *Wiley series in the psychology of crime, policing and law* (2nd ed.). Chichester, England: John Wiley & Sons Ltd.

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Wiley series in the psychology of crime, policing and law. Detecting deception: Current challenges and cognitive approaches* (pp. 205–229). Wiley Blackwell: Southern Gate, Chinchester, West Sussex, UK.

Vrij, A., Fisher, R. P., Mann, S. A., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, *5*(1-2), 39–43. https://doi.org/10.1002/jip.82

Vrij, A., Granhag, P. A., Mann, S. A., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, *20*(1), 28–32. https://doi.org/10.1177/0963721410391245

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253–265. https://doi.org/10.1007/s10979-007-9103-y

Vrij, A., Mann, S. A., Leal, S., & Fisher, R. P. (2010). 'Look into my eyes': Can an instruction to maintain eye contact facilitate lie detection? *Psychology. Crime & Law*, *16*(4), 327–348. https://doi.org/10.1080/10683160902740633

Vrij, A., & Winkel, F. W. (1992). Crosscultural police-citizen interactions: The influence of race, beliefs, and nonverbal communication on impression formation. *Journal of Applied Social Psychology*, *22*(19), 1546–1559. https://doi.org/10.1111/j.1559-1816.1992.tb00965.x

Watkins, I. J., & Martire, K. A. (2015). Generalized linear mixed models for deception research: Avoiding problematic data aggregation. *Psychology, Crime & Law*, *21*(9), 821–835. https://doi.org/10.1080/1068316X.2015.1054384

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*(1), 41–69. https://doi.org/10.1017/S0272263101001024

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1–59). New York, London: Academic Press. https://doi.org/10.1016/S0065-2601(08)60369-X

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, *6*(2), 105–114. https://doi.org/10.1007/BF00987286

How to cite this article: Volz S, Reinhard M-A, Müller P. Why don't you believe me? Detecting deception in messages written by nonnative and native speakers. *Appl Cognit Psychol*. 2020;34:256–269. https://doi.org/10.1002/acp.3615