AMERICAN PSYCHOLOGICAL ASSOCIATION
American Psychology-Law Society
Law and Human Behavior


ISSN: 0147-7307
https://doi.org/10.1037/lhb0000539

# Improving Juror Assessments of Forensic Testimony and Its Effects on Decision-Making and Evidence Evaluation

Devon E. LaBat[1], Deborah Goldfarb[1, 2], Jacqueline R. Evans[1], Nadja Schreiber Compo[1], Cassidy J. Koolmees[1], Gerald LaPorte[3], and Kevin Lothridge[3]
[1] Department of Psychology, Florida International University
[2] College of Law, Florida International University
[3] Global Forensic and Justice Center, Florida International University



***Objective:*** We explored whether an educational forensic science informational (FSI) video either alone or with specialized jury instructions would assist mock jurors in evaluating forensic expert testimony. ***Hypotheses:*** We predicted that the FSI video would help participants distinguish between low-quality and high-quality testimony, evidenced by lower ratings of the testimony and the expert when the testimonial quality was low compared with when it was high. ***Method:*** Jury-eligible adults ($N$ = 641; $M_{age}$ = 38.18 years; 77.4% White; 8.1% Latino/a or Hispanic; 50.1% male) watched a mock trial and were randomly assigned to a no-forensic-evidence control condition or to a test condition (i.e., participants either watched the FSI video before the trial or did not and either received specialized posttrial instructions or did not). In the test conditions, a forensic expert provided low-quality or high-quality testimony about a latent impression, and participants rated the expert, their testimony, and the forensic evidence. All participants rendered verdicts. ***Results:*** The presence of the FSI video interacted with testimonial quality on ratings of the expert and forensic testimony: In the video-present condition, participants rated the expert in the low-quality testimony condition lower than did participants in the high-quality testimony condition (between-condition differences for credibility: $d$ = –0.52, 95% confidence interval [CI] [–0.78, –0.27]; trustworthiness: $d$ = –0.67, 95% CI [–0.92, –0.42]; knowledgeability: $d$ = –0.54, 95% CI [–0.80, –0.29]). The pattern was the same for the expert's testimony (between-condition differences for convincingness: $d$ = –0.41, 95% CI [–0.66, –0.16]; validity: $d$ = –0.60, 95% CI [–0.86, –0.35]; presentation quality: $d$ = –0.51, 95% CI [–0.76, –0.25]). Participants' ratings in the video-absent condition did not differ on the basis of testimonial quality ($ds$ = –0.07–0.11). The ratings of the print evidence and verdicts were unaffected. Specialized jury instructions had no effect. ***Conclusion:*** The FSI video may be a practical in-court intervention to increase jurors' sensitivity to low-quality forensic testimony without creating skepticism.

***Public Significance Statement***
Informational videos may be a useful tool for decreasing jurors' misinterpretations of forensic evidence in criminal trials. Such videos can help jurors detect suboptimal forensic expert testimony without increasing jurors' overall skepticism of the forensic evidence in general.

*Keywords:* juror decision making, forensic science, expert testimony, juror evaluations of evidence, forensic evidence

*Supplemental materials:* https://doi.org/10.1037/lhb0000539.supp

Lora Levett served as Action Editor.

Devon E. LaBat https://orcid.org/0000-0003-1014-5978
Deborah Goldfarb https://orcid.org/0000-0002-3769-8137
Jacqueline R. Evans https://orcid.org/0000-0002-2044-1853
Nadja Schreiber Compo https://orcid.org/0000-0002-1755-2552
Kevin Lothridge https://orcid.org/0000-0001-5055-4480

This research was funded by the National Institute of Justice (NIJ-2019-15645) to awardees Kevin Lothridge (PI), Deborah Goldfarb (Co-PI), Jacqueline R. Evans (Co-PI), and Nadja Schreiber Compo (Co-PI). All statements in this publication are those of the authors and not those of the funder. The authors thank the participants and Florida International University's Global Forensic and Justice Center. The authors have no conflicts of interest to disclose.

The data were previously disseminated via conference presentations at the 2023 American Psychology-Law Society conference and the 2023 Global Forensic and Justice Center 12th Annual Forensic Science Symposium. The study materials and data are available on the open science framework at https://osf.io/r98pj/.

Devon E. LaBat played a lead role in data curation, formal analysis,

*continued*

Forensic science testimony can suffer from data presentation errors, and forensic experts may overstate the precision and probative value of the evidence in court (Eastwood & Caldwell, 2015; Gould et al., 2014). These testimonial errors have important implications, as jurors typically give substantial weight to forensic evidence when coming to a verdict (Cooper et al., 1996; Cutler et al., 1990; Lieberman et al., 2008). Indeed, inappropriate forensic testimony or improper interpretation of forensic testimony has contributed to a substantial number of wrongful convictions (Acker & Redlich, 2019). B. L. Garrett and Neufeld (2009) found that 60% of the DNA exoneration cases they assessed included invalid testimony by a prosecution forensic expert witness. Further, in 35% of cases, the forensic expert interpreted nonprobative information as inculpatory. B. L. Garrett and Neufeld (2009) also found that forensic experts presented testimony including statistical probabilities (in 3.6% of cases) or nonnumerical probabilities (in 15% of cases) that lacked scientific or empirical support. It is thus possible that forensic experts present evidence in a way that might mislead jurors as to its probative value. Thus, although experts should avoid data presentation errors, jurors must be able to detect errors when they are present and appropriately weigh the forensic evidence when rendering a verdict.

## Jurors' Comprehension of Forensic Experts' Testimony

Jurors often struggle to understand forensic evidence, particularly the highly technical nature of the evidence (Martire et al., 2013). Nevertheless, jurors are generally overconfident in their ability to comprehend such testimony (McQuiston-Surrett & Saks, 2009) and place heavy weight on forensic evidence (B. L. Garrett, Crozier, et al., 2020). This is particularly problematic in light of research finding that expert testimony heavily influences jurors' decisions (Cooper et al., 1996; Eastwood & Caldwell, 2015). Cooper et al. (1996) investigated the effect of expert testimony complexity and strength of the expert's credentials and found that jurors were more persuaded by an expert with strong credentials than an expert with moderately strong credentials, but only when the expert's testimony was highly complex.

Consistent with Cooper et al. (1996), Koehler et al. (2016) found that jurors tend to rely heavily on an expert's experience when evaluating fingerprint and bite mark evidence and when rating defendant guilt. Specifically, their results revealed that although mock jurors viewed evidence from scientifically validated techniques as stronger than those not scientifically validated, guilt ratings were unaffected. However, the experience of the expert significantly impacted jurors' ratings of the evidence strength such that forensic evidence presented by a highly experienced expert was viewed as stronger than evidence presented by a less experienced expert. On the basis of these findings, the authors concluded that "jurors use the background and experience of an expert as a proxy for the value of the evidence the expert provides" (Koehler et al., 2016, p. 410).

Furthermore, research suggests that jurors may ignore the limitations of forensic science present in expert witness testimony. In one study involving testimony regarding microscopic hair examination, jurors were informed about the limitations of the examination either during the defense's cross-examination of the expert or during posttrial jury instructions (McQuiston-Surrett & Saks, 2009). Participants largely ignored these limitations when assessing the expert's testimony, and jurors' knowledge of the limitations did not yield a change in jurors' judgments of the defendant's guilt and comprehension of the forensic testimony. In sum, the literature suggests that jurors tend to rely on an expert's level of experience in coming to conclusions about the evidence the expert presented while largely ignoring interventions that highlight the limitations of the procedures behind such testimony. The literature thus paints a concerning picture of jurors' ability not only to understand complex testimony but also to accurately interpret and evaluate that testimony when rendering a verdict.

## The Department of Justice's Uniform Language for Testimony and Reports

An additional consideration in how jurors evaluate forensic testimony is the language used to convey the evidence and examinations. In response to calls to standardize the language used in expert testimony (B. L. Garrett & Neufeld, 2009) and to use more cautious language when expressing examination conclusions (National Research Council, 2009), the U.S. Department of Justice (DOJ) began formulating the Uniform Language for Testimony and Reports (ULTR) for various forensic science disciplines in 2016. The latent print ULTR, which is specific to fingerprint examination, was the first ULTR document officially approved by the DOJ in 2018 (Cole, 2018). The footwear print and tire print ULTRs largely echo the latent print ULTR's guidelines but contain one additional guideline compared with the latent print ULTR. The main goal of the latent print evidence ULTRs was thus to standardize the language and terminology examiners use to reduce overstatements or invalid statements in forensic testimony and reports (B. L. Garrett, Scurich, et al., 2020), resulting in five guidelines stipulating statements that forensic experts should not say (United States Department of Justice, 2018; see Table 1).

investigation, visualization, writing–original draft, and writing–review and editing; a supporting role in conceptualization; and an equal role in methodology and resources. Deborah Goldfarb played a lead role in conceptualization, funding acquisition, project administration, and writing–review and editing; a supporting role in formal analysis; and an equal role in methodology, resources, and supervision. Jacqueline R. Evans played a lead role in conceptualization, funding acquisition, project administration, and writing–review and editing and an equal role in methodology, resources, and supervision. Nadja Schreiber Compo played a lead role in conceptualization and writing–review and editing; a supporting role in funding acquisition and project administration; and an equal role in methodology, resources, and supervision. Cassidy J. Koolmees played a supporting role in conceptualization and writing–review and editing. Gerald LaPorte played a supporting role in validation, writing–original draft, and writing–review and editing. Kevin Lothridge played a supporting role in conceptualization, project administration, and writing–review and editing.

The data are available at https://osf.io/r98pj/.

The experimental materials are available at https://osf.io/r98pj/.

Correspondence concerning this article should be addressed to Devon E. LaBat, Department of Psychology, Florida International University, 11200 SW 8th Street, Miami, FL 33199, United States. Email: dlabat@fiu.edu

**Table 1**
*Department of Justice (DOJ) Uniform Language for Testimony and Reports (ULTR) Guidelines and Violations*

| Violation number | DOJ ULTR guideline | Expert's statement in the low-quality trial testimony |
|---|---|---|
| 1 | An expert should not use phrases such as "reasonable degree of scientific certainty," "reasonable scientific certainty," unless required to do so by law. | No statement related to this violation was included in the trial. |
| 2 | An expert should not claim that two prints came from the same person to the exclusion of all other persons or use the words "individualize" or "individualization." | "[The latent print] did not come from any other source." |
| 3 | An expert should not give a conclusion about prints that includes a numerical or statistical degree of probability. | No statement related to this violation was included in the trial. |
| 4 | An expert should not talk about the number of print examinations they have done in their career as proof that their conclusion is correct. | "In my years of experience, I have examined thousands of latent prints. I therefore have no doubt in the accuracy of my conclusion." |
| 5 | An expert should not state that a forensic print examination is perfect, is mistake-free, or has a zero-error rate. | "[The method of latent print comparison and identification] is simply without error." |

*Note*. We numbered violations in accordance with how they are referenced on open science framework in the "Full Fingerprint Script" and "Full Footwear Print Script" documents at https://osf.io/r98pj/.



For the present study, we used the five guidelines shared by all three ULTRs.[1] Hereafter the fingerprint, footwear print, and tire print guidelines will be referred to collectively as the *print evidence ULTRs*.

## The Impact of Testimony Language on Jurors' Evaluations of Forensic Evidence

To investigate whether such guidelines for testimony and report language are effective, B. L. Garrett, Scurich, et al. (2020) tested whether the language used by a firearm examiner impacted participants' verdicts and their ratings of the expert. The examiner indicated either that there was a match between two firearm cartridge casings or that the examination was inconclusive. The authors found that participants were not sensitive to the variations in language and that the specific language used generally did not impact guilty verdicts, ratings of the likelihood that the defendant was the person who fired the gun, or ratings of the credibility and reliability of the firearm analysis.

Furthermore, B. Garrett and Mitchell (2013) explored whether the specific language used in fingerprint testimony impacted evaluations of the evidence. They found that regardless of whether the language used to describe a fingerprint comparison was vague or more detailed, mock jurors' ratings of the likelihood that the defendant committed the crime and ratings of the probability that the defendant left their fingerprints at the scene were unaffected. The studies by Garrett and colleagues (B. L. Garrett, Scurich, et al., 2020; B. Garrett & Mitchell, 2013) highlight a key point: Despite the ULTR guidelines' aim to keep examiners from overstating their conclusions, changing the language alone does not appear to impact trial outcomes. The present study built on this line of research by exploring mock jurors' evaluations of expert testimony regarding latent prints (rather than firearms) that varied in quality on the basis of the ULTR guidelines. Additionally, the present study extended B. L. Garrett, Scurich, et al. (2020) work by testing two potential interventions to increase jurors' sensitivity to testimonial quality without creating overall skepticism.

## Sensitivity Versus Skepticism

It is imperative that any intervention used to improve jurors' assessments of forensic testimony increases jurors' sensitivity to the testimonial quality without creating a general skepticism toward the science. *Sensitivity* in the present study refers to the desirable outcome, whereby a juror applies the information they learn about appropriate forensic testimony to make appropriate judgments about the quality of the expert's testimony (i.e., distinguishing between low-quality and high-quality testimony). Conversely, *skepticism* is an undesirable outcome, whereby a juror disbelieves all of the expert's testimony, regardless of its quality (Devenport et al., 1997).

Both the elaboration likelihood model (ELM; Petty & Cacioppo, 1986) and the heuristic-systematic model (Chaiken, 1980) have been theorized to explain skepticism and sensitivity effects. The ELM suggests that there are two routes to persuasion: (a) the central route, in which an individual carefully considers the quality of information, and (b) the peripheral route, in which an individual relies on cues or heuristics to evaluate information more easily (Petty & Cacioppo, 1986). Similarly, the heuristic-systematic model suggests that people evaluate an argument either via (a) systematic processing, whereby they actively assess the validity and quality of a message in a way that is cognitively effortful, or (b) heuristic processing, whereby individuals rely on cognitive cues, or heuristics, resulting in less effortful processing of the argument (Chaiken, 1980). Cutler et al. (1990) applied the ELM to expert testimony related to eyewitness identifications. They proposed that in this context sensitivity requires more effortful processing (central route) because individuals must evaluate eyewitness evidence on the basis of the expert testimony. However, skepticism causes jurors to use heuristic cues to disregard the evidence overall. In the context of jurors' perceptions of forensic testimony, Kovera and Levett (2015) posited that when expert testimony is complex, jurors may have a difficult time understanding it, thus leading them to rely on

[1] In the present study, we used the latent print ULTR published on July 24, 2018. However, the DOJ has since updated the ULTR documents. The updated documents have slightly different language than the 2018 version that was used in the present study. As cited in the reference section, the 2018 version can be accessed via The Wayback Machine archival database. We have also uploaded a PDF version to OSF, https://osf.io/r98pj/. The updated documents can be accessed at https://www.justice.gov/olp/uniform-language-testimony-and-reports#:~:text=These%20documents%2C%20known%20as%20%E2%80%9CUniform,when%20drafting%20reports%20or%20testifying.

peripheral or heuristic processing rather than central or systematic processing. Applied to the present study, it may be that educating jurors on appropriate or inappropriate expert testimony via an informational video may assist their evaluations of the testimony such that they can engage in more effortful, systematic processing as opposed to relying on heuristics.

## Improving Juror Comprehension of Forensic Testimony

### Informational Videos

Recently, courts have started to create informational videos to explain important trial topics and procedures to jurors prior to a trial. For example, the Western Washington District Court (2021) created an implicit bias video to help jurors fulfill their role as objective fact finders. The video explains the concept of unconscious bias and how it may play a role in court. Such videos may be useful because they provide a consistent communication of information via which jurors can organize, understand, and apply case-relevant information. Similarly, we developed a video about forensic expert testimony that included reason-based arguments to explain why certain statements violated the DOJ's ULTR guidelines (forensic science informational [FSI] video). Previous research suggests that reason-based instructions help jurors determine the reliability of eyewitness testimony (B. L. Garrett et al., 2023). The present study extended this research by investigating the impact of such reason-based arguments on mock jurors' decisions in a forensic science instructional video. The video was also designed to provide mock jurors with a framework with which they could more easily evaluate the quality of forensic expert testimony and thus further avoid heuristic (or peripheral) processing. We tested the FSI video in a previous study (LaBat et al., 2023) and found that it helped increase jurors' identification of statements that violated the ULTR guidelines. However, the video was tested neither in the context of a full trial nor with varying quality of testimony to investigate whether it influences skepticism. These were both aims of the present study.

### Specialized Jury Instructions

Another way to improve juror comprehension is via jury instructions. Research on the usefulness of such an intervention is mixed. In terms of generally improving jurors' decision making related to evidence and different types of expert testimony, some research suggests that specialized and/or case-specific jury instructions may improve jurors' decision making (Ribbers & Henneberg, 2018), whereas other research casts doubt on their usefulness (Bornstein & Greene, 2011; Eastwood & Caldwell, 2015; Jones et al., 2020; Papailiou et al., 2015). The eyewitness literature has largely found that when presented with jury instructions highlighting the limitations of eyewitness testimony or the variables that can affect identification reliability, jurors tend to become more skeptical of the eyewitness evidence overall, regardless of the evidence's strength or weakness (Dillon et al., 2017; Papailiou et al., 2015). However, other research suggests that jury instructions that are simplified and evidence relevant may be useful in increasing juror comprehension (Devine et al., 2001). In the present study, we thus aimed to systematically examine the usefulness of jury instructions that address evaluations of forensic expert testimony specifically while varying the quality of the testimony across differing types of forensic evidence.

## The Present Study

Our aim in the present study was to determine how to improve jurors' sensitivity to the quality of forensic evidence testimony without increasing their skepticism of all forensic testimony. Specifically, we aimed to address the potential benefits of instructional videos on mock jurors' evaluations of forensic science experts and testimony and to determine the effect of such a video in conjunction with, and compared with, specialized jury instructions.

We aimed to investigate the efficacy of (a) an informational video on forensic science presented at the beginning of a trial, (b) a specialized jury instruction about forensic expert testimony presented after the trial, or (c) a combination of the two. We thus had three research questions:

1. Do pretrial instructional videos and/or specialized post-trial jury instructions improve mock jurors' ability to distinguish between a forensic expert providing low- and high-quality testimony?
2. Relatedly, will our interventions improve mock jurors' awareness of the limitations of low-quality testimony without creating general skepticism toward forensic science?
3. Will the improvement in jurors' assessments as a result of viewing the instructional video and/or receiving specialized posttrial jury instructions translate to differences in case outcomes, as measured by verdicts?

Specifically, we hypothesized that (a) participants in the FSI video-absent condition would rate the expert and the latent print testimony higher than would participants in the FSI video-present condition and (b) this effect would interact with testimonial quality. We predicted that without exposure to the FSI video, participants would rate the expert and their testimony similarly regardless of testimonial quality, but after exposure to the FSI video, participants would rate the expert and the testimony more highly when the testimony was of high (vs. low) quality. In this way, we expected that the video would help increase mock jurors' sensitivity to testimonial quality. We hypothesized that the video would avoid creating a skepticism effect because it would help mock jurors evaluate the expert's credibility, resulting in increased elaboration of the testimony, evidenced by lower ratings of the expert and their testimony in the low-quality testimony condition compared with the high-quality testimony condition.

Several additional analyses were exploratory. First, analyses of the posttrial jury instruction were exploratory in nature because of the mixed results in the literature on their efficacy in sensitizing jurors to testimony without causing overall skepticism of the evidence (Devine et al., 2001; Dillon et al., 2017; Papailiou et al., 2015). Second, exploratory analyses tested whether participants' ratings of the evidence itself were impacted by our manipulations. Third, the control condition was included to provide a comparison of what jurors' ultimate case decisions would be in the absence of any forensic science testimony compared with our test conditions with forensic science testimony. No hypotheses were made as to this comparison.

# Method

## Design

A 2 (FSI video: present vs. absent) × 2 (posttrial jury instruction: standard vs. specialized) × 2 (forensic testimonial quality: low vs. high) × 2 (latent print type: fingerprint vs. footwear print) + 1 (control: no forensic science testimony, no instructions regarding forensic science evidence, no FSI video) between-participants design was employed, with participants randomly assigned to conditions. Latent print type was used as a covariate and was included as a means of stimulus sampling.

## Participants

A power analysis using G*Power (Erdfelder et al., 1996) revealed that 351 participants would be needed to generate sufficient power (.8) to detect a small-to-medium two-way interaction ($f = .15$). Because recent research suggests that G*Power underestimates the necessary sample size (Porter et al., 2020; see also Giner-Sorolla, 2018) and to ensure sufficient power to detect effects among the experimental conditions (excluding the control group), we increased the sample size. The final total sample consisted of 641 U.S. jury-eligible adults after we excluded duplicate responses, participants who did not finish the study, and participants who answered more than one attention, comprehension, and/or manipulation check question incorrectly ($n = 53$). Excluding the control group ($n = 132$), we had a final sample for the experimental conditions ($n = 509$) that was 45% greater than the original projected sample size.

All participants were recruited via Prolific, an online research platform. Participants were compensated $9.75 for completing the study. Those who answered all attention check, comprehension check, and manipulation check questions correctly or only one incorrectly received a bonus of $10.25, earning a total of $20.

The final sample consisted of 50.1% male-identifying participants, 47.1% female-identifying participants, 1.2% nonbinary participants, and <1% transgender male, transgender female, agender, genderfluid, or genderqueer participants. The majority of participants were White (77.4%), followed by Black or African American (8.1%), East Asian (5.0%), biracial or multiracial (4.1%), South Asian or Indian (2.7%), other (1.1%), American Indian or Indigenous or Alaska Native (<1%), and Native Hawaiian or other Pacific Islander (<1%). Additionally, 8.1% of participants were Latino/a or Hispanic. The mean age of the sample was 38.18 years ($SD = 12.71$, range = 18–77). The majority of participants identified as slightly to extremely liberal (58.0%), followed by slightly to extremely conservative (22.5%), and moderate (19.5%). The majority of participants also reported their highest level of education as a bachelor's degree (42.6%), followed by high school graduate or general education development (31.5%), a master's degree (11.9%), an associate's degree (10.1%), a PhD or professional degree (3.0%), and some high school (0.9%).

## Materials

### *Trial Video*

The mock criminal trial involved a defendant who was accused of theft by larceny. Multiple case fact patterns derived from real trials were pilot-tested (*Castellon v. State*, 2009; *United States v. Haight,* 2015). The video we created was roughly 40 min long but varied depending on participant's assigned condition. Except in the control condition, the trial featured latent prints that were found at the crime scene; in addition, a forensic expert testified that the prints were similar to the defendant's prints, and the defendant could not be excluded (the specific language used varied across testimonial quality conditions, as described below). To increase the external validity of the trial video, we adapted the forensic expert testimony from a real trial (*United States v. Henry,* 2015) and edited it with the help of forensic experts. Further, real forensic experts played the role of the forensic expert in this case. We randomly assigned each participant to see one of four different forensic experts as a means of stimulus sampling. Each of the four experts provided all types of testimony, that is, there were four versions of (a) the low-quality fingerprint testimony, (b) the high-quality fingerprint testimony, (c) the low-quality footwear impression testimony, and (d) the high-quality footwear impression testimony. In this way, we reduced the chances that our results were driven by a particular forensic expert and avoided confounding the specific expert with a particular condition. Additionally, green screen technology replicated the courtroom setting and actors appeared in relevant attire (e.g., robe for the judge; see https://osf.io/r98pj/, for screenshots of the trial video).

Participants first heard pretrial jury instructions taken from the Judicial Council of California Criminal Jury Instructions (Judicial Council of California, 2019). California jury instructions are written specifically to be "legally accurate and understandable to the average juror" (p. ix) and thus were used in this study to minimize the chance that findings could be attributed to confusion stemming from complex jury instructions.

Participants next watched opening statements from the prosecutor and defense attorney. The prosecution then presented its case, which included (a) testimony by a police officer who testified about a chase with the suspect and (b) testimony by a forensic expert. The latter included two manipulations: latent print type and testimonial quality. Specifically, the trial included either fingerprint evidence or footwear print evidence. The testimony was similar in the two conditions; only the specific terms used to describe the examination differed between the two print types (e.g., in the fingerprint condition, the expert talked about the friction ridge, but in the footwear print condition, the expert talked about class characteristics). Examples of the trial transcripts can be found at https://osf.io/r98pj/. The testimonial quality was either high or low and is described in further detail below. Both the police officer and the forensic expert were cross-examined by the defense attorney. The control condition did not include any forensic expert testimony.

After the prosecution concluded the presentation of their evidence, the defense presented its case, which consisted of testimony by an alibi witness, including a cross-examination by the prosecutor and redirect examination by the defense attorney. Both sides gave closing arguments, the prosecution first, followed by the defense. Finally, depending on their condition, participants watched the judge read either standard posttrial jury instructions only or the standard instructions with specialized jury instructions (described further below). The standard jury instructions specific to expert witness testimony were adapted from Judicial Council of California Criminal Jury Instruction No. 332 (Judicial Council of California, 2019).

**Forensic Expert Testimony.** The quality of the forensic expert's testimony was manipulated such that the expert either

violated several of the ULTR guidelines (low-quality testimony) or did not violate any of the guidelines (high-quality testimony). To this end, we developed and pilot-tested five statements, each corresponding to one of the five guidelines. Three statements (i.e., violations) were selected for inclusion in the low-quality testimonial condition (see Table 1). This manipulation of testimonial quality refers only to the forensic expert's testimony. We did not manipulate the quality of the forensic evidence itself.

### FSI Video

The FSI video explains the basic process of latent print analysis and the specific types of inappropriate testimony that should not be provided by forensic experts. Specifically, the FSI video incorporates all five DOJ ULTR guidelines for latent print examiner testimony (United States Department of Justice, 2018; see Table 1) and explains why certain statements are not appropriate. The video, which can be found at https://osf.io/r98pj/, is approximately 4.5 min long and includes a narrator explaining the aforementioned information, photographs of different types of prints and related images, and text examples of guideline violations. We aimed to liken the FSI video both in structure and general intent to the Western Washington District Court's implicit bias video.

### Specialized Posttrial Jury Instructions

The specialized jury instructions appeared in addition to the standardized posttrial jury instructions and consisted of similar information to the FSI video. The transcripts of both the standard and the specialized jury instructions can be found at https://osf.io/r98pj/.

### Dependent Measures

**Ratings of the Expert: Witness Credibility Scale.** Participants completed Brodsky et al.'s (2010) Witness Credibility Scale (WCS); the scale assesses juries' and judges' evaluations of an expert witness's credibility in court. This measure consists of 20 items rated on 10-point Likert scales that load highly onto four subscales: Confident, Trustworthy, Knowledgeable, and Likeable. The Confident subscale included ratings of how self-assured, well-spoken, confident, poised, and relaxed the participants found the expert. The Trustworthy subscale included ratings of how truthful, trustworthy, honest, dependable, and reliable the participants found the expert. The Knowledgeable subscale included ratings of how logical, informed, wise, educated, and scientific the participants found the expert. Finally, the Likeable subscale included ratings of how kind, friendly, pleasant, respectful, and well-mannered participants found the expert. Brodsky et al. (2010) tested their scale and found its overall internal reliability was high ($\alpha = .95$) as were its subscales (Confident: $\alpha = .89$; Trustworthy: $\alpha = .93$; Knowledgeable: $\alpha = .86$; Likeability: $\alpha = .86$). They found that the reliability remained stable across six different studies and found the WCS to be generalizable across different types of expert witnesses. Our data showed similar results (overall credibility: $\alpha = .96$; Confident: $\alpha = .89$; Trustworthy: $\alpha = .96$; Knowledgeable: $\alpha = .91$; Likeability: $\alpha = .90$).

**Ratings of the Testimony.** Participants rated the forensic expert's testimony on the following scales: (a) convincingness (1 = *unconvincing*, 10 = *convincing*), (b) validity (1 = *invalid*, 10 = *valid*), (c) presentation quality (1 = *poorly presented*, 10 = *well presented*), (d) complexity (1 = *simple*, 10 = *complex*), and (e) ease of understanding (1 = *difficult to understand*, 10 = *easy to understand*).

**Ratings of the Print Evidence.** Participants rated the print evidence on the following scales: (a) "How weak/strong was the print evidence that was presented against the defendant?" (1 = *extremely weak*, 10 = *extremely strong*), (b) "How would you rate the quality of the print evidence that was presented against the defendant?" (1 = *extremely poor*, 10 = *extremely good*), and (c) "How useful was the print evidence that was presented against the defendant in deciding your verdict?" (1 = *extremely useless*, 10 = *extremely useful*). These three dependent measures will hereafter be referred to as *evidence strength*, *evidence quality*, and *evidence usefulness*, respectively.

**Final Case Decisions.** Participants made a verdict decision, either finding the defendant guilty or not-guilty. They also rated their confidence in their decision (1 = *not at all confident*, 10 = *completely confident*). Next, they indicated how likely it was that the defendant was guilty or not-guilty on a 10-point bipolar scale (–5 = *extremely likely he IS NOT guilty*, 5 = *extremely likely he IS guilty*). Note that although all participants, including control participants, completed these final case decision measures, control participants did not complete any of the previously described measures because their trial stimulus did not include forensic evidence.

## Procedure

At the beginning of the study, potential participants answered screening questions ensuring that they met the criteria for jury eligibility (U.S. citizens, 18 years or older, and fluent in English) and that they had a device with functioning video and audio capabilities. Participants then read and electronically signed a consent form, thereby indicating that they agreed to participate. If they did not consent, they were dismissed from the survey. All study procedures were approved by Florida International University's Institutional Review Board as well as the funder's review process. After consenting, participants answered demographics questions and questions about prior experience as a juror, and they rated how convincing they found different types of evidence (e.g., DNA, bite marks, police testimony). Then all participants watched their assigned version of the mock trial. In the first clip, the judge presented the pretrial jury instructions; next, participants in the video-present condition watched the FSI video. Control participants never saw the FSI video. Then participants watched the witness testimony (in line with their assigned condition), closing arguments, and posttrial instructions (presented by the judge). Each participant was randomly assigned to receive only standard posttrial jury instructions or to receive specialized posttrial jury instructions in addition to the standard instructions. Ten attention check and manipulation check questions were embedded between trial video clips and throughout the dependent measures. All attention and manipulation check questions, including a description of where each question appeared in the survey, can be found at https://osf.io/r98pj/.

After the trial, all participants (except for the control group) rated the expert's credibility, the expert's testimony, and the print evidence. All participants (including the control group) then made their final case decisions. Participants had access to a digital written copy of the jury instructions when making their case decisions

(either only standard instructions or standard and specialized instructions) to simulate deliberations, during which jurors can have a copy of the jury instructions.

Next, participants answered additional attention check questions and rated how convincing they found different types of evidence (the same questions that were asked at the beginning of the study). Last, participants received debriefing information and compensation.

## Results

First, the control group (participants who did not receive any forensic expert testimony or view the FSI video) was compared with the experimental groups via a chi-square test and two one-way analyses of variance to determine whether participants made different case decisions as a function of the presence of forensic science testimony. The remaining analyses did not include the control participants, resulting in an adjusted sample size of 509. Next, analyses of covariance (ANCOVAs) tested the effects of the FSI video and specialized posttrial jury instructions on jurors' evaluations of (a) the expert's credibility via the WCS, (b) ratings of the expert testimony, and (c) ratings of the print evidence. Specifically, we ran a series of 2 (FSI video: present vs. absent) × 2 (posttrial jury instruction: standard vs. specialized) × 2 (forensic testimonial quality: low vs. high) × 2 (latent print type: fingerprint vs. footwear print) ANCOVAs, with latent print type serving as the covariate, and applied Bonferroni corrections to account for multiple comparisons, as described below. Finally, a binary logistic regression and two ANCOVAs were run to explore the effects of the independent variables on mock jurors' ultimate case decisions. Only significant results will be discussed below, but full results can be found in the tables in each section. Means and standard deviations for each dependent measure by participants' overall condition can be found in Supplemental Table S1.

### Control Group Analyses

First, the chi-square test explored whether participants' overall condition (control, only the FSI video, only the specialized posttrial jury instructions, both the video and the jury instructions, and neither) affected their verdict decisions. Results revealed that there was no difference in verdict decisions across conditions, $\chi^2(4) = 1.80$, $p = .77$, $V = .05$.

Next, the two analyses of variance tested the effect of participant condition on ratings of verdict confidence and guilt likelihood ratings, yielding no significant differences between the control condition and the experimental conditions, $F(4, 636) = 1.09$, $p = .36$, $\eta_p^2 = .01$, and $F(4, 636) = 0.56$, $p = .69$, $\eta_p^2 < .01$, respectively.

### Participants Who Viewed the FSI Video Rated the Expert Providing Low-Quality Testimony Lower Than the Expert Providing High-Quality Testimony

ANCOVAs were run for the overall credibility score and each of the four WCS subscales (critical $p \leq .01$). See Table 2 for the full WCS ANCOVA results and Supplemental Table S2 for means and confidence intervals (CIs) of the interaction effects described below.

There was a main effect of testimonial quality on participants' ratings of expert trustworthiness; specifically, participants in the high-quality testimony condition rated the expert as more trustworthy ($M = 8.25$, 95% CI [8.03, 8.47]) than participants in the low-quality testimony condition ($M = 7.68$, 95% CI [7.46, 7.89]). This main effect was qualified by an interaction between FSI video and forensic testimonial quality on ratings of expert trustworthiness. There was also an interaction between FSI video and forensic testimonial quality for the overall ratings of the expert's credibility as well as the knowledgeability and confidence subscales (see Figure 1). In the FSI-absent condition, participants' ratings of the expert's overall credibility, $t(256) = 0.89$, $p = .38$, $d = 0.11$, 95% CI [–0.13, 0.36]; trustworthiness, $t(256) = 0.17$, $p = .87$, $d = 0.02$, 95% CI [–0.22, 0.27]; and knowledgeability, $t(256) = 0.86$, $p = .39$, $d = 0.11$, 95% CI [–0.14, 0.35], did not differ depending on the testimonial quality condition. However, participants in the FSI-present condition rated the expert in the low-quality testimony condition as less credible, $t(249) = –4.15$, $p < .001$, $d = –0.52$, 95% CI [–0.78, –0.27]; trustworthy, $t(249) = –5.31$, $p < .001$, $d = –0.67$, 95% CI [–0.92, –0.42]; and knowledgeable, $t(249) = –4.31$, $p < .001$, $d = –0.54$, 95% CI [–0.80, –0.29], than the expert in the high-quality testimony condition. Participants' ratings of the expert's confidence did not significantly differ depending on whether they were in the low-quality testimony condition or the high-quality testimony condition in either the FSI-absent condition, $t(256) = 1.86$, $p = .06$, $d = 0.23$, 95% CI [–0.01, 0.48], or the FSI-present condition, $t(249) = –1.73$, $p = .09$, $d = –0.22$, 95% CI [–0.47, 0.03]. There were no significant predictors of participants' ratings of the expert's likeability ($ps \geq .03$).

### The FSI Video Improved Juror Sensitivity to Testimonial Quality

Ratings of the testimony's convincingness, validity, presentation quality, complexity, and the ease of understanding the testimony were analyzed via a series of five ANCOVAs. A Bonferroni correction was applied to account for multiple comparisons, resulting in a critical $p$ value of .01. See Table 3 for all testimony ANCOVA results and Supplemental Table S3 for means and CIs of the interaction effects described below.

There was a main effect of testimonial quality on the testimony validity ratings and presentation quality ratings; specifically, participants in the high-quality testimony condition rated the testimony's validity ($M = 7.43$, 95% CI [7.14, 7.72]) and presentation quality ($M = 7.81$, 95% CI [7.54, 8.09]) as higher than did participants in the low-quality testimony condition ($M = 6.64$, 95% CI [6.36, 6.93] and $M = 7.22$, 95% CI [6.95, 7.49], respectively). These main effects were qualified by an interaction between FSI video and testimonial quality. There was also an interaction between FSI video and testimonial quality on ratings of the convincingness of the testimony (see Figure 2). In the FSI-absent condition, participants' ratings of the testimony's validity, $t(256) = –0.58$, $p = .56$, $d = –0.07$, 95% CI [–0.32, 0.17]; presentation quality, $t(256) = –0.15$, $p = .88$, $d = –0.02$, 95% CI [–0.26, 0.23]; and convincingness, $t(256) = 0.59$, $p = .56$, $d = 0.07$, 95% CI [–0.17, 0.32], did not differ depending on whether they were in the low-quality testimony condition or the high-quality testimony condition. In contrast, participants in the FSI-present condition rated the testimony validity, $t(249) = –4.78$, $p < .001$, $d = –0.60$, 95% CI [–0.86, –0.35]; presentation quality, $t(249) = –4.01$, $p < .001$, $d = –0.051$, 95% CI [–0.76, –0.25]; and convincingness, $t(249) = –3.22$, $p < .001$, $d = –0.41$, 95% CI [–0.66, –0.16], lower in

**Table 2**
*Analysis of Covariance Results for Ratings of the Expert*

| Dependent measure and between-participants effects | Sum of squares | *df* | *F* | $\eta_p^2$ | *p* |
|---|---|---|---|---|---|
| Overall credibility | | | | | |
| FSI video | 8.19 | 1 | 4.28 | .01 | .04 |
| Posttrial jury instruction | 2.95 | 1 | 1.54 | <.01 | .22 |
| Forensic testimonial quality | 10.50 | 1 | 5.48 | .01 | .02 |
| Latent print type | 3.88 | 1 | 2.03 | <.01 | .16 |
| FSI Video × Posttrial Jury Instruction | 9.75 | 1 | 5.09 | .01 | .03 |
| FSI Video × Forensic Testimonial Quality | 23.16 | 1 | 12.09 | .02 | .001 |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 1.02 | 1 | 0.53 | <.01 | .47 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 1.71 | 1 | 0.89 | <.01 | .35 |
| Error | 957.69 | 500 | | | |
| Confidence | | | | | |
| FSI video | 8.63 | 1 | 3.59 | .01 | .06 |
| Posttrial jury instruction | 0.97 | 1 | 0.40 | <.01 | .53 |
| Forensic testimonial quality | 0.002 | 1 | <0.01 | <.01 | .98 |
| Latent print type | 5.32 | 1 | 2.22 | <.01 | .14 |
| FSI Video × Posttrial Jury Instruction | 9.26 | 1 | 3.86 | .01 | .05 |
| FSI Video × Forensic Testimonial Quality | 14.94 | 1 | 6.22 | .01 | **.01** |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 0.98 | 1 | 0.41 | <.01 | .52 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 0.80 | 1 | 0.34 | <.01 | .56 |
| Error | 1,200.49 | 500 | | | |
| Trustworthiness | | | | | |
| FSI video | 14.02 | 1 | 4.51 | .01 | .03 |
| Posttrial jury instruction | 7.01 | 1 | 2.25 | <.01 | .13 |
| Forensic testimonial quality | 41.55 | 1 | 13.36 | .03 | **<.001** |
| Latent print type | 7.02 | 1 | 2.26 | <.01 | .13 |
| FSI Video × Posttrial Jury Instruction | 17.34 | 1 | 5.57 | .01 | .02 |
| FSI Video × Forensic Testimonial Quality | 43.49 | 1 | 13.98 | .03 | **<.001** |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 0.64 | 1 | 0.20 | <.01 | .65 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 1.22 | 1 | 0.39 | <.01 | .53 |
| Error | 1,555.25 | 500 | | | |
| Knowledgeability | | | | | |
| FSI video | 5.76 | 1 | 2.33 | .01 | .13 |
| Posttrial jury instruction | 4.83 | 1 | 1.95 | <.01 | .16 |
| Forensic testimonial quality | 13.91 | 1 | 5.63 | .01 | .02 |
| Latent print type | 0.59 | 1 | 0.24 | <.01 | .63 |
| FSI Video × Posttrial Jury Instruction | 11.31 | 1 | 4.58 | .01 | .03 |
| FSI Video × Forensic Testimonial Quality | 30.43 | 1 | 12.31 | .02 | **<.001** |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 0.97 | 1 | 0.39 | <.01 | .53 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 0.94 | 1 | 0.38 | <.01 | .54 |
| Error | 1,236.04 | 500 | | | |
| Likeability | | | | | |
| FSI video | 5.60 | 1 | 2.45 | .01 | .12 |
| Posttrial jury instruction | 1.08 | 1 | 0.47 | <.01 | .49 |
| Forensic testimonial quality | 8.01 | 1 | 3.51 | .01 | .06 |
| Latent print type | 4.65 | 1 | 2.04 | <.01 | .15 |
| FSI Video × Posttrial Jury Instruction | 3.68 | 1 | 1.61 | <.01 | .21 |
| FSI Video × Forensic Testimonial Quality | 10.72 | 1 | 4.70 | .01 | .03 |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 1.62 | 1 | 0.71 | <.01 | .40 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 5.09 | 1 | 2.23 | <.01 | .14 |
| Error | 11,400.56 | 500 | | | |

*Note*. Boldface indicates that effects met the adjusted critical value for significance, $p = .01$. FSI = forensic science informational.



the low-quality testimony condition compared with the high-quality testimony condition. Note that these interactions align with what was found for several of the WCS ratings.

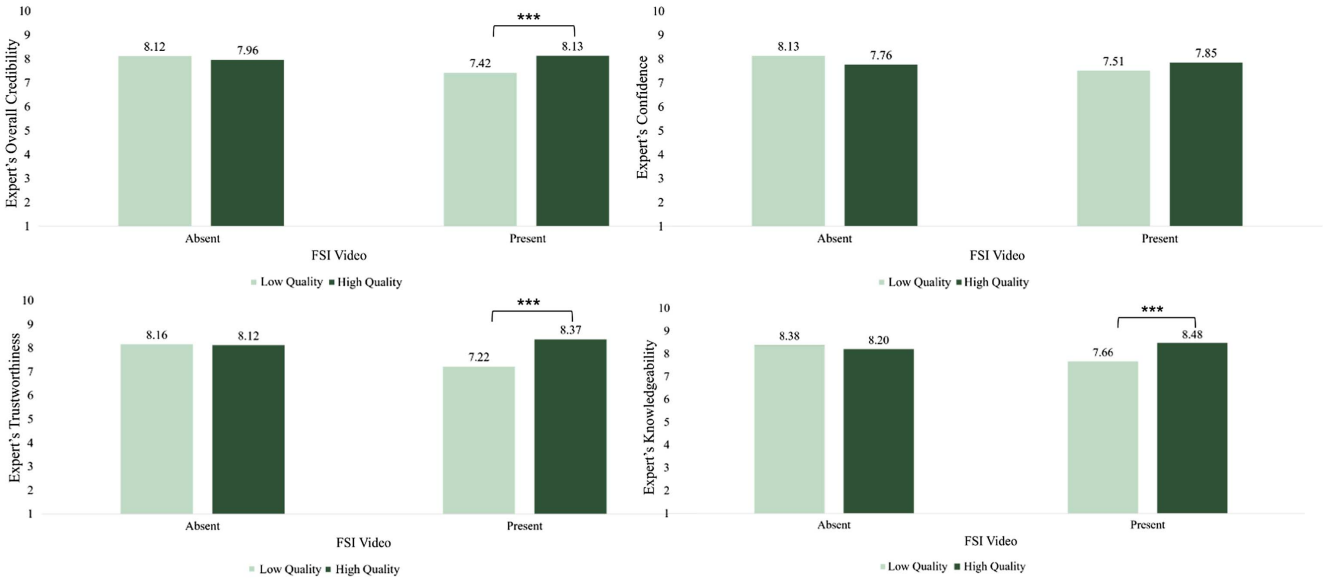
Additionally, there was a main effect of latent print type on participants' ratings of testimony presentation quality, complexity, and the ease or difficulty of understanding the testimony. Participants rated the footwear print testimony as better presented ($M$ = 7.81, 95% CI [7.69, 7.93]), less complex ($M$ = 4.95, 95% CI [4.82, 5.08]), and easier to understand ($M$ = 8.27, 95% CI [8.16, 8.38]) than the fingerprint testimony ($M$ = 7.22, 95% CI [7.08, 7.36]; $M$ = 5.76, 95% CI [5.63, 5.89]; $M$ = 7.79, 95% CI [7.67, 7.91], respectively). No other effects were significant.

## Ratings of the Print Evidence Were Unaffected by the Manipulations

Ratings of evidence strength, quality, and usefulness were analyzed via a series of three ANCOVAs. A Bonferroni correction

**Figure 1**

*Interactions Between Forensic Science Informational (FSI) Video and Testimonial Quality on Participants' Ratings of the Forensic Expert*




*Note*. All of the FSI Video × Testimonial Quality interactions shown were significant ($ps \leq .01$). Asterisks indicate that the *t* tests comparing the testimony conditions were significant (*** $ps \leq .001$). See the online article for the color version of this figure.





was applied to account for multiple comparisons, yielding a critical *p* value of .017 (see Supplemental Table S4, for all print evidence ANCOVA results).

There was a main effect of latent print type on participants' ratings of evidence strength, $F(1, 500) = 20.10$, $p < .001$, $\eta_p^2 = .04$, and evidence quality, $F(1, 500) = 31.03$, $p < .001$, $\eta_p^2 = .06$; specifically, participants rated the footwear print evidence as stronger and of higher quality ($M = 4.94$, 95% CI [4.79, 5.09] and $M = 6.42$, 95% CI [6.27, 6.57], respectively) than the fingerprint evidence ($M = 3.96$, 95% CI [3.82, 4.10] and $M = 5.22$, 95% CI [5.08, 5.36], respectively), perhaps reflecting that participants found the footwear testimony easier to understand (see Sneyd et al., 2020). There were no significant predictors of participants' ratings of the usefulness of the evidence in coming to a verdict, $ps \geq .04$.

## Final Case Decisions Were Unaffected by the Manipulations

For verdict decisions, a binary logistic regression tested the effects of the independent variables on participants' verdict decisions (see Supplemental Table S5, for the full model results). For both verdict confidence and guilt likelihood ratings, ANCOVAs were conducted with latent print type as the covariate. A Bonferroni correction was applied to account for multiple comparisons, yielding a critical *p* value of .017. See Supplemental Table S6 for the full ANCOVA results.

Overall, 81.1% of participants found the defendant not-guilty. The logistic regression model was not statistically significant, $\chi^2(1) = 0.06$, $p = .81$, Nagelkerke's $R^2 = .02$. Thus, no effects are interpreted. There were no significant predictors of participants' confidence in their verdicts or ratings of guilt likelihood ($ps \geq .051$).

## Discussion

The present study is an important first attempt at examining whether instructional videos and posttrial jury instructions increase juror sensitivity to forensic testimonial quality. Instructional videos have been used by the courts (e.g., implicit bias videos), but little is known about their effectiveness, especially in the context of forensic expert testimony. A pretrial instruction video increased mock jurors' sensitivity to the quality of forensic evidence testimony, rather than increasing skepticism of the testimony, relative to mock jurors' evaluations without any such interventions. In contrast, specialized jury instructions had no impact on evaluations.

## Did the Interventions Improve Mock Jurors' Ability to Distinguish Between a Forensic Expert Providing Low- Versus High-Quality Testimony?

In the present study, when presented with an informational video, mock jurors showed sensitivity but not skepticism to testimonial quality, with lower ratings of low-quality testimony than of high-quality testimony. These results suggest that sensitizing jurors to forensic expert testimony standards prior to trial may improve jurors' subsequent assessment of such testimony. Additionally, participants' high ratings of the expert and their testimony in the video-absent condition regardless of testimonial quality is aligned with previous research that suggests that mock jurors rely heavily on an expert's experience when evaluating the evidence and use an

**Table 3**
*Analysis of Covariance Results for Ratings of the Testimony*

| Dependent measure and between-participants effects | Sum of squares | *df* | *F* | $\eta_p^2$ | *p* |
|---|---|---|---|---|---|
| Convincingness | | | | | |
| FSI video | 0.14 | 1 | 0.02 | <.01 | .88 |
| Posttrial jury instruction | 0.55 | 1 | 0.09 | <.01 | .77 |
| Forensic testimonial quality | 20.77 | 1 | 3.22 | .01 | .07 |
| Latent print type | 36.79 | 1 | 5.71 | .01 | .02 |
| FSI Video × Posttrial Jury Instruction | 10.73 | 1 | 3.22 | <.01 | .20 |
| FSI Video × Forensic Testimonial Quality | 43.52 | 1 | 6.75 | .01 | **.01** |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 4.81 | 1 | 0.75 | <.01 | .39 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 1.82 | 1 | 0.28 | <.01 | .60 |
| Error | 3,224.61 | 500 | | | |
| Validity | | | | | |
| FSI video | 3.66 | 1 | 0.68 | <.01 | .41 |
| Posttrial jury instruction | 5.58 | 1 | 1.03 | <.01 | .31 |
| Forensic testimonial quality | 78.33 | 1 | 14.48 | .03 | **<.001** |
| Latent print type | 14.52 | 1 | 2.68 | .01 | .10 |
| FSI Video × Posttrial Jury Instruction | 20.21 | 1 | 3.74 | .01 | .05 |
| FSI Video × Forensic Testimonial Quality | 43.66 | 1 | 8.07 | .02 | **.01** |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 2.48 | 1 | 0.46 | <.01 | .50 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 2.28 | 1 | 0.42 | <.01 | .52 |
| Error | 2,705.18 | 500 | | | |
| Presentation quality | | | | | |
| FSI video | 9.17 | 1 | 1.89 | <.01 | .17 |
| Posttrial jury instruction | 0.58 | 1 | 0.12 | <.01 | .73 |
| Forensic testimonial quality | 44.77 | 1 | 9.23 | .02 | **.003** |
| Latent print type | 41.51 | 1 | 8.56 | .02 | **.004** |
| FSI Video × Posttrial Jury Instruction | 10.00 | 1 | 2.06 | <.01 | .15 |
| FSI Video × Forensic Testimonial Quality | 36.10 | 1 | 7.44 | .02 | **.007** |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 15.68 | 1 | 3.21 | .01 | .07 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 5.11 | 1 | 1.05 | <.01 | .31 |
| Error | 2,424.42 | 500 | | | |
| Complexity | | | | | |
| FSI video | 0.45 | 1 | 0.10 | <.01 | .75 |
| Posttrial jury instruction | 4.24 | 1 | 0.93 | <.01 | .34 |
| Forensic testimonial quality | 9.05 | 1 | 1.99 | <.01 | .16 |
| Latent print type | 87.04 | 1 | 19.15 | .04 | **<.001** |
| FSI Video × Posttrial Jury Instruction | 2.92 | 1 | 0.64 | <.01 | .42 |
| FSI Video × Forensic Testimonial Quality | 0.31 | 1 | 0.07 | <.01 | .80 |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 1.38 | 1 | 0.30 | <.01 | .58 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 2.79 | 1 | 0.59 | <.01 | .44 |
| Error | 2,272.11 | 500 | | | |
| How easy or difficult the testimony was to understand | | | | | |
| FSI video | 2.14 | 1 | 0.58 | <.01 | .45 |
| Posttrial jury instruction | 5.96 | 1 | 1.61 | <.01 | .21 |
| Forensic testimonial quality | 9.70 | 1 | 2.62 | .01 | .11 |
| Latent print type | 27.25 | 1 | 7.36 | .01 | **.01** |
| FSI Video × Posttrial Jury Instruction | 7.19 | 1 | 1.94 | <.01 | .16 |
| FSI Video × Forensic Testimonial Quality | 5.62 | 1 | 1.52 | <.01 | .22 |
| Posttrial Jury Instruction × Forensic Testimonial Quality | 0.31 | 1 | 0.77 | <.01 | .77 |
| FSI Video × Posttrial Jury Instruction × Forensic Testimonial Quality | 0.04 | 1 | 0.01 | <.01 | .92 |
| Error | 1,852.51 | 500 | | | |

*Note*. Boldface indicates that effects met the adjusted critical value for significance, $p = .01$. FSI = forensic science informational.



expert's years of experience as a metric for the accuracy of their conclusions (Koehler et al., 2016).

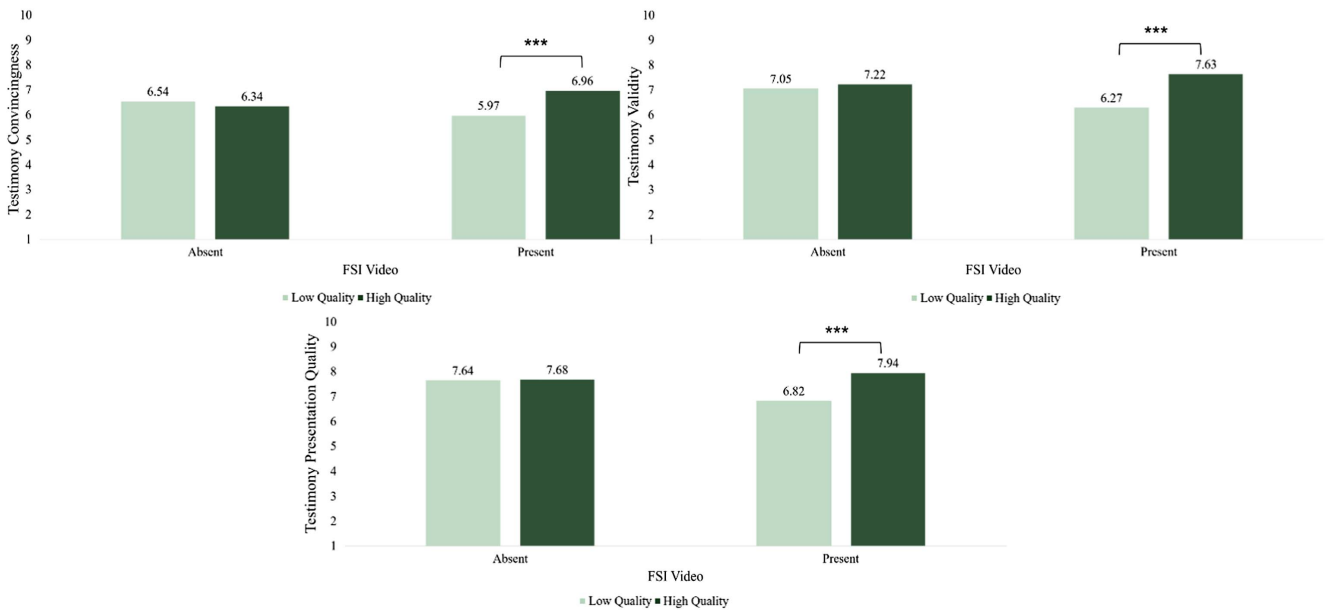
## Did the Interventions Avoid Creating a General Skepticism Toward Forensic Science?

The present study's findings also suggest that with the help of the FSI video, participants were able to evaluate the quality of the testimony more appropriately without either discounting the evidence altogether or inflating the importance of the evidence. Therefore, the FSI video appeared to increase mock juror sensitivity to forensic expert testimony without appearing to increase skepticism of the science overall.

We did not manipulate the quality of the forensic evidence, only the quality of the expert's testimony (e.g., the prints matched and the forensic evidence was of the same quality in all conditions). Participants' ratings of the print evidence may reflect that there was no difference in the evidence's strength or quality. It is promising that participants' ratings of the evidence did not change on the basis of testimonial quality alone, given that testimonial quality is not necessarily indicative of evidence quality. That is, it is important that targeting evaluations of the expert's testimony does not unduly

**Figure 2**

*Interaction Between Forensic Science Informational (FSI) Video and Testimonial Quality on Participants' Ratings of the Validity and Presentation Quality of the Forensic Testimony*




*Note*. All of the FSI Video × Testimonial Quality interactions shown were significant ($ps \leq .01$). Asterisks indicate that the *t* tests comparing the testimony condition were significant ($^{***}ps \leq .001$). See the online article for the color version of this figure.



ONLINE COLOR

influence evaluations of the evidence itself. More research, however, is needed to explore how the relationship between *evidence* quality and testimonial quality impacts the utility of the FSI video.

## Do Improved Assessments of the Testimony Also Lead to Differences in Case Outcomes?

Our findings align with previous research showing that simply altering the language used in forensic testimony to be more cautious and to keep examiners from overstating their conclusions does not significantly impact trial outcomes. B. L. Garrett, Crozier, et al. (2020) found that participants did not differentiate between the more cautious language suggested by the DOJ ULTR guidelines and language that violated these guidelines. Our study expanded on these previous findings and discovered that despite the lack of impact on trial outcomes, there was nonetheless an effect of the language used in the forensic testimony on ratings of the forensic expert and their testimony. This illustrates the potential danger of forensic experts describing their findings in ways that may be misleading, inaccurate, or misinterpreted. It also provides empirical support for the concern that language that violates the ULTR can impact jurors' evaluations of the testimony.

Importantly, the presence of the instructional video did not impact ratings of the latent print evidence or verdicts. This aligns with Koehler et al.'s (2016) study, which found that although mock jurors viewed evidence from scientifically validated techniques as stronger than those not scientifically validated, guilt ratings were unaffected. Additionally, the results of the present study seem to indicate a skew toward a not-guilty verdict regardless of condition. Thus, our specific case facts may have yielded a ceiling effect, rendering it difficult to determine whether the manipulations would have had a measurable effect on verdict decisions outside of the context of the chosen case materials. We pilot-tested two sets of case facts, and although we found that the present study's facts led to more variability in verdicts, the present study found little variability in verdicts. However, it is important to note that the selected case was derived from real-world cases in which the defendant was found guilty and in which the latent print evidence was consistent with the defendant's prints. It is surprising that, given these facts, participants tended to find the defendant not-guilty.

## Policy Implications

On one hand, our findings suggest that video interventions, such as our FSI video, can be useful for jurors' interpretations of forensic testimony. Mock jurors adjusted their ratings of the expert in line with the quality of the testimony they gave with the aid of the FSI video. Further, the video did not shift verdict decisions. This suggests that the forensic evidence in this trial was not pivotal to participants' ultimate case decisions. Alternatively, it is possible that verdicts were unaffected because they are a function of case or evidence strength, which were not manipulated in this study (e.g., low-quality testimony does not necessarily equate to low-quality evidence). Thus, a pretrial instructional video may be a useful tool in providing jurors with additional knowledge that can assist them in evaluating the testimony they hear. It is particularly encouraging that the FSI video did not create a skepticism effect in this study. Thus, we believe the courts would be open to a video intervention of this kind, as similar videos are being used in the Western Washington District Court for implicit bias (Su, 2020). Furthermore,

our video was less than 5 min long. In this way, instructional videos may be a practical intervention to aid jurors in real trials while providing consistency across courtrooms.

On the other hand, our data do not suggest that the specialized posttrial jury instructions have any practical effects on mock jurors' ratings of either a forensic expert, a forensic expert's testimony, latent print evidence, or ultimate case decisions. These findings align with previous research suggesting that jury instructions related to forensic expert testimony often do not impact jurors' ultimate case decisions (Eastwood & Caldwell, 2015; Jones et al., 2020; Papailiou et al., 2015). However, contrary to the prior findings showing that jury instructions can create skepticism in jurors (Dillon et al., 2017), we did not find a skepticism effect produced by our jury instructions. Furthermore, our findings do not lend support to the idea that these jury instructions bolster or impede the utility of the FSI video (or vice versa). Therefore, although the current results do not necessarily support the implementation of specialized, case-specific forensic science jury instructions, they do not suggest there is danger in using them either.

## Limitations

The present study lacked jury deliberations. Previous research has suggested that jury decision-making studies without deliberations are less ecologically valid and thus less generalizable to real-world contexts (Nuñez et al., 2011). However, numerous other efforts were made to increase the ecological validity here, including that participants could review jury instructions while making their verdict decisions in an effort to provide somewhat of a proxy to the deliberation process, in which jurors may refer to the instructions while making their decisions. Although this proxy is not perfect, it does help to increase the overall ecological validity.

Finally, our sample was largely Western, educated, industrialized, rich, and democratic (WEIRD; Henrich et al., 2010). In their study coining this term, Henrich et al. (2010) found that WEIRD populations (particularly American undergraduate populations) tend to be outliers in terms of factors such as cognition, motivation, and behavior. The implication is that findings from studies that use WEIRD populations are not generalizable worldwide. The present study focused on U.S. jury-eligible adults, and thus, we intend to generalize these results only to the U.S. population. However, it is a limitation that our sample was highly educated and slightly liberal leaning and thus not entirely representative of the U.S. population.

## Future Directions

Future research should determine the possible interaction between the overall case evidence strength and the FSI video on jurors' decisions. Given the limitations in the present study regarding the high acquittal rate, it is important to investigate how jurors adjust their evaluations of forensic experts and forensic testimony in light of weak and strong case evidence. Future research should also investigate whether the FSI video's efficacy in aiding jurors in assessing forensic evidence is dependent on the strength of the forensic evidence in a case. This would also help to address whether the lack of effects on the verdict decisions in the present study was a product of the manipulations or due to the case facts used. Thus, future research should test the limits of instructional videos and examine ways to improve such interventions (e.g., exploring the utility of interactive videos) and explore the interaction between testimonial quality and evidence or case strength in the presence of the FSI video.

It is also important to determine at what point in the trial the FSI video should be presented to jurors and whether the timing impacts the effectiveness of the intervention. These considerations may become particularly important in cases that take place over several days or weeks. Relatedly, the present study created specialized posttrial jury instructions. Previous research has found that the timing of the jury instructions can affect jurors' decisions (e.g., Cruse & Browne, 1987; Heuer & Penrod, 1989; Ingriselli, 2015; Kassin & Wrightsman, 1979; Smith, 1991). Thus, future research should explore the effects of the timing of these interventions' presentation on jurors' evaluations of the testimony, the evidence, and their final case decisions.

## Conclusion

The presentation of forensic evidence in court has contributed to wrongful convictions via several mechanisms, including errors during forensic expert testimony and jurors' lack of understanding and evaluation of such testimony. The FSI video created for the present study is a promising start to assist jurors in comprehending and assessing forensic expert testimony with an in-court intervention. The present findings suggest that an instructional video can be a useful tool for helping jurors determine when testimony is of low quality. Although more research is needed, the present study provides support for the use of informational videos in matters of forensic expert testimony.

## References


Acker, J. R., & Redlich, A. D. (2019). *Wrongful convictions: Law, science, and policy* (2nd ed.). Carolina Academic Press.

Bornstein, B. H., & Greene, E. (2011). Jury decision making: Implications for and from psychology. *Current Directions in Psychological Science*, *20*(1), 63–67. https://doi.org/10.1177/0963721410397282

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law*, *28*(6), 892–907. https://doi.org/10.1002/bsl.917

*Castellon v. State*, 302 S.W.3d 568 (Tex. Ct. App. 2009). https://casetext.com/case/castellon-v-state-3

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Cole, S. A. (2018). A discouraging omen: A critical evaluation of the approved uniform language for testimony and reports for the forensic latent print discipline. *Georgia State University Law Review*, *34*, 1103–1128.

Cooper, J., Bennett, E. A., & Sukel, H. L. (1996). Complex scientific testimony: How do jurors make decisions? *Law and Human Behavior*, *20*(4), 379–394. https://doi.org/10.1007/BF01498976

Cruse, D., & Browne, B. A. (1987). Reasoning in a jury trial: The influence of instructions. *The Journal of General Psychology*, *114*(2), 129–133. https://doi.org/10.1080/00221309.1987.9711063

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1990). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Social Psychology*, *20*(14), 1197–1207. https://doi.org/10.1111/j.1559-1816.1990.tb00400.x

Devenport, J. L., Penrod, S. D., & Cutler, B. L. (1997). Eyewitness identification evidence: Evaluation commonsense evaluations. *Psychology, Public Policy, and Law*, *3*(2–3), 338–361. https://doi.org/10.1037/1076-8971.3.2-3.338

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law*, *7*(3), 622–727. https://doi.org/10.1037/1076-8971.7.3.622

Dillon, M. K., Jones, A. M., Bergold, A. N., Hui, C. Y., & Penrod, S. D. (2017). Henderson instructions: Do they enhance evidence evaluation? *Journal of Forensic Psychology Research and Practice*, *17*(1), 1–24. https://doi.org/10.1080/15228932.2017.1235964

Eastwood, J., & Caldwell, J. (2015). Educating jurors about forensic evidence: Using an expert witness and judicial instructions to mitigate the impact of invalid forensic science testimony. *Journal of Forensic Sciences*, *60*(6), 1523–1528. https://doi.org/10.1111/1556-4029.12832

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Garrett, B., & Mitchell, G. (2013). How jurors evaluate fingerprint evidence: The relative importance of match language, method information, and error acknowledgment. *Journal of Empirical Legal Studies*, *10*(3), 484–511. https://doi.org/10.1111/jels.12017

Garrett, B. L., Crozier, W. E., & Grady, R. (2020). Error rates, likelihood ratios, and jury evaluation of forensic evidence. *Journal of Forensic Sciences*, *65*(4), 1199–1209. https://doi.org/10.1111/1556-4029.14323

Garrett, B. L., Crozier, W. E., Modjadidi, K., Liu, A. J., Kafadar, K., Yaffe, J., & Dodson, C. S. (2023). Sensitizing jurors to eyewitness confidence using "reason-based" judicial instructions. *Journal of Applied Research in Memory and Cognition*, *12*(1), 141–157. https://doi.org/10.1037/mac0000035

Garrett, B. L., & Neufeld, P. J. (2009). Invalid forensic science testimony and wrongful convictions. *Virginia Law Review*, *95*(1), 1–97.

Garrett, B. L., Scurich, N., & Crozier, W. E. (2020). Mock jurors' evaluation of firearm examiner testimony. *Law and Human Behavior*, *44*(5), 412–423. https://doi.org/10.1037/lhb0000423

Giner-Sorolla, R. (2018, January 24). *Powering your interaction. Approaching significance: A methodology blog for social psychology*. https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/

Gould, J., Carrano, J., Leo, R. A., & Hail-Jares, K. (2014). Predicting erroneous convictions. *Iowa Law Review*, *99*(2), 471–522.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Heuer, L., & Penrod, S. D. (1989). Instructing jurors: A field experiment with written and preliminary instructions. *Law and Human Behavior*, *13*(4), 409–430. https://doi.org/10.1007/BF01056412

Ingriselli, E. (2015). Mitigating jurors' racial biases: The effects of content and timing of jury instructions. *The Yale Law Journal*, *124*(5), 1690–1745.

Jones, A. M., Bergold, A. N., & Penrod, S. (2020). Improving juror sensitivity to specific eyewitness factors: Judicial instructions fail the test. *Psychiatry, Psychology and Law*, *27*(3), 366–385. https://doi.org/10.1080/13218719.2020.1719379

Judicial Council of California. (2019, March). *Judicial council of California criminal jury instructions*. LexisNexis. https://www.courts.ca.gov/partners/documents/calcrim-2019-edition.pdf

Kassin, S. M., & Wrightsman, L. S. (1979). On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology*, *37*(10), 1877–1887. https://doi.org/10.1037/0022-3514.37.10.1877

Koehler, J. J., Schweitzer, N. J., Saks, M. J., & McQuiston, D. E. (2016). Science, technology, or the expert witness: What influences jurors' judgments about forensic science testimony? *Psychology, Public Policy, and Law*, *22*(4), 401–413. https://doi.org/10.1037/law0000103

Kovera, M. B., & Levett, L. M. (2015). Jury decision making. In B. L. Cutler & P. A. Zapf (Eds.), *APA handbook of forensic psychology, Vol. 2. Criminal investigation, adjudication, and sentencing outcomes* (pp. 271–311) American Psychological Association. https://doi.org/10.1037/14462-010

LaBat, D. E., Goldfarb, D., Evans, J. R., Schreiber Compo, N., Koolmees, C. J., LaPorte, G., & Lothridge, K. (2023). *A forensic science informational video can help mock jurors evaluate forensic expert testimony* [Manuscript submitted for publication].

Lieberman, J. D., Carrell, C. A., Miethe, T. D., & Krauss, D. A. (2008). Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychology, Public Policy, and Law*, *14*(1), 27–62. https://doi.org/10.1037/1076-8971.14.1.27

Martire, K. A., Kemp, R. I., Watkins, I., Sayle, M. A., & Newell, B. R. (2013). The expression and interpretation of uncertain forensic science evidence: Verbal equivalence, evidence strength, and the weak evidence effect. *Law and Human Behavior*, *37*(3), 197–207. https://doi.org/10.1037/lhb0000027

McQuiston-Surrett, D., & Saks, M. J. (2009). The testimony of forensic identification science: What expert witnesses say and what factfinders hear. *Law and Human Behavior*, *33*(5), 436–453. https://doi.org/10.1007/s10979-008-9169-1

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*. National Academies Press.

Nuñez, N., McCrea, S. M., & Culhane, S. E. (2011). Jury decision making research: Are researchers focusing on the mouse and not the elephant in the room? *Behavioral Sciences & the Law*, *29*(3), 439–451. https://doi.org/10.1002/bsl.967

Papailiou, A. P., Yokum, D. V., & Robertson, C. T. (2015). The novel New Jersey eyewitness instruction induces skepticism but not sensitivity. *PLOS ONE*, *10*(12), Article e0142695. https://doi.org/10.1371/journal.pone.0142695

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty & J. T. Cacioppo (Eds.), *Communication and persuasion* (pp. 1–24). Springer. https://doi.org/10.1007/978-1-4612-4964-1_1

Porter, C. N., Morrison, E., Fitzgerald, R. J., Taylor, R., & Harvey, A. C. (2020). Lie-detection by strategy manipulation: Developing an asymmetric information management (AIM) technique. *Journal of Applied Research in Memory and Cognition*, *9*(2), 232–241. https://doi.org/10.1016/j.jarmac.2020.01.004

Ribbers, E., & Henneberg, M. L. (2018). Evidentiary instructions improve mock juror assessment of feature-comparison evidence. *The International Journal of Evidence & Proof*, *22*(3), 262–288. https://doi.org/10.1177/1365712718783427

Smith, V. L. (1991). Impact of pretrial instruction on jurors' information processing and decision making. *Journal of Applied Psychology*, *76*(2), 220–228. https://doi.org/10.1037/0021-9010.76.2.220

Sneyd, D., Schreiber Compo, N., Rivard, J., Pena, M., Stoiloff, S., & Hernandez, G. (2020). Quality of laypersons' assessment of forensically relevant stimuli. *Journal of Forensic Sciences*, *65*(5), 1507–1516. https://doi.org/10.1111/1556-4029.14495

Su, A. (2020). A proposal to properly address implicit bias in the jury. *Hastings Women's Law Journal*, *31*(1), 79–100.

United States Department of Justice. (2018, September 24). *Uniform language for testimony and reports for the forensic latent print discipline*. https://web.archive.org/web/20181030065745

*United States v. Haight*, 15-CR-088, Documents 93-107 (D.D.C. August 13, 2015).

*United States v. Henry*, 14-CR-10319 Day 1 Trial Transcript, Document 145 (D. Mass July 20, 2015).

Western Washington District Court. (2021, May 12). *Unconscious bias* [Video]. YouTube. https://www.youtube.com/watch?v=hdjBbfdRLkA

Received December 12, 2022
Revision received June 8, 2023
Accepted June 10, 2023 ■