Lay Interpretation

of the Questioned Document Examiner's

Verbal Conclusion Scale

BY

EMILY MIDORI NAKAMOTO
B.A., University of Southern California, 2017

THESIS

Submitted as a partial fulfillment of the requirements
for the degree of Master of Science in Forensic Science
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:

AK Larsen Jr. PhD, Chair and Advisor, Pharmaceutical Sciences
Dr. Ashley Hall, Pharmaceutical Sciences
Dr. Donald Waller, Pharmaceutical Sciences
Charles Steele M.S, Purdue University Northwest

Dedicated to my parents, Rod and Elsie,
And my partner Ted.
All of whom supported my fervent data collection
Through these wild Chicago winters.
(-40°F)

Brrrr…
Data…
Brrrr…
Data…

And to my cats, Yam and Rainy,
Who take pride in ensuring that I'm awake
At the crack of dawn
Every day.

Thank you

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor, Dr. Larsen, of the University of Illinois at Chicago. He has been incredibly encouraging and gone above and beyond to help me tackle every hurdle along the way. Even his office, which he stocks full of chocolates and at least 15 plants, is special and will be remembered as a warm work space. I can't thank him enough.

I would also like to thank the experts who advised and cultivated this research project. Professor and committee member Dr. Ashley Hall, statistics advisor and committee member Charles Steele, committee member Dr. Waller, and Psychology Professor Dr. Alex Demos. Additionally, the University of Illinois at Chicago for awarding my research the funding it needed.

I'm also very grateful to Tina Griffin for helping me through the one of the most challenging parts of this research study: starting from scratch. She introduced and taught me how to utilize the resources available. Tina is a hero!

Dr. Ted Hsu deserves a big shout-out for editing my thesis through its numerous, lengthy drafts. Moreover, Ted always made sure I was having fun in Chicago with salsa dancing, running, and baking.

My siblings, Brynn and Scott, have been wonderful cheerleaders throughout my studies. Their optimism has motivated me to keep pushing and striving for more. Similarly, my aunts, uncles, and cousins have offered such support. To Uncle Brian, Uncle Doug and Aunty Kris, Raymond, Aunty Patricia, Duncan, Cameron, Kira, Reni, Colemans, Bonnells, Nakashimas… Love you!

To my friends, I couldn't have done it without you. Katrina, Kallista, Steph, Tran, Kris, Jim, Corinne, Ryan, Vaibhav, Sharon, Dave, Dean, Janice, and more (you know who you are)- thank you for all the awesome mail, hilarious messages, and overall support. It's finally time to celebrate!

Finally, I must express my profound gratitude to my parents for their unwavering support and encouragement. From moving to a new city to wrapping up this thesis project, I can absolutely say that this accomplishment would not have been possible without them. Thank you and love you!

EMN

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

When 83-year-old Modine Wise was assaulted in 1989, her failing eyesight hindered her ability to later identify the perpetrator. Attaining justice relied upon the forensic scientist's analysis and testimony on the two pieces of forensic evidence available, a rape kit and two hairs found on Wise's bed. FBI-trained forensic scientist, Elinos Whitlock, compared these hairs to those of suspect Timothy Bridge and ultimately testified to a "strong identification" in which it was "likely" that the hairs originated from Bridge[1]. Following Whitlock's confident testimony, innocent Bridge was wrongfully sentenced and served 25 years in prison before DNA analysis of the bodily fluid evidence exonerated him. This case is a harrowing example of what can happen when a forensic scientist misrepresents the strength of the evidence and thus misleads a jury. It is a forensic scientist's duty to remain objective, employ validated scientific techniques, and relay conclusions to triers-of-fact in a clear, standardized manner. The research presented herein aims to investigate the conclusion scale terminology used by Questioned Document Examiners in an effort to improve communication between jurors and expert witnesses. By assessing juror opinions for specific phrases, expert witnesses will be better equipped to avoid gross errors such as the overstatement of evidence leading to Bridge's wrongful conviction.

A single criminal trial can require expert testimonies from multiple pattern disciplines in Forensic Science such as Firearms, Latent Fingerprints, and Questioned Documents. In court, pattern experts discuss the relevant evidence and then relay their final conclusion and how confident they are in their findings. Forensic Scientists in each of the pattern disciplines relay their opinions through disparate verbiage, vetted by their particular consensus standards. For example, Questioned Documents currently abide by conclusion terminology standards developed

by the Scientific Working Group for Forensic Document Examination (SWGDOC). In contrast, Latent Fingerprints adheres to consensus body standards vetted by their Scientific Working Group for Friction Ridge Analysis. With each discipline using different conclusion terminology in court, the variation may mislead jurors such as in the case of Timothy Bridge. Jurors may falsely believe that certain evidence conclusions are far more important than others, simply due to differing word choice[2-5]. With this in mind, the federally funded Organization of Scientific Area Committee (OSAC) Forensic Document Examination subcommittee is explicitly requesting further research validating the conclusion scale used by Forensic Document Examiners.

Assessing juror interpretation of conclusion scale terminology requires a preliminary investigation into the many potential confounds. It is known that there are a number of factors that influence triers-of-fact in addition to testimonial content such as contextual information, the expert's gender, and speech style[6-11]. The conclusion scale is also at risk of inducing a "weak evidence effect", which would further separate juror and expert witness understanding[2,12]. Examination of the confounds and terminology already in use will ultimately help conclude what is the most effective way to relay an evaluative opinion.

After a thorough literature review of the many influencing factors, this research project will assess jury-eligible participant's opinions on Questioned Document Examiner's conclusion scale through the use of a well-developed survey and subsequent data analysis. The specific aim of this research will be to evaluate if jurors hold a proportional amount of weight behind each of the nine conclusive phrases in comparison to the Questioned Document Examiner's intended meaning. If this is not the case, data must be analyzed to pinpoint which particular words are causing the miscommunication. We hypothesize that participants will not understand the terminology in alignment with the conclusion scale. In order to test this hypothesis, the following

aims and goals must be met: Aim 1: Develop research and demographic surveys with Goal 1: Design questions and answer choices and Goal 2: Determine how and where to find a wide sampling of jury-eligible participants. Aim 2: Survey validation via pilot studies with Goal 1: Administer pilot studies and Goal 2: Statistical analyses, interpretation, and revision. Aim 3: Final survey administration with Goal 1: Publish finalized and approved survey and Goal 2: Collect and interpret data. Each of these aims and goals are detailed and explained below in the "Materials and Methods" section of this thesis.

In regards to the broader impact of this thesis research, improving communication between experts and triers-of-fact could assist all future criminal trails. It would be ideal for forensic science in the United States to have one validated scale in use by not only the Questioned Documents discipline, but by the entire field of Pattern Evidence (Latent Fingerprints, Toolmarks, Blood Spatter, and Firearms). This would promote widespread standardization and thus eliminate confusing, field-specific variations presented in court. It is critical that a jury understands the evidence conclusions, to what degree of confidence the conclusions were made, and how much weight they should hold to the specific testimony. However, contextual information affects lay person perception of the conclusion scale terms and thus researchers should specifically investigate one forensic discipline's verbal scale within its own context. Researchers have begun seeking alternative methods to verbal conclusion scales such as likelihood ratios and membership functions, but forensic verbal scales still require further investigation. Ultimately, a jury's comprehension of the expert's testimony directly affects the outcome of many criminal cases and therefore the judicial system's ability to identify the guilty and exonerate the innocent.

II. CONCEPTUAL FRAMEWORK AND RELATED LITERATURE

A. Court Cases that Defined Expert Witness Admissibility

In 1923, James Frye appealed his second-degree murder conviction in what was to become a foundational case for the basis of expert witness admissibility. Before further discussion of my thesis research that aims to evaluate specific verbiage used by expert witnesses, it's imperative to understand the court rulings that determined the role and limitations of expert witnesses. Frye v. United States and the subsequent Daubert v. Merrell Dowell Pharmaceuticals Inc. case laid out the specificities defining who may deem a person as an "expert witness" and the standards they must meet. These cases are relevant to Forensic Scientists as it is their duty to not only analyze the criminal evidence but also relay their expert opinion via court testimony when needed.

After confessing to his second-degree murder charge, Frye then retracted his statement and instead wanted the chance to declare his innocence in front of the jury while strapped to a lie detector machine for all to see. The significance of this case is now known as the "Frye standard of general acceptance". In the conclusion of Frye's case, the Federal Appellate Court ultimately ruled against use of the systolic blood pressure deception test and its operator's expert testimony. This set the precedent that expert testimony must be based on scientific techniques that have been generally accepted by the scientific community[13]. The systolic blood pressure deception test had not yet been peer-reviewed nor fully accepted. The requirement for peer reviewed papers are a clear manifestation of the Frye standard; however, the expert testimony must also be relevant, material, and competent[13]. In other words, the techniques used and the following testimony must be of value to the case, reliable, and help the triers-of-fact better understand the issue at hand.

Soon, the Federal Rules of Evidence and the Daubert v. Merrell Dow case would better define the Frye Standard along with a few alterations.

While the Frye Standard was clear about the community acceptance (e.g., peer review) needed for scientific techniques used in testimony, it remained somewhat unclear in regards to who may be an expert and the methodologies behind the peer reviewed techniques. In 1975, the Federal Rules of Evidence (FRE) Rule 702 provided more guidelines. Rule 702 states that the witness must be an expert by knowledge, skill, experience, training, or education and that the testimony must be a product of reliable principles and methods[14]. Daubert v. Merrell Dow Pharmaceuticals, Inc. further added to the specificities when Merrell Dow was accused of selling medications that caused birth defects. As there were unpublished pharmacological studies indicative of birth defects *in vitro* and *in vivo* animal studies, the issue was whether or not those data met the Frye Standard of general acceptance. With FRE Rule 702 support, the Supreme Court ruled that the admissibility of expert testimony does not depend only on Frye's general acceptance and conclusions generated by scientific experiments, but rather on the methodology and principles as well. Thus, the preliminary data of sound methodology was allowed to serve as evidence against Merrell Dow. Furthermore, this case determined that trial judges are the final arbiter or "gatekeeper" on admissibility of evidence and acceptance of an expert witness per trial[15]. For each court case in which a forensic scientist is called to testify, the judge reviews their qualifications and reports before deciding whether or not the person may be deemed an expert.

Following the mentioned updates and alterations, Joiner v. General Electric (1997), Kumho Tire v. Carmichael (1999), and United States v. Melendez Diaz (2009) further refined the rules regarding expert witnesses. These three cases are now known as the "Daubert Trilogy". Joiner v. General Electric and Kumho Tire v. Carmichael made refinements including the claim

that there's no need to draw a distinction between "scientific" and "technical" knowledge, additional support for the judge as gatekeeper with an "abuse of discretion" clause, and clarification that the judge must focus on methodology and techniques rather than on the conclusions found[16,17]. Most relevant to forensic scientists was the precedent set by the third case of the Daubert trilogy, United States v. Melendez Diaz. This case cited the 6th amendment of the Bill of Rights, and stated that since the accused have rights to confront witnesses against them, the forensic scientist who worked the evidence may not submit an affidavit but instead be present at trial, aside from certain exceptions[18,19]. Since this trial concluded, analysts now receive many subpoenas and there is a decreased risk of misinforming the jury because the analysts themselves are most often speaking for their own work. As these alterations to the Frye standard came to rule, the precedent set by the Daubert v. Merrell Dow case was set as the federal guideline and as a result, states are allowed to choose which ruling to follow for their particular state jurisdictions[13]. Currently, around 16% of states in the U.S adhere to Frye standards whereas 78% of states follow Daubert guidelines[20]. The remaining 6% of states judge expert witness testimony via their own individualized approach.

In summary, the role of an expert witness, either under the Frye standard or Daubert ruling, is to bring knowledge to the trier-of-fact in order to help the case come to a fair sentencing. Forensic scientists give a voice to the physical evidence by following techniques accepted within the appropriate scientific community. Each judge per trial will assess the qualifications and methodologies of the forensic scientist and then decide whether they may serve as an "expert witness" or not. Furthermore, the United States v. Melendez Diaz case mandates that the forensic scientist be present at the trial when necessary. Expert witnesses are a

critical part of the justice system as they are needed to explain their analytical process/knowledge to the jury and the accused must be given the chance to question their methods and results.

B. Expert Testimony Factors Affecting Jurors

While it's important to understand the court cases that set precedent for expert witness admissibility by a judge, it is equally necessary to understand how jurors perceive an expert's qualifications. My thesis research will examine juror understanding of specific words used by the experts, but first previous literature on additional expert witness effects must be reviewed. For example, the expert's personality, presentation style, gender and even speech hesitations have been shown to affect juror's opinions on the case evidence presented. Researching these effects is critical because trials by jury rely on the jury's full understanding of the evidence so they may reach a fair conclusion. Additionally, the Constitution of the United States Amendment VII states that, "no fact tried by a jury, shall be otherwise re-examined in any Court of the United States…"[19] Thus, it is very uncommon for a judge to deem a particular jury incompetent and as a result, initial jury conclusions stand with finality.

1. Juror Judgement of Expert Witness Qualifications

The prosecutor and defense offer the jury polarizing arguments and expert witnesses who are meant to corroborate their differing theories. Litigators may fear that the juries will be too gullible when listening to the opposing argument, however numerous studies have shown that juries are quite aware of the adversarial process[21]. Jurors exhibit skepticism of expert testimony and after hearing all arguments, they do not explicitly mention how superficial factors such as an expert's personality or other physical attributes affect their opinion as many researchers hypothesized they would[22,23]. However, it is known that there are many factors implicitly affecting jurors despite their outward denial of such influence. Specifically, when the prosecutor

and defense both have engaging expert witnesses offering contradictory statements, the jurors begin to rely more heavily on their general impressions of the experts testifying[6]. Ivković and Hans assessed the breadth of external influencing factors by examining 269 jurors and recording 55 semi-structured interviews. They questioned jurors about how they decided upon a verdict and asked them to discuss the factors that affected their decision. The juror's responses were then analyzed and roughly split into two main categories: discussion on the personal aspects of the expert and comments about the testimonial content. Results showed that 33.2% of the comments discussed the expert's credibility (e.g., presentation style, clarity, enthusiasm), 31.1% were about the expert's credentials and motives (e.g., institutional affiliation and sympathy), 20.6% of comments voiced general impressions of the expert (e.g., attitude and personality) and only 15% of comments were about testimonial content[21]. This research article highlighted the fact that there are numerous sources of expert witness influence that reach far beyond the content of testimony alone.

2. Gender of Expert and Complexity of Testimony

Numerous published articles have documented gender differences in which women as experts are rated less favorably than men[24], however more current research has honed in on which specific situations exacerbate the differences[25]. A review of this literature is necessary to understand potential biases perceived during forensic science expert testimony for differing case subjects (e.g., homicides, crimes against children, sexual assaults) and the relation between gender and testimony complexity.

In regards to domain-related gendered effects, it has been shown that jurors are influenced by the expert's gender when the case subject involves battered women. When the expert witness is female, the female jurors were more likely to believe the battered women's

claims of self-defense against her abuser. Male jurors responded to the female expert witnesses with an even more pronounced gendered effect of believing the defendant when compared to male response to male expert testimony[7,26]. Furthermore, there are gendered effects when court cases fall under the domain of child abuse. Both male and female jurors more strongly supported the abused children's self-defense claims when the expert testifying was a woman[10]. It has been speculated that these effects arise due to social constructs of traditional gender roles and stereotyping. In the domain of child abuse, jurors may assume that women are better at understanding a child's needs and therefore are more qualified as an expert in this subject. Domain specific gendered effects are biases that must be taken into consideration when trying to understand how jurors are affected by factors other than testimonial content alone.

Expert witnesses are tasked with relaying highly technical information to a lay audience. Interestingly, the gender of the expert and the complexity of their testimony are intertwined. Research suggests that jurors respond more favorably to expert witnesses who speak in alignment with stereotyped language expectations. For example, it has been shown that jurors prefer male experts testifying with either complex or simple language and yet are less persuaded by female experts using complex language[11]. These results are partially explained by the fact that the participants held gender related language expectations in which complex testimonies were seen as masculine in contrast to simpler, more understandable testimony being associated with femininity. Jurors perceived a female witness speaking in complex terms to be gender-incongruent and therefore less effective[11]. Additional research on this topic revealed surprising results. It was expected that when male and female experts relayed their findings with simple terms, that jurors would accept the information equally. However, it was instead found that in this situation, mock jurors prefer the female's simple testimony over their male counterparts[8]. This gendered bias

related to testimony complexity can be best decreased if all experts relay their findings as clearly as possible to the jury[27] and by studies that specifically research effective word choice.

3. Expert and Translator's Speech Hesitations

As there are many ways to phrase a sentence, different expert witnesses utilize varying linguistic approaches when testifying in court. These variations are exacerbated when witnesses require a translator because of language barriers. Previous literature has shown that jurors' estimation of the defendant's culpability is reliant on whether the expert spoke with either "powerful" or "powerless" speech. Powerful speech was characterized as blunt statements made in active voice whereas powerless speech included filler words such as "I guess" and "I'd really appreciate it if", question intonations, and polite phrases in passive voice. Results clearly showed that jurors rate witnesses who use powerful speech as far more credible, believable, and trustworthy than those who use powerless speech[28]. However, when further investigated, researchers found a dramatic cultural difference in juror reaction to powerless and powerful speech, because these aspects are interpreted differently. For example, it was found that powerless speech is considered a positive attribute in Latin America as the wording is politer. Contrastingly, Anglo Americans strongly prefer powerful speech as their culture positively supports concise language. This was demonstrated by having mock jurors listen to separate testimonies with a controlled difference, which created either a powerless or powerful testimony, followed by opinion assessment. Hispanic participants found no difference between the two testimonies whereas Anglo Americans found the powerless speech testimony to be incompetent and untrustworthy[9] . These findings were further supported by research showing that jurors are more likely to judge the expert witness positively when the testimony was spoken in their

dominant language[29]. Current research has taken these research aims, gone forward, and investigated the nuances.

The topic of expert speech hesitations and its effects related to translators is still extremely relevant as many court trials use translators throughout the entire process. Research by Mendoza and colleagues honed in on this issue by creating an expert testimony control (with speech hesitations and without) and used Spanish-speaking, bilingual (Spanish and English), and English-speaking participant groups. Results show that jurors were impartial to whether or not the testimony was translated (both did not include any speech hesitations). This was interpreted by the fact that the three participant groups were equally likely to convict the suspect after both an expert witness and a translator gave the testimony[30]. However, there were cultural differences relating to the witness' speech hedges and hesitations. Bilingual and English-dominant jurors found testimony with hesitations to be less trustworthy, which is in agreement with previous literature[9,29], whereas Spanish-dominant jurors were unaffected by hesitations as their culture favors polite speech forms. What separates this newer study from previous research is that the bilingual juror effects are in alignment with English-dominant jurors; it's speculated that the American bilingual jurors are more comfortable and familiar with English preference for concise verbiage and therefore find this style to be more persuasive. Additionally, this study found that the translator's speech hesitations and hedges did not impact verdict sentencing by any of the three participant groups[9]. This research is critical as 96.8% of court interpreter requests in the United States are for Spanish translation[9,31]. Thus, the effects of the translator are less than what was emphasized by previous literature, but the difference in powerless and powerful speech by the expert witness remains an influencing factor on jurors' opinions.

C. <u>Conclusion Scales</u>

1. Administrative Divisions Cause Differing Conclusion Scales

Forensic science is categorized into numerous disciplines, each of which are under the administration of the National Institute of Standards and Technology (NIST) Organization of Scientific Area Committees (OSAC). OSAC aims to provide standardization and ensure that all forensic analyses are executed with reliability and reproducibility. Specifically, Scientific Working Groups (SWG's) under OSAC oversight are in charge of validating and vetting the specific language that their forensic experts are allowed to use in courtroom testimony. These words constitute the "conclusion scale" of phrases that are relied on by expert witnesses to explain their degree of confidence in their evidence conclusion to a jury. Since SWGs are convened for each forensic science discipline and are charged with standardizing language for their section, a large issue arises from the differences between the SWG disciplines. For example, the SWG for Friction Ridge Analysis, Study and Technology guides their experts to use the conclusion "individualized" in order to state the highest degree of confidence such as in the testimony, "the print was *individualized* with the suspect"[32]. Contrastingly, the SWG for Questioned Documents (SWGDOC) guides their experts to use the word "identification" to express the highest degree of confidence in a conclusion such as in the statement, "the questioned document was *identified* as the handwriting of John Doe"[33]. The issue is that this differing verbiage between forensic disciplines may confuse jurors who need to weigh the relative importance of the evidence presented. Many criminal trials involve multiple pieces of evidence from the various disciplines, which cause jurors to hear differing expert testimonies one after the other. The lack of cohesion is a problem that requires attention and ideally these verbal scales must be researched within the correct context.

2. Questioned Document Examiner's Conclusion Scale

My thesis research is studying the Questioned Document Examiner's conclusion terms to seek insight on lay interpretation. As dictated by SWGDOC, there are nine terms for Questioned Document Examiner expression of evidence conclusions. These phrases include the following: identification, strong probability, probable, indications, no conclusion, not indicated, probably did not, strong probability did not, and elimination[33]. Currently, there are no validation studies available to the public that assess these conclusive phrases.

3. Phrases to Avoid

Some researchers are in favor of standardizing verbiage, but warn against the use of specific words that were common place in forensic science testimony, but are now avoided. It has been noted that the words "consistent with" and "match" should not be used[34]. For example, the phrase "consistent with" has been used in testimony such as, "the handwriting sample was consistent with the suspect, John Doe," but this word lacks relevance. Most lawyers and scientists used this phrase to imply an opinion of strong support, however it confuses jurors because it could additionally be said that, "the handwriting sample was also consistent with John Doe's neighbor." This phrase can too easily mispresent the causality of the evidence presented. Use of the word "match" has also been discouraged because it has multiple meanings. "Match" can either mean that two objects share a characteristic (e.g.,, same material or other class characteristic) or that the two items came from the same source. Furthermore, it has been argued that "match" is overly conclusive and can cause the jury to hold excessive weight unto the testimony. Conclusion scale verbiage has evolved from problematic wording like "match" and "consistent", but currently used and newly proposed words must be carefully considered[34-36].

D. National Academy of Science (NAS) 2009 Report

In 2009, the National Research Council of the National Academies published an infamous critique of forensic science, *Strengthening Forensic Science in the United States: A Path Forward*[36]. This report brought many issues to light such as the lack of error rates associated with Pattern Evidence disciplines, forensic science origin and current ties to law enforcement, rigor of quality assurance guidelines and many more. While this report immediately sparked discussion and some polarizing commentary, this NAS report comprehensively critiqued current conclusion terminology and provided useful recommendations to help resolve these issues.

In regards to reporting and testifying standards, the NAS report mirrored critiques discussed previously about the pitfalls of certain phrases. For example, it's noted that "matched," "consistent with," and "similar in all respects tested" are terms that can "have a profound effect on how the trier of fact in a criminal or civil matter perceives and evaluates evidence"[36]. The article highlights the impact of conclusion terminology and how forensic disciplines have not yet reached a consensus to define these phrases or proceed with standardization. It is recommended that validation studies, such as this thesis research, be conducted in order to establish standardized terminology. Additionally, they emphasize that all forensic scientists should be mandated to use standardized evidence conclusion scales as a prerequisite to attaining laboratory accreditation and certification[36]. If followed, these recommendations would help minimize miscommunications between forensic scientists themselves and triers-of-fact.

E. Previous Research on Probabilistic Terminology

Standardizing words to be similarly interpreted is an extremely difficult task due to intra-individual and inter-individual variation as well as overall uncertainty. The conclusion scale is a list of evaluative terms meant to represent degrees of uncertainty ranging from "vague

uncertainty" to "precise uncertainty". Individual variation and uncertainty are thoroughly discussed in the following sections as these factors provide insight on what contributes to a successful conclusion scale.

1. Verbal Versus Numerical Probabilities

When assessing the best way to relay uncertainty to jurors, it's necessary to consider communicating with either probabilistic terminology or numerical values. For example, a forecaster could either say, "There's a 3 in 4 chance it will rain," or "It's highly likely it will rain." Research shows that people prefer to communicate their uncertainties with probabilistic terminology, but receive uncertainty information via numerical values[5,37]. Probabilistic language is defined as evaluative terms meant to qualitatively discuss a range of uncertainty. Some research articles conclude that numerical probabilities and verbal terms have a reasonable degree of correspondence in which no significant differences are observed[12,38]. However, there is contrasting research that indicates numerical probabilities are more precise and better at keeping people on the same page[2,37,39]. Many researchers have investigated probabilistic words/numbers and their differing opinions are still very much in contention.

The support for use of numerical probabilities is largely due to intra and inter-individual variations. A considerable body of literature has demonstrated that individuals have a large lexicon for describing uncertainty and favor specific words over others[2-5]. This causes people to both use and interpret probabilistic verbiage differently from one another, which is known as inter-individual variation. The implication is that this could be a prominent issue when one expert witness is trying to describe their uncertainty to a group of jurors interpreting words differently from one another. Intra-individual variation is seen when the participants are asked to assign a numerical value (0 to 1) of certainty to a list of probabilistic terms. Results of this task

show that most synonymous terms are assigned slightly different values, but some have discrete, precise values[4,40]. For example, the phrases "it's a toss up" and "even odds" are always rated at 0.5. In conclusion, some researchers insist on the use of numerical probabilities because they are precise and avoid issues of inter and intra-individual variation that arise from people's different lexicons and preferences.

In favor of probabilistic verbiage, it has been noted that verbal statements allow for greater nuances to be expressed. Budescu and Wallsten clarified how, "rich semantic structure allows one to convey not only approximate location and degree of imprecision, but also relative weights over levels of uncertainty within an implied range, and perhaps also other aspects of the communicator's knowledge or opinions beyond degrees of uncertainty"[4]. Additionally, research has indicated that juries disliked numerical reasoning and that it encouraged them to either dismiss the evidence entirely or cling to it above all else[34,41]. There are also well received examples of probabilistic verbiage in use for mass understanding such as the terms used by weather forecasters to indicate the likelihood of specific weather (e.g., slight chance, chance, likely). While probabilistic verbiage use is successful for weather forecasters, this thesis research needs to analyze how jury-eligible participants understand the Questioned Document Examiner's probabilistic verbiage.

2. Pattern Evidence Analysis Struggles to Give Precise Uncertainty

Pattern evidence disciplines include Firearms/Toolmarks, Friction Ridge Analysis, and Questioned Documents, which all rely on the expert's processing and ultimate visual examination of unique markings. After forming a conclusion, the expert then announces their findings via testimony, but struggles to speak of their precise uncertainty. Precise uncertainty is defined as the uncertainty due to quantifiable, random variation. For example, the forensic DNA

scientists can analyze an individual's profile generated from a blood drop and then calculate the precise uncertainty in regards to the likelihood of finding this particular profile in the population. This can be done by knowing allele frequencies in the population and then quantifying the random variation between groups, which is known as random match probability[42]. Forensic pattern evidence scientists on the other hand cannot give the trier-of-fact a precise uncertainty associated with their conclusion because there are no frequency statistics to be calculated. For example, a Questioned Document Examiner may compare a ransom note to the suspect's handwriting and conclude that there is a strong probability that the suspect wrote the document. However, there is no known probability numeral of uncertainty because there is no database of handwriting and they cannot examine all other handwriting samples in the world. Thus, the examiner cannot testify to a precise uncertainty in their conclusion. Pattern evidence disciplines cannot relay uncertainty probabilities via numbers, but instead can discuss vague uncertainties via verbal expressions.

Since pattern evidence scientists struggle with precise uncertainty, it has been suggested that error rates should be included in testified conclusions[36]. The infamous case, *Mayfield v. United States*, brought this issue to center attention when US citizen Brandon Mayfield was wrongly identified by FBI examination of his prints against those found on explosives involved in the Madrid terrorist bombing[43]. Prior to this case, latent print examiners generally claimed that they only testified to identifications when they were 100% confident and therefore the error rate must be zero. However, the Mayfield case made it clear that although there is no quantified error rate at the moment, it is more than zero[44]. The NAS 2009 Report addressed this issue by recommending that pattern disciplines further support the conclusions and methods with, "meaningful scientific validation, determination of error rates…reliability testing to explain the

limits of the discipline"[36]. Research is currently examining proficiency tests and the Analysis Comparison Evaluation Verification (ACE-V) methodology used by pattern disciplines, which will help tackle the issue of error rates and better communication of expert findings to triers-of-fact[45]. As this is addressed, it's important to keep in mind while investigating conclusive phrases and potentially drafting new conclusion scales.

3. Probability Phrase Context Dependence

Previous research indicates that there are widely variable inter and intra-individual differences in regards to probability term interpretation. Investigating this further, researchers have studied whether or not these individual differences for certain words remain stable over varying contexts, such as discussions on medical diseases versus weather forecast. Researchers took probability phrases and embedded them into paragraphs of differing contexts such as severe illness or hostile activity risks. Specifically, the paragraphs used the same probability verbiage (dependent variable) to either discuss the chance of rain versus the chance that a sport related twisted ankle was either sprained or broken (independent variables). Participants were asked to describe the likelihood of events and results showed that inter and intra-individual variance fluctuated greatly depending on context[39]. Indeed, several investigations support the notion that this large variation is due to individuals imposing their own judgements and prior beliefs on the contexts rather than focusing on the presented probabilities alone and because the context paragraphs included vague terms about the situation, which may have caused additional participant speculation[4,5,46]. For example, the National Weather Service (NWS) uses certain terms to describe the probability of precipitation such as "slight chance" or "likely." When these standardized terms were applied to the likelihood that a sports-related twisted ankle was either sprained or broken, participants interpreted the probabilistic terms very differently[47]. These

results highlight the fact that if this thesis research validates the Questioned Document conclusion scale verbiage, the terms cannot be assumed to work for all pattern evidence disciplines. New studies would be necessary for the differing contexts (e.g.,, friction ride analysis and firearms/toolmarks).

4. Weak Evidence Effect

When probability terms are used to express opinions ranging from no chance to strongest likelihood, there is an interesting weak evidence effect that occurs for low-supported opinions. It has been shown that when an expert endorses weak or limited support, participants surprisingly interpret this as negative support rather than simply weakly positive support. This weak evidence phenomenon has been observed in many different contexts by numerous researchers[12,48]. The Questioned Document conclusion scale is a list of evaluative terms used by forensic scientists to help relay uncertainty information to triers-of-fact and its range includes negative endorsement as well as positive endorsement. For example, the conclusion scale includes "strong probability did not" and "strong probability did." While investigating this conclusion scale, it will be important to note if there are weak evidence effects for both the weakly supportive conclusion, "indicated," and its negative counterpart, "not indicated"[33].

5. Likelihood Ratio

While this thesis research is investigating the verbal conclusion scale of Questioned Document Examiners (QDE), some people advocate that forensic scientists move away from probabilistic terms and instead move towards membership functions and likelihood ratios. Currently, examiners in the United States pattern disciplines are expressing their opinions by using the conclusion scale terms to relay their confidence/uncertainty. In comparison, a likelihood ratio (LR) estimates the amount of support given for a proposed statement. For

example, an expert could use a verbal or numerical likelihood ratio to support the prosecution's theory or conversely support the defense's statement. There is a growing amount of literature as researchers aim to understand lay perception of likelihood ratios. Results have shown that LRs in pattern evidence suffer from both the weak evidence effect and contextual influence, and that participant's interpretation revealed substantial undervaluing of the expert's intended meaning[12,49]. Despite mixed results, likelihood ratios were endorsed and standardized by the forensic science representative body of the UK and Ireland, Association of Forensic Science Providers, and others[35,50,51]. Researchers continue to investigate this topic further[52] and have even begun to explore new ways of quantifying probability terms (e.g., membership functions)[53].

F. Previous Literature on Lay Interpretation of Verbal Conclusion Scales

While there has been a shift towards likelihood ratios and other alternative methods for expert witnesses to express their opinions, forensic verbal conclusion scales have yet to be adequately reviewed. The few available papers on this topic have concluded that trace evidence verbal scales, in general, are not precise enough to be recommended for court use. However, each of these studies sheds light on differing nuances, has unique limitations, and utilizes slightly varied methods; it is necessary to review these works before conducting further conclusion scale research.

Each of the forensic conclusion scale research projects employed a similar design and answer choice format. Participants were given an expert witness testimonial statement, which used a conclusion scale verbal term, and then were assessed for their opinions of the statement[12,49,53-56] . All of the research projects presented participants with a sliding scale answer choice with descriptive ends; without discrete numbers or options, the sliding scale avoids pre-binning participants' answers. For data analysis, each of these researchers superimposed numbers

onto the sliding scale in order to quantify participant answers. However, each of these papers used different measurements (e.g., 0-10, 0-20) and were either equidistant or logarithmic. This thesis research project will be the first to use a 0-8 Likert scale answer choice. The differences and thought process for choosing to use a Likert scale opposed to a sliding scale is thoroughly documented in the "Discussion on the Materials and Methods" section below.

Most researchers directly analyzed lay interpretation of the verbal terms via the sliding scale methodology, but researchers have also utilized indirect assessment[12]. This involved giving participants details of a mock case, asking for their prior-belief of the defendant's guilt, exposing them to a shoe mark expert testimony, and then re-evaluating their belief of guilt. Participant interpretation of the verbal term was indirectly evaluated by their belief change. While this methodology is rooted in Bayesian theory, our thesis research has chosen to directly assess participant understanding in an effort to avoid any potential biases that may arise from the participant exposure to too many details of the mock case.

Previous literature on this topic has also utilized different participant samples[12,49,53-56] . For example, researchers have used convenience samples of undergraduate students, online sampling via Amazon Mechanical Turk, and particular groups of people such as legal professionals. Undergraduate students are a group of people similar in age and educational level and thus may not be the best representation of a typical jury. One study speculated that participants of a higher education may have had prior knowledge of forensic science and/or evidential proceedings and possibly subjected the verbal terms to even more scrutiny than the average juror[54].

Despite varying methodology, research on forensic verbal terms have shown similar results concluding that pattern evidence conclusion scales are not well understood. For example,

participants' understanding of the conclusion scale was varied and that terms of lesser support were surprisingly perceived as being weighted more heavily. In contrast, terms of higher support were perceived to have less weight[54]. A weak evidence effect was also noted in which a verbal term of weakly positive support was interpreted as weakly negative by participants[12]. Multiple researchers concluded that participants' responses were generally in alignment with the conclusive terms (e.g., increasing weight given to more supportive verbiage); however, the range of responses per term greatly overlapped and thus the terms were imprecisely and inaccurately understood. Furthermore, data revealed that participants struggled to differentiate the verbal phrases, "strong" "very strong" and "extremely strong" and as a result, the verbal scale terms were poorly resolved[55]. Ultimately, results of the previous literature highlight critical issues with lay interpretation of conclusion scales and conclude that research efforts should continue to investigate forensic science expression of uncertainty.

Previous papers on this topic are generally in agreement that verbal scales are not well understood, but each project had its own unique limitations and methodology. No research project is without limitations and each of these papers have greatly contributed to a future of effective communication between triers-of-fact and experts. This thesis project aims to build on this literature, address the limitations, and specifically research the United States Questioned Document Examiner's conclusion scale in the context of questioned document evidence. Specifically, we have pooled a large, diverse participant sample, directly measured participants' verbal term interpretation, offered minimal mock case information, and have employed an eight-point Likert scale answer choice.

G. Amazon Mechanical Turk and Selection Strategy

The online platform, Amazon Mechanical Turk (MTurk), has been used to poll participants across the United States for this thesis research. MTurk is an online platform from which many individuals are paid to complete tasks such as research surveys. The use of MTurk over traditional sampling methods, such as convenience samples or college campuses, was decided based on numerous factors. Deciding factors included breadth of demographics, attention rates, and the benefits of online sampling. Furthermore, factors such as monetary compensation for participants, average MTurk wages, and associations between earning potential and character misrepresentation all influenced the use of MTurk and selection strategies[57-79].

1. Online vs. Traditional Sampling Methods

As many people in the United States have Internet access, it is becoming more common for researchers to consider using online surveys in place of traditional sampling methods. With this transition, an examination of differing sampling results has become an important task. Numerous papers have published on this topic and conclude that online sampling, MTurk included, results in an extremely similar sampling pool in comparison to that of in-person methods[64,71,77-79]. Reported attitudes, interests, ethnic diversity, and socio-economics of online-sampled participants have even been reported as more diverse than the typical convenience sample[78]. Furthermore, multiple benefits of online sampling have been highlighted. Online surveying can accumulate a larger sample size in a much shorter amount of time, question randomization and skip logic can be used, and there is lessened observer bias[71]. The well-known Hawthorne effect refers to the way in which research subjects respond differently when they are aware that the investigator is present and evaluating them. While the specifics of Hawthorne's original study have been debated, the significant effects of observer bias have since been demonstrated in a variety of circumstances[69,70]. Fortunately, with online surveys, participants are

not influenced by observer bias from the physical presence of the research investigator. Additionally, online survey's question randomization and skip logic abilities have proven to be important advantages as this research project relies on these particular user-friendly features. The use of online survey platform MTurk is beneficial to this research project in a variety of ways, however it comes with its own unique set of problems worth reviewing.

2. <u>Inattention Rates</u>

While online surveys come with benefits, such as question randomization, lack of observer bias, and a diverse participant pool, they suffer from greater inattentive response potential. In support of this, research has shown that people taking online surveys are prone to multitasking and checking their cellphones, both of which would not happen in a controlled, in-person study[59,68]. Studies on MTurk emphasize that a survey's collected data averages 15-20% inattentive responses, which have been determined via instructional manipulation checks defined in the immediately following section of this thesis[74]. This percentage size is concerning, however, literature has shown that this is very similar to inattention rates from traditional sampling methods[63,74]. Online survey-takers may have the greater opportunity to multi-task, but inattention rates remain the same compared to convenience and college samples. Some papers even suggest that Mturkers are slightly more attentive than participants from traditional samples[66,72,73]. This is partially explained by the situation that MTurkers are people who consistently take surveys, unlike college students who may only take a handful of surveys for one psychology class. In other words, MTurkers are a group of people who progress and have trained to focus on inattentive question checks and subtle details in the survey's directions[68,72]. As traditionally sampled participants and Mturkers are similarly inattentive, researchers have honed in on and debated methods to check participant inattention.

3. Instructional Manipulation Checks

Survey responses can be difficult to label as "inattentive" because the response may be indicative of a poorly-worded question, participant's honest attempt, or true inattention. After considerable research, this research project will not be checking for inattention. The most common and widely accepted method of assessing inattention is via instructional manipulation checks (IMCs), in which the survey sets up a question, but then instructs the participant to select a particular answer even though it may not be the intuitively correct answer choice[57,58]. By using IMCs, a researcher can assess whether the participant is paying attention to the questions or merely filling in responses. IMCs are widely accepted, however there is contrasting literature that claims IMCs can cause a priming effect on participants, after which participants may approach the rest of the survey with systematic thinking in preparation for other tricky questions[72,73]. Furthermore, attention ebbs and flows, but IMCs only assess a participant's attention at one specific moment[72]. With the pros and cons of IMCs weighed, this research project will not use attention checks, however steps have been taken to limit inattentive issues and are discussed below.

4. Guarding Against Inattention: Brevity, Value, and Monetary Compensation

The choice to not include attention checks for this research survey was based on the literature discussing reasons why people do pay attention to surveys. For example, it has been shown that participants pay more attention to a survey if they see value in the research[63]. We considered using a preface message discussing the relevance of the project and how language in courts affects criminal sentences for this research project. However, this was decided against because of the potential bias it could introduce[80]. Instead, the research survey has been titled, "Forensic Science Graduate Research Survey," in an effort to emphasize how this survey is

academically based, without revealing biasing details. Results of previous literature also indicate that fewer survey questions are directly related to attention levels[61,62,81]. As the research survey is only expected to take approximately 6 minutes, with a total of 8 research questions followed by a simple demographic questionnaire, inattention is not expected to be a significant factor. In conclusion, participant inattention is a problem seen not only in online surveys, but also convenience and college samples. This research survey has been kept short and given a title that showcases its greater impact in an effort to guard against inattention and thus will not be assessed via IMCs also due to cautioning literature on the topic.

Inattention is a problematic behavior shared by MTurkers and traditionally sampled participants alike, but research shows that there are a concerning number of Mturkers who misrepresent their identity and consumer information in order to qualify for surveys and therefore more earning opportunities[75]. Moreover, there is no validation from Amazon Mechanical Turk that could guarantee participants have entered their correct demographic information. However, researchers have identified the specific motives behind character misrepresentation and thus, we can minimize the issue as best as possible for this research project. Character misrepresentation arises due to economic motivation, pre-screening questionnaires, and rare populations[67,75]. To lower economic motive, the research project's roughly 6-minute survey will be priced at $1/survey completion, which is in alignment with average Mturk wages. Average MTurk pay is $7.25/hour. There are many researchers paying $11/hour but a majority of low paying tasks brings the average down to $7.25/hour[60]. Offering a wage in between $7.25 and $11.58 per hour, averages to $0.16/minute ($0.16/min x 6 min = $0.94), which we are rounding up to $1. By averaging $7.25/hour and $11/hour, we are

acknowledging that $7.25/hour is too low for MTurkers, while also avoiding the upper wage limit, which could cause unwanted character misrepresentation due to economic motive.

By keeping the price within the average range, rather than higher paying, economic motive is reduced. Chandler and Paolacci in 2017 have shown that deception rates drop from 23-80% to 0-4% when there is no monetary motive for MTurker impersonation[67]. Our research study will not be collecting individualizing identifiers, and therefore MTurkers could possibly re-take the survey, which would impact the data. Keeping the price near the average pay rate will hopefully minimize this potential issue by not calling attention to this research survey for monetary reasons. To further dissuade MTurkers from re-taking the survey, the survey-building software's "prevent ballot box stuffing" feature has been employed. This does not rely on an MTurker's unique identifier, but instead prevents a participant's multiple responses by leaving a cookie on their browser after they first submit a response. Specifically, if a re-taking attempt is made, the survey software will recognize the cookie and subsequently block following submissions. This feature certainly serves as a worthy deterrent, however, it is not infallible. Tech savvy people may circumvent the feature by either clearing their browser's cookies or using a new browser/computer.

In regards to pre-screening questionnaires, MTurkers mispresent themselves to pass these questions in order to earn money for the screened survey[67,75]. This is primarily an issue with surveys screening for rare populations, such as people with very uncommon diseases[75]. While this research project is not seeking a rare population, it will best avoid the issue of character misrepresentation by not screening for jury eligible participants, but instead, paying for and collecting all data. Unusable data from non-jury eligible participants will simply not be analyzed. By allowing all MTurkers to access the paid research survey despite differing

eligibility, there will be no need for character misrepresentation. Data is skewed when participants misrepresent their identity and habits, however, the common motives of rare populations, high economic value, and pre-screening questions have been greatly minimized for this research project.

III. MATERIALS AND METHODS

To assess jury-eligible participants' understanding of Questioned Document Examiner's conclusive terms, the materials and methods of this project have built upon literature discussed in the previous section, and have been thoroughly pilot tested. The overall protocol includes the development of a research and demographic survey, pilot testing, secondary pilot testing, and data analyses. Comprehensive information on the rationale and literature behind the research and demographic survey development can be found in the "Discussion on Materials and Methods" section of this thesis below. Additionally, Institutional Review Board (IRB) approval has been obtained for the use of human subjects.

A. Survey Design and Administration

Our research survey first provides a Questioned Document Examiner's testimonial opinion and then asks the survey-taker to relay their understanding of the statement. The nine testimonial opinions were randomized such that a participant would only be presented with four of the nine possible questions. As previously discussed in the "Conclusion Scale" section of the literature review, the nine phrases used in the testimonial opinions ranged from negative to positive terms (e.g., eliminated, no conclusion, strong probability, identified). Answer choices are in a horizontal form of a nine-point Likert scale. The survey then provides background information on forensic science testimony and lists all nine conclusion scale terms in order of confidence that the suspect wrote the document (e.g., identified to eliminated). This contextual information is followed by another set of four randomized research questions out of the nine possible, which also have a horizontal nine-point Likert scale answer format. The nine demographic questions immediately follow the research survey.

The demographic survey collects the participants' general information inclusive of jury eligibility, age, education level, gender, ethnicity, employment status, childhood and current residence, as well as native language. The question's answer choices include both multiple choice and fill-in-the-blank options.

The surveys have been built via Qualtrics software and administered by publication on the online platform, Amazon Mechanical Turk (mturk.com). Participants may choose to complete the survey for $1 compensation.

B. Sample

1. Selection Criteria

In order for a participant's response to be included in the data set, they must be meet jury-eligibility criteria as dictated by United States rules. Therefore, selection criteria required participants to be: "a United States citizen, at least 18 years of age, adequately proficient in English, reside primarily in a judicial district for at least one year, have no disqualifying mental or physical conditions, not currently subject to felony charges, and have never convicted a felony" as quoted from USCourts.gov.

2. Size and Participant Exclusion

Sample size was determined by running a power analysis for both a one-sample t-test and a two-sample t-test. The power analysis for the one-sample t-test (power 0.8) utilized an alpha of 0.005 due to a Bonferroni adjustment for the 9 required t-tests. Results indicated that a minimum of 337 participants be used per data set. The power analysis for the two-sample t-test (power 0.8, alpha 0.05) calculated that a minimum of 175 participants be pooled per verbal term.

Participants (n=649) were sampled via Amazon Mechanical Turk, however participants' data were excluded based on jury eligibility. Moreover, outliers were removed based on

participant's averaged absolute inaccuracy of the verbal terms; those who scored over 2 standard deviations away from the sample's cumulative average were excluded (n=57). Thus, a total of 592 participants' data were analyzed and, due to the power analysis results, this project has surpassed the minimum participant requirements for sufficient power and effect size.

C. Data Analysis

Data will be collected through the survey-building and analysis platform, Qualtrics. After data are exported from Qualtrics as an Excel sheet, data will be analyzed via Statistical Analysis System (SAS) software and Excel. Mean absolute error, averaged error, standard deviation, and standard error will be calculated per participant, verbal term, and set of questions in order to compare intended QDE terminology definitions with lay interpretation. Regarding error, participants were asked to read the expert testimony blurbs and then translate their interpretation of the term into a numerical format based on a Likert scale of least confidence to most confidence that the suspect wrote the document (0-8). If the participant received the verbal term of the highest confidence, "identified", which has been assigned an "8" out of the 0-8 answer options, but they responded with a "5", then their inaccuracy error would be -3 with an absolute error of 3.

An independent sample t-test will be used to assess any significant differences between participants' responses to question sets 1 and 2 (without and with context) based on mean absolute error. Additionally, averaged error per verbal term will be analyzed via a one sample t-test (Bonferroni adjusted alpha value of 0.005) against the correct answer (averaged error of 0). This will determine which terms were related to significantly inaccurate responses. Independent sample t-tests will also be used to analyze absolute error per verbal term for its no context and

contextual counterpart. Results from this unpaired t-test will highlight if contextual information significantly changes participants' interpretation for specific terms.

IV. RESULTS

A. Demographic Information

Of the 592 included participants, 73.3% identified as Caucasian, 11.2% Black or African American, 6.5% Hispanic or Latino, 5.9% Asian, 2.5% multiracial, 0.3% American Indian or Alaska Native, and 0.2% Native Hawaiian or Pacific Islander. Many of the participants indicated English as their native language (98.5%) and had a minimum of some college education (88%). The sample included more males (59.8%) than females (39.8%), with one individual identifying as non-binary (0.2%). Participant's average age was 38 (range 19-74, sd 11.3). The vast majority (89.8%) of participants indicated that they were employed either for wages or self-employed. Regarding residence, roughly 83% of participants grew up and currently reside in an urbanized area, ~3.5% in an urban cluster, and ~12.5% in a rural area. Demographic specifics can be found in Appendix F.

B. Research Survey: Verbal Term Accuracy

Participants read an expert's testimonial blurb and then indicated their understanding of it via a Likert scale (0-8) answer choice. The data set was analyzed for the mean response per verbal term for both sets of four research questions, which included sets without and with context (e.g., exposure to the full Questioned Document Examiner conclusion scale). Results for the first set of questions, without context (Figure 1A), show that the average response for "identified" (8) was 6.45 (sd 1.52), "strong probability" (7) was 5.91 (sd 1.35), "probable" (6) was 5.18 (sd 1.43), "indicated" (5) was 5.94 (sd 1.57), "no conclusion" (4) was 2.74 (sd 2.03), "not indicated" (3) was 2.73 (sd 2.19), "probably not" (2) was 2.64 (sd 1.84), "strong probability did not" (1) was 2.68 (sd 2.11), and "eliminated" (0) averaged 2.05 (sd 2.50).

Figure 1A

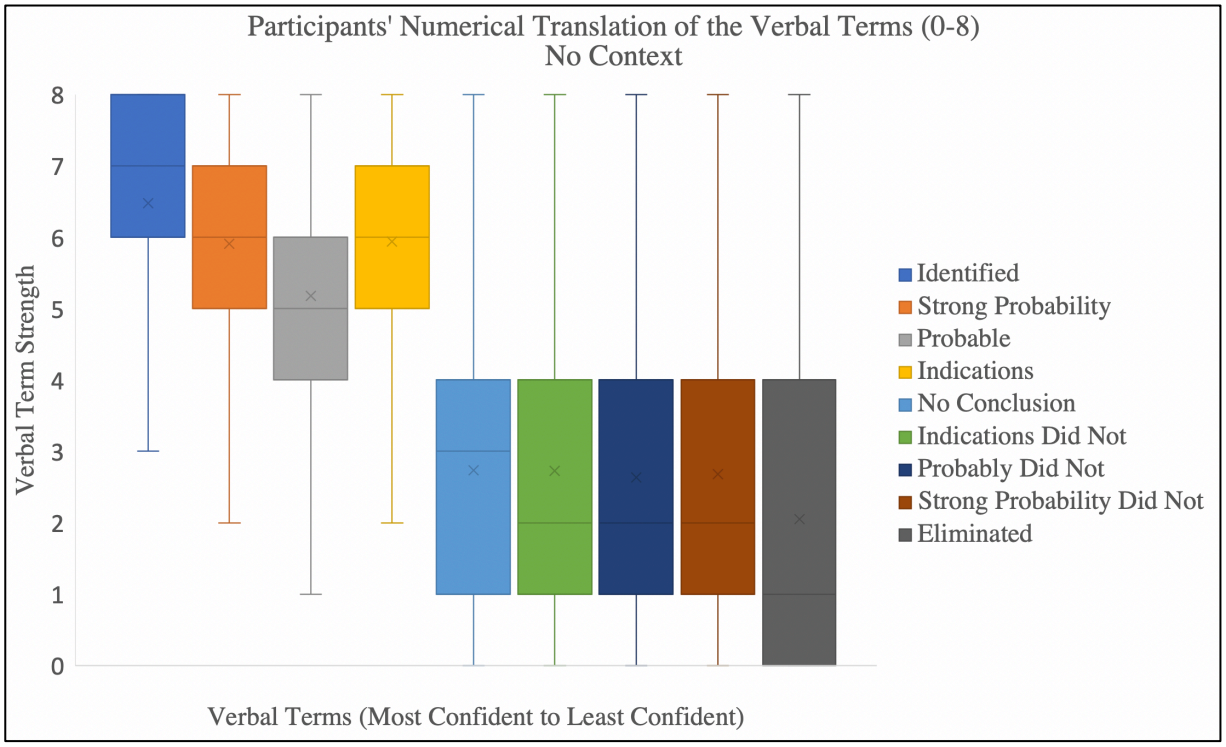


Fig. 1A Boxplots show participants' mean responses per verbal term without context. The y axis displays the correct responses in order (8-0) from "identified" to "eliminated"

Results for the second set of questions, with contextual information of the QDE conclusion scale (Figure 1B), show that participants' average response for "identified" (8) was 6.97 (sd 1.45), "strong probability" (7) was 6.15 (sd 1.21), "probable" (6) was 5.36 (sd 1.24), "indicated" (5) was 5.90 (sd 1.49), "no conclusion" (4) was 3.23 (sd 1.99), "not indicated" (3) was 2.92 (sd 2.14), "probably not" (2) was 2.75 (sd 1.84), "strong probability did not" (1) was 3.05 (sd 2.42), and "eliminated" (0) averaged 1.63 sd (2.34). Figure 1A and 1B are visual representations of these raw participant responses.

Figure 1B

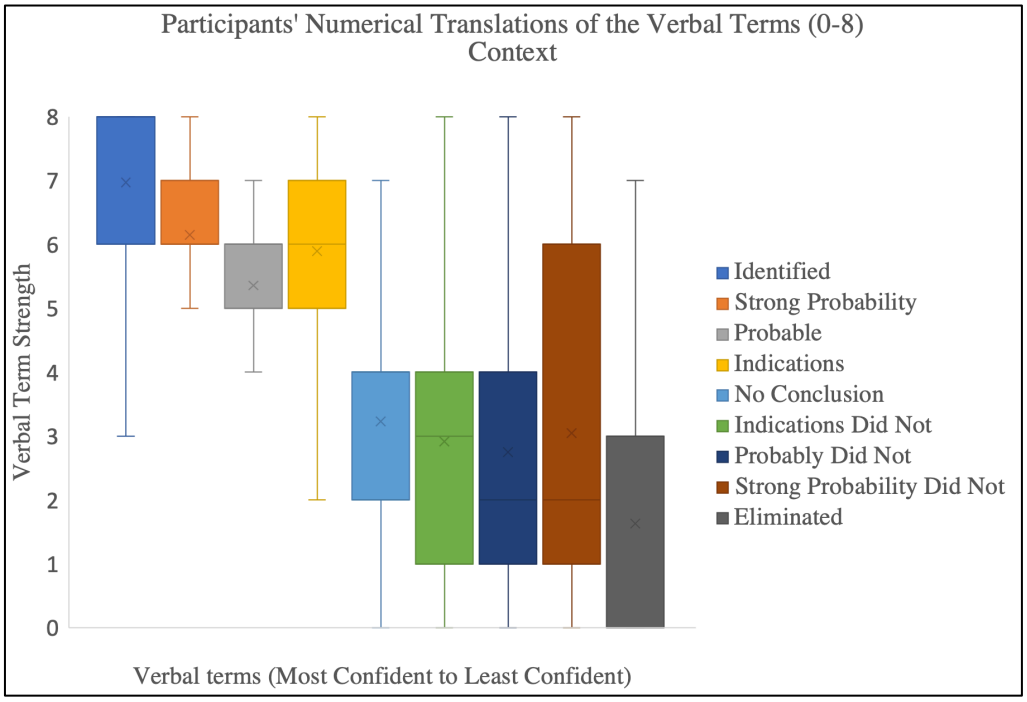


Fig. 1B Boxplots show participants' mean responses per verbal term with context. The y axis displays the correct responses in order (8-0) from "identified" to "eliminated"

The averaged inaccuracy of the participant's score per verbal term, without and with context, can be seen in Figures 2A and 2B respectively. By showcasing the averaged inaccuracy, over and undervaluing of the terms is readily apparent. For example, the term "identified" (8) received an averaged response undervaluing of -1.53 (sd 1.53). A one sample t-test was run per verbal term against zero, in order to see if the over or undervaluing was significantly different than the correct answer (averaged inaccuracy of 0). Running numerous t-test poses a large (36%) risk of type I error. However, the Bonferroni correction was implemented and significance was only determined if $p<0.005$.

One sample t-test results show that for verbal terms without context, "identified" was $t(242)=15.56$, $p< 3.00x10^{-38}$, 1.00, "highly probable" was $t(274)=13.42$, $p< 7.14 \times 10^{-32}$, 0.84, "probable" was $t(276)=9.56$, $p< 7.22 \times 10^{-19}$, 0.575, "indicated" was $t(255)=-9.47$, $p< 1.90 \times 10^{-18}$, -

0.594, “no conclusion” was t(281)=10.44, p< 9.24 x 10^{-22}, 0.623, “not indicated” was t(253)=1.98, p< 0.049, 0.125, “not probable” was t(266)=-5.63, p< 4.58 x 10^{-8}, -0.345, “strong probability did not” was t(266)=-12.92, p< 6.24 x 10^{-30}, -0.820, and “eliminated” was t(250)=-12.96, p< 1.02 x 10^{-29}, -0.820. All term response errors were significantly inaccurate aside from responses to term “not indicated”.

One sample t-test results for the verbal terms with contextual information include the following: “identified” was t(258)=11.43, p< 1.00 x 10^{-24}, 0.711 “highly probable” was t(259)=11.34, p< 1.90 x 10^{-24}, 0.705, “probable” was t(268)=8.44, p< 2.02 x 10^{-15}, 0.516, “indicated” was t(260)=-9.65, p< 4.90 x 10^{-19}, -0.599, “no conclusion” was t(276)=6.47, p< 4.47 x 10^{-10}, 0.389, “Not indicated” was t(278)=0.646, p< 0.519, 0.039, “not probable” was t(242)=-6.33, p< 1.17 x 10^{-9}, -0.407, “strong probability did not” was t(266)=-13.77, p< 6.28 x 10^{-33}, -0.845, and “eliminated” was t(257)=-11.19, p< 6.38 x 10^{-24}, -0.698. All were significantly inaccurate except for responses to term “not indicated”.

Figure 2A

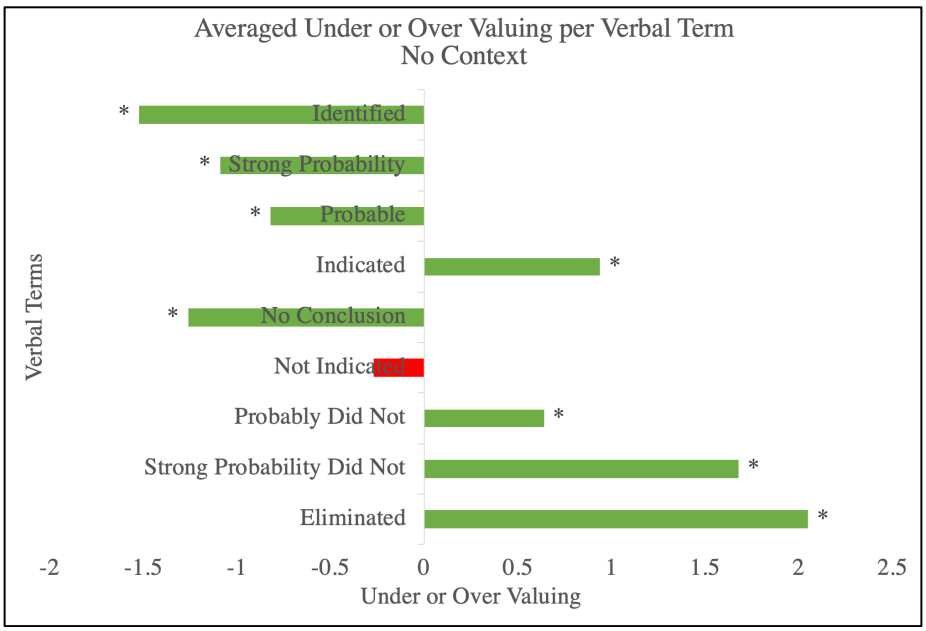


Fig. 2A Participants’ over and undervaluing per verbal term without context via mean response error. *p<0.005

Figure 2B

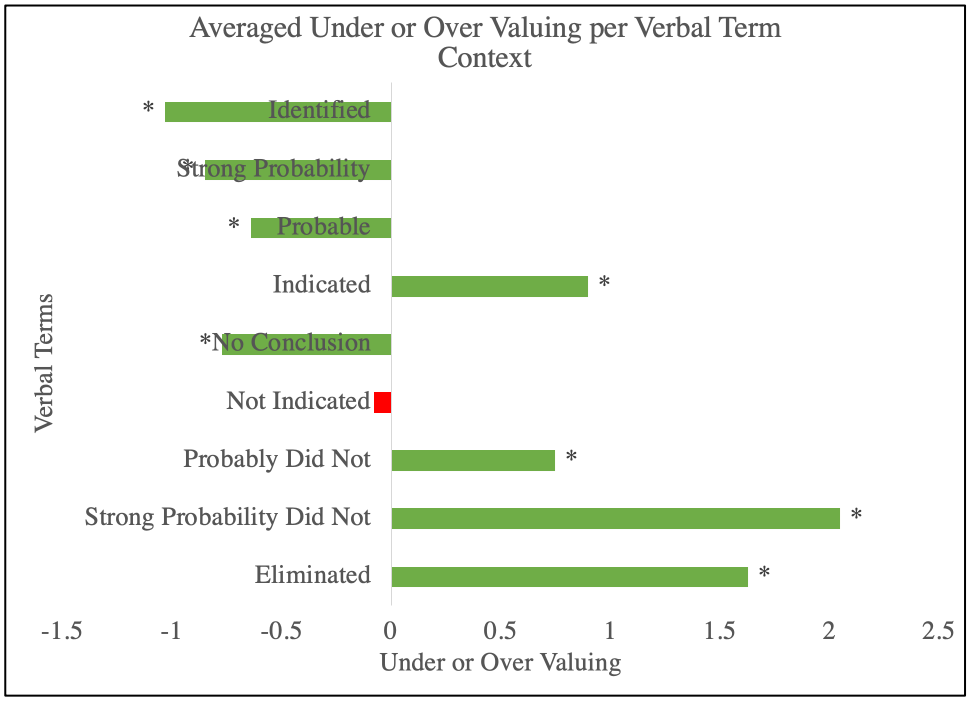


Fig. 2B Participants' over and undervaluing per verbal term with context via mean response error. *p<0.005

Absolute error was also calculated per verbal term and per data set/participant. The averaged absolute error between the first and second set of research questions (without and with context) was then assessed with an independent sample t-test (alpha=0.05). Without context, the averaged absolute inaccuracy was 1.62 (sd 0.89). With context, the averaged absolute inaccuracy was 1.41 (sd 1.00). The independent sample t-test ($p < 0.0001$) indicated a significant difference in verbal term score accuracy between the two question sets as seen in Figure 3.

Figure 3

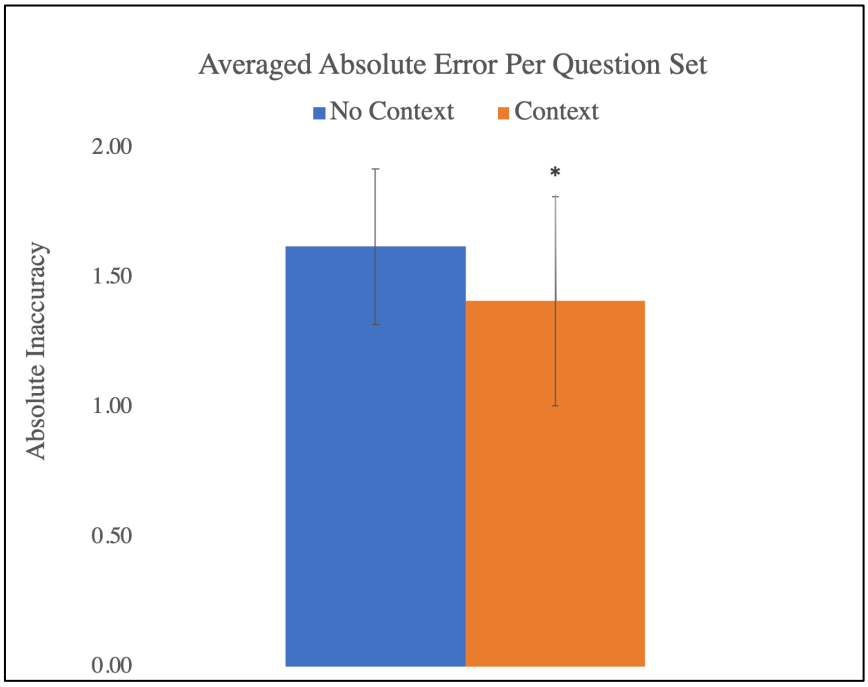


Fig. 3 Mean absolute error comparison between groups, without and with context, via an independent sample t-test. *p<0.0001

While Figure 3's independent sample t-test highlights the overall significant difference between the questions sets, Figure 4 represents the data above in an absolute inaccuracy break down per verbal term both with and without context. Averaged absolute error per verbal term, without context, results included: 1.55 (sd 1.53) for "identified", 1.24 (sd 1.22) for "strong probability", 1.22 (sd 1.10) for "probable", 1.49 (sd 1.05) for "indicated", 1.85 (sd 1.51) for "no conclusion), 1.91 (sd 1.10) for "not indicated", 1.46 (sd 1.33) for "probably not", 1.85 (sd 1.93) for "strong probability did not", and 2.04 (sd 2.50) for "eliminated". Averaged absolute error per verbal term, with context, results included: 1.55 (sd 1.53) for "identified", 1.24 (sd 1.22) for "strong probability", 1.22 (sd 1.10) for "probable", 1.49 (sd 1.05) for "indicated", 1.85 (sd 1.51) for "no conclusion), 1.91 (sd 1.10) for "not indicated", 1.46 (sd 1.33) for "probably not", 1.85 (sd 1.93) for "strong probability did not", and 2.04 (sd 2.50) for "eliminated".

An independent t-test was used to compare absolute error responses of each verbal term, without context, to its contextual counterpart. Due to the fact that 9 independent t-tests were run, the Bonferroni adjusted alpha=0.005. Results are: “identified” (p=0.0002, t=3.75), “highly probable” (p=0.0081, t=2.67), “probable” (p=0.0016, t=3.2), “indicated” (p=0.2852, t=1.07), “no conclusion” (p=0.0223, t=2.29), “not indicated” (p=0.0531, t=1.94), “probably did not” (p=0.3682, t=1.82), “strong probability did not” (p=0.0693, t=1.82), and “eliminated” (p=0.0291, t=2.19). The significant differences between the two sets of question’s verbal terms were only seen for terms “identified” and “probable”.

Figure 4

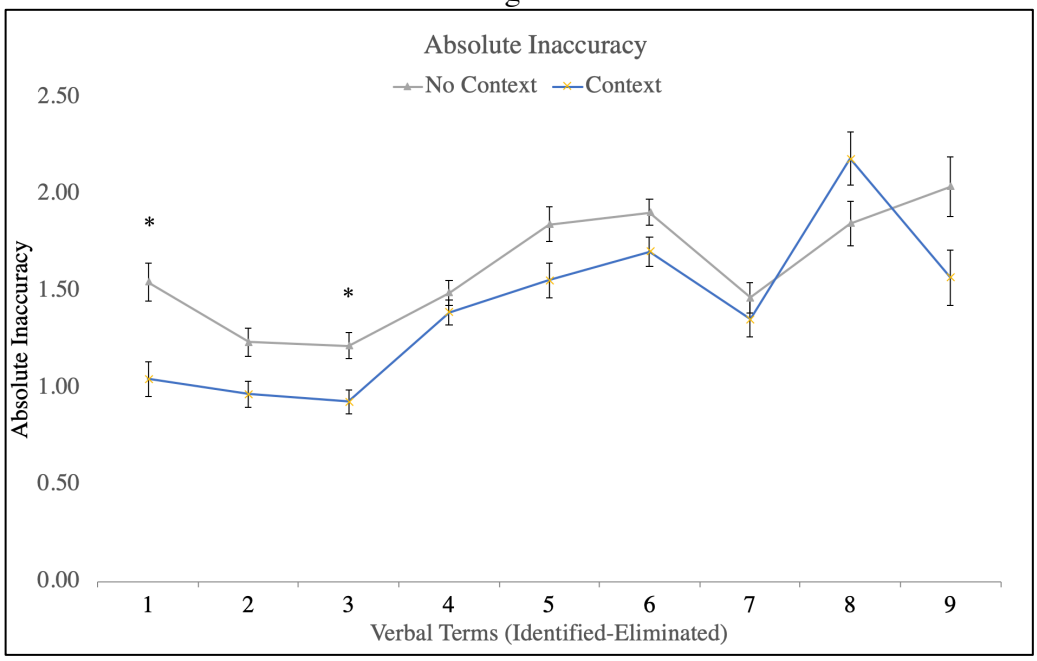


Fig. 4 Mean absolute error per verbal term displayed both without and with context. *p<0.005 indicates that mean absolute error for a particular term was significantly different to the mean response for that same term following contextual information.

A Pearson’s correlation between the verbal term responses’ absolute inaccuracy without context (NC1-NC9) and its contextual counterpart (C1-C9) is shown in Figure 5A and 5B as heatmaps. Figure 5A displays the Pearson correlation p values that have been color coded as detailed in the legend. A red tone signifies highly significant interactions (p<0.0001) whereas

blue represents the least significant (p>0.7501). A general trend emerges in which exposure to the highly confident terms, both positive and negative, (e.g., identified, strongly probable, probable, probably did not, strong probability did not, eliminated) were significantly correlated with like terms. Figure 5B displays the Pearson correlation r values which have also been color coded and are detailed in the legend. Red tones indicate positive correlations (r > 0.4996) whereas blue tones represented negative correlations (r < -0.2504). The general trends mirror those of Figure 5A, but assign a positive or negative correlation per significant interaction. Each significant correlation is positive albeit one divergence (NC4 and C4's significant negative correlation).

Figure 5A

Pearson correlation p values

Research questions Set 1 (No Context) with Set 2 (Context)

	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9
NC 1	0.000	0.000	0.002	0.009	0.664	0.285	0.062	0.343	0.000
NC 2	0.000	0.000	0.000	0.409	0.360	0.035	0.018	0.169	0.045
NC 3	0.023	0.006	0.001	0.819	0.022	0.344	0.563	0.421	0.649
NC 4	0.871	0.431	0.298	0.000	0.162	0.207	0.982	0.312	0.016
NC 5	0.011	0.024	0.330	0.391	0.000	0.000	0.951	0.430	0.014
NC 6	0.584	0.231	0.363	0.754	0.028	0.001	0.013	0.640	0.024
NC 7	0.006	0.205	0.244	0.776	0.279	0.000	0.000	0.000	0.000
NC 8	0.010	0.011	0.021	0.412	0.079	0.000	0.000	0.000	0.000
NC 9	0.001	0.081	0.046	0.414	0.162	0.011	0.000	0.001	0.000

Legend

P-value ranges

0.000

0.250

0.500

0.750

0.982

Fig. 5A Heatmap displaying Pearson correlation p values between verbal terms, without context (NC1-NC9) and with context (C1-C9). Highly significant correlations (p<0.0001) are colored red whereas the most insignificant correlations (p>0.982) are a deep blue.

Figure 5B

Pearson correlation r values

Research questions Set 1 (No Context) with Set 2 (Context)

	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9
NC 1	0.753	0.424	0.29	-0.25	-0.041	-0.103	0.183	0.09	0.354
NC 2	0.496	0.453	0.438	0.077	0.078	-0.188	0.223	0.126	0.181
NC 3	0.202	0.246	0.306	0.021	0.205	0.085	0.055	0.072	0.041
NC 4	-0.015	-0.075	0.095	0.357	0.13	0.117	0.002	0.096	-0.225
NC 5	-0.231	-0.199	-0.088	0.075	0.404	0.333	-0.006	0.073	-0.226
NC 6	0.052	-0.115	-0.087	0.03	0.199	0.323	0.24	0.043	0.222
NC 7	0.256	0.116	0.108	-0.026	0.099	0.314	0.719	0.435	0.527
NC 8	0.229	0.235	0.211	0.075	0.16	0.325	0.604	0.408	0.421
NC 9	0.312	0.167	0.186	-0.078	0.132	0.241	0.579	0.323	0.662

Legend

Pearson's r value ranges

-0.250
0.000
0.250
0.500
0.753

Fig. 5B Heatmap displaying Pearson correlation r values between verbal terms, without context (NC1-NC9) and with context (C1-C9). The most positive r values (r>0.735) are colored red whereas the most negative r values (r<-0.250) are a deep blue.

V. DISCUSSION ON MATERIALS AND METHODS

A. General Survey Features

The research and demographic surveys utilized both employ similar survey features, such as back buttons and a lack of answer enforcement. A back button allows the survey-taker to revisit previous pages and therefore keeps the survey instructions and contextual information accessible. By adding a back button, we aim to subtly reduce the potential perception that the survey requires participant's rote memorization. The survey-designing platform, Qualtrics, also allows questions to be enforced, which results in participants being unable to move forward in the given survey if the previous question was left unanswered. However, answer enforcement breeches ethical guidelines. The survey's IRB approval maintains that participants must be allowed to discontinue the survey at any time. Thus, questions will not be enforced in order to stay aligned with these ethical terms.

B. Research Survey Development

The research survey, meant to assess lay interpretation of QDE's conclusion scale terminology, was developed to best minimize biases and increase clarity. The research survey's design of four randomized questions, contextual information, and additional four randomized questions can be referenced in the Materials and Methods section above. This section discusses the rationale behind the survey's overall development including prefaces, contextual information, bias prevention, and answer format.

There are a total of nine QDE conclusive terms, however this survey only offers sections of 4/9 randomized terms, because of survey fatigue and contextual effects[4,5,39,46,47,82,83]. If participants were able to view all nine terms, this could cause participants to weight their

opinions of the terms relative to the others and thus coerce ideal responses. Furthermore, exposure to all nine terms would negate any effort to examine contextual effects based on knowledge of the full conclusion scale. Aside from these issues, odds of survey fatigue would increase with the addition of more questions[82,83]. Thus, the survey only exposed participants to sets of four questions, with and without context, for a total of eight questions.

The survey's written information and questions (e.g., preface, contextual information blurb, and QDE testimony scripts) were kept as concise as possible to avoid biases from extraneous information. As thoroughly explained in the literature review, an expert witnesses' age, gender, and presentation style all influence juror-interpretation of criminal evidence[7,10,11,21,24,26,84]. Moreover, probability term interpretation changes based on the context of its use [47]. Therefore, the script that describes a Questioned Document Examiner's final conclusion on the questioned document evidence does not include extraneous contextual information such as the type of crime the evidence came from (e.g., robbery, murder, arson). Intentional contextual information describing the QDE's nine-term conclusion scale is introduced between the first and second set of research questions. All research survey questions also maintained similar structure and emphasis as a result of Pilot 1, which can be referenced below for a more detailed explanation.

Question randomization and skip logic were employed to make the survey user-friendly and avoid question-order bias. For this research project, question-order bias would arise if all participants received the research questions in the same order. By having one of the QDE terms always first for example, participant responses to the subsequent terms could be influenced. To effectively disrupt question-order bias, the research questions have been randomized[85]. Skip logic enabled participants to smoothly transition between the first set of questions, contextual

information, and secondary set. Specifically, skip logic allows for the development of one master survey, from which participants view the differently randomized sets of 4/9 questions rather than creating hundreds of randomized surveys to avoid question-order bias. This is one of the great advantages of online surveying in comparison to printed methods.

During the development of the research survey, it was crucial to choose the best answer format to assess lay interpretation of QDE terminology. The debate was ultimately between a Likert and sliding scale. Following a statement of beliefs, attitudes, or opinions, Likert scales require the participant to 'point and click' an answer from a range of words or numbers. In comparison, sliding scales require participants to 'drag and drop' a cursor in response to a statement. Previous research on lay interpretation of forensic science expert witness verbiage have employed the use of sliding scales, however the pros and cons of sliding scales were weighed for this research study before following suit[12,49,53-55,86] . One of the benefits of sliding scales is the way in which participants are free to drag and drop the cursor at any point along the scale. In other words, they are not automatically pre-binned into specific categories, as is the case when Likert scales are used. Research has shown that participants find that the sliding scale best assesses their true opinions[87]. The sliding scale can remain unmarked until data collection is completed and numbers can then be superimposed on the scale for analysis. The cons of a sliding scale include the fact that it requires the participant to have additional technology such as Javascript, requires more participant effort, and can confound nonresponse estimation[86,88]. If the participant does not have software, such as Javascript for example, already downloaded on their personal devices (e.g., laptop they're using to complete an online survey), then the sliding scale questions will not load and thus be inaccessible. Additionally, having the participant point, click the slider icon, drag it, and then release their button involves more effort than a Likert scale's

point and click method, and indeed previous work has shown that sliding scale questions are related to higher instances of survey break-off[89]. Moreover, the handle often starts on a valid position on the scale, which is the reason non-response becomes difficult to estimate[86,88]. While the benefits of a sliding scale may outweigh its faults for the purpose of this research project, UIC has banned the use of slider questions due to its inaccessibility for those with certain disabilities. Ultimately, this project developed a Likert scale for the research question's answer choices for a variety of reasons. The Likert scale herein is user-friendly on a variety of devices (cell phone, tablet, computer), non-verbal, accepted by UIC, and allows for clear nonresponse assessment[86,88].

C. Demographic Survey Development

A demographic survey is a crucial aspect of a research project that allows the investigator to define the sampled population and interpret the data accordingly. Specifically, demographic information can be categorized and confirm to what extent the results may be generalized[80]. By categorizing participant's information, interactions with demographic effects will reveal confounds, limitations, and/or insight. This section of the thesis will discuss the demographic survey's placement and reasoning behind the chosen questions.

The demographic survey immediately follows the research questions in order to cater to typical survey fatigue patterns and avoid both priming and stereotype biases. It's important for the participants to answer the research questions first before survey fatigue sets in and potential for break-off rises[82,83,89]. Moreover, the demographic survey not only requires less effort for participants to answer since the questions simply inquire about themselves, but also this review of their personal information can cause stereotypic bias if research questions were to follow.

For example, researchers Blascovich et al showed that when people are primed with stereotypes, significant blood pressure changes were detected during the following cognitive task in comparison to control group's stable blood pressure[90]. Lastly, demographic questions could skew data due to priming effects, which is thoroughly discussed in the Conceptual Framework and Related Literature section above[80].

Interpretation of QDE conclusion terminology may be influenced by a participant's English skills. The demographic survey covers participant's native language as well as childhood and current residence. Residence data is being collected and subsequently assigned into the following categories as dictated by the United States Census: Urbanized Areas (>50,000 people), Urban Clusters (2,500-50,000 people), and Rural (areas not classified as urban). This research project has thus been designed to include current residence and native language as potential interactors with our primary data set.

In addition to English and residential information, the demographic survey was developed to define personal characteristics of the participants. Questions include jury eligibility, age, education level, gender (female, male, or nonbinary), ethnicity, and employment status. Survey pilot testing resulted in linguistic improvement of these questions and these details can be found in the "Survey Validation" section below. While the demographic survey aims to understand who the participants are, it must do so without unnecessarily breeching privacy. For example, a question of income was not included, because income and education level are indicators of socioeconomic status (SES) and thus two measures of SES would be unnecessary for this project[91]. An understanding of QDE terminology may be more closely related to education level rather than income and research has also shown that the longer a demographic survey is, the more frustrated participants may feel and perceive a violation of privacy[80]. This

project's demographic survey has been thoughtfully developed and has successfully defined the sampling pool.

D. Survey Validation

1. Pilot Testing

After the research and demographic surveys were constructed, participants were recruited (n=9) for initial feedback. The purpose of this small pilot study was to assess the question semantics and general survey quality. The participants consisted of a nonrandom sample, which biased the demographics towards highly educated people of Asian and White ethnicities. Regardless, feedback inspired a modification of the research aims and also addressed outstanding issues.

Pilot data revealed that three people found the follow up question, "From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that the suspect wrote the document?" to be less than ideal because the follow up question was always posed in the affirmative position regardless of which of the nine phrases, positive or negative, was used beforehand. Unfortunately, highlighting the negative or positive term being researched (e.g., eliminated, no conclusion, strong probability), in an effort to clarify the question, would introduce unacceptable priming bias. Specifically, emphasizing the researched term could activate participant's particular concepts revolving the research term, which reduces the question's cognitive challenge and may cause them to answer in a particular way[80]. However, the survey has been altered as a result of this feedback and now the follow up question's phrase, "…that the suspect wrote the document," has been bolded. This avoids bias by not emphasizing the conclusion scale terminology in question, but instead bolding question semantics that remain identical for each of the nine iterations.

One pilot participant stated that the survey used passive language and suggested a switch to active voice. While some may prefer active voice, this is not applicable for each of the nine versions. In this research situation, passive voice serves to maintain structural similarity across each of the nine iterations. For example, if the survey question using the conclusion scale phrase "identified" were posed in active voice, it would read clearly as, "Based on my analysis of the document, *I have identified* the evidence as the suspect's handwriting". While this is clear, the survey question using conclusion scale term "indications" would forcibly become overly complicated in an effort to maintain structural similarity. The "indications" research question would then read as, "Questioned Document Examiner: 'Based on my analysis of the document, *it is my opinion that it is indicated that* the evidence is the suspect's handwriting'" By using passive voice instead, no conclusion term misguidedly sounds overconfident in comparison to the others and participants can focus solely on the necessary components of the sentences. Therefore, no changes were made to the question's language in response to the request for active voice. Please view the attached survey, located in Appendix A, to see the particular wording and sentence structure used.

Regarding the demographic questions that immediately follow the research survey, there were multiple changes made after pilot testing. One of the questions asked about the survey-taker's employment status by offering two answer choices: employed or unemployed. It was noted that this misguidedly encouraged retired participants to mark "unemployed." Research on demographic survey questions was then thoroughly reviewed and thus the answer choices have changed to include more options[80,92,93]. Additionally, the demographic question designed to assess highest level of education achieved has changed. The three answer choices listing master's, professional, and doctorate degrees were collapsed into a singular answer choice,

"Advanced degree (e.g., master's, professional, doctorate)". For the purpose of our research aims, it is unnecessary to divide participants within the category of an advanced degree since group effects are not expected to differ between those degrees.

Fewer answer options may also lessen potential survey fatigue[82,83].

A review of the pilot data also inspired a change to the research aims. At this stage of pilot testing there were only four randomized research questions of the nine possible. No Questioned Document conclusion scale contextual information was provided nor followed up by an additional set of four randomized research questions. With this preliminary survey design, the research aim was to determine how jury-eligible participants understand the particular conclusive scale terminology used in court testimony. Pilot data revealed that participants were surprisingly very accurate in their understanding with an average inaccuracy of -0.02. However, the largest discrepancy between participant understanding and QDE intended terminology meaning was an average of -2.0, which is over two standard deviations away (sd = 0.897) from an accurate answer for the conclusion scale term "probably not". These data inspired a change to the research aims to now assess contextual information effects. In order to achieve this, the survey was modified to include contextual information and an additional set of four randomized research questions as discussed in the section, "Design Overview," above. These changes make it possible to determine if participant understanding of terms implicated to be difficult by pilot data, such as "probably not", improve with context or not.

The pilot study offered insight and as a result, the research survey has improved. While some changes were minor, such as altering the wording for two demographic questions, larger modifications to the experimental design were inspired as well. The effects of all the alterations have since been analyzed via secondary pilot testing with a larger number of participants pooled.

2. Secondary Pilot Testing

Following the first pilot study, the research aims, survey, and demographic questions were altered and then tested via secondary pilot testing. While the initial pilot study was intended to receive feedback on the survey's semantics and general quality, the purpose of secondary pilot testing was to assess the newly made changes and pool a larger sample for data analysis. Similar to pilot one, the participants (n= 41) were pooled via nonrandom sampling, which was not representative of a typical jury pool. Despite this, valuable information was obtained, such as contextual effects and participant's average accuracy to the Questioned Document Examiner's conclusion scale.

Analysis of the data showed that participants were very accurate in understanding the intended meaning of the QDE's conclusion terminology. The mean absolute inaccuracy (absolute difference between participant's answer and QDE's conclusion term ranking) was 1.61 (sd= 1.02) for the first set of four randomized questions. Participant's mean absolute inaccuracy after viewing the contextual information, 1.25 (sd=0.89), was not significantly different from the initial "No Context" set of 4 randomized questions (independent t-test, p=0.105).

Figure 6

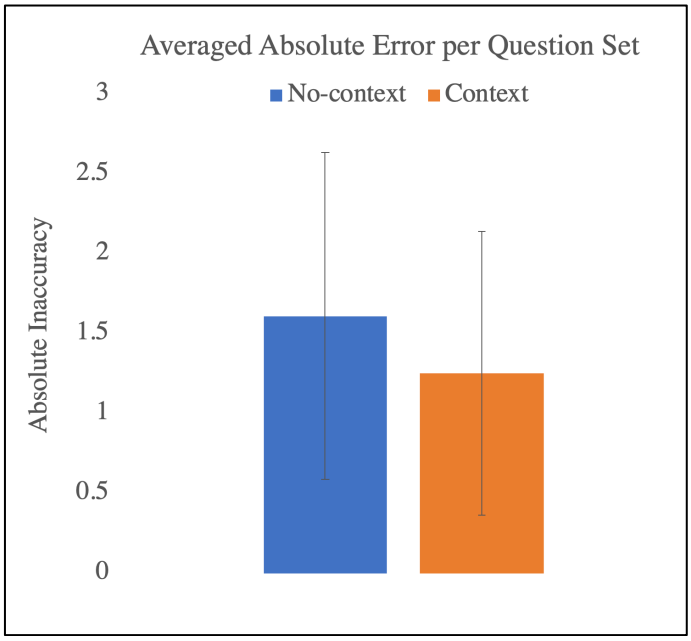


Fig. 3 Mean absolute error comparison between groups, without and with context, via an independent sample t-test.

While there was no statistically significant difference between the No Context and Context scores, the p=0.10 supported a trend indicating the potential benefit of contextual information on participant understanding. It is possible that when random sampling protocols are followed with an appropriately large sample size, significant contextual effects may emerge. Another possible interpretation of these data is that a "ceiling effect" is being observed because participant's scores are of such high accuracy without context that there is minimal room for improvement with added contextual information. However, this is not a concerning issue because these data support an ideal situation in which jury-eligible participants understand QDE conclusion terms highly accurately without context.

Feedback gathered from secondary pilot testing indicated that minimal changes to the survey were necessary. Only 1 of 41 participants was not able to answer the questions and stated that the survey was unclear. Additionally, one participant suggested the addition of an "Other"

answer option for demographic questions inquiring about ethnicity. After much consideration, this amendment was not made because it would cause participants in the “Other” category to fall into “statistical limbo”[93]. In this instance, statistical limbo refers to the issue in which data from participants selecting the "other" answer choice cannot be analyzed further for possible group effects and thus their unusable answers are said to be in "limbo". Our survey will continue to offer ethnicity options in alignment with the United States Census Bureau. Pilot data revealed that three participants chose not to enter their geographic information when asked in which city, state, and country they live. This may be because participants feel that this information is too individualizing, therefore the survey has been altered to ask for county, state, and country instead. The question was not removed entirely, because it serves as a working part of regional difference assessment. Overall, the changes made to the survey following secondary pilot testing was very minimal and without large foreseeable implications that would require a third pilot test.

Secondary pilot testing allowed for minor improvements, an assessment of all survey changes made following initial pilot testing, and a preliminary look into data analysis. The vast majority of participants overtly claimed that the survey instructions/questions were clear and this sentiment was supported by the mean absolute inaccuracy of 1.61 (sd= 1.02) without context and 1.25 (sd=0.89) with context. The results and feedback from secondary pilot testing implicated the validity of the survey.

3. Institutional Review Board

The finalized survey was submitted to UIC’s Institutional Review Board (IRB) for approved use of human participants. In response to the IRB application, several minor revisions were required. The survey instructions now clearly indicate that the survey is voluntary and may be exited at any time. Furthermore, researcher and IRB contact information have been added

(e.g., principal investigator, IRB, and thesis advisor emails). It has also been determined that all data collected from this project will be discarded after one year post-analysis. Following these revisions, IRB determined that the research project, (Protocol # 2019-0952) "meets the criteria for exemption as defined in the U.S. Department of Health and Human Services Regulations for the Protection of Human Subjects [45 CFR 46.104(d)]."

VI. DISCUSSION ON THE RESULTS

This study aimed to asses lay interpretation of Questioned Document Examiner's testimony in court through two sets of questions, without and with context. Participants responded to the QDE verbal terms by indicating their interpretation via a Likert scale (0-8) from least to most confidence that the suspect wrote the document. We hypothesized that the jury-eligible sample would not perceive the verbal terms in alignment with the expert's intended meaning. In agreement with our hypothesis, the data showed high variability and significant inaccuracy by both an under and overvaluing of the terms.

When participants' absolute inaccuracy was averaged per set of questions, a two sample t-test showed a significant difference between the means. While this difference is statistically significant, the responses to the two sets of questions display large error overlap. Furthermore, with high variability in responses and jury groupthink effects, it is possible that the inaccuracy in responses with and without context may be indistinguishable in practice. The data was then analyzed via a series of independent t-tests (Bonferroni adjusted alpha=0.005). Eight out of the nine terms improved in accuracy following contextual information with two terms displaying significant changes. "Strong probability did not" served as the exception as it was the one term that participants answered more incorrectly following the QDE scale information. However, the terms "identified" and "probable" were the two terms in which participants significantly increased their score accuracy following contextual information. Between the sets of questions, the absolute inaccuracy responses followed very similar trends. This implies that a relative understanding of the conclusion scale terms was not greatly shifted by exposure to the full scale. Regardless, the responses to the second set of questions, following contextual information of the

full QDE scale, were significantly different than those without. As this information could quickly be stated to the jury, it may be beneficial to have QDE's integrate this topic into their testimony. Alternatively, the court system could introduce a neutral party educator to brief the triers-of-fact on Questioned Document analyses and conclusion scale terminology prior to the evidential testimony. The use of a neutral party educator has been shown as an effective method by which to discuss DNA and family mediation in courts throughout the United States[94] .

Although data showed a significant increase in verbal term comprehension following contextual information, the participants' responses without context are more representative of interpretations made in court today because jurors are currently not briefed with exposure to the QDE conclusion scale prior to expert witness testimony. Response to the high confidence terms (e.g., identified, strong probability, probable) follow the expected trend of lessening confidence, however each of these terms were significantly undervalued by participants. This may be due to jurors' hesitance to accept a polarizing opinion without additional details of the mock case. Similarly, participants greatly undervalued the term with the least confidence that the suspect wrote the document (eliminated). The participants were solely asked to interpret the confidence level of the final conclusive testimonial blurb, but they may have been reading the conclusive sentence within their own concepts of criminal trials (e.g., jury duty, tv exposure). Therefore, when faced with a confirmatory term (e.g., identified, eliminated), they may be responding with lenience as one testimonial blurb is seemingly not enough to convince beyond a reasonable doubt.

The one sample t-tests comparing averaged response error per verbal term to the correct answer error (0) highlighted that participants were significantly inaccurate for eight out of the nine terms. In other words, participants' over and undervaluing of the terms were indeed

significantly incorrect aside from interpretation of the term "indicated". While the terms (e.g., probably, indicated) inherently allow for varied interpretations, these data suggest that lay interpretation of the conclusion scale is too varied for certain circumstances such as courtroom usage.

The implication of Figure 5A and 5B's Pearson correlation heatmap is that highly inaccurate responses for terms at the ends of the scale (identified, highly probable, probable, not probable, strong probability did not, eliminated) were strongly correlated with similarly inaccurate responses to like terms despite contextual information of the QDE scale. This can be seen via both the Pearson p values ($p<0.001$), which were color coded red to signal strong significance, and positive r values indicating similar responses. In conclusion, the Pearson correlation heat map further supports the notion that participants typically undervalue strongly conclusive words regardless of information to guide them otherwise.

When Martire and colleagues studied forensic conclusion scales in 2013, they found a weak evidence effect, in which weakly endorsed terms resulted in participants responding in an opposing fashion[12]. The Questioned Document Examiner's conclusion scale ranges from positive to negative endorsement and thus has potential for two weak evidence effects observed for this QDE conclusion scale. However, weakly positive endorsement term "indications" was in fact slightly overvalued by participants and its negative counterpart, "not indicated" was the most accurately interpreted term of the full scale (without and with context). This implies that these particular verbal terms are effective at communicating the expert's intended meaning despite the issue surrounding weak endorsement. This may be due to the fact that "indicated" and "not indicated" have concrete definitions in comparison to the other conclusion scale terms, such as "probably" and "strong probability". These probability terms are more vague and highly

dependent on the context in which they are used. In other words, the QDE conclusion scale is composed of evaluative terms in which some are more precisely perceived and therefore more easily understood.

Data from this research is in agreement with all previous literature on forensic conclusion scales in regards to the base statement that the verbal terms are too imprecise for recommended use with a lay audience[12,49,53-56]. As other papers have noted, there is wide intra- and inter-individual variation, differing lexicons, and preferences for certain words over others[2-5,40]. These factors make it difficult for an expert witness to relay a particular confidence level via verbal terms to a group of people (triers-of-fact). While the data agrees with previous literature in the overall conclusion of imprecision, the conclusion scale is unsuccessful for different reasons than previously described. For example, Arscott et al noted that the upper end of the scale (strong, very strong, extremely strong) may not be effective since participants rated these terms similarly. In contrast, the QDE upper end (identified, strong probability, probable) followed the general intended trend, but the negative terms (eliminated, strong probability not, not probable) received plateaued responses similar to Arscott's upper scale[55]. In other words, Arscott et al found the strongly positive endorsement terms to be the least reliable, whereas these data suggest that the negatively endorsed terms resulted in the most variability.

Great efforts were taken to ensure the efficacy of the research questions as well as the sampling of a large (n=592) and diverse group of jury eligible participants across the country; However, this research is not without limitations. The testimonial blurb offered to participants was quite minimal and discussed hand writing analyses. Handwriting analyses is easily understood by a lay audience without extensive background information required, but participants' concepts surrounding this particular kind of analyses could possibly influence

verbal term interpretation. For example, one participant noted that they do not trust handwriting analyses and thus may have been influenced and resulted in them undervaluing each term. It could also be said that the Likert scale (0-8) answer choices may limit participants' nuanced differences. However, this answer choice format was based on literature discussing the efficacy of 7 to 9 choices and the negative effects associated with offering too many choices (e.g 0-25). Alternatively, an unnumbered sliding scale could have been used, but this was specifically prohibited by the University of Illinois at Chicago. Participants were also pooled from Amazon Mechanical Turk which cannot validate that each user has input their true demographic information. Our research project is not without limitations, but we have accounted for each of them as best as possible.

VII. CONCLUSION

The role of an expert witness is to interpret evidence in an unbiased manner and then provide the triers-of-fact with information that will help the trial come to a just conclusion. Thus, it is critical for the expert witness to be consistently and clearly understood. SWGDOC has approved the conclusion scale evaluated by this research study and it is currently used in courts today. Based on previous literature, we hypothesized that the verbal terms would not be interpreted in agreement with the experts intended meaning. These data supported the hypothesis as participants' responses showed high variability, a general trend of undervaluing, and poor distinction between items at the negative end of the scale. As the conclusion scale consists of evaluative language (e.g., probably not), varied interpretations are to be expected, however the variability displayed in this data set exceeds acceptable limits and displays significant potential for miscommunication. Providing a lay audience with a brief overview of the nine conclusion scale terms prior to a final testimonial statement technically resulted in a significantly improved verbal term interpretation. However, the general understanding per term followed an extremely similar trend despite contextual information and the large error between the groups overlapped. Together, these factors imply that the statistically significant difference seen between the groups may not be overly robust. This is further supported by the strong absolute inaccuracy correlations between the sets of questions (with and without context) for the terms at the ends of the scales.

Results from this thesis research are in alignment with previous literature and suggest that the current Questioned Document Examiner conclusion scale is not well understood by a jury-eligible lay audience. As many researchers are looking into a new SWGDOC conclusion scale, possibly including a numerical translation such as likelihood ratios, it may be worthwhile to first

evaluate the implementation of a such contextual information prior to testimony. For example, an assessment of neutral party educators vs. the Questioned Document Examiner briefing the conclusion scale could provide insight on triers-of-fact understanding. As OSAC works to refine an ideal conclusion scale, we hope the data from this research project has provided a guiding direction.

APPENDICES

Appendix A

Forensic Science Graduate Survey

Start of Block: Block 1

Survey Instructions:

Questioned Document Examiners are Forensic Scientists who both analyze criminal evidence, such as forged or altered documents, and then testify in court when needed. When testifying, a Questioned Document Examiner explains their analyses of the evidence to the triers-of-fact (e.g., judge and jury) and concludes by stating their final opinion on the matter.

In the following questions, you will be provided with a Questioned Document Examiner's final testimonial statement and asked for your opinion on the statement. Please read carefully. These research questions are followed by a short demographic survey.

This survey is entirely voluntary and you may exit at any time.

Research Contact Information:
Principle Investigator- Emily Nakamoto, enakam3@uic.edu
Faculty Advisor- Dr. Albert K. Larsen, larsena@uic.edu
Human Subject Compliance- University of Illinois at Chicago, Institutional Review Board, uicirb@uic.edu

End of Block: Block 1

Appendix A (continued)

Start of Block: Block 2 (Randomized; 4 of the following 9 questions will be presented)

Questioned Document Examiner: “Based on my analysis of the document, the evidence has been identified as the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, there is a strong probability that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, it is probable that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Appendix A (continued)

Questioned Document Examiner: “Based on my analysis of the document, it is indicated that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, there is no conclusion in regards to the evidence being the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, it is not indicated that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Appendix A (continued)

Questioned Document Examiner: “Based on my analysis of the document, the evidence is probably not the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, there is a strong probability that the evidence is not the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, the evidence has been eliminated as the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

End of Block: Block 2

Appendix A (continued)

Start of Block: Block 3

When Questioned Document Examiners testify in court, they must use specific words to express their confidence to the triers-of-fact (e.g., judge and jury). They must use one of the following terms, which are ordered from most confident to least:

- Identified
- Strong probability
- Probable
- Indications
- No conclusion
- Indications did not
- Probably did not
- Strong probability did not
- Elimination

After reading this contextual information, please continue the survey.

End of Block: Block 3

Appendix A (continued)

Start of Block: Block 4 (Randomized; 4 of the following 9 questions will be presented)

Questioned Document Examiner: “Based on my analysis of the document, the evidence has been identified as the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, there is a strong probability that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, it is probable that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Appendix A (continued)

Questioned Document Examiner: “Based on my analysis of the document, it is indicated that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, there is no conclusion in regards to the evidence being the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, it is not indicated that the evidence is the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Appendix A (continued)

Questioned Document Examiner: “Based on my analysis of the document, the evidence is probably not the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, there is a strong probability that the evidence is not the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

Questioned Document Examiner: “Based on my analysis of the document, the evidence has been eliminated as the suspect's handwriting”

From a scale 0 (Least Confident) to 8 (Most Confident), how confident are you that **the suspect wrote the document**?

0	1	2	3	4	5	6	7	8

End of Block: Block 4

Appendix A (continued)

Start of Block: Block 5

We would like to know if you qualify for jury service. Are you...

- a United State citizen at least 18 years old
- residing primarily in a judicidal district and have been for one year
- adequately proficient in English
- without a disqualifying mental or physical condition
- someone who has never committed a felony

Q1
Based on this description, are you jury eligible?

- ○ Yes, I am eligible
- ○ No, I am not eligible

Q2 What is your age?

_____

Appendix A (continued)

Q3 What is the highest degree or level of school you have completed?

- ○ Less than a high school diploma
- ○ High school degree or equivalent (e.g., GED)
- ○ Some college, no degree
- ○ Associate degree (e.g., AA, AS)
- ○ Bachelor's degree (e.g., BS, BA)
- ○ Advanced degree (e.g., Master's, Professional, or Doctorate)
- ○ Other _____

Q4 What is your gender?

- ○ Female
- ○ Male
- ○ Non binary

Appendix A (continued)

Q5 What is your ethnicity?

- ○ American Indian or Alaska Native
- ○ Asian
- ○ Black or African American
- ○ Hispanic or Latino
- ○ Multiracial
- ○ Native Hawaiian or Pacific Islander
- ○ White

Q6 Which of the following best describes you?

- ○ Employed for wages
- ○ Self-employed
- ○ Out of work
- ○ A homemaker
- ○ A student
- ○ Military
- ○ Retired
- ○ Unable to work

Appendix A (continued)

Q7 Is English your native language?

- ○ Yes
- ○ No, my native language is: _____

Q8 Where did you grow up?

- ○ County _____
- ○ State _____
- ○ Country _____

Q9 Where is your current residence?

- ○ County _____
- ○ State _____
- ○ Country _____

End of Block: Block 5

Appendix B

April 30, 2019 (corrected October 4, 2019)

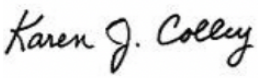
Dear Emily Nakamoto:

I am pleased to tell you that you have been selected to receive an Award for Graduate Research in the amount of **$1850.00** for your project, "Validation of Questioned Document Examiner Conclusion Scale." The amount of your award may be less than you requested in your proposal, due to budget constraints. No matching funds were requested.

The award is intended to recognize outstanding researchers among UIC graduate students, to enhance the quality of your research, and to assist in your progress toward completion of your advanced degree. I speak both for the Graduate College and the Graduate Faculty of the University when I say that the recipients represent the excellence of our programs and students, and they will make outstanding contributions to learning and society.

Make sure to review the enclosed guidelines outlining the terms of this award.
If you have any questions concerning your award please contact Benn Williams, our Fellowships and Awards Coordinator (bwilli7@uic.edu, 312-413-2389). Again, please accept my congratulations.

Sincerely,



Karen Colley
Dean
Kc/bw

Appendix B (continued)

Award for Graduate Research and the Deiss Award in Biomedical Graduate Research

Spring 2019 Guidelines

1. Recipients must be currently enrolled in a UIC graduate degree program during the time they hold the award.
2. The Graduate College will transfer funds in the amount of the award to the home department of the award recipient. The department will administer the funds to the student (see #9).
3. Awards are designed to allow students to take advantage of unique opportunities to further their research, and to aid progress toward their degree. Effective Fall 2018, the Graduate College is delineating three spending categories: research travel, summer research stipend, and materials/supplies.
4. Funds may not be used for routine expenses or those covered by other sources, such as grants of a major advisor. Examples of expenses that will not be allowed are: journal subscriptions, books, professional society dues, computers for routine use, travel to professional society meetings or general conferences.
5. Examples of allowable expenses are: stipend; travel to archives; travel to confer with collaborators or with distinguished researchers who can make an unusual contribution to the student's research project; expenses related to attending specialist conferences directly in the student's research area; expenses of conducting surveys; expenses for performances or exhibitions; expenses for extraordinary laboratory materials; access to databases or libraries; significant and unusual photocopying expenses (e.g., of archival materials; or special software essential to research.
6. Award funds must be expended substantially in accord with the budget submitted with the proposal. Significant variations from that budget must be approved in advance by the Graduate College. Under no circumstances will the Graduate College assume responsibility for expenses in excess of the amount of the award.
7. Equipment purchased with Awards for Graduate Research funds will remain the property of the student's home department, for the primary use of the award recipient and other graduate students.
8. Expenses for Spring 2019 awards must be incurred between June 1, 2019 and June 30, 2020.
9. Funds will be transferred to your department for distribution. **If the department processes the funds as an award, rather than as a reimbursement for research/equipment/etc., the following may apply:**
    a. The stipend portion of the award is taxable income, although the University will not withhold federal or Illinois income taxes. **International students**, please check with the Office of International Students. Due to differences in the agreements between individual countries and the U.S., students from **SOME BUT NOT ALL** foreign countries will be taxed. Therefore, international students should contact the payroll office at payinq@uillinois.edu or 866-476-3586.
    b. **Students who are obtaining loans from Financial Aid**: Note that obtaining an award will become part of your financial aid package and potentially affect your

Appendix B (continued)

financial aid eligibility. If you have already received a refund from Federal Loans for the current year, you may be responsible for **RETURNING THIS REFUND** to UIC. Consult with the Financial Aid office if your department processes as an award and you have a federal loan.

10. Research Award recipients **must submit a short report** within six months after the end of the grant period, but no later than the date that all degree requirements are completed.

Please indicate acceptance of these terms by signing and dating the form and converting it to a PDF. Use this naming convention: AGR19Sp_Lastname_Firstname, i.e., AGR19Sp_Doe_Jane.pdf. Keep a copy for your records and send a copy electronically to Benn Williams via email (bwilli7@uic.edu) by May 14, 2019. Failure to do may result in forfeiture of the award.

X **I accept the award** ____ **n/a This award includes a match**

____ **I decline this award**

The monies should be transferred to

FOAP: 200250-440000-xxxxxx-440162

Emily Nakamoto 10/6/2019

(Emily Nakamoto) **UIN:** 670358617 **Date**

Appendix C

Variances and Covariances Covariance / Row Var Variance / Col Var Variance / DF									
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**
C1	1.701287129 1.999603960 2.548316832 100	0.762218891 1.684257871 1.401499250 115	0.268966379 1.898512686 0.932258468 126	-0.025634726 2.329402129 1.157903358 110	-0.515406162 2.110644258 2.341736695 119	0.084602785 2.259623260 1.144635545 110	0.528335832 2.711244378 1.567841079 115	0.617886179 2.097560976 3.464135851 123	1.203683738 1.857053010 8.010871518 105
C2	0.708340284 1.407923269 1.974061718 109	0.541262564 0.962793158 1.482630929 106	0.299813408 1.131280821 1.303212049 122	-0.076817889 1.162081725 0.900176963 111	-0.357921512 1.367126938 2.355741279 128	-0.149291076 1.297414512 1.294412010 109	0.173479561 1.319470161 1.667996012 118	0.525937031 1.284257871 3.888155922 115	0.455713094 1.183402836 6.277231026 109
C3	0.416666667 0.917631918 2.248310811 111	0.398623641 0.740400185 1.119303724 131	0.331559220 0.903973013 1.300974513 115	0.106083187 1.159802195 1.073025335 121	-0.124193548 0.892645161 2.241935484 124	-0.104176904 1.038820639 1.373464373 110	0.157685064 1.341807910 1.582862524 117	0.466101695 1.271186441 3.844893890 118	0.457271364 1.098350825 5.517841079 115
C4	-0.360342091 0.976547346 2.120260977 106	0.120763336 1.347627468 1.824638963 116	0.024242424 1.044490358 1.266666667 120	0.407821782 1.139405941 1.145346535 100	0.114908628 1.097096923 2.119997687 131	0.033535165 1.163251048 1.101692284 113	-0.037245407 1.130465746 1.757584390 118	0.166253444 1.275895317 3.819421488 120	-0.194922195 0.970679771 6.388861589 110
C5	-0.094965675 2.338977879 2.304347826 114	0.146700031 2.295485351 1.530080284 138	0.323225806 2.183870968 1.142967742 124	0.192823460 1.939581491 1.131005010 116	0.871911337 2.017441860 2.304990310 128	0.320552771 2.299959355 1.128912072 121	0.185606061 2.102066116 1.666666667 120	0.450482094 2.010743802 3.921763085 120	0.444806707 2.003726129 5.666278528 113
C6	-0.208630649 1.667686035 2.446653075 108	-0.275301587 1.830603175 1.173650794 125	0.111923510 1.630671166 1.073490814 126	0.167173693 1.761263219 1.156960742 117	0.605469163 1.511530731 2.190680207 137	0.487532175 1.653072716 1.378297941 111	0.555992844 1.574478235 1.996660704 129	0.847276688 1.632897603 4.157516340 135	0.802948403 1.739066339 6.409828010 110
C7	0.449402539 2.281926811 2.630227782 103	0.418918919 2.162725225 1.629343629 111	0.090171990 1.982800983 1.334643735 110	0.003448276 2.119640180 1.226011994 115	-0.013116308 2.079804046 2.390486726 112	0.372332922 2.290424969 1.046199965 106	1.354875283 1.990929705 1.781900639 98	1.613549832 1.864053751 3.829339306 94	2.153941104 2.023628990 6.845353553 106
C8	0.326169406 5.116782554 2.567319848 112	0.345938375 5.323249300 1.410014006 119	0.164124016 5.274606299 0.993540846 127	0.217920354 5.143963338 1.001738306 112	0.233893557 4.646988796 2.224089636 119	0.107031568 5.421284379 1.156211895 121	1.200910433 4.901328740 1.558009350 127	1.669435791 4.803392177 3.488750433 107	1.794079794 5.152831403 5.994851995 111
C9	1.188251718 5.393470790 2.094931271 96	0.503065440 5.268425963 1.468745835 122	0.120698413 6.269269841 1.386984127 125	-0.596646483 5.590746778 1.260363298 113	-0.789873968 5.721932493 2.143488338 117	0.591757091 6.606129831 1.075004759 102	1.485794131 4.904285049 1.616208663 113	1.868345135 5.583005507 3.521308681 123	3.434418458 5.277496839 5.093710493 112

Appendix D

Participants' responses to the two sets of research questions for verbal terms, identified-eliminated (8-0), without context followed by responses to the survey's contextual counterpart.

	8	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
1	7		6					1	1			7		1			1	0
2		7	6		0	1					7	6	8		0			
3	7	7		6		6					8	7		7		1		
4	8					0	4	2		8		5					6	0
5	7	6	5			7					5			7	7			6
6	2				0		0		0	3			3		3		3	
7	6		7	6			3				6		4		5		2	
8	6			5			3		2		7	5	6					1
9		6		6				2	1	8					3	2		0
10			4	6		5			5			6		6		5	6	
11				7	4			2	0	8			6			2	1	
12	5					1	3		0					4	4	3		0
13	7		3	6					0	8	7		7					0
14		7		5	2			1		8						2	1	0
15				8		0		1	0			6		4		2	1	
16	5				6	4		5			5	4	4			6		
17	7	8		6					8	7		6	5			6		
18				7		1	3		0		7	6			2	2		
19		4				2	2		0		7	5	8			2		
20	7	6	5			3						5	7	0		1		
21		7	6				2	1		8		7	5	4				
22	7		2		1			0		8	6		6	3				
23			3			1		4	0	8		6	5		3			
24		7	5	8		0					7	4		0		2		
25	8		5	8		1				8			8		1		1	
26	7				1		2	1		8		6		4	4			
27			8			7	6	6		8			6		5	6		
28	8						2	1	0		7		5		3			0
29			7				1	1	1	8	7		8			1		
30			5		0		1		0	8		6		4				0
31	7	5	5			2						5		7		2	4	
32		4		3	4	2				5		4	3					2
33		7	7		6	8				6		7			7			6
34	4	3				0			0		5			4	0		1	

Appendix D (continued)

35		6	3	4	5									5	6		4	4
36			5				3	6	4			3	4		7			5
37				4	4			4	4		4		4		4			4
38		7			2	1	1					6		4		3		0
39				6	2	1		2		8		6	8					0
40			0		0	0			0			4		0		4		7
41						5	3	7	0				5	0	0			0
42			5	4		4		3			5			4			2	1
43	6			4	1				1					7		1	5	0
44		6			0			0	0	8	6			0	0			
45		6		8		0		1		8				4	3			0
46	8	7			0	0				8				0		0	0	
47	7			6	2		4			6	4		6			2		
48	7	4		7	6						6		6		4	5		
49		6		6			4		1		6	5	5					1
50		7	6	8		2				8				4	3		1	
51	7		6				2		1	8		5			2		1	
52	7		6		4	1					6		6			2	1	
53			6	5			1	7		8	7	6			2			
54	7		5		6			5			6		5		5	6		
55		6		7		2	2					6			1		2	2
56	8	7		8			1							4	3	2		0
57	4				4		5	1		6	7		6	1				
58	7		6	7				1		7				2			1	0
59	4		2				1		0		4		3			1	1	
60					4	4	4		4	6	4			4				6
61	8			8			1		1	8	5	4					6	
62			4		0			2	0		6	4				2		0
63		5	5			1		2				5	4			2		0
64	5	6			4				4	5	6				3	3		
65	6		4			2			0		5		7	0	0			
66			0		0	0	0				8	6			1			0
67			6	7	4				1			6	7	4				0
68		8					2	7	0					0		2	3	0
69			6	5				3	1	6					1	4		4
70			4			5	4		4		4	5			5		4	
71	8		7			1			0	7		6	7		0			

Appendix D (continued)

72	4	3			1			3					6	1			5	0
73	7		5				8	6		4		6			3			4
74		7	5		1				1		7	7				1		1
75				8	0		0		0			6	8			3	2	
76	6					4		3	1	7		5	4				2	
77			3	4	0			5		8	7				3			0
78				6	2	2			0		7	6				2		0
79	8	4	4	8								2		0		7	8	
80	4			4				4	2		4			4	4	4		
81		6			4		1		1			6	7			2	2	
82			6		6	2	5			7	7				3		2	
83				5	3	1		1		6			5		3		1	
84				7		3		1	0	8		7			1		1	
85					2	2	2		2		6	5	3				1	
86		7			4	3	1				7	7			3		0	
87				6	2			1	2			6		4	2		1	
88		4			4	3		4			5	3	5		5			
89			4		1	6	3			1		3			4		2	
90	7		5				3		1	8	6		7	4				
91		6	6		4	6				8	6		5	4				
92	7				4			2	4		7				3	2	1	
93	6					2	3	2		8			5		2		7	
94	3					2	2		1			4				2	1	1
95	4		4	4		4				8				0	0		0	
96		7	5	7	0								7	0	0		0	
97		7	6	8			1								0	1	1	0
98	4					4	4		4	5		5			5	5		
99	4	4	4			4					7				4	1		0
100			7	7	6	6				6			6			6		7
101		7	5			0		1		8	7				0		1	
102	8	7	7		4					8				3		2		2
103		6			2	4		3				6	5	4		3		
104		2	3		0			2				2		0	1	1		
105	8			7		0	2				6		7				1	0
106			6		2	1			3		7		6			1		0
107				4	5	4		4		5	5				5			4
108	8		4		3			2		8	7			3			1	

Appendix D (continued)

109		7	4					1	0	8					1		7	0
110	7	7						1	1	7			7	4		2		
111		6		6	2			1		8					3		1	0
112			7	7	6	6					6			7	5			6
113	3						3	3	3			3		3	3	3		
114		5	5			1	1			7				2			2	0
115			3	4			4		5		4		5	5		3		
116			5	8			2	1		8		5			3			2
117		4	5				0		0		6		7	0			0	
118	8		6	8		0					6				2	2		0
119						4	4	2	8	8					2	3		0
120				7			2	5	0		6				1	2	2	
121			7			6	0		0	8				5		1	1	
122	4	4			4	4				4	4				4		4	
123				7	4	2		1					8	4	3	2		
124			5	6		6	5			5	5						5	5
125	8	6					1	1				5			1	1		0
126		7				1		0	6	8	6			3			0	
127			2	6	0				0		6		7			0		0
128		7			2		2	1		7	6			1		1		
129				5	6		3	2				5	7	1		4		
130	2			5		6	4			1		5		6			3	
131		7	5		4			1			7	6		4		2		
132			6		5	6			7		6	7	5					5
133	8			8	4				0	8			7	4		1		
134	7				0		2	2		8				4		3		0
135		7	4		0	1								0	0	1		0
136		8			7	6			7	7				6	7	8		
137	8	6	4		0					8		6		0	0			
138	7	5	7			1				8			6		1			0
139		5	5					3	0	7	5	5					2	
140	6		5		2				1		7	5			2			1
141		7			4	1	2					6	7	4	1			
142		7		7				1	1			6			2		2	0
143			6	7		0		0		7	7			4		0		
144	8				4			3	0			5	6			4	1	
145				8	5	8	5			4		6			4			4

Appendix D (continued)

146	8		4	5	5					5	7	4						8
147		3		3				4	4	3			4		3		2	
148			8		4			0	0	8		6					7	0
149	4			5	4		5						3	3	4		3	
150	8	7	6		1							6	4				1	1
151	7					4		2	0		7		5		3	2		
152	3		3		3			3		3	3				2	2		
153			7	7	2	1				8			6			1	6	
154					4		4	1	4	8	8	7			1			
155	8				5		3	1		8		5					3	0
156						4	3	1	0		7		6	4				0
157	8	2		6		0						7		2		1	6	
158	7			7		7		7		7	7	7				6		
159		7	6					7	6				3	2	1	1		
160	4	3		8			3			5	6						7	5
161			6	5			3		1			5	6		3	2		
162			4	8	0		1			8	7				2	2		
163	7	6	3				1			7			5		3			0
164			6		3	2			2				6	2	3	3		
165	3	6		2		6				5	6		4		3			
166		6	5		0		4				7				1	2		0
167			4		4		4	4		4		4	4				4	
168				5	6	4	6				5	7		4				5
169		5	5			1		2		8				4			1	0
170		6		7	7			7		7	6		7	7				
171	5					2		5	1				5		2		4	6
172	7	6			4	2				7		5		4		1		
173		7		6	0		0			8			7	1			0	
174		6	4				1	2					8			1	2	0
175	7		6					2	1			5	7			3	2	
176		6			2			1	0				7			3	6	0
177	4					5		3	6	6	5				4	2		
178			5		6	6	5						6	5	6		6	
179	8	7				0		0					8	0	0			0
180						1	1	0	1			0		0		1		1
181				1			1	1	1	3	3						4	3
182		5	5				6	1				6		1		1	1	

Appendix D (continued)

183						2	2	1	1	3				6			4	5
184				6	0		2		0		7	6	5		3			
185	7						6	6	5	6		7	6	5				
186	8	5				0	2			8	6		8					0
187		7		6	0			1				6	5	3				0
188	8	7	7	5						8				6	4	4		
189	8		3				2		7		5	3	8	2				
190	6				4			3	2		6	5			4		2	
191			6		4	1	3				6				1	2		1
192			5		1	1	1						7		1	2	2	
193	7			6			3	7			6				2	2	3	
194	7	5		5		1				7				3		3		1
195		6			3	0		1			7		5		3		1	
196			7	6			6		6	7	7			7		7		
197		7	6	6					6	6	4		4	5				
198			8		6	7	7				5		7		6		6	
199	6	7	5				1				7	5					1	0
200	6		7				7	8		7			8			7	7	
201				0			0	4	0			0		0	0	0		
202				0			0	0	0	8			8	3				0
203	7				3		7	7			4			7	6	8		
204	8	6	5					3		8					3		2	0
205		6	7			3			0			7		1	2		6	
206	8	5		8					0				8		4	2		0
207	7				6		6		7			7	6		6			7
208	7						3	1	0		7				5		2	0
209	4		4		1	2				6		5				0	1	
210			2	3	4		5					3	2		3			3
211	7		5	7	0						6	5		0				0
212			6				1	1	1	7			8		1			1
213	7	6			2		0					5	5	4	2			
214		4	5	4			4			6	6	2	6					
215	7	6				8	6					6			6	7	8	
216		8			4			0	0	8	7	7			3			
217				7	0	0		0			7		5				1	0
218		7	5	6					0			6	7	2		1		
219				6		0	0	0		8	7	6		4				

Appendix D (continued)

220				5	4			4	5			4		4		3		4
221		5	4	3					0						0	0	0	0
222	7		6	7		1					5	5	6			1		
223		6	5		0			0			7	6			4		1	
224		7	4		2				0			5		2			6	0
225	5			5			3		2				6		2	3	2	
226						7	3	2	0	5				2	7		4	
227	8	4		7		0							4	0		1	5	
228		4	5	5				2				7	8			1		8
229	8	6				0		1		8	6		6					0
230		6		4		5		5		6				5	5			5
231		7				2	3		0				5	4		2	1	
232				5	3	5		5			4			5	5			5
233	5				4			2	7	5						2	1	0
234		2			4	4	5			5			0	7		0		
235	8		6				1	2			7	6				2	1	
236	6				4	2			1			6	7			7		2
237	7						4	2	1					4	4		2	1
238	4			4			4	4				6		4	3			2
239	5	5			0		0			8	2		7		5			
240	4	4				4		3				4	5	4		4		
241	7			6			3	0				4			0		2	0
242			3	3	3			5		4	4	3						3
243		7	5	8	2									4	2	3		0
244	6		3		1		1				7	6					1	0
245	8		8	7			7				8				7	6		7
246			5	7				1	0		7	6		4				0
247		6		7				2	0	8		6				3	2	
248			7		1	2	3						8	4	2		1	
249		5			4	4	3			7	5			2			1	
250		6		5	0		3					7	8	0			2	
251	7		5	6		1				8	7						1	0
252		6			4	1		1			7			4	3		1	
253	7				0			1	2		6			1	0	1		
254	5				0	1		1				5		0	1		7	
255				6	4	1		2		8	6	6		4				
256	4						3	3	4		6			2		3		4

Appendix D (continued)

257		6		8	1			6		8		3	0		1			
258		5			3		3	2						3	3		2	1
259				6	1		2	1				4		2		3		1
260	8						2	6	0		6		8	4			7	
261	8		6			2			0				8			2	6	0
262		6	7			1		2					5	4		2		0
263	5			5		5		5				5	6		5			6
264		5				7	6		5	6	6			5				6
265		5		7		1			0	8		5		4			1	
266	8		8	8				4		8		7	8	8				
267		6		6				6	7	7		6			7	6		
268		6		8			2		0		7		7			1		0
269			6		3		1	2		8			8		1			0
270	7	4					3	2		8		5		2			1	
271		6		5	0		0			8	7	5				2		
272	8		6				2	4				6	8		2		0	
273		7	7		6			6			6		7		6	5		
274		4		8	0			2				4		2		1		0
275	5	4				3		5				5	5				5	4
276	7		5	6		2				7	5	6				3		
277			4	4		5	4			5					3	3	5	
278		6		6	1				1	7		5	7				6	
279				6	4		1		2	8	3			1			2	
280			6			6	1		8	8		7					7	8
281	7	7	8	8							8	6			8	7		
282				4	8		0		6					6	5	4	8	
283	7	6			4	3						5		4	3		6	
284		5	7	7	0					8		5					0	0
285	8		7				1		0		7	5					1	0
286	3				3	3	2						6	0	6		4	
287	7	6	5		2					7		5			1		6	
288	8	5					2		0		7			4	3			0
289	6			5			0	6			7	6	6				0	
290			5		2	2	2				5					3	3	1
291		5	4	5	4					4				5			3	6
292			5		0	1	0			8			6			2	0	
293					4	4	3	1		8	7				3	3		

Appendix D (continued)

294		6	5			1			1		6	5		1			1	
295			6	7	4			5				5	6	4				0
296	5				6	5	4					1	4			3		4
297	5	1			1	0					4		5	0				0
298	8		7				1		0	8			5	4			2	
299	6	4	3	7						8				0			2	0
300				4	1		1	7				5	7		0			0
301		7		5	0			1			6			4	1			0
302		6		4		4			7	7		3	6	2				
303		5	6	5				6					5	3	2	1		
304				7	4	3	2				6		7	4		2		
305				4	4		2		2		6			3			7	2
306		5		7				7	6				6		5	5		6
307		6	6	4		2						6		1	5			0
308			8		0			1	0		7				0	3	1	
309		7	5		4		1			8		7				2		0
310			6			6	2		6	8			5			3	1	
311				5	0			0	0		5		2		0			0
312	7			6			7		8				6		7	5	8	
313				5		4	3	2			7		4			2		1
314		5		6				3	2	6	5			2		1		
315	8	5	2	7										4	0	2	1	
316			6	7				1	1		7	6		2				0
317	7				4	1			0	8	7				1	2		
318		7			1	1			0			6	6	1			1	
319	8	6			1		2						8		0	2	1	
320		7	5			1			0	7		6		4				0
321		7			0			1	6		7	5					2	0
322	6		7				5	3		6		7		4	5			
323		7			4		2		0			6			1	3		0
324	7			6			4	5			5	4		3			2	
325				7	3		3		1		7	6			2			0
326					7	7	8		7		8		7		7	7		
327			6		7			6	5		5			7	6		5	
328	8			8	3				0	8	7	6		3				
329			6	7	4			1		7	6		4	4				
330	6			6				2	2	8		6		4			7	

Appendix D (continued)

331		6	4				3		5		6			1	3			1
332			6	7			3	1		8			6	4	2			
333	7	6					2		4		6		7		3		2	
334	8					0	2	1		8	7			0	0			
335	4		3			3		4		8	8				8		8	
336		6				0	4		0	8		6				2		0
337	8	4			0	0				8		4	8					0
338		6	6			0	2				7				1	1	1	
339	6	5				7		7			6		7	6			7	
340		6	2				0	0				6	7		0			0
341	8		6		0			7		8	7	6		1				
342		7	6		4				0	8	7		7			1		
343				7		0	4	2			7	5		0		3		
344	5				3	4		6		5	5			5				6
345	7					0	3		0	8				2	1	4		
346		7	6		3		4						5	3	5		6	
347		6	6		5				7	7	6		6					8
348		4				4	4	3			5				5	5		5
349			1	1		1			1		6		5			1		1
350				6	4	3			2		7	6		4	3			
351		5	4		4		3			7				4		2		1
352		7	5	8	0					8	7				4	3		
353	8			6	3			1			4	3			4	5		
354			4	7		1		0		8					1	3	1	
355		6	8		5				6			5		7		6		6
356		8			4	0	2					6	8				2	0
357		8		6	2		0				8			5		2	0	
358	6		5	5			1			7	7					2	1	
359	5		4			4			4		4			4		4		4
360		8		8		0		0		8		7	8		0			
361		1		1			0	0		2			2			0		0
362		7	6		2				0	8		6				3		0
363			5	4	1		0						6	2			7	0
364		4	6	8			0				8			4	0		2	
365	4			4		6			3			4				4	5	5
366		2	5		5			2					6		3	4	4	
367	7			6		4			6		6			6		3	4	

Appendix D (continued)

368			4				1	2	0		5		6	0		1		
369	7	6			4	1						5	8	0			6	
370	7			6	4				1	7	5		6					1
371	4	4	4		4					5			5	3		4		
372		6				6	7	6				7	6				6	6
373	7	4	3	7						8	6	5		0				
374	4		2		3				3			3	3	3			4	
375	7			6		2			0			7	6	4		1		
376				7	1			2	1			6			1	3	1	
377		6				5		6	6		6	6	6	5				
378	8			7		1	2					6		1		2	0	
379				4		5	4		5	6			5		4			4
380			5		4		1	0			5		6	4			2	
381	8				0		0		0	8	8						0	0
382					5		5	5	5	4				4	5	5		
383		5	6	7			5			5				4		7		3
384	7		2		5	6				7	4					3	2	
385		7			1		2		1		7	6					1	1
386		4		7		6	7					7		4	3			1
387	7			8	5				0					7	2	0	7	
388	7	6			2				0		6			2	1			0
389	7	3					1		4		5	4		1			6	
390		8			4	2		0			8		6			2	0	
391				6		2	3	1		8			5	4			1	
392		6	5			2		1						4		2	1	0
393	7	7	4					1					5	4	3		7	
394			4	7				4	4		8		4	1				5
395			5	8	0		3				6			1	0			0
396	8		7		0			7			7			2	0	1		
397	3			5	4			4			4	5				5	5	
398	6	6						3	2	6	6	6	5					
399		8		7		7		5			8	5					6	7
400		6	5	8				0		4	7		8				0	
401				4		2		4	2		6	5	7					1
402	7		6				3	1				6	6			2		8
403		6			6			5	6	6				5	7			6
404	7		6		4				6			6			3		2	0

Appendix D (continued)

405			6	7			2		1		6				3		2	0
406	6	5	7						8	8		7	8		6			
407				1		0		1	2	7		7					6	6
408	3			6		4		6				2	8			6		4
409	6	7				7		6				6	7	6	5			
410		8	6		0	1					7				1		1	1
411	7	6				1			0	7				0	0			0
412			5	6	4			1		7		6	5				1	
413		7	6	6			3			8		5					6	0
414	6	7			4			2			7		6		2			1
415	3	4			5	4				5				4	4	4		
416		6	5			7			8	8				8		6	7	
417			6	8			1	2			6			4	0			0
418	7				2	3			1			5	6		1	2		
419		7		5		3		1			7	6				2	1	
420			4	4		5		4		4			5	3			4	
421		7			3	5		2				4	7		2		2	
422		5		6	0		3			6					2	3		2
423	7	6					3	1		7	7				2	3		
424	7	5						2	1				5	4	3			0
425		8	5		5		7					8		7	6			7
426			6		1	1			0	8				4		2	7	
427	3				4		6	5				4		5		4	3	
428	8	6	5		2					8	7	6		4				
429	6			6		1			1		3			3	3	4		
430	7					4		2	1		7	7	6			1		
431	6			6	1			1				6		1	2	2		
432		8				5	6	6			7			5	8			7
433	8	6	4		1						7		3				1	0
434		7		6	2			1				6	5	4				0
435		8		8	0	4						6	5		3		7	
436	5			5		5		5				5	4	5			4	
437		7	6		2	3					7		6		2		2	
438			8	7			6	7				5		7	7		6	
439	7	6						2	1				5	4			6	1
440	8	5		6		2				7		5		4			3	
441		6		7			1	1		8			7	3	1			

Appendix D (continued)

442	6				1			1	0	7				1	0			0
443		6	5		1				0			5	6	4			7	
444	6	4	4					3		8	7			4				0
445	7	6					7		7	7	6		6				6	
446			6		2	1		1			7	5	5				1	
447		6	7	6					6		7	7			7			6
448			7			6	7	6		6	7				6		7	
449	7				3			8	1			7			0		0	1
450	7	6		8	0									0	8	1	2	
451					3	2	3		1		7	6		3			2	
452			6	5	3				1				6	3		2	6	
453	8	7	5				1				7	4			1			0
454	8					1		1	0	8	7	7						0
455	8						2	1	0	8	7			4				0
456			6		3	0	1			7				3		2	1	
457			6			0	2		0	8		6		4			1	
458			4		1		1		0	8		5	8					0
459	7					7	2	1				6	3	7	7			
460	8	7		5	4									4		2	1	0
461	8		6			4		2		8					4	1		0
462		7		6	0			0		6	7				0	0		
463				5	3	2			1			4	5				0	2
464	8			6			2	0			7			4	2	1		
465	6	6			3	3					7	5					7	3
466		6		8		3		1		8	7	6			3			
467		7		7	8							7	6	7	8			
468	4		5	4					3		3	5				4	3	
469	7		5			2	2			8			7	3			7	
470	8		5	6			2					6	8		3			0
471					4	1	3	2		8			5		3		2	
472	7	6					6	2		7			7	0				0
473		4	4	4	2								4	3			5	4
474					4		3	2	0				6	4		2	1	
475			4		3		4	3		7		4	6		7			
476		8	6			0			0			5		0		0		0
477	7		6			2	1						7	1		0	6	
478		5				4	4		5	5	5	4						3

Appendix D (continued)

479	5			6	3			4		6			5		7	5		
480			5		0	2			0	7					0	3		0
481	8		6		6				1				5			5	4	1
482	8		5				3	1		6	7	4						0
483			7				2	2	3	7			7		4			0
484			4			4	4	4				4	4	4	4			
485		6				4		5	2				6	5	5			4
486	3	4		5				3		4				3	3			5
487			5	5	2		2					6			1	1		0
488			4		6	5	1				5		5				5	4
489			5		1			0	2	8		6	8		4			
490	7			7	3	0				7	6			3		1		
491			4		5	4		5		5	5			4	6			
492	4	4						4	4			4	4			4		4
493	8	7	5	8								5		2		3		0
494		7		8			1		0	8				3	8		7	
495					1	2	3		1		5			3	3			1
496	6				0		1	2			5		6				1	0
497				5	1			0	0	7					1		1	0
498		6	5				2	1			7					2	1	0
499			5	8				1	0	8			6		2			0
500		8			4	6	1						6	1	2			1
501	6	6	5	4								4			5		6	4
502	6			6	4	4						5		6		5	4	
503	7	6	8		4						7		6	6				3
504		6		6	7				7	6	7				6	7		
505		5			0	2	4			7		6			5	5		
506		7	4		1			2		7	6	6						3
507				5	0	0	2						5	2	1	2		
508			7		2	0		2		8			6	4	2			
509		7	5				0		0		7	6		0			0	
510	8			5	1			8		8	7	4			1			
511			4		5		2	2			7		8				1	0
512	8			4		0		0		8	5						3	0
513	8				2		1		1	8		6		1	1			
514		6		7	0	1					7	6				2	1	
515		5		6			2		0	6		5		3			2	

Appendix D (continued)

516				6		2	1		0		6		7				1	0
517	8			7		0	3				6		7	4			2	
518			4				4	4	4		4	4	4		4			
519			4		5		5	4		5	6				6	3		
520		7		8		2	1					6			1	3	7	
521		7		6	2				7		7		6		1		1	
522		7	5	8					0					4		2	0	0
523			7	6		6			7		6				7	7	6	
524		6	5	6			1						5		2	1		5
525	6	6					2		2		7		5	4			1	
526			5	5	7			7		7				4		5		6
527		7			1			2	1			7	6		2			1
528			7		4	2	1						7		3		1	0
529			6		0			7	3	7			8		0			2
530				5			1	2	2						3	3	3	4
531	7	7	6				1			7	6	6	5					
532	8	7					3	1				5	8				2	0
533		2		6	0				0	8	7		6		1			
534	8		4		1	4				8			7	1	1			
535		7	6			2	1			8				3		4	1	
536	7				0	1		1						4	3		7	0
537			5		2	1			0	8			6		2			0
538		6	4	7		0					7	6	5	3				
539			5	6			1	2		8	7				3			0
540	8		3	6			0				8	6	7				0	
541			7	8		2		6				6	8				7	0
542		8			0	0			0		8			4		2	0	
543		7	5				4		8	8	7	7					7	
544	5		6		5				6	5	6						6	6
545	8	6			4		2			8		6	8				2	
546	5	3	5			3					2		4	6		6		
547		7	7				4		0	8			5		2	1		
548		5	5		5			5			5			2	2	2		
549			7		1		1	1		8	7			3	5			
550		7	3	4				3				3		6		3	8	
551			6			2	6	2		8			8	1	0			
552				6		2	5		7	8	7				2		7	

Appendix D (continued)

553		6		7	0		1			8					2	2	7	
554	7	6	4	7							6	6			2	2		
555	7	5				1	4				5	4	7	0				
556					4	3	2	3			1		3	4				0
557	8			5	0				0	8		5	6		1			
558	4			5	6		4					5	6		5		3	
559			7		6	5			6				6	6	7		5	
560		5	4	4				4			5	4			4		3	
561		8	7	7			1			7	7			0	1			
562		7	7		4		5					7	7	6		5		
563				6		2	1		0	8		7				2		0
564				7		1	1	4		8		4					6	0
565		6			4		4	2		8		7	7	0				
566		6	5				3		1			5		2	4		3	
567		6		5	3				1			6	5		3	2		
568					4		1	1	0		7	6		4	0			
569		6	5	7			3				7	6					1	1
570				7	5	7			6	4	8		7	6				
571		6	4				2		7	4			3	2		1		
572	8	8	8				0			8			4		0	0		
573		6			4	2	1			8	7	5	6					
574		7			3			1	0			6	5	4				0
575		4	5			3	2					5			2		6	0
576					2	0		1	1	7	6		7				1	
577	4	4				4		4							6	6	5	4
578	8	7					1	1			6	6			1			0
579		6			4		2		0	8		6		4			1	
580		7		6	0	1				8			7			2	7	
581	4			5	4		3				6	4		5		6		
582		6		7	4	4					6				1	2		0
583		7	5		2			2		8	5	5					3	
584		6		7		7			5	7		5				7	5	
585		7				0		1	0	8		6				2		0
586	4		6			4		6					6			6	4	4
587	4		2				2		5	5				0		1	1	
588		8	4	6					1	8		6	5		3			
589			6				3	2	0				7	3	3		3	

Appendix D (continued)

590	7		5	2					2					3		1	0	1
591			5			4	5		4	4	6			4			6	
592	7	4	3						0			4		4	3			3

Appendix E

Participant's demographic information (exclusive of outliers and those who are not jury eligible)

	N	%
Jury Eligibility		
Eligible	592	100
Not Eligible	0	0
Age		
< 23 years	26	4.4
24-41 years	380	64.4
42-54 years	128	21.7
55-73 years	55	9.3
73 years +	1	0.2
Education		
< High School Diploma	5	0.8
High School Degree or Equivalent (GED)	66	11.2
Some College, No Degree	115	19.5
Associate's Degree	66	11.2
Bachelor's Degree	268	45.4
Advanced Degree	70	11.9
Other	0	0.0
Gender		
Female	235	39.8
Male	353	59.8
Non-Binary	1	0.2
Ethnicity		
American Indian or Alaska Native	2	0.3
Asian	35	5.9
Black or African American	66	11.2
Hispanic or Latino	38	6.5
Multiracial	15	2.5
Native Hawaiian or Pacific Islander	1	0.2
White	432	73.3

	N	%
Employment		
Employed for Wages	450	76.4
Self-Employed	79	13.4
Out of Work	12	2.0
A Homemaker	20	3.4
A Student	9	1.5
Military	0	0.0
Retired	15	2.5
Unable to Work	4	0.7
English as Native Language		
Yes	575	98.5
No	9	1.5
Childhood Residence		
Urbanized Area	432	82.9
Urban Cluster	16	3.1
Rural	68	13.1
Not United States	5	1.0
Current Residence		
Urbanized Area	439	84.3
Urban Cluster	19	3.6
Rural	63	12.1
Not United States	0	0.0

CITED LITERATURE

1. *Timothy Bridges v. City of Charlotte, et al.*, 564 N.E (2017)

2. Nakao MA, Axelrod S. Numbers are better than words: Verbal specifications of frequency have no place in medicine. *The American Journal of Medicine*. 1983;74(6):1061-1065. https://www.sciencedirect.com/science/article/pii/0002934383908197. doi: 10.1016/0002-9343(83)90819-7.

3. Budescu D, Wallsten T. Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*. 1985;36(3).

4. Budescu D, Wallsten T. *Decision making from a cognitive perspective*. Volume 32 ed. Academic Press; 1995:275-313.

5. Brun W, Teigen K. Verbal Probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*. 1988;41(3).

6. Vidmar N, Diamond S. Juries and Expert Evidence, 66 *Brooklyn Law Review* 1121 (2001).

7. Schuller RA, Cripps J. Expert evidence pertaining to battered women. *Law Hum Behav*. 1998;22(1):17-31. https://www.jstor.org/stable/1393994. doi: 1025772604721.

8. Schuller RA, Terry D, McKimmie B. The impact of expert testimony on jurors' decisions: Gender of the expert and testimony complexity. *Journal of Applied Social Psychology*. 2005;35(6):1266-1280. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-1816.2005.tb02170.x. doi: 10.1111/j.1559-1816.2005.tb02170.x.

9. Susan Berk-Seligson. *The Bilingual Courtroom*. The University of Chicago Press; 2002. http://data.theeuropeanlibrary.org/BibliographicResource/3000087289627.

10. Swenson, R. A., Nash, D. L., & Roos, D. C. (1984). Source credibility and perceived expertness of testimony in a simulated child-custody case. *Professional Psychology: Research and Practice, 15*(6), 891-898. http://dx.doi.org/10.1037/0735-7028.15.6.891

11. McKimmie BM, Newton SA, Schuller RA, Terry DJ. It's not what she says, it's how she says it: The influence of language complexity and cognitive load on the persuasiveness of expert testimony. *Psychiatry, Psychology and Law*. 2013;20(4):578-589. doi: 10.1080/13218719.2012.727068.

12. Martire K, Kemp R, Watkins I, Sayle M, Newell B. The expression and interpretation of uncertain forensic science evidence: verbal equivalence, evidence strength and the weak evidence effect. *Law and Human Behavior*. 2013;37.

13. *Frye v. United States*, 293 F. 1013 (D.C Cir. 1923)

14. Fed. R. Evid. 702

15. *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S 579 (1993)

16. *General Electric Co. v. Joiner*, 522 U.S. 136 (1997)

17. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999)

18. *Melendez-Diaz v. Massachusetts*, 557 U.S 305 (2009)

19. "The Constitution of the United States," Amendment 6.

20. Morgenstern, M. (2018, August 09). Daubert v. Frye – A State-by-State Comparison. Retrieved from https://www.theexpertinstitute.com/daubert-v-frye-a-state-by-state-comparison/

21. Ivković SK, Hans VP. Jurors' evaluations of expert testimony: Judging the messenger and the message. *Law & Social Inquiry*. 2003;28(2):441-482. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1747-4469.2003.tb00198.x. doi: 10.1111/j.1747-4469.2003.tb00198.x.

22. Anthony Champagne, Daniel W. Shuman, Elizabeth Whitaker. Expert witnesses in the courts: An empirical examination. *Judicature 76*. 1992.

23. Daniel W. Shuman, Elizabeth Whitaker, Anthony Champagne. An empirical examination of the use of expert witnesses in the courts- part II: A three city study. *Jurimetrics*. 1994;34.

24. Swim J, Borgida E, Maruyama G, Myers DG. Joan McKay versus john McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*. 1989;105(3):409-429. https://search.proquest.com/docview/1296896406. doi: 10.1037/0033-2909.105.3.409.

25. Eagly AH, Makhijani MG, Klonsky BG. Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*. 1992;111(1):3-22. doi: 10.1037//0033-2909.111.1.3.

26. Schuller RA, Mckimmie BM, Janz T. The impact of expert testimony in trials of battered women who kill. *Psychiatry, Psychology and Law*. 2004;11(1):1-12. http://www.tandfonline.com/doi/abs/10.1375/pplt.2004.11.1.1. doi: 10.1375/pplt.2004.11.1.1.

27. Schuller, Regina A., Terry, Deborah J. and McKimmie, Blake (2001). The impact of an expert's gender on jurors' decisions. *Law & Psychology Review 25* 59-79.

28. O'Barr, William M. Linguistic evidence: Language, power, and strategy in the courtroom. US: Academic Press; 1982. http://catalog.hathitrust.org/Record/000148781.

29. Stephan C, Stephan WG. Habla Ingles? The effects of language translation on simulated juror Decisions. *Journal of Applied Social Psychology*. 1986;16(7):577-589. doi: 10.1111/j.1559-1816.1986.tb01160.x.

30. Norma A Mendoza, Harmon M Hosch, Bruce J Ponder, Victor Carrillo. Well...ah...: Hesitations and hedges as an influence on jurors' decisions. *Journal of Applied Social Psychology*. 2000;30(12):2610. https://search.proquest.com/docview/1873342287. doi: 10.1111/j.1559-1816.2000.tb02452.x.

31. Court Interpreters Deliver Justice in All Languages. (2017, August 10). Retrieved from https://www.uscourts.gov/news/2017/08/10/court-interpreters-deliver-justice-all-languages

32. Document #10 Standards for Examining Friction Ridge... (2013, April 27). Retrieved February, 2019, from https://www.nist.gov/document/swgfastexaminations-conclusions20130427pdf.

33. Riordan WM, Gustafson JA, Fitzgerald MP, Lewis JA. Forensic document examination. In: *Forensic science*. Chichester, UK: John Wiley & Sons, Ltd; 2012:225-251. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118373880.ch9. 10.1002/9781118373880.ch9.

34. Robertson B. *Interpreting evidence: Evaluating forensic science in the courtroom*. Second edition ed. Chichester, West Sussex, UK: Wiley; 2016:55-67. https://onlinelibrary.wiley.com/doi/book/10.1002/9781118492475.

35. Evett I. Expressing Evaluative Opinions: A position statement. (2011). *Science & Justice, 51*(1), 1-2. doi:10.1016/j.scijus.2011.01.002

36. National Research Council. 2009. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12589.

37. Karelitz TM, Budescu DV. You say "probable" and I say "likely". *Journal of Experimental Psychology: Applied*. 2004;10(1):25-41. https://www.ncbi.nlm.nih.gov/pubmed/15053700. doi: 10.1037/1076-898X.10.1.25.

38. Mullet E, Rivet I. Comprehension of verbal probability expressions in children and adolescents. *Language and Communication*. 1991;11.

39. Budescu D, Weinburg S, Wallsten T. Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*. 1988;14.

40. Fillenbaum S, Wallsten TS, Cohen B, Cox JA. Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *American Journal of Psychology*. 1991(104):35-60.

41. G. Jackson, et al., Communicating the results of forensic science examinations, in: Cedric Neumann, Anjali Ranadive, David H. Kaye (Eds.), Final Technical Report for NIST Award 70NANB12H014, 2015, 2015; Penn State Law Research

42. Casillas S, Barbadilla A. Molecular population genetics. *Genetics*. 2017;205(3):1003-1035. https://www.ncbi.nlm.nih.gov/pubmed/28270526. doi: 10.1534/genetics.116.196493.

43. *Mayfield v. United States*, 599 F.3d 964 (2010)

44. Saks MJ, Faigman DL. Failed Forensics: How forensic science lost its way and how it might yet find it. *Annual Review of Law and Social Science*. 2008;4(1):149-171. doi: 10.1146/annurev.lawsocsci.4.110707.172303.

45. Stevenage S, Pitfield C. Fact or fiction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis. *Forensic science international*. 2016.

46. Hichy Z, Gerges MHH, Santisi G. Interpretation of probability phrases in intergroup relations effects of context and prior beliefs. *Procedia - Social and Behavioral Sciences*. 2013;82:257-260. https://www.sciencedirect.com/science/article/pii/S1877042813013220. doi: 10.1016/j.sbspro.2013.06.255.

47. Wallsten T, Fillenbaum S, Cox J. Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*. 1986:571-587. http://dx.doi.org/10.1016/0749-596X(86)90012-4.

48. Fernbach PM, Darlow A, Sloman SA. Asymmetries in predictive and diagnostic reasoning. *Journal of experimental psychology. General*. 2011;140(2):168-185. https://www.ncbi.nlm.nih.gov/pubmed/21219081. doi: 10.1037/a0022100.

49. Martire, K.A., Kemp, R.I., Sayle, M., Newell, B.R. On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*. 2014;240:61-68. https://www.clinicalkey.es/playcontent/1-s2.0-S0379073814001364. doi: 10.1016/j.forsciint.2014.04.005.

50. De Kinder J, Olsson T. Expressing evaluative opinions: A position statement. *Science and Justice*. 2011;51(1):1-2. https://www.clinicalkey.es/playcontent/1-s2.0-S1355030611000037. doi: 10.1016/j.scijus.2011.01.002.

51. Association of Forensic Science Providers. Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice*. 2009;49(3):161-164. https://www.clinicalkey.es/playcontent/1-s2.0-S1355030609001087. doi: 10.1016/j.scijus.2009.07.004.

52. Biedermann A, Champod C, Willis S. Development of European standards for evaluative reporting in forensic science. *The International Journal of Evidence & Proof*. 2017;21(1-2):14-29. https://journals.sagepub.com/doi/full/10.1177/1365712716674796. doi: 10.1177/1365712716674796.

53. Martire KA, Watkins I. Perception problems of the verbal scale: A reanalysis and application of a
membership function approach. *Science and Justice*. 2015;55.
http://www.sciencedirect.com/science/journal/13550306.

54. Mullen C, Spence D, Moxey L, Jamieson A. Perception problems of the verbal scale. *Science and Justice*. 2014;54(2):154-158. https://www.clinicalkey.es/playcontent/1-s2.0-S1355030613001160. doi: 10.1016/j.scijus.2013.10.004.

55. Arscott E, Morgan RM, Meakin GE, French JC. Understanding forensic expert evaluative evidence: A study of the perception of verbal expressions of the strength of evidence. *Science and Justice*. 2017.
https://www.openaire.eu/search/publication?articleId=od_____322::c2dfaf6b43cf2b8683d9ea920fcf9bcf.

56. Sjerps M, Biesheuvel DB. The interpretation of conventional and bayesian verbal scales for expressing expert opinion: A small experiment among jurists. *Forensic Linguistics*. 1999;6(2):214-227. doi: 10.1558/sll.1999.6.2.214.

57. Kane J, Barabas J. No harm in checking: Using factual manipulation checks to assess attentiveness in experiments. *American Journal of Political Science*. 2019.
doi.org/10.1111/ajps.12396

58. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*. 2009;45(4):867-872. https://www.sciencedirect.com/science/article/pii/S0022103109000766. doi: 10.1016/j.jesp.2009.03.009.

59. Clifford S, Jerit J. Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Experimental Political Science*. 2014. doi: 10.1017/xps.2014.5.

60. Hara K, Adams A, Milland K, Savage S, Callison-Burch C, Bigham J. A data-driven analysis of workers' earnings on amazon mechanical turk. *Proceedings of the 2018 CHI Conference on human factors in computing systems*. Apr 21, 2018:1-14.
http://dl.acm.org/citation.cfm?id=3174023. doi: 10.1145/3173574.3174023.

61. Berry DT, Wetter MW, Baer RA, Larsen L. MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*. 1992;4(3):340-345. doi: 10.1037//1040-3590.4.3.340.

62. Baer R, Ballenger J, Berry D, Wetter M. Detection of random responding on the MMPI-A. *Personality Assessment*. 1997:139-151.

63. Rogelberg SG, Fisher GG, Maynard DC, Hakel MD, Horvath M. Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods*. 2001;4(1):3-25. https://journals.sagepub.com/doi/full/10.1177/109442810141001. doi: 10.1177/109442810141001.

64. Landers RN, Behrend TS. An inconvenient truth: Arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Industrial and Organizational Psychology*. 2015;8(2):142-164. https://search.proquest.com/docview/1721577447. doi: 10.1017/iop.2015.13.

65. Galesic M, Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opin Q*. 2009;73(2):349-360. https://www.jstor.org/stable/25548084. doi: 10.1093/poq/nfp031.

66. Klein RA, Ratliff KA, Vianello M, et al. Investigating variation in replicability. A "many labs" replication project. *Social psychology*. 2014;45(3):152. http://www.fachportal-paedagogik.de/fis_bildung/suche/fis_set.html?FId=1082035. doi: 10.1027/1864-9335/a000178.

67. Chandler JJ, Paolacci G. Lie for a dime. *Social Psychological and Personality Science*. 2017;8(5):500-508. https://journals.sagepub.com/doi/full/10.1177/1948550617698203. doi: 10.1177/1948550617698203.

68. Chandler J, Mueller P, Paolacci G. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behav Res*. 2014;46(1):112-130. https://www.ncbi.nlm.nih.gov/pubmed/23835650. doi: 10.3758/s13428-013-0365-7.

69. Dhar S, Tansek R, Toftey E, Kaye K. Observer bias in hand hygiene compliance testing. *The Society for Healthcare Epidemiology of America*. 2010;31.

70. Adair JG. The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*. 1984;69(2):334-345. doi: 10.1037//0021-9010.69.2.334.

71. Schillewaert N, Meulemeester P. Comparing response distributions of offline and online. *International Journal of Market Research*. 2005;47(2):163-178. doi: 10.1177/147078530504700203.

72. Hauser, D. J., and Schwarz, N. (2015). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. SAGE Open 5, 1–6. doi: 10.1177/2158244015584617

73. Hauser DJ, Ellsworth PC, Gonzalez R. Are manipulation checks necessary? *Frontiers in psychology*. 2018;9:998. https://www.ncbi.nlm.nih.gov/pubmed/29977213. doi: 10.3389/fpsyg.2018.00998.

74. Fleischer A, Mead AD, Huang J. Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*. 2015;8(2):196-202. https://search.proquest.com/docview/1721577480. doi: 10.1017/iop.2015.25.

75. Wessling KS. MTurk character misrepresentation. *Journal of consumer research*. 2017;44(1):211-230. http://www.econis.eu/PPNSET?PPN=1009498452.

76. Hauser D, Schwarz N. Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*. 2016;48(1):400-407. https://www.ncbi.nlm.nih.gov/pubmed/25761395. doi: 10.3758/s13428-015-0578-z.

77. Huff C, Tingley D. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research and Politics*. 2015. https://journals-sagepub-com.proxy.cc.uic.edu/doi/pdf/10.1177/2053168015604648. doi: 10.1177/2053168015604648.

78. Casler K, Bickel L, Hackett E. 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*. 2013;29(6):2156-2160. https://www.sciencedirect.com/science/article/pii/S074756321300160X. doi: 10.1016/j.chb.2013.05.009.

79. Necka E.A., Cacioppo S, Norman G.J., Cacioppo J.T., Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PLoS One*. 2016;11(6):e0157732. https://www.ncbi.nlm.nih.gov/pubmed/27351378. doi: 10.1371/journal.pone.0157732.

80. Dobosh M. Survey: Demographic Questions. *The SAGE encyclopedia of communication research methods*. Thousand Oaks: SAGE Publications, Inc; 2018.

81. Krosnick, J. A. (1999). Survey research. Annu. Rev. Psychol. 50, 537–567. doi: 10.1146/annurev.psych.50.1.537

82. Porter SR, Whitcomb ME, Weitzer WH. Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*. 2004;2004(121):63-73. doi: 10.1002/ir.101.

83. Fulton BR. Organizations and survey research. *Sociological Methods & Research*. 2018;47(2):240-276. https://journals.sagepub.com/doi/full/10.1177/0049124115626169. doi: 10.1177/0049124115626169.

84. Schuller RA, Terry D, McKimmie B. The impact of expert testimony on jurors' decisions: Gender of the expert and testimony Complexity1. *Journal of Applied Social Psychology*. 2005;35(6):1266-1280. doi: 10.1111/j.1559-1816.2005.tb02170.x.

85. Holyk GG. Context Effect. *The SAGE encyclopedia of survey research methods*. Thousand Oaks: SAGE Publications, Inc; 2011.

86. Buskirk TD, Sanders T, Michaud J. Are sliders too slick for surveys? An experiment comparing slider and radio button scales for smartphone, tablet and computer based surveys. *Methods, Data, Analyses*. 2015;9(2):229-261. https://search.datacite.org/works/10.12758/mda.2015.013. doi: 10.12758/mda.2015.013.

87. Cape P. Slider scales in online surveys. Paper presented at the CASRO Panel Conference, Feb 2-3, 2009, New Orleans, LA.

88. Funke F. A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*. 2016;34(2):244-254. https://journals.sagepub.com/doi/full/10.1177/0894439315575477. doi: 10.1177/0894439315575477.

89. Funke F, Reips U, Thomas RK. Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*. 2011;29(2):221-231. https://journals.sagepub.com/doi/full/10.1177/0894439310376896. doi: 10.1177/0894439310376896.

90. Blascovich J, Spencer SJ, Quinn D, Steele C. African Americans and high blood pressure: the role of stereotype threat. *Psychological Science*. 2001;12(3):225-229.

91. Marchant G. Demographics *The SAGE encyclopedia of survey research methods*. Thousand Oaks: SAGE Publications, Inc; 2018:486-488.

92. Hughes JL, Camden AA, Yangchen T. Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research*. 2016;21(3):138-151. doi: 10.24839/2164-8204.JN21.3.138.

93. Durand C. Surveys and society. *The SAGE handbook of survey methodology*. 2016:57-66. https://search.credoreference.com/content/entry/sageukxuzo/surveys_and_society/0.

94. State of Illinois circuit court of cook county. (n.d.). 13.4 pre-trial phase. http://www.cookcountycourt.org/FORATTORNEYSLITIGANTS/RulesoftheCourt/ReadLocalRule/tabid/1139/ArticleId/2277/-13-4-Pre-Trial-Phase.aspx.

VITA

EDUCATION

University of Illinois at Chicago, Chicago, IL, FEPAC accredited

- Forensic Science M.S, 2019
- *Summa cum laude*
- Relevant Coursework: Physical Pattern Evidence, Drug Analysis and Toxicology, Biological and Trace Evidence, Expert Witness Testimony and Courtroom Demeanor

University of Southern California, Los Angeles, CA

- Neuroscience B.A, 2017
- Drawing and Psychology minor
- *Magna cum laude*
- Relevant Coursework: Neurobiology, Cognitive Neuroscience, Advanced Drawing, Systems Neuroscience, Behavioral Genetics, Organic Chemistry I & II

RESEARCH EXPERIENCE:

Graduate Research Assistant (Aug 2018 – Dec 2019) Chicago, IL
University of Illinois at Chicago

Investigating juror understanding of the specific verbiage used in Questioned Document Examiner's expert testimony as dictated by the existing SWGDOC conclusion scale. Jury eligible participants (n=592) reviewed a short testimonial script and then completed an evaluative survey. Conclusion scale findings will aid OSAC as they are currently drafting a new scale to utilize in court.

- Built surveys through Qualtrics software
- Analyzed data via Statistical Analysis Software (SAS)

Research Assistant (May 2014 – May 2017) Los Angeles, CA
University of Southern California

Conducted research experiments to measure and analyze the interaction between ghrelin signaling in higher-order brain regions to mediate cognitive and motivational based aspects of feeding behavior.

- Assisted with surgical cannula implantations
- Handled the Sprague-Dawley rats to perform neuropharmacological injections
- Executed behavioral paradigms
- Managed food intake and body weight measurements
- Data collection and analysis to garner insights about the neuroendocrine system changes
- Performed brain tissue extractions and brain sectioning for immunohistochemical processing
- Assisted in neurobiological assays, such as Western blots and immunohistochemistry

PUBLICATIONS

1. Suarez A.N., Hsu T.M., Liu C.M, Noble E.E., Cortella A.M., **Nakamoto E.M.**, Hanh J.D., Lartigue G., Kanoski S.E.(2017). Gut vagal sensory signaling regulates hippocampus function through multi-order pathways. Nature Communications.

2. Hsu T.M., Hahn J.D., Konanur V.R., Noble E.E., Suarez A.N., Thai J, **Nakamoto E.M.**, Kanoski S.E. (2015).
Hippocampus ghrelin signaling mediates appetite through lateral hypothalamic orexin pathways. eLife.

SCIENTIFIC PRESENTATIONS:

1. **Nakamoto E.M.**, Hsu T.M., Konanur V.R., Kanoski S.E. (2016). Hippocampus ghrelin signaling mediates appetite through lateral hypothalamic orexin pathways. USC Undergraduate Research Symposium (Annual Meeting).

TRAINING

- Received polarized light microscopy training by McCrone Institute in (2019)

PROFESSIONAL ORGANIZATIONS

- American Academy of Forensic Science (AAFS) Student Affiliate (February 2019-Present)

HONORS AND AWARDS:

- UIC Award for Graduate Research (May, 2019)
- Finalist in the Three Minute Thesis spoken word competition for UIC Masters/PhD graduate students (Mar, 2019)
- Stanley & Betty Jane Susina Scholarship awarded for scholarly work (Feb, 2019)
- USC Renaissance Scholar (Awarded to those who received >3.5 GPA in two widely disparate fields)
- USC Dornsife Dean's List (Fall 2014, Spring 2015, Fall 2015, Spring 2016, Fall 2016, Spring 2017)
- Provost Undergraduate Research Fellowship recipient (Summer 2015, Spring 2016, Spring 2017)
- Student Opportunities for Academic Research (SOAR) Fellowship recipient (Fall 2014)

TEACHING EXPERIENCE:

- **USC Interaxon President** (February 2015-May 2017)
    - President of the executive council (May 2015- May 2017)
        - Coordinated meetings between Interaxon and neighboring schools
        - Created and presented neuroscience lectures at neighboring elementary, middle, and secondary schools to encourage awareness about the field of neuroscience

    - Coordinated Neurocamp for underserved high school students to further their exposure to science
        - Led dissections and laboratory tours

ACTIVITIES:

- **USC Neuroscience Executive Committee** (August 2016-May 2017)
    - Served with the executive heads of the Neuroscience undergraduate program and neuroscience advisors to provide insight on the curriculum and effective communication with the undergraduates

WORKING REFERENCES
Linked to refworks
DO NOT DELETE

1. Rudolf D, Peterson D, Newbold M. Timothy bridges v. city of charlotte, et al. . 2017.

2. Nakao MA, Axelrod S. Numbers are better than words: Verbal specifications of frequency have no place in medicine. *The American Journal of Medicine*. 1983;74(6):1061-1065. https://www.sciencedirect.com/science/article/pii/0002934383908197. doi: 10.1016/0002-9343(83)90819-7.

3. Budescu D, Wallsten T. Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*. 1985;36(3).

4. Budescu D, Wallsten T. *Decision making from a cognitive perspective*. Volume 32 ed. Academic Press; 1995:275-313.

5. Brun W, Teigen K. Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*. 1988;41(3).

6. Vidmar N, Diamond SS. Juries and expert evidence. . 2001.

7. Schuller RA, Cripps J. Expert evidence pertaining to battered women. *Law Hum Behav*. 1998;22(1):17-31. https://www.jstor.org/stable/1393994. doi: 1025772604721.

8. Schuller RA, Terry D, McKimmie B. The impact of expert testimony on jurors' decisions: Gender of the expert and testimony complexity. *Journal of Applied Social Psychology*. 2005;35(6):1266-1280. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-1816.2005.tb02170.x. doi: 10.1111/j.1559-1816.2005.tb02170.x.

9. Susan. Berk-Seligson. *The bilingual courtroom*. The University of Chicago Press; 2002. http://data.theeuropeanlibrary.org/BibliographicResource/3000087289627.

10. Swenson RA, Nash DL, Roos DC. Source credibility and perceived expertness of testimony in a
simulated child-custody case. *Professional Psychology: Research and Practice*. 1984;15.

11. McKimmie BM, Newton SA, Schuller RA, Terry DJ. It's not what she says, it's how she says it: The influence of language complexity and cognitive load on the persuasiveness of expert testimony. *Psychiatry, Psychology and Law*. 2013;20(4):578-589. doi: 10.1080/13218719.2012.727068.

12. Martire K, Kemp R, Watkins I, Sayle M, Newell B. The expression and interpretation of uncertain forensic science evidence: verbal equivalence, evidence strength and the weak evidence effect. *Law and Human Behavior*. 2013;37.

13. VAN ORSD EL, Associate Justice. FRYE v.UNITED STATES. 293 F. 1013 (D.C.. cir 1923). .

14. Sultan L, Mandel M, Phillips SS, Forth RF. Federal rules of evidence. . 2017.

15. DAUBERT, et al. v. MERRELL
DOW PHARMACEUTICALS, INC. . 1992.

16. United States. Supreme Court. GENERAL ELECTRIC CO.et al.v.JOINER et ux. . 1997.

17. AJ Grudzinskas J. Kumho tire co., ltd. v. carmichael. . 1999;27(3):482.

18. Melendez-diaz v. massachusetts, 557 U.S. 305 (2009). .

19. *Civil rights in the united states*. United States: .
http://catalog.hathitrust.org/Record/004574344.

20. Morgenstern M. Daubert v. frye – A state-by-state comparison. . 2017.

21. Ivković SK, Hans VP. Jurors'Evaluations of expert testimony: Judging the messenger and the message. *Law & Social Inquiry*. 2003;28(2):441-482.
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1747-4469.2003.tb00198.x. doi: 10.1111/j.1747-4469.2003.tb00198.x.

22. Anthony Champagne, Daniel W. Shuman, Elizabeth Whitaker. Expert witnesses in the courts: An empirical examination. *Judicature 76*. 1992.

23. Daniel W. Shuman, Elizabeth Whitaker, Anthony Champagne. An empirical examination of the use of expert witnesses in the courts- part II: A three city study. *Jurimetrics*. 1994;34.

24. Swim J, Borgida E, Maruyama G, Myers DG. Joan McKay versus john McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*. 1989;105(3):409-429.
https://search.proquest.com/docview/1296896406. doi: 10.1037/0033-2909.105.3.409.

25. Eagly AH, Makhijani MG, Klonsky BG. Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*. 1992;111(1):3-22. doi: 10.1037//0033-2909.111.1.3.

26. Schuller RA, Mckimmie BM, Janz T. The impact of expert testimony in trials of battered women who kill. *Psychiatry, Psychology and Law*. 2004;11(1):1-12.
http://www.tandfonline.com/doi/abs/10.1375/pplt.2004.11.1.1. doi: 10.1375/pplt.2004.11.1.1.

27. Regina A. Schuller, Deborah Terry, Blake McKimmie. The impact of an expert's gender on jurors' decisions. . 2001.

28. O'Barr, William M. *Linguistic evidence: Language, power, and strategy in the courtroom*. US: Academic Press; 1982. http://catalog.hathitrust.org/Record/000148781.

29. Stephan C, Stephan WG. Habla ingles? the effects of language translation on simulated juror Decisions1. *Journal of Applied Social Psychology*. 1986;16(7):577-589. doi: 10.1111/j.1559-1816.1986.tb01160.x.

30. Norma A Mendoza, Harmon M Hosch, Bruce J Ponder, Victor Carrillo. Well...ah...: Hesitations and hedges as an influence on jurors' Decisions1. *Journal of Applied Social Psychology*. 2000;30(12):2610. https://search.proquest.com/docview/1873342287. doi: 10.1111/j.1559-1816.2000.tb02452.x.

31. uscourts.gov. Court interpreters deliver justice in all languages. https://www.uscourts.gov/news/2017/08/10/court-interpreters-deliver-justice-all-languages. Updated 2017.

32. 1. Preamble 1.1. Friction ridge impression examinations are conducted examiners using the Analysis, Comparison, Evaluation, et al. Document #10 standards for examining friction ridge impressions and resulting conclusions (latent/tenprint). .

33. Riordan WM, Gustafson JA, Fitzgerald MP, Lewis JA. Forensic document examination. In: *Forensic science*. Chichester, UK: John Wiley & Sons, Ltd; 2012:225-251. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118373880.ch9. 10.1002/9781118373880.ch9.

34. Robertson B. *Interpreting evidence: Evaluating forensic science in the courtroom*. Second edition ed. Chichester, West Sussex, UK: Wiley; 2016:55-67. https://onlinelibrary.wiley.com/doi/book/10.1002/9781118492475.

35. Evett I. Expressing evaluative opinions: Position statement. *Science and Justice*. (51qa).

36. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council. Strengthening forensic science in the united states: A path forward. . 2009:13-190,.

37. Karelitz TM, Budescu DV. You say "probable" and I say "likely". *Journal of Experimental Psychology: Applied*. 2004;10(1):25-41. https://www.ncbi.nlm.nih.gov/pubmed/15053700. doi: 10.1037/1076-898X.10.1.25.

38. Mullet E, Rivet I. Comprehension of verbal probability expressions in children and adolescents. *Language and Communication*. 1991;11.

39. Budescu D, Weinburg S, Wallsten T. Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*. 1988;14.

40. Fillenbaum S, Wallsten TS, Cohen B, Cox JA. Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *American Journal of Psychology*. 1991(104):35-60.

41. Steven D. Schwinn, Ruthann Robson. Jackson et al. on communicating the results of forensic science examinations. Constitutional Law Prof Blog Web site. http://scholar.aci.info/view/1449d89f838141d0104/1513b42444500010003. Updated 2015.

42. Casillas S, Barbadilla A. Molecular population genetics. *Genetics*. 2017;205(3):1003-1035. https://www.ncbi.nlm.nih.gov/pubmed/28270526. doi: 10.1534/genetics.116.196493.

43. Letter D, McIntosh S, Rosenthal E. Mayfield v. united states. . 2009.

44. Saks MJ, Faigman DL. Failed forensics: How forensic science lost its way and how it might yet find it. *Annual Review of Law and Social Science*. 2008;4(1):149-171. doi: 10.1146/annurev.lawsocsci.4.110707.172303.

45. Stevenage S, Pitfield C. Fact or friction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis. *Forensic science international*. 2016.

46. Hichy Z, Gerges MHH, Santisi G. Interpretation of probability phrases in intergroup relations effects of context and prior beliefs. *Procedia - Social and Behavioral Sciences*. 2013;82:257-260. https://www.sciencedirect.com/science/article/pii/S1877042813013220. doi: 10.1016/j.sbspro.2013.06.255.

47. Wallsten T, Fillenbaum S, Cox J. Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*. 1986:571-587. http://dx.doi.org/10.1016/0749-596X(86)90012-4.

48. Fernbach PM, Darlow A, Sloman SA. Asymmetries in predictive and diagnostic reasoning. *Journal of experimental psychology. General*. 2011;140(2):168-185. https://www.ncbi.nlm.nih.gov/pubmed/21219081. doi: 10.1037/a0022100.

49. Martire, K.A.|Kemp, R.I.|Sayle, M.|Newell, B.R. On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*. 2014;240:61-68. https://www.clinicalkey.es/playcontent/1-s2.0-S0379073814001364. doi: 10.1016/j.forsciint.2014.04.005.

50. De Kinder J, Olsson T. Expressing evaluative opinions: A position statement. *Science and Justice*. 2011;51(1):1-2. https://www.clinicalkey.es/playcontent/1-s2.0-S1355030611000037. doi: 10.1016/j.scijus.2011.01.002.

51. Association of Forensic Science Providers. Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice*. 2009;49(3):161-164. https://www.clinicalkey.es/playcontent/1-s2.0-S1355030609001087. doi: 10.1016/j.scijus.2009.07.004.

52. Biedermann A, Champod C, Willis S. Development of european standards for evaluative reporting in forensic science. *The International Journal of Evidence & Proof*. 2017;21(1-2):14-29. https://journals.sagepub.com/doi/full/10.1177/1365712716674796. doi: 10.1177/1365712716674796.

53. Martire K, Watkins I. Perception problems of the verbal scale: A reanalysis and application of a
membership function approach. *Science and Justice*. 2015(10). http://lib.cqvip.com/qk/85265X/201710/72747566504849554948484955.html.

54. Mullen C, Spence D, Moxey L, Jamieson A. Perception problems of the verbal scale. *Science and Justice*. 2014;54(2):154-158. https://www.clinicalkey.es/playcontent/1-s2.0-S1355030613001160. doi: 10.1016/j.scijus.2013.10.004.

55. Arscott E, Morgan RM, Meakin GE, French JC. Understanding forensic expert evaluative evidence: A study of the perception of verbal expressions of the strength of evidence. . 2017. https://www.openaire.eu/search/publication?articleId=od_____322::c2dfaf6b43cf2b8683d9ea920fcf9bcf.

56. Sjerps M, Biesheuvel DB. The interpretation of conventional and bayesian verbal scales for expressing expert opinion: A small experiment among jurists. *Forensic Linguistics*. 1999;6(2):214-227. doi: 10.1558/sll.1999.6.2.214.

57. Kane J, Barabas J. No harm in checking: Using factual ManipulationChecks to assess attentiveness in experiments. *American Journal of Political Science*. 2019.

58. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*. 2009;45(4):867-872. https://www.sciencedirect.com/science/article/pii/S0022103109000766. doi: 10.1016/j.jesp.2009.03.009.

59. Clifford S, Jerit J. Is there a cost to convenience? an experimental comparison of data quality in laboratory and online studies. *Experimental Political Science*. 2014. doi: 10.1017/xps.2014.5.

60. Hara K, Adams A, Milland K, Savage S, Callison-Burch C, Bigham J. A data-driven analysis of workers' earnings on amazon mechanical turk. *Proceedings of the 2018 CHI Conference on human factors in computing systems*. 2018:1-14. http://dl.acm.org/citation.cfm?id=3174023. doi: 10.1145/3173574.3174023.

61. Berry DT, Wetter MW, Baer RA, Larsen L. MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*. 1992;4(3):340-345. doi: 10.1037//1040-3590.4.3.340.

62. Baer R, Ballenger J, Berry D, Wetter M. Detection of random responding on the MMPI-A. *Personality Assessment*. 1997:139-151.

63. Rogelberg SG, Fisher GG, Maynard DC, Hakel MD, Horvath M. Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods*. 2001;4(1):3-25. https://journals.sagepub.com/doi/full/10.1177/109442810141001. doi: 10.1177/109442810141001.

64. Landers RN, Behrend TS. An inconvenient truth: Arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Industrial and Organizational Psychology*. 2015;8(2):142-164. https://search.proquest.com/docview/1721577447. doi: 10.1017/iop.2015.13.

65. Mirta Galesic, Michael Bosnjak. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opin Q*. 2009;73(2):349-360. https://www.jstor.org/stable/25548084. doi: 10.1093/poq/nfp031.

66. Klein RA, Ratliff KA, Vianello M, et al. Investigating variation in replicability. A "many labs" replication project. *Social psychology*. 2014;45(3):152. http://www.fachportal-paedagogik.de/fis_bildung/suche/fis_set.html?FId=1082035. doi: 10.1027/1864-9335/a000178.

67. Chandler JJ, Paolacci G. Lie for a dime. *Social Psychological and Personality Science*. 2017;8(5):500-508. https://journals.sagepub.com/doi/full/10.1177/1948550617698203. doi: 10.1177/1948550617698203.

68. Chandler J, Mueller P, Paolacci G. Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behav Res*. 2014;46(1):112-130. https://www.ncbi.nlm.nih.gov/pubmed/23835650. doi: 10.3758/s13428-013-0365-7.

69. Dhar S, Tansek R, Toftey E, Kaye K. Observer bias in hand hygiene compliance testing. *The Society for Healthcare Epidemiology of America*. 2010;31.

70. Adair JG. The hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*. 1984;69(2):334-345. doi: 10.1037//0021-9010.69.2.334.

71. Schillewaert N, Meulemeester P. Comparing response distributions of offline and online. *International Journal of Market Research*. 2005;47(2):163-178. doi: 10.1177/147078530504700203.

72. Hauser D, Schwarz N. It's a trap! instructional manipulation checks prompt systematic thinking on "Tricky" tasks. . . doi: 10.1177/2158244015584617.

73. Hauser DJ, Ellsworth PC, Gonzalez R. Are manipulation checks necessary? *Frontiers in psychology*. 2018;9:998. https://www.ncbi.nlm.nih.gov/pubmed/29977213. doi: 10.3389/fpsyg.2018.00998.

74. Fleischer A, Mead AD, Huang J. Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*. 2015;8(2):196-202. https://search.proquest.com/docview/1721577480. doi: 10.1017/iop.2015.25.

75. Wessling KS. MTurk character misrepresentation. *Journal of consumer research*. 2017;44(1):211-230. http://www.econis.eu/PPNSET?PPN=1009498452.

76. Hauser D, Schwarz N. Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behav Res*. 2016;48(1):400-407. https://www.ncbi.nlm.nih.gov/pubmed/25761395. doi: 10.3758/s13428-015-0578-z.

77. Huff C, Tingley D. "Who are these people?" evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research and Politics*. 2015. https://journals-sagepub-com.proxy.cc.uic.edu/doi/pdf/10.1177/2053168015604648. doi: 10.1177/2053168015604648.

78. Casler K, Bickel L, Hackett E. 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*. 2013;29(6):2156-2160. https://www.sciencedirect.com/science/article/pii/S074756321300160X. doi: 10.1016/j.chb.2013.05.009.

79. Elizabeth A Necka, Stephanie Cacioppo, Greg J Norman, John T Cacioppo. Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PLoS One*. 2016;11(6):e0157732. https://www.ncbi.nlm.nih.gov/pubmed/27351378. doi: 10.1371/journal.pone.0157732.

80. Dobosh M. *The SAGE encyclopedia of communication research methods*. Thousand Oaks: SAGE Publications, Inc; 2018.

81. Krosnick JA. *Survey research*. London [u.a.]: Sage; 1999.

82. Porter SR, Whitcomb ME, Weitzer WH. Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*. 2004;2004(121):63-73. doi: 10.1002/ir.101.

83. Fulton BR. Organizations and survey research. *Sociological Methods & Research*. 2018;47(2):240-276. https://journals.sagepub.com/doi/full/10.1177/0049124115626169. doi: 10.1177/0049124115626169.

84. Schuller RA, Terry D, McKimmie B. The impact of expert testimony on jurors' decisions: Gender of the expert and testimony Complexity1. *Journal of Applied Social Psychology*. 2005;35(6):1266-1280. doi: 10.1111/j.1559-1816.2005.tb02170.x.

85. Holyk GG. *The SAGE encyclopedia of survey research methods*. Thousand Oaks: SAGE Publications, Inc; 2011.

86. Buskirk TD, Sanders T, Michaud J. Are sliders too slick for surveys? an experiment comparing slider and radio button scales for smartphone, tablet and computer based surveys. *Methods, Data, Analyses*. 2015;9(2):229-261. https://search.datacite.org/works/10.12758/mda.2015.013. doi: 10.12758/mda.2015.013.

87. Cape P. Slider scales in online surveys. . 2009.

88. Funke F. A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*. 2016;34(2):244-254. https://journals.sagepub.com/doi/full/10.1177/0894439315575477. doi: 10.1177/0894439315575477.

89. Funke F, Reips U, Thomas RK. Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*. 2011;29(2):221-231. https://journals.sagepub.com/doi/full/10.1177/0894439310376896. doi: 10.1177/0894439310376896.

90. Blascovich J, Spencer SJ, Quinn D, Steele C. African americans and high blood pressure: The role of stereotype threat. *Psychological Science*. 2001;12(3):225-229.

91. Marchant G. *Demographics*. ; 2018:486-488.

92. n this editorial, we will encourage authors to information is to accurately describe their sample. rethink, update the demographic questions It is important to accurately describe a sample for they use in their research surveys. We argue the following reasons. First, et al. Jennifer L. hughes,* abigail A. camden, and tenzin yangchen agnes scott college. .

93. Durand C. Surveys and society. *SAGE handbook of survey methodology*. 2016:57-66. https://search.credoreference.com/content/entry/sageukxuzo/surveys_and_society/0.

94. State of illinois circuit court of cook county. (n.d.). 13.4 pre-trial phase. http://www.cookcountycourt.org/FORATTORNEYSLITIGANTS/RulesoftheCourt/ReadLocalRule/tabid/1139/ArticleId/2277/-13-4-Pre-Trial-Phase.aspx.