EVALUATING THE INFLUENCE OF DAUBERT'S CROSS-EXAMINATION SAFEGUARD ON ATTORNEYS' AND JURORS' JUDGMENTS ABOUT SCIENTIFIC EVIDENCE

by

JACQUELINE L AUSTIN

A dissertation submitted to the Graduate Faculty in Psychology in
partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted by the
Graduate Faculty in Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy

Margaret Bull Kovera

Date	Chair of Examining Committee

Maureen O'Connor

Date	Executive Officer

Brian Cutler

Deryn Strange

Steven Penrod

Bradley McAuliff

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

EVALUATING THE INFLUENCE OF DAUBERT'S CROSS-EXAMINATION SAFEGUARD ON ATTORNEYS' AND JURORS' JUDGMENTS ABOUT SCIENTIFIC EVIDENCE

by

JACQUELINE L AUSTIN

Adviser: Margaret Bull Kovera

The Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc* clarified that federal trial judges were to serve as evidentiary gatekeepers for scientific evidence, evaluating scientific reliability when determining admissibility. When judges fail at gatekeeping and admit unreliable expert testimony, the Court expresses faith in the ability of cross-examination to reveal the reliability of testimony for jurors. For cross-examination to function as the Court intends, attorneys must recognize scientific flaws and craft cross-examination questions that expose scientific threats. Moreover, these scientifically informed cross-examinations must act as a form of scientific training for jurors. I conducted two studies to empirically examine the Court's assumptions regarding cross-examination. In Study One, 95 attorneys read a trial summary that contained expert testimony regarding an intelligence test. I varied the validity (presence v. absence of experimenter bias threat) and reliability (moderate v. high reliability indices on test-retest, inter-observer, and internal consistency scores) of the intelligence test. Attorneys provided lower ratings of scientific quality when the test was unreliable but did not craft cross-examination questions designed to expose the low reliability indices of the scientific test. Attorneys did not provide lower ratings of scientific quality when the intelligence test was invalid; however, a proportion of attorneys did craft cross-

examination questions to expose the validity threat. In Study Two, I again varied the reliability and validity of the intelligence test and whether the cross-examination educated jurors about the study's flaws (scientifically informed vs. naïve). Either a judge or an attorney conducted the scientifically informed cross-examinations. Scientifically informed cross-examinations did not assist jurors with evaluating scientific reliability or validity. These studies suggest that cross-examinations may not function as a safeguard against flawed scientific evidence. Although some attorneys may be able to meet the Court's expectations, cross-examination may be an ineffective method of providing methodological training for jurors.

Acknowledgements

This research was generously supported by grant from the National Science Foundation (SBE # 115525). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

I express extreme gratitude to my committee members, Margaret Bull Kovera, Steven Penrod, Deryn Strange, Brian Cutler, and Bradley McAuliff, for their involvement and commitment to this project. Your feedback during my defense shaped what is presented below. I am quite flattered to have such esteemed researchers agree to work on this project with me. I also cannot thank you all enough for your understanding when time restricted my lofty intentions to include judges as well. Your "pragmatic" approach (to borrow a word from Brian) saved Margaret and Deryn from a dealing with an over panic-stricken graduate student so for that I send their thanks on my behalf. I also want to thank Larry Wrightsman for introducing me to this field and his mentorship at KU.

I owe a special thanks to my advisers during graduate school, Margaret Kovera and Deryn Strange. To my adviser, Margaret, I want first thank you for taking a chance on me by accepting me as your student. If not for you, I might have pursued life plan B (bartending in Mexico) at the age of 22 and irreparably damaged my relationship with my family. In all seriousness though, thank you for the many professional opportunities you provided to me over the course of my graduate career. You possess the skills of an excellent researcher and I am so grateful that you took the time to impart your knowledge with me. I would also like to say thank you for taking the time to improve my writing. As a professor myself, I recognize and appreciate the time you spent helping me to improve my skills. Finally, I would like to say thank you for being that home away from home when I needed a warm meal, a hug, or some time to play video games with a substitute younger sibling. You've got an amazing family and I appreciate that you all welcomed me as a part of it.

To my adviser, Deryn, I want to first thank you for conducting research and writing grants with me that did not automatically align with your established line of research. You truly are a selfless adviser. I required you to learn new literatures that intersected with my interests and I started this process before you even arrived – could I have been a more demanding student? I'm also so grateful for your encouragement and direction as I struggled to transition between social and cognitive theories. Your time will hopefully pay off soon, but regardless of our funding status I'm really proud of that line of research that we created. I think it's innovative and integrative in a field that is primarily dominated by social theories. Needless to say, I'm glad we're in the field together and I hope we can continue our professional collaborations. But beyond the academic life, I also want to simply thank you for your friendship.

I am also incredibly grateful to my parents for their endless supply of encouragement. To my dad, you opened up so many doors I never even knew existed before you came into my life. Thank you for loving me like I was your own daughter and thank you for never letting me settle on anything less than you believed I could achieve. To my mother, words cannot express how you've shaped me and contributed to my success. I do want to thank you for listening to me when I was [insert adjective here] (i.e., always). You're my rock and I love you.

Finally, I must acknowledge the role my friends played because it really took a village. Alexis, thank you for providing me with the survival skills to live in NYC and showing me all NYC had to offer. You made my experience unforgettable. More broadly, to the class of 2030, I'm thankful I was a part of the best cohort. To Sarah Greathouse, Caroline Crocker, and Julia Kennard, thank you for your mentorship and friendship. Although life took us down separate career paths, I'm ecstatic to share in all of your amazing life adventures. To Karima – I know we only had a year together in lab, but I'm so excited you've joined this field and so happy we've become friends. To Lindsey, cheers to our successes – may we have many more! To Chris, thanks for being a reprieve from school. I like that you can make my face hurt from smiling too much.

I also appreciate the hard work of Megan Gilmartin, Navila Abbas, and Nikoleta Despodova for their assistance on this project. You guys helped make the impossible happen.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: THE USE OF EXPERTS IN LEGAL PROCEEDINGS

Since the 14^{th} century, courts have sought the knowledge of experts to assist in rendering appropriate legal judgments (Wigmore, 1978). In court, psychologists may provide social framework testimony in which they may describe previously collected social science research to help triers of fact (i.e., judges and jurors) evaluate an issue within a case (Monahan & Walker, 1988). For example, the courts may seek a social or cognitive psychologist to educate the judge or jury about factors known to influence eyewitness memory. This social framework expert may review the scientific literature documenting the reality of wrongful eyewitness identifications and then elaborate on factors known to influence a witnesses' accuracy such as issues of procedural suggestiveness (e.g., lineup composition) and issues related to witness reliability (e.g., the weak relationship of witness confidence and accuracy). Other examples of social framework evidence include expert testimony on the effects of the situational factors and police behaviors that may increase a suspect's propensity to confess (Blandón-Gitlin, Sperry, & Leo, 2011), the factors that contribute to a hostile work environment (Kovera, McAuliff, & Hebert, 1999) or the effects of age and interview format on a child's vulnerability to suggestion (McAuliff & Duckworth, 2010).

Alternatively, psychologists may be called to provide social fact evidence. That is, they may testify about research that was conducted purposely to address a specific issue within the trial (Monahan & Walker, 1988). For example, an expert may opine about whether pretrial publicity has caused bias within a venue. To make the argument for or against bias, the expert would not only discuss the relevant psychological literature about the effects of pretrial publicity on juror decision making, but would also opine about bias observed within a particular venue and provide recommendations to remedy the bias (Kovera & Cutler, 2013). An eyewitness expert offering social fact evidence may conduct an experiment to directly test whether a lineup was biased in favor of choosing a particular suspect by running an empirical study and presenting the results of that study (Faigman & Monahan, 2005). A clinical psychologist may offer an assessment of a defendant's competency, mental status, criminal responsibility, future risk, or intellectual functioning based on research and evaluations of the defendant (Melton, Petrila, Poythress, & Slobgin, 2007).

Research shows that expert testimony can influence jurors' verdicts and evaluations of trial evidence. In one study, jurors who heard expert testimony about factors that can diminish an eyewitness's accuracy (e.g., the presence of high stress) rated the eyewitness to be less accurate and reliable and discussed the eyewitness's identification longer during deliberation than did those juries who did not hear expert testimony (Hosch, Beck, & McIntyre, 1980). In another study, expert testimony that addressed the weak to moderate correlation between confidence and accuracy in eyewitness identifications reduced jurors' reliance on confidence as a predictor of accuracy (Fox & Walters, 1986). In general, the presentation of an eyewitness expert may sensitize jurors to witnessing factors that may influence an eyewitnesses' identification (Cutler, Dexter, & Penrod, 1989; Cutler, Penrod, & Dexter, 1989). Furthermore, jury and juror decisions are influenced by other forms of psychological expert testimony as well such as testimony on rape trauma syndrome (Brekke & Borgida, 1988), child sexual abuse (Kovera, Levy, Borgida, & Penrod, 1994), sexual harassment (Kovera et al., 1999), and battered woman syndrome (Schuller, 1992; Schuller, McKimmie, & Janz, 2004). Thus, social framework testimony influences jurors' decision making.

Although influential on jurors' decisions, the use of social framework evidence in courts is a controversial issue (Melton et al., 2007). Some scholars question whether the findings from the types of laboratory simulations that social scientists create can generalize beyond the lab to reflect phenomena in the real world. For example, do eyewitness tasks in the laboratory mimic the conditions and emotions that are faced by real eyewitnesses (Flowe, Finklea, & Ebbesen, 2009)? Would the factors that induce false confessions in laboratory research also produce real world confessions considering the consequences of confessions made in the lab are considerably less severe? Other psychologists caution against the use of clinical predictions in a legal setting due to the highly unreliable nature of fitting an individual's behavior to group tendencies (Cooke & Michie, 2010). Faust and Ziskin (1988) argued in *Science* that clinical opinions are not reliable or valid enough to warrant use in legal proceedings. Other critics suggest that courts use extra care in evaluations of psychological testimony because of the inherently unreliable nature of the field (Holly, 1999).

Irrespective of concerns about the reliability of expert evidence, courts have long recognized the need for experts. Jurors often require the assistance of experts to evaluate other testimony presented at trial, but not all expert evidence will be helpful or admissible. Expert testimony that improves the juror decision-making process should be admitted, as long the expert's testimony is reliable. However, questions about the reliability of science (especially psychological science) and thus the admissibility of scientific evidence, consistently confront the courts (Melton et al., 2007).

CHAPTER 2: LEGAL BACKGROUND ON THE ADMISSIBILITY OF EXPERT TESTIMONY

Beginning in 1923, federal courts evaluated the admissibility of scientific testimony using the *Frye* rule: scientific evidence should be admitted if it is generally accepted within the relevant scientific community (*Frye v. United States*, 1923). This rule originated from a case in which James Frye was on trial for murder and his attorney wanted to present the results of a systolic blood pressure test, a predecessor to the polygraph, in support of his innocence. The trial court did not permit this testimony. Frye appealed and argued the expert's testimony was improperly excluded. The DC Court of Appeals, however, agreed with the trial court and stated that because systolic blood pressure was not generally recognized as a method of lie detection within the scientific community, the trial court had met the appropriate grounds for exclusion. *Frye* established the general acceptance rule of evaluating expert testimony in federal courts and placed the responsibility of evaluating the admissibility of expert testimony squarely on the shoulders of the relevant scientific community.

In the late 1960's the Chief Justice of the Supreme Court and Congress worked together to create a series of new rules designed help judges determine whether to admit evidence at trial, referred to as the Federal Rules of Evidence (FRE). The FRE included new rules governing the admissibility of expert testimony. These rules specify that expert evidence must be relevant to an issue in the trial (FRE 403), must not be within the common understanding of the juror, and must assist jurors with understanding a disputed issue in the trial (FRE 702). Further, in the Advisory Committee notes, the committee members specified that: "The rule is broadly phrased. The fields of knowledge which may be drawn upon are not merely to the “scientific” or “technical” but extend to all “specialized” knowledge. Similarly, the expert is viewed, not in a narrow sense, but as a person qualified by “knowledge, skill, experience, training or education.” Thus, within the scope of the rule are not only experts in the strictest sense of the word, e.g., physicians, physicists, and architects, but also the large group sometimes called “skilled” witnesses, such as bankers or landowners testifying to land values” (Federal Rules of Evidence Handbook, 2000, p. 104). In essence, the FRE called on judges to evaluate all forms of expert testimony – whether it is specialized, technical, or scientific—when determining admissibility. One shortcoming of the FRE was that the drafters did not address the *Frye* rule, which led to some confusion among judges in federal courts about whether the *Frye* standard or the FRE should govern the admissibility of expert testimony (Sanders, Diamond, Vidmar, 2002). This confusion was addressed in the 1993 *Daubert v. Merrell Dow Pharmaceuticals, Inc* decision in which the U.S. Supreme Court clarified the standards for the admissibility of scientific evidence for federal courts by deeming the FRE the appropriate test judges should use when determining whether to admit scientific evidence or not.

In the *Daubert* case, Jason Daubert, Eric Schuler and their parents sued Merrell Dow Pharmaceuticals for marketing a drug, Bendectin, given to mothers for morning sickness. Daubert and Schuler suffered from serious birth defects and proffered testimony from eight experts to demonstrate the link between their birth defects and Bendectin; they presented findings from *in vitro* and *in vivo* animal studies, pharmacological studies on the chemical structure of Bendectin showing that Bendectin had a chemical structure that mimicked other drugs known to cause birth defects, and a meta-analysis composed of epidemiological studies that compared the rate of birth defects among children whose mothers had and had not ingested Bendectin during pregnancy. Merrell Dow argued for a summary judgment and submitted an affidavit of an epidemiologist to the effect that no published studies had documented a statistically significant association between Bendectin and the cited birth defects. The District Court granted Merrell Dow's motion for summary judgment on the grounds that the research methodologies cited by the plaintiffs were not generally accepted and therefore violated the *Frye* standard. The 9th Circuit affirmed the District Court's ruling, again on grounds that the research proffered by the plaintiff experts was not generally accepted.

Appealing to the U.S. Supreme Court, the plaintiffs argued that the lower courts improperly applied the *Frye* rule when determining admissibility and stated that the Federal Rules of Evidence should govern the admissibility of expert testimony. The U.S. Supreme Court granted certiorari, overturned *Frye's* general acceptance test as the singular rule for granting expert testimony in federal courts, and stated that the more recently adopted Federal Rules of Evidence superseded *Frye*. In general, the Court found *Frye's* general acceptance test to be too inflexible and incompatible with the Federal Rules of Evidence, which were designed to encourage the admission of testimony that could aid jurors' decisions (Sanders, et al., 2002).

Under *Daubert*, scientific evidence must meet two criteria must for admission into court: relevance and reliability. Specifically, expert evidence must be relevant to an issue in the trial to be admissible (Federal Rule of Evidence 403). Further, the evidence must not be within the common understanding of the juror and the evidence must assist jurors with understanding a disputed issue in the trial (Federal Rule of Evidence 702). With regard to the reliability criterion, the Court provided a list of four non-exhaustive factors that Courts could use in their determination of reliability: whether the data can be falsified, whether the data have been subjected to peer review or publication, whether the known or potential error rates are available, and general acceptance of the findings in the relevant scientific community (*Daubert v. Merrell Dow Pharmaceuticals,* 1993). Federal Rule of Evidence 702 was amended in 2000 to mirror the reliability factors in *Daubert* by stating that expert testimony should be admitted "provided that (1) the testimony is sufficiently based upon reliable facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case" (Federal Rule of Evidence 702, 2000).

Daubert placed the responsibility of determining whether scientific evidence is reliable and thus admissible squarely on the shoulders of the trial judge. Indeed, the *Daubert* decision urged federal judges to evaluate expert testimony by proposing factors for judges to consider in their assessments of scientific reliability rather than by simply allowing judges to evaluate the expert's or scientific field's conclusions. Essentially, this decision conveyed that judges are to be evidentiary "gatekeepers" for scientific testimony. *General Electric Co v. Joiner* (1997) reiterated judges' gatekeeping role and stressed that judges must determine admissibility by evaluating reliability. Furthermore, in his concurring opinion, Justice Breyer encouraged judges to seek help from scientists when fulfilling their "gatekeeping" role by appointing reputable experts to assist with the evaluation of scientific evidence.

The *Daubert* decision only addressed the admission of scientific expert testimony, but was mute on the requirements for non-scientific testimony. Therefore, judges could avoid the responsibility of evaluating evidentiary reliability by categorizing an expert's testimony as "specialized" rather than "scientific." Debates ensued in lower courts regarding whether certain types of non-scientific expert testimony were subject to the admissibility requirements set forth in *Daubert*. For example, a panel of Fifth Circuit judges opined that the *Daubert* factors did not pertain to clinical medical testimony; however, when the full Fifth Circuit reviewed the case *en banc*, they reversed the panel's decision and stated that *Daubert* factors may pertain to clinical medical testimony (*Moore v. Ashland Chemical, Inc*., 1997; 1998). The Supreme Court resolved these underlying disagreements about the application of *Daubert* in *Kuhmo Tire Co. v. Carmichael* (1999).

The *Kumho* case evolved from a serious car crash caused by a tire blowout. The plaintiffs maintained that the tire blowout was either caused by a defect in the tire's design or was a manufacturing defect and proposed expert testimony to support their claim. The trial court excluded the testimony on grounds that the testimony did not meet the standards established in the FRE or the reliability factors specified in *Daubert*. The Supreme Court disagreed with the trial court and instructed judges that the factors listed in the *Daubert* decision do not apply to the evaluation of nonscientific testimony (*Kumho Tire Co. v. Carmichael*, 1999). Although ruling that the factors did not apply, the Court stressed judges' obligation to evaluate specialized and technical knowledge. In determinations of reliability for specialized and technical knowledge, the Court emphasized that the four factors listed in Daubert are not exhaustive and gave trial judges a wide range of discretion in their determination of reliability Emulating the federal courts, many state courts in the U.S. have adopted admissibility standards based on *Daubert* and its progeny.

CHAPTER 3: EVALUATING SCIENTIFIC QUALITY AND APPLYING *DAUBERT*

Judges' Ability to Evaluate Scientific Quality and Apply *Daubert*

Given the *Daubert* decision, judges must evaluate whether expert testimony should be admitted. Judges may make a good faith effort to distinguish between "good" and "bad" expert testimony, but how successful are judges at evaluating scientific evidence? Judges may be required to evaluate many types of scientific evidence (e.g., medical, economic, psychological, mechanical). Inevitably, some topics may be easier for them to evaluate than others. Further, the question remains as to whether judges are able to apply the *Daubert* factors in their evaluations of scientific evidence in the absence of scientific training?

Judges may lack the skills necessary to detect flaws in scientific research (Gatowski et al., 2001; Kovera & McAuliff, 2000; Manuto & O'Rourke, 1991) because law school curricula are not designed to train future lawyers and judges to evaluate scientific concepts (Lehman, Lempert, & Nisbett, 1988). Indeed, first year law students have a low level of understanding of statistics and research methodology. Further, unlike psychology graduate students, third year law students do not show improvement over the course of their graduate studies in their ability to apply statistical and methodological rules to everyday events or scientific studies (Lehman et al., 1988).

Surveys of judges suggest that judges have little to no knowledge of social science methods, making it unlikely that they are capable of evaluating empirical methods (Manuto & O'Rourke, 1991). A national survey of state trial court judges found that although 91% of the state judges believed that their role of gatekeeper was appropriate, only 52% stated that they felt their education had sufficiently prepared them to perform these evaluations (Gatowski, et al., 2001). Although the judges felt that the four *Daubert* factors would guide their decision to admit or deny expert testimony, they lacked the ability to operationalize and understand concepts central to *Daubert's* criteria for evaluating reliability such as error rate and falsifiability. Only 6% of judges understood the concept of falsifiability and only 4% understood what is meant by an error rate. Judges were more likely to understand peer review (71%) and general acceptance (81%).

In one study designed to answer the question of whether judges have the ability to make competent judgments about the reliability and validity of expert evidence, Florida state judges read a summary of an expert's testimony in a hostile work environment case (Kovera & McAuliff, 2000). The judges read one of eight versions of the expert's proposed testimony that varied in its internal validity (valid, missing control group, confound, or experimenter bias) and peer reviewed status (published in a peer review journal or not). Following the case facts, judges ruled on whether they would admit the testimony in court, provided justifications for their decision, and evaluated the quality of the expert testimony. Overall, the admission rate of the science did not vary as a function of study validity. When rejecting the testimony, judges rarely mentioned the existing internal validity threat in their justification for rejecting the testimony; only 8% of judges identified the missing control group, 13% mentioned the confound, and 9% cited experimenter bias as a reason for excluding the testimony. Thus, judges were not generally sensitive to variations in scientific validity such as the presence of a control group or the presence of experimenter bias and were not sensitive to factors such as peer review; however, judges with some prior scientific training demonstrated more sensitivity to factors affecting validity. Specifically, judges with some prior training rated valid research as more legally admissible than invalid science. In contrast, judges without scientific training rated research plagued by a confound as more legally admissible than did scientifically trained judges (Kovera & McAuliff, 2000). Overall, studies examining judges' ability to apply the *Daubert* factors suggest that judges may lack the skills necessary to differentiate between valid and flawed scientific testimony without training and therefore may be ineffective gatekeepers resulting in the admission of unreliable expert evidence.

***Daubert's* Effect on the Admission of Expert Testimony**

How did *Daubert* affect admissibility decisions? Thirty-eight percent of state trial court judges from *Daubert* states who participated in the 1998 survey reported that the *Daubert* decision had not changed their role in admitting expert evidence in contrast to only 10% who were unsure whether their role had changed since *Daubert* (Gatowski et al., 2001). Further, there was little consensus among judges who did offer an opinion about *Daubert's* intent for admissibility decisions: approximately 32% opined that *Daubert* was intended to raise the threshold of admissibility, 23% opined that *Daubert* was intended to lower the threshold of admissibility, and 36% opined that *Daubert* was not intended to raise or lower the threshold of admissibility decisions but was intended to give judges a set of criteria with which to evaluate evidence; the remaining judges stated that they were unsure of *Daubert's* intent (Gatowski, et al., 2001).

Thus, judges lack a shared understanding of the purpose of *Daubert,* but did *Daubert* affect admissibility decisions? Are judges' decisions to admit expert evidence more lenient or more restrictive since *Daubert*? The Federal Judicial Center conducted a survey of 303 federal court district judges and compared self-report data from two surveys: one collected pre-*Daubert* (collected in 1991) and one collected post-*Daubert* (collected in 1998) to examine the effects of *Daubert* on admissibility decisions. Post-*Daubert,* judges reported that they excluded more evidence than before *Daubert* (Krafka, Dunn, Johnson, Cecil, & Miletich, 2002). According to the survey from 1991, 75% of trial judges admitted expert testimony in their last civil trial. In 1998, this rate dropped to 59%. In addition, judges self-reported that they were more critical of expert testimony. When providing their reasons for excluding evidence, judges only cited validity issues about 18% of the time. For example, in fewer than 2% of cases did judges cite validity issues regarding the falsifiability of data or whether the known or potential error rates were unknown or large. Judges primarily gave legal reasons for excluding the testimony (e.g., the testimony was irrelevant (47%), the expert was not qualified (42%), or the testimony would not aid the trier of fact in their determination of liability (40%)). Judges rarely discussed *Daubert's* specific reliability factors in their decisions to exclude, which may suggest that judges do not apply these factors in their determinations of admissibility (Krafka et al., 2002).

Although self-report data suggest that judges exclude more evidence now than before *Daubert (*Krafka et al., 2002), a content analysis of 693 criminal case decisions sampled from both federal and state appellate court between the years of 1988 and 1998 revealed that judges admitted expert testimony pre and post-*Daubert* at comparable rates (Groscup, Penrod, Studebaker, Huss, & O'Neil, 2002). Post-*Daubert,* judges spent more time discussing *Daubert* and FRE 702 and spent less time discussing *Frye's* general acceptance test in court opinions than discussions pre-*Daubert*; however, judges rarely addressed the four specific *Daubert* factors in their opinions post-*Daubert*, which again suggests that judges do not understand the *Daubert* factors or may be unable to apply the factors when making admissibility decisions (Groscup et al., 2002).

An analysis of federal district court opinions in civil trials between 1980 and 1999 suggests that *Daubert* has indeed raised the admissibility threshold for expert testimony (Dixon & Gill, 2002). However, tighter standards do not mean that only bad science is excluded from the courtroom (Dixon & Gill, 2002) as judges are disinclined to admit psychological expert testimony regardless of scientific validity (Kovera & McAuliff, 2000). Thus, it is likely that judges sometimes exclude valid science (Dixon & Gill, 2002; McAuliff, 2009). Of course, the exclusion of valid science is a concern because jurors may not receive the necessary information to render an appropriate decision, but the admission of invalid science is also a concern because invalid science may influence juror decision making.

Attorney and Juror Ability to Evaluate Scientific Quality

Once expert testimony is admitted, the duty of evaluating scientific quality becomes the responsibility of attorneys and jurors. If attorneys raise questions about the scientific validity of an expert's research, they will do so in a motion to exclude the evidence or during the cross-examination of the expert. However, for attorneys to raise questions about threats to validity, they must be sufficiently knowledgeable about scientific methods to detect violations. Do attorneys successfully recognize flawed science? To answer this question, researchers recruited a sample of attorneys from across the United States specializing in employment law to read a summary of an expert's testimony in a hostile work environment case (Kovera & McAuliff, 2002). The attorneys read one of eight versions of the expert's proposed testimony that varied in its internal validity (valid, missing control group, confound, or experimenter bias) and peer reviewed status (published in a peer review journal or not). The researchers asked the attorneys to rate the scientific quality of the expert's study. Validity did not impact attorneys' ratings of scientific validity and reliability; however, general acceptance of the proposed testimony did affect attorneys' ratings of scientific evidence. Indeed, attorneys provided higher ratings of scientific quality when the testimony was generally accepted within the scientific field than when it was not, irrespective of the study's validity. In addition, the researchers also asked the attorneys whether they would file a motion to exclude the testimony if they were playing the opposing attorney, 95% said 'yes'. This decision was not affected by scientific validity; rather most attorneys stated that they would routinely file to exclude expert testimony that did not favor their case as part of their typical trial strategy. Of the attorneys who stated they would object to the testimony, merely 5% objected to the admission of the scientific study because of the study's validity. Further, only half of the 5% who objected on issues of validity raised the appropriate validity flaw during their objection. Thus, attorneys may not be able to assist judges with evaluations of scientific quality (Kovera & McAuliff, 2002). Presumably judges and attorneys possess similar abilities in evaluating scientific quality given that neither attorneys nor judges receive scientific training, which suggests that jurors may be presented with unreliable testimony and forced to determine its value toward influencing perceptions of evidence strength and verdict.

The Supreme Court expressed some concern with jurors' use of expert testimony in that "scientific proof may in some instances assume a posture of mystic infallibility in the eyes of a jury" (*United States v. Addison*, 1974, p. 744). Can jurors critically evaluate scientific evidence and assign appropriate weight to that testimony when rendering their verdict? Most laypeople, like judges and attorneys, are not trained in scientific reasoning and struggle with the concepts necessary to evaluate scientific quality. For example, when attempting to judge the probability of an outcome, people are not sensitive to base-rate information (Kahneman & Tversky, 1973). People, without proper instruction, do not recognize that results from small samples are more likely to be erroneous than results from large samples (Fong, Krantz, & Nisbett, 1986; Tversky & Kahneman, 1974). Further, they do not consider issues of sample representation when determining whether to generalize from a sample to the population (Hamill, Wilson, & Nisbett, 1980).

In the legal setting, research shows that mock jurors struggle when tasked with scientific evaluations. Indeed, mock jurors are prone to ignoring quantitative descriptions as compared to qualitative ones (Saks & Kidd, 1980). Although mock jurors may be sensitive to some internal validity threats such as missing control groups (McAuliff & Duckworth, 2010; McAuliff & Kovera, 2008; McAuliff, Kovera, & Nunez, 2009), they are unable to recognize more sophisticated threats to internal validity such as confounds, or experimenter bias – threats commonly present in research (McAuliff & Duckworth, 2010; McAuliff, et al., 2009). Overall, laypeople generally and mock jurors more specifically are unable to recognize most flaws in scientific research, making them susceptible in the courtroom to the influence of invalid science. Additionally, mock jurors are less critical of scientific evidence when it is presented in the context of a trial than when they review it outside of a trial context, which suggests that jurors assume evidence admitted at trial is valid (Schweitzer & Saks, 2009). Perhaps jurors are aware that judges review and admit all scientific testimony before it reaches the jury. If jurors are aware of judges' "gatekeeping role," then jurors may believe that any testimony that has been deemed admissible is valid and thus jurors may assign persuasive value to the testimony. Given that expert testimony can influence the decision-making process of jurors, it is disconcerting that judges may admit unreliable science into evidence and jurors may also fail to identify its flaws.

CHAPTER 4: LEGAL SAFEGUARDS AGAINST UNRELIABLE EXPERT TESTIMONY

The Court acknowledged that judges occasionally may evaluate invalid science favorably and admit it into evidence but also highlighted three procedural safeguards already in place in courtrooms to remedy the problem and protect jurors from the untoward influence of invalid science; "vigorous cross-examination, presentation of contrary evidence, and careful instruction on burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence" (*Daubert v. Merrell Dow Pharmaceuticals*, 1993, p. 576). Do cross-examination, use of opposing experts, and the presentation of judicial instructions help jurors distinguish between valid and flawed scientific evidence?

Instruction on the Standard of Proof

Although there are no published empirical tests of whether judicial instructions will assist jurors with the evaluation of scientific testimony, jurors do not comprehend judicial instructions (Reifman, Gusick, & Ellsworth, 1992). Judges often create judicial instructions by piecing together sentences from appellate court decisions (Elwork, Sales, & Alfini, 1977). This practice ensures that the instructions are legally accurate, but fails to consider laypersons' abilities to comprehend their meaning. Indeed, existing judicial instructions do not improve jurors' comprehension of the law (Elwork et al., 1977). A survey of citizens called for actual jury duty found that jurors did not understand half of the instructions that the judge provided at trial (Reifman et al. 1992). In particular, jurors struggled with the concept of burden of proof. Less than one third of the jurors serving on criminal cases in this sample knew which side was responsible for proving the case beyond a reasonable doubt (Reifman et al., 1992).

Jurors' lack of comprehension affects the efficacy of judicial instructions. For example, jurors do not follow judicial instructions to disregard inadmissible pretrial publicity (Kramer, Kerr, & Caroll, 1990) nor do they follow instructions to disregard inadmissible evidence (Sue, Smith, Caldwell, 1973; Thompson, Fong, & Rosenhan, 1981). Further, jurors often fail to follow judicial instructions limiting the use of a defendant's prior record (Tanford & Cox, 1988; Wissler & Saks, 1985). Testimony on a defendant's prior record is permitted in court to help jurors assess a witnesses' credibility but should not affect perceptions of guilt, blame, or perceptions of the defendant's harmful propensity. However, prior record evidence affects juror perceptions of these factors even in the presence of limiting instructions (Tanford & Cox, 1988; Wissler & Saks, 1985). Thus, many judicial instructions do not function as the Court intends.

Beyond jurors' lack of understanding and application of law is a more intractable problem: *Daubert* articulated that "careful instruction on the burden of proof" should aid juror understanding, but instructions on the burden of proof are not designed to help educate jurors about issues of reliability (Kovera, McAuliff, & Russano, 2002). Judicial instructions on the burden of proof may affect the decision criterion that jurors use (e.g., influence how confident they must be in the defendant's guilt to find him or her guilty), but they are not likely to assist jurors with the task of evaluating scientific reliability and validity.

Potentially, instructions that are designed to educate jurors about scientific reliability may assist jurors tasked with evaluating the quality of an expert's testimony;

however, no published research addresses whether extended judicial instructions on scientific validity can educate jurors. Other research has examined whether educational judicial instructions can educate jurors on topics normally reserved for expert testimony (e.g., Cutler, Dexter, & Penrod, 1990). For example, the *Telfaire* instruction is designed to educate jurors about eyewitness reliability. Although mock jurors rated the instruction as useful, trial judgments did not differ between participants who did or did not receive the instruction (Cutler et al., 1990; Greene, 1988; Ramirez, Zemba, & Geiselman, 1996).

Cutler et al. (1990) presented participants with either the *Telfaire* instruction, expert testimony, or no additional instructions and examined whether experts or judicial instructions improved juror sensitivity to issues of eyewitness reliability. The *Telfaire* instructions did not affect mock jurors' verdicts, their sensitivity to the factors affecting eyewitness reliability, or their skepticism regarding eyewitness identifications. Alternatively, expert testimony reduced mock jurors' guilty verdicts, improved juror sensitivity, and increased juror skepticism regarding the eyewitness's identification (Cutler, et al. 1990).

In the first of two studies from another laboratory, jurors heard evidence that a witness had good or bad visibility while observing a crime. They also heard the *Telfaire* instruction either before the evidence, after the evidence, or both before and after the evidence (Ramirez et al., 1996). Instructions presented before the evidence sensitized jurors to variations in witnessing conditions. In contrast, when the instruction was presented after the evidence, jurors rendered fewer not guilty verdicts regardless of the witnessing conditions suggesting that jurors' skepticism toward the eyewitness increased

(Ramirez et al., 1996). As written, *Telfaire* instructions do not operate as the Court intended.

In the second study, some mock jurors heard modified judicial instructions that mimicked the information that would be presented during expert testimony. By examining the modified instructions, the researchers could examine whether it was the *Telfaire* instructions specifically or judicial instructions more generally that were ineffective at sensitizing jurors to variations in the quality of witnessing conditions. Thus, mock jurors either heard the *Telfaire* instruction, modified judicial instructions, or no instructions. Orthogonal to the manipulation of instruction type, jurors heard testimony from an eyewitness who viewed the crime under conditions that varied in quality to examine whether jurors became sensitive to factors known to influence the accuracy of eyewitness identifications. Again, the *Telfaire* instruction increased skepticism, but not sensitivity to factors relating to eyewitness reliability (Ramirez et al., 1996). The modified jury instructions had no effect, although it is certainly possible that further modifications to the instructions might yield instructions that increase juror sensitivity to factors affecting eyewitness reliability (Ramirez et al., 1996).

More research is needed on modified judicial instructions to determine whether judicial instructions can be an effective method of educating jurors about variations in scientific validity before conclusions can be drawn, but for now it seems that judicial instructions will not provide jurors with the necessary tools to evaluate scientific evidence unless they can be written in a manner that jurors can comprehend so they many appropriately apply the instruction. Juror comprehension improves with rewritten judicial instructions (e.g., Charrow & Charrow, 1979; Elwork et al., 1977; English &

Sales, 1997; Lieberman & Sales, 1997; Severance, Greene, Loftus, 1984; Severance & Loftus, 1982). For example, juror comprehension improved in one study by simply rewriting standard instructions to use more common and concrete words, fewer negations and less legal jargon, sentences with a reduced length, and a more logical flow of legal concepts (Elwork, Sales, & Alfini, 1982). Other researchers cite similar success with these techniques (Charrow & Charrow, 1979; Severance & Loftus, 1982; Severance et al., 1984). It is unknown whether modified judicial instructions would be a potential safeguard against jurors' reliance on unreliable expert testimony.

The Presentation of Opposing Experts

The second safeguard to help jurors recognize unreliable scientific evidence proposed in *Daubert* is the presentation of contrary evidence, often in the form of opposing experts. Several studies have tested the effects of opposing experts on the juror decision making process (Cutler & Penrod, 1995; Griffith, Libkuman, & Poole, 1998; Raitz, Greene, Goodman, & Loftus, 1990). In one study, for example, undergraduate participants read one of four versions of a civil case concerning a daughter's repressed memories of sexual abuse by her father (Griffith et al., 1998). Participants either read the testimony of no expert, an expert who testified for the plaintiff, an expert who testified for the defendant, or opposing experts. Participants rated the defendant as more likely to have abused his daughter and rated the plaintiff's case to be stronger when they heard only the plaintiff's expert or opposing experts than when they did not hear expert testimony. Similarly, participants judged the defendant's case to be weaker when the plaintiff's expert testified as compared to all other conditions. Although undergraduate participants perceived trial evidence differently depending on whether an opposing expert was present, the presence of an opposing expert did not affect the compensation participants awarded to the plaintiff.

In another study, researchers examined the effect of opposing expert testimony by presenting mock jurors with either no expert, an expert for the plaintiff, or opposing experts in an age discrimination civil suit (Raitz et al., 1990). Community member participants awarded higher damage awards in conditions with expert testimony, but damage awards did not differ between the opposing expert and plaintiff expert only conditions. In another study of a simulated age discrimination case, participants did not differentially award damages when they heard opposing experts or just plaintiff expert testimony (Greene, Downey, & Goodman-Delahunty, 1999). Opposing expert testimony affected credibility ratings: jurors rated the plaintiff's expert as more credible when the expert was unopposed by another expert (Greene et al., 1999). Again, this study demonstrates that the presence of opposing expert testimony may affect evidence evaluations but may not affect important trial judgments.

Although these studies provide information about how opposing experts affect juror judgments, they do not tell us whether opposing experts might improve jurors' sensitivity to the reliability and validity of scientific evidence. To answer this question, researchers examined whether experts increase jurors' sensitivity to issues affecting the reliability of eyewitness identifications by varying both the witnessing conditions of the eyewitness (good or bad) and whether jurors heard no expert, the testimony of one eyewitness expert, or the testimony of two eyewitness experts refute one another (Cutler & Penrod, 1995). The defense expert discussed reliability issues for eyewitnesses under different viewing conditions. The prosecution expert highlighted the limitations of the research presented by the defense expert. The use of one expert sensitized jurors to witnessing and identification conditions, but the use of opposing experts increased skepticism about the eyewitness identification and desensitized jurors to the witnessing and identification conditions (Cutler & Penrod, 1995). In another study, participants viewed a trial with evidence of an eyewitness identification, in which the presence of biased lineup administration and eyewitness expert testimony (none, defense expert only, or opposing experts) was manipulated (Devenport & Cutler, 2004). The addition of an opposing expert reduced the credibility of the original defense expert and also reduced ratings of psychology as science. Thus, opposing experts may have the unintended effect of increasing distrust in expert testimony regardless of scientific validity and may negatively impact the credibility of the field providing the testimony (Devenport & Cutler, 2004).

Cutler and Penrod (1995) and Devenport and Cutler (2004) help us understand the effects of presenting opposing experts to jurors, but they were not designed to test if opposing experts can help jurors differentiate between valid and invalid science. Only two studies varied both the presence of an opposing expert and the scientific quality of the initial expert testimony (Levett & Kovera, 2008; 2009). The manipulation of the initial expert's scientific validity is necessary to examine whether opposing experts can be effective and influential at educating jurors about the scientific methods employed. The researchers manipulated the validity of the first expert (valid, missing control group, confound) and presence and type of opposing expert testimony (none, present/not address initial expert's methods, present/address initial expert's methods) in a child sexual abuse case (Levett & Kovera, 2008). When there was an opposing expert, mock jurors rendered more verdicts consistent with the testimony of the opposing expert regardless of scientific validity. In addition, they found that the presence of opposing experts failed to help mock jurors identify two types of internal validity flaws (absence of a control group and counterbalancing of interview items) present or absent in the expert's research. Thus, there was support for a skepticism effect but not a sensitivity effect (Levett & Kovera, 2008).

In another study, jury eligible community members viewed trials that varied the validity of the plaintiff's expert testimony (valid/control group vs. invalid/no control group) and the presence and type of opposing expert testimony (none, present/did not address initial expert's methods, present/addressed initial expert's methods) in a sexual harassment case (Levett & Kovera, 2009). When there was no opposing expert testimony, the mock jurors rated the defendant to be more liable and the initial expert as more competent and qualified than when an opposing expert also testified. Consequently, this study corroborates other studies that demonstrated skepticism effects with opposing expert testimony. However, the researchers also found slight evidence for a sensitivity effect. The presence of an opposing expert who addressed scientific methodology sensitized jurors to the absence of a control group. Specifically, the mock jurors rated research that was valid (i.e., had an appropriate comparison group) to be more valid than invalid research (i.e., did not have an appropriate comparison group) and the mock jurors rated the expert who presented valid science to be more competent than the expert who presented invalid science. Although the mock jurors rated scientific validity accurately when the opposing expert addressed validity, opposing expert testimony did not affect the main decisions in the case such as the likelihood that the plaintiff faced a hostile work environment or verdict (Levett & Kovera, 2009). Overall, studies on the presentation of opposing experts suggest that opposing experts consistently increase skepticism but either fail to educate jurors about unreliable scientific evidence admitted into court or fail to influence trial decisions enough to influence key decisions. Obviously, these findings do not support the Court's faith in the presentation of contrary evidence as a safeguard.

Cross-examination

The final safeguard proposed to protect jurors from invalid science is cross-examination. The Court has strong faith in the ability of cross-examination to reveal the reliability of testimony. Wigmore (1974) dubbed cross-examination "the greatest legal engine ever invented for the discovery of truth" (p. 29). Supreme Court justices have reiterated Wigmore's statement in written opinions (*Maryland v. Craig*, 1990; *Lilly v. Virginia*, 1999; *California v. Green*, 1970).

What effect does cross-examination have on jurors' ability to detect witness accuracy? Berman, Narby, and Cutler (1995) examined how cross-examinations that point out inconsistencies by an eyewitness on peripheral or central details of a crime affect jurors' decisions. When the eyewitness was inconsistent on central details (details about the perpetrator), mock jurors rated the eyewitness as less credible and rendered significantly less guilty verdicts than when the eyewitness was inconsistent on peripheral details (details about the crime scene). Wells, Lindsay, and Ferguson (1979) examined whether leading or non-leading cross-examination questions could help undergraduate mock jurors discriminate between accurate and inaccurate eyewitness identifications. Mock jurors who heard the leading cross-examination reported lower confidence in their ability to differentiate between accurate and inaccurate identifications. However, mock jurors believed accurate witnesses more when the assistant performed a leading cross-examination as compared to a non-leading cross-examination. Participants also believed inaccurate witnesses more when the assistant performed a non-leading cross-examination than a leading cross-examination. These two studies collectively demonstrate that mock jurors use information learned during cross-examination in their evaluations of eyewitnesses, however, it does not appear that cross-examinations consistently sensitize jurors to eyewitness consistency or accuracy.

Cross-examinations may affect jurors' perceptions of expert witnesses differently than other witnesses. Undergraduate participants read a rape trial in which the cross-examination of the victim or psychological expert began with an attack on their credibility or they did not read an attacking question (Kassin, Williams, & Saunders, 1990). In response to a credibility-damaging question, the witness provided an affirmative response, a denial, or the question was withdrawn after the prosecuting attorney objected to the question. Regardless of the expert's response to the credibility-attacking question, participants rated the expert as less credible. The victim's credibility, however, was not affected by their response. Thus, attacking questions were only effective at reducing the credibility of the expert witness. From the data it is unclear why the attacking questions were only effective in reducing expert witness credibility, but it is possible that jurors felt more capable of evaluating the victim's testimony and depended on heuristic cues when evaluating the expert's testimony to ease the decision making process (Cooper, Bennett, & Sukel, 1996; Kovera et al., 1999).

As for the ability of cross-examination to affect juror perceptions of expert evidence quality, cross-examination strength (strong vs. weak) did not affect undergraduate mock jurors' perceptions of evidence quality or verdict (Kovera et al., 1994), but other research demonstrates that cross-examinations that attack the content of the expert's testimony may affect juror perceptions of evidence quality. Reliability based cross-examination that attacked the content and reliability of expert testimony on future dangerousness decreased mock jurors' assessments of dangerousness and decreased mock jurors' perceptions of the expert more than a cross-examination that just attacked the general credibility of the expert (Krauss & Sales, 2001). Thus, cross-examinations may influence jurors' perceptions of evidence quality when the content of the cross-examination is geared toward addressing methodological issues but because this study did not systematically manipulate evidence quality, we do not know if the cross-examination influenced evaluations appropriately.

In a study designed to test whether cross-examination can help educate jurors about scientific validity, undergraduate students heard opposing experts testify in a civil liability suit (Salerno & McCauley, 2009). The plaintiff expert presented research that employed high quality methodology (i.e., valid science); the defense expert presented research that lacked an appropriate control condition, suffered from an inadequate sample size, and had not been peer reviewed or published (i.e., invalid science). Mock jurors either heard the experts get cross-examined by the attorneys or they did not. The cross-examination of the plaintiff's expert focused on the expert's credibility. The cross-examination of the defense's expert focused on the scientific validity of the expert's methodology. Mock jurors who heard cross-examinations provided higher culpability ratings and rated the defense expert as less credible. The authors suggest from these data that cross-examination may have the ability to sensitize jurors to scientific validity (Salerno & McCauley, 2009); however, this study did not fully cross type of cross-examination (attacking credibility vs. attacking scientific quality) with scientific validity (valid vs. invalid) and did not measure participants' perceptions of scientific quality. Thus, it is unclear whether cross-examinations that attack scientific validity merely reduce ratings of expert credibility and ratings of culpability regardless of the validity of the study's methodology.

To examine whether cross-examination can help sensitize jurors to variations in research quality, undergraduate students playing the role of mock jurors heard either valid or invalid science from a psychological expert on sexual harassment and witnessed either a scientifically naïve or scientifically informed cross-examination (Kovera et al., 1999). A naïve cross-examination focused on the heuristics of the expert or their qualification to testify on the material. A scientifically informed cross-examination focused on the research methods that the expert employed and was designed to assist the juror in evaluation of the quality of expert evidence. Additionally, the researchers manipulated general acceptance of the research and the ecological validity of the research to identify the instances in which jurors are likely to rely on heuristics to aid them in their evaluations of the expert's testimony rather than the evidence quality. Scientifically informed cross-examinations did not assist participants with the task of evaluating scientific evidence. Instead, mock jurors relied on heuristic cues for evidence evaluation (Kovera et al., 1999).

To examine whether a more extensive scientifically informed cross-examination may be effective at educating jurors about scientific validity, undergraduate students playing the role of mock jurors heard expert testimony on the reliability of eyewitness identifications that was either valid or invalid (i.e., the evidence either contained or did not contain a control group) and heard either an extensive scientifically informed cross-examination or a non-scientific cross-examination that merely attacked the expert's credentials (Austin & Kovera, under review). Participants who heard the extensive scientifically informed cross-examination differentiated between valid and invalid science, giving higher validity ratings to the valid science, whereas participants who heard the scientifically naïve cross-examinations gave uniformly high ratings for scientific validity, regardless of the study's validity. Although these findings represent the first evidence that cross-examination can be an effective safeguard against unreliable scientific evidence, there is also evidence that a missing control group may be an easier validity threat for mock jurors to recognize than more sophisticated threats such as confounds and experimenter bias (McAuliff & Duckworth, 2010; McAuliff et al., 2009). Indeed, participants in this study rated science as less valid when it was missing a control group without the assistance of a scientifically informed cross-examination; however, the scientifically informed cross-examination simply increased the influence of the validity manipulation on participants' ratings of the scientific evidence (Austin & Kovera, under review). In Kovera et al. (1999), the researchers manipulated validity by varying the study's construct validity. Potentially, cross-examination may not be able to increase juror sensitivity to variations in construct validity because the concept is more difficult to grasp than a missing control group. Thus, it remains unclear whether cross-examination can educate jurors about sophisticated threats to scientific validity (e.g., construct validity, confounds, experimenter bias, statistical conclusion validity). Moreover, it is possible that although college student participants were sensitive to variations in the presence of a control group (Austin & Kovera, under review), similar results may not be obtained with a more representative sample that lacks the methodological training that undergraduate psychology majors have had.

Overall, *Daubert's* safeguards require further empirical research to examine the effectiveness of combating unreliable expert testimony. Methodological training outside the courtroom improves lay peoples' ability to provide scientific answers to real world problems (Fong et al., 1986), but can *Daubert's* safeguards provide the necessary methodological training for jurors? The Supreme Court intended for the three proposed safeguards to act as a form of scientific training for jurors. What we do not know is why some forms of methodological training – specifically the types of training that are provided in the safeguards set forth in *Daubert* – may be ineffective. To answer this question we turn to social cognitive research on persuasion to help understand and predict the juror decision making process.

CHAPTER 5: PROCESSING MODELS OF PERSUASION

The social-cognitive literature on persuasion may provide a framework for understanding how jurors process expert evidence. The Elaboration Likelihood (ELM; Petty & Cacioppo, 1986) and the Heuristic-Systematic (HSM; Chaiken, 1980; Chaiken, Lieberman, & Eagly, 1989) models of persuasion are relevant to the process of juror decision-making. Both models posit that people intend to form accurate opinions after hearing a persuasive argument but that people may use different cognitive processes to form their opinion depending on the level of effortful processing in which they are willing or capable of engaging (Chaiken et al., 1989; Petty & Cacioppo, 1986).

Specifically, people process incoming arguments in one of two ways: either by evaluating the quality of an argument or by relying on mental shortcuts and decision-rules to guide evaluations of the argument. Central (ELM) or systematic (HSM) processing is an effortful process through which the target of the persuasive message carefully analyzes the quality of the arguments being made. People engaging in this type of processing will be most persuaded by well-developed, high quality arguments (Petty & Cacioppo, 1984; Petty, Cacioppo, & Goldman, 1981; Leippe & Elkin, 1987). Jurors employing this processing style may, for example, focus on the quality of the research being presented.

Peripheral (ELM) or heuristic (HSM) processing, on the other hand, is a processing strategy in which the target of the persuasion uses mental shortcuts in argument evaluation (Chaiken, 1980; Petty et al., 1981). For example, people employing this type of processing may be persuaded by an argument that has many points rather than an argument with few points, irrespective of the argument quality (Petty & Cacioppo, 1984).

Research on Peripheral/Heuristic Processing

Student participants read a persuasive message that either argued for or against the implementation of comprehensive exams. The message contained either 3 or 9 arguments that were either weak or strong. Further, the researchers manipulated personal involvement of the student by leading the student participants to believe that the comprehensive examinations would either be implemented in the next year and thus would affect whether they had to take a senior comprehensive examination prior to graduation (high involvement) or the implementation of the exams would take place in the next 10 years (low involvement). Long arguments were more persuasive to the students in the low involvement group regardless of message quality. Alternatively, participants in the high involvement group were more persuaded by high quality arguments regardless of quantity. In summary, message length is one heuristic cue that people may rely in their evaluations of persuasive messages when employing a low level processing strategy (Petty & Cacioppo, 1984).

People employing a low level processing strategy may also be persuaded by messages that are agreed upon by many people in a group (i.e. messages that have consensus; Axsom, Yates, & Chaiken, 1987; Harkins & Petty, 1981; McAuliff & Kovera, 2008; Maheswaran & Chaiken, 1991; Martin, Hewstone, & Martin, 2007). In a study that used a modified version of the comprehensive oral examination study, when outcome relevance was low (i.e., the students would not be affected by the implementation of the comprehensive exams), opinions regarding the implementation of the exams were more persuasive when there was high consensus (9 of 11 students agreed) as compared to low consensus (2 of 11 students agreed; Martin et al., 2007). Further, the researchers found no effect of argument quality on student opinions in the low outcome relevance condition. However, in the high outcome relevance conditions, message quality affected student opinions (Martin et al., 2007). Therefore, in addition to message length, consensus is another heuristic cue that people may rely on when evaluating persuasive messages. In the legal setting, while evaluating an expert's testimony, jurors may rely on this heuristic when relying on information such as whether conclusions from a study have obtained general acceptance or whether the study has been published within a peer reviewed journal in their evaluation of evidence quality (McAuliff & Kovera, 2008).

Source cues such as expertise, attractiveness, or likability may also be influential in evaluating persuasive messages. These variables can affect persuasiveness, irrespective of argument quality (Chaiken & Maheswaran, 1994; Wood & Kallgren, 1988). For example, participants listened to a transfer graduate student give a talk at his new school (Wood & Kallgren, 1988). The transfer student's likability varied based on whether the student complimented the faculty and students at the new university (i.e., likeable) or criticized the new university, faculty, and students for being inferior (i.e., unlikeable). Some participants recalled fewer of the graduate student's opinions and behaviors in a later recall task. The researchers classified these students as engaging in peripheral processing. In addition, the researchers found that these students also rated the likeable source as more persuasive than the unlikable source, suggesting that source characteristics such as likability may affect those who engage in lower levels of message evaluation (Wood & Kallgren, 1988).

Message related cues may also encourage jurors to engage in peripheral processing. For example, the representativeness heuristic may ease jurors' decision-making task (Tversky & Kahneman, 1974), because people relying on a representative heuristic may be more persuaded by an argument when factors such as sample participants, experimental methods, or stimulus materials in the experiment emulate real world conditions (Kovera et al., 1999; McAuliff & Kovera, 2008). For example, jurors were less persuaded by a study with undergraduate students rather than workers from an occupational setting to which the expert wanted to generalize her research (Kovera et al., 1999).

When evaluating the scientific quality of an expert, systematic processing that involves an evaluation of the quality of the expert's evidence is obviously ideal. In general, both HLM and ELM posit that people generally desire to make accurate decisions because they aspire to hold attitudes and beliefs that are correct (Chaiken, 1980; Petty & Cacioppo, 1986). However, a few conditions must be met for people to process arguments systematically. For one, a person must be both motivated to process the argument and able to process the argument to engage in systematic (HSM) or central (ELM) processing (Chaiken, 1980; Petty & Cacioppo, 1986). In the legal setting, researchers have generally assumed that jurors are motivated to make accurate judgments and process scientific evidence systematically, (Cecil, Lind, & Bermant, 1987; Cooper et al., 1996; Kovera et al., 1999). However, several factors can influence personal motivation to engage in systematic processing such as personal relevance to the topic (Petty & Cacioppo, 1981), personal responsibility (Chaiken et al., 1989), or an individual's Need for Cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996).

Factors Affecting Motivation

There are individual differences in people's propensity to engage in effortful processing, which can be measured using the Need for Cognition (NFC) scale (Cacioppo & Petty, 1982). Low NFC individuals lack motivation to engage in effortful, cognitive thinking. For example, they may rely on message or source related cues to evaluate information and form opinions. Alternatively, high NFC individuals form opinions by seeking and thinking about information (Cacioppo et al., 1996). In terms of the ELM and HSM, high NFC individuals rely more on systematic/central methods of processing information whereas low NFC individuals rely more on heuristic/peripheral information when evaluating argument quality. Specifically, people who are high in NFC are better able to differentiate between strong and weak arguments (Cacioppo, Petty, Kao, & Rodriguez, 1986; Smith & Petty, 1996).

In the legal setting, an individual's NFC level is relevant to evaluating expert evidence. For example, high and low NFC jurors differ in their sensitivity to a missing control group (McAuliff & Kovera, 2008). High NFC mock jurors rendered more verdicts in the direction of the expert's testimony and gave more positive evaluations of scientific quality when the expert presented a study with a valid control group. Alternatively, low NFC jurors' verdicts and evaluations of trial evidence did not differ based on scientific validity (McAuliff & Kovera, 2008). Thus, individual differences exist in jurors' motivation to evaluate scientific quality.

Other variables may also influence a jurors' motivation to engage in systematic processing. For example, who hires the expert and who examines the expert may influence whether jurors attend systematically to the information presented by the expert.

In the adversarial court system used in the United States, experts may be either hired by an attorney or by the court as a court-appointed expert. In the inquisitorial system of justice, employed in many European countries, the expert is always appointed by the judge. Another distinction between the systems is who questions the expert. In adversarial courts, attorneys are usually responsible for the questioning of the expert. Alternatively, in the inquisitorial system, the presiding judge is responsible for gathering evidence and questioning witnesses (Anderson & Otto, 2003; Collett & Kovera, 2003; Fulero & Turner, 1997). In the inquisitorial system, the judge requests an approved expert to deliver an impartial report on a requested topic. Following the presentation of the expert's report, clarifying questions may be asked by the judge and follow-up questions may be asked by the attorney when necessary. Attorney questioning is frequently short because the judge's job is to ask all relevant questions. Truth is thought to arise simply from questioning the people most involved in the disputed issue. Conversely, adversarial trial technique manuals instruct attorneys to attack the opposing side during cross-examination, to get the witness to admit facts that support their case, and to try and discredit the witness (Mauet, 2007). Thus, the goals of the people conducting the cross-examination differ significantly.

Past research has examined the influence of adversarial, attorney-hired experts compared to inquisitorial, court-appointed experts (Brekke, Enko, Clavet, & Seelau, 1991) on juror and jury decisions. To emulate the adversarial system, the researchers introduced the expert as a witness for the prosecution and the expert was questioned by attorneys. To emulate the inquisitorial system, the researchers introduced the expert as a court-appointed witness and the expert was questioned by the judge. Deliberating jurors recalled more information from the expert's testimony in the adversarial condition. Testimony elicited by judges from court-appointed experts may lower juror motivation to process information systematically because court-appointed experts and judges are seen as more credible and neutral than attorneys or attorney hired experts because attorneys possess a motive to reduce witness credibility and an attorney hired expert could be a hired gun (Brekke, et al., 1991). Thus, adversarial methods may encourage jurors to engage in systematic processing. However, this study did not include a condition in which the attorney cross-examined a court-appointed expert (a condition in which the judge cross-examined an attorney-hired expert is not realistic as this does not reflect inquisitorial practice), thus it is difficult to determine whether it is the judge or the court-appointed status or both that reduce juror motivation to engage in systematic processing.

The person conducting the cross-examination may affect a jurors' propensity to engage in central/systematic processing because jurors may be more motivated to process information elicited from a cross-examination if they feel engaged during the witness's examination. In Brekke et al. (1991), jurors rated the adversarial trial as more interesting than the inquisitorial trial. This could be due, in part, to the conversational nature of adversarial examinations rather than the mere reporting that occurs in inquisitorial courts. People may be more likely to systematically process information they deem interesting (e.g., attorney conducted cross-examinations). Furthermore, judges, unlike attorneys, are held in high regard (Finkelman, 2005). Because jurors may deem judges as more credible and truth-seeking, jurors may have less motivation to process the expert's testimony because they may reason heuristically that the testimony is fair and balanced.

Austin and Kovera (under review) examined the influence of adversarial (both court-appointed and defense hired) and inquisitorial expert methods on jury decision making and found that expert type did not impact verdict nor did expert type aid jurors with the task of evaluating scientific validity. Notably, this study lacked a recognition or recall memory test to examine whether expert type influenced the amount of information recalled, which may be a more sensitive indication of systematic versus peripheral processing than either verdict or perceptions of evidence strength. Further, only 71% of participants in their study only noticed the manipulation of who had hired the expert; thus, it is possible that the identification of the manipulation was too weak to influence jurors.

Factors Affecting Ability

High motivation does not guarantee that systematic/central processing will occur (Chaiken, 1980; Petty & Cacioppo, 1986). A person must also possess the requisite knowledge necessary to process and attend to the argument. Therefore, despite jurors' motivations to process an expert's testimony, they may lack the ability to reason scientifically (McAuliff & Duckworth, 2010; McAuliff et al., 2009). For example, participants were more persuaded by communicator expertise when the testimony was complex because they were unable to process the information centrally (Petty et al., 1981). Moreover, participants engage in mental shortcuts and avoid processing arguments centrally when time does not allow participants to engage in effortful processing (Mackie & Worth, 1989).

Another factor that may affect juror ability to attend to information is related to sheer amount of testimony that jurors are required to attend to during a trial. Our cognitive abilities are limited. Once our limited supply of cognitive resources is spent, performance on subsequent tasks decreases (Baddeley, Scott, Drynan, & Smith, 1969; Conway & Engle, 1996; Murdock, 1965). For example, in one study participants sorted cards into one pile, two piles by color, or four piles by suit while simultaneously being read lists of common English words. As the sorting task became more difficult, participants' recall of English words decreased (Murdock, 1965).

Type of witness may also affect whether jurors engage in systematic processing. Arguably, expert testimony may be more difficult for jurors to comprehend than other lay witnesses due to the scientific jargon and specialized knowledge used by experts (Cecil, Has, & Wiggins, 1991). Thus expert testimony may require more cognitive resources than would other witnesses. An initial study designed to explore the conditions under which mock jurors would engage in systematic/central processing for expert testimony manipulated both testimony complexity and expert credentials in a product liability case (Cooper et al., 1996). Complex testimony contained scientific jargon whereas simple testimony described the same information in layman's terms. The highly credible expert was the editor of a journal, held advanced degrees, had published many peer reviewed, relevant journal articles, and was employed at a prestigious university. The low credibility expert stemmed from and worked at unknown institutions, had only published a few articles, and held no other academic positions. When the expert's testimony was complex, mock jurors were more persuaded by an expert with high credentials. In contrast, jurors relied on the content of the expert's testimony when the testimony was easy to understand, suggesting that jurors rely on peripheral/heuristic source related cues for persuasiveness when they are unable to understand the information during trial but will engage in systematic/central processing when the topic is comprehensible. Further, a follow-up study manipulating both testimony complexity and level of pay/frequent court appearances found that when the testimony was complex, mock jurors relied on heuristic cues for persuasiveness (i.e., they rated the expert as less likable, less believable, and less convincing when the expert appeared to be a hired gun). However, when the testimony was presented in language that was easier to comprehend, mock jurors were not affected by any of the peripheral cues (Cooper & Neuhaus, 2000).

In summary, the central/systematic versus peripheral/heuristic processing can help to predict which characteristics of the testimony and of the juror will influence jurors' propensity to engage in a systematic evaluation of testimony. More specifically, these theories may be especially useful for understanding the conditions in which jurors may critically evaluate expert testimony on issues of reliability and validity.

CHAPTER 6: CONCLUSIONS

Expert testimony is often necessary to aid jurors in understanding scientific, technical, or otherwise complex evidence. Expert testimony has a persuasive influence on jurors' evaluations of trial testimony (e.g., Brekke & Borgida, 1988; Kovera, et al. 1994; 1999; Schuller, 1992; Schuller, et al., 2004). Because expert testimony influences juror decisions, the use of experts in court is often a controversial issue. At the forefront of these controversies are concerns about the quality of the proffered testimony. Although Daubert clarified judges' role as gatekeepers of scientific evidence, judges may lack the skills necessary to evaluate scientific quality (Gatowski et al., 2001; Kovera & McAuliff, 2000; Manuto & O'Rourke, 1991) and may admit testimony and opinions derived from invalid science into evidence (Kovera & McAuliff, 2000; McAuliff, 2009).

When judges fail to recognize flawed science, attorneys may aid the judge in determining scientific quality by filing a motion to exclude the testimony. For attorneys to successfully assist judges, attorneys must be sufficiently knowledgeable about scientific methodology to identify threats to validity and reliability, but attorneys may fail to identify methodological issues in scientific studies (Kovera & McAuliff, 2002). Jurors, without training, also struggle with scientific concepts and evidence evaluation (Kovera et al., 1999; Levett, 2008, 2009; McAuliff & Duckworth, 2010; McAuliff & Kovera, 2008; McAuliff et al., 2009).

To combat the effects of invalid science, the Supreme Court argued that three safeguards would assist jurors with their evaluations of the scientific reliability of expert evidence: "vigorous cross-examination, presentation of contrary evidence, and careful instruction on burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence" (*Daubert v. Merrell Dow Pharmaceuticals* 1993, p. 576). Instructions on burden of proof were not designed with the intent to help educate jurors about issues of reliability (Kovera et al., 2002). Opposing experts increase skepticism and either fail to educate jurors about unreliable scientific evidence admitted into court or fail to influence trial decisions enough to influence key decisions (Levett & Kovera, 2008; 2009). Cross-examination has demonstrated promise for educating jurors about scientific validity (Austin & Kovera, under review). Recall that jurors who heard extensive scientifically informed cross-examinations differentiated between valid and invalid science, giving higher validity ratings to the valid science, whereas jurors who heard the scientifically naïve cross-examinations gave uniformly high ratings for scientific validity, regardless of the study's validity. Again, although cross-examination assisted jurors in evaluating scientific quality, verdicts were not affected.

Several questions remain unanswered about the ability of cross-examination to increase juror sensitivity to variations in scientific methodology. For example, can cross-examination educate jurors about sophisticated threats to scientific validity beyond missing control groups? Can cross-examinations educate jurors about reliability? Reliability in measurement (i.e., consistency) is a necessary aspect of valid measurements. No previous research has examined whether *Daubert's* safeguards can educate jurors about scientific reliability, and issues of reliability are problematic for some fields in psychology (Hagen, 1997; Holly, 1999; Faust & Ziskin, 1988). Can laypeople without training in scientific methodology, unlike introductory psychology students, be educated by scientifically informed cross-examinations? Although a scientifically informed cross-examination should provide jurors with the ability to evaluate scientific quality, it may not provide jurors with the adequate motivation. Are there factors that influence juror motivation to attend expert testimony during trial?

The legal role of the person who cross-examines the expert witness may influence jurors' motivation to engage in systematic processing of expert testimony. Jurors may be more motivated to process information learned from a cross-examination if they feel engaged during the witness's examination. Adversarial trials are more interesting than inquisitorial trials (Brekke et al., 1991). Further, jurors may implicitly trust information from a judge because judges are neutral trial players compared to attorneys who are hired to advocate for their witness. Thus, jurors may feel less motivation to process information during a judge-conducted cross-examination and consequently may rate the expert's testimony of high quality, irrespective of the actual quality.

In evaluating whether the behavioral assumptions underlying *Daubert* are valid, one must assess whether cross-examination is effective at sensitizing jurors to variations in the quality of expert evidence. Another important question is to examine whether attorneys are able to perform their duties specified by the Court. Specifically, can attorneys identify validity threats and develop questions to highlight issues of scientific validity and reliability for jurors? Although attorneys may not be able to identify threats to validity (Kovera & McAuliff, 2002), no research has examined whether they possess the skills to identify threats to reliability and to develop a line of questioning that addresses those threats.

CHAPTER 7: OVERVIEW

To answer these questions, I conducted two studies to empirically examine the Court's assumptions about cross-examination. In the first study, I recruited 95 attorneys to read the fact pattern from a civil trial that included expert testimony. This trial summary was based on the testimony from the appeals court transcript of *Hoffman v. Board of Education of the City of New York* (1978). According to the fact pattern, the plaintiff was placed in remedial classes based upon a determination by the defendant's certified psychologist that the child was of reduced mental capacity. I varied both the validity and reliability of the intelligence test the expert administered to the plaintiff. It is important to note that psychological tests can vary in both validity and reliability. Validity refers to the how well an instrument measures what it intends to measure (e.g., I.Q. tests measure cognitive ability, personality tests measures personality characteristics, scales measures weight) and reliability refers to consistency of measurement. Although people often use these terms interchangeably (e.g., the Supreme Court in *Daubert* used the term reliability when referring to validity), reliability and validity are distinct concepts and a reliable measure does not imply that measure is also valid. It is conceivable that a measurement tool may consistently get the same number, but can be invalid (e.g., the I.Q. test consistently gives the same score, but does not really measure cognitive capacity; the bathroom scale consistently reads 5 pounds less than what it really should). For this study, I created three versions of the intelligence test. The test was either valid and reliable, invalid due to experimenter bias but reliable, or the test had average reliability indices on the full scale internal consistency score, the test/retest score, and the inter-observer reliability scores. Thus, attorneys read that the expert administered a valid and reliable, invalid but reliable, or unreliable intelligence test to the plaintiff. Based on the expert testimony and the accompanying trial facts, attorneys answered questions about whether they would file a motion to exclude the testimony, rated the quality of the scientific testimony, and wrote cross-examination questions to assess the validity and reliability of the expert's testimony.

ELM and HLM posit that both motivation and ability must be present for people to attend to argument quality. Arguably, attorneys should be motivated to process scientific quality because they are ethically bound to zealously defend their client's interests. Although Kovera and McAuliff (2002) found that validity did not impact attorneys' ratings of scientific validity and reliability, I predicted that attorneys would provide higher ratings of scientific quality when the intelligence test administered was valid and reliable than when it was invalid but reliable or unreliable based on the Court's assumptions that judges are capable of performing this role and judges and attorneys receive the same educational classes. I also predicted attorneys would provide scientific justifications for moving to exclude the expert's testimony when the intelligence test suffered from a threat to validity or low reliability and would provide legal justifications or cite a trial strategy when the intelligence test was valid and reliable. Further, I predicted that the attorneys' cross-examination questions would address the relevant validity and reliability issue when the study was invalid or unreliable.

In Study Two, I evaluated the ability of cross-examination to educate jurors about a sophisticated threat to scientific validity (experimenter bias). Jurors watched a simulated videotaped trial based on the testimony from an actual case *Hoffman v. Board of Education* (1978). During the trial, jurors heard testimony from an expert about an intelligence test she administered the plaintiff. I manipulated the validity and reliability of the intelligence test in the same way that I manipulated these variables in Study 1 (valid and reliable, invalid but reliable, unreliable). In addition, I manipulated cross-examination type. Two-thirds of the sample heard a scientifically informed cross-examination designed to help jurors understand concepts of reliability and validity and to apply these concepts to the test administered by the expert. Half of the scientifically informed cross-examinations emulated testimony from adversarial courts and half emulated testimony from inquisitorial courts. In the scientifically informed inquisitorial conditions, the judge asked the expert to provide a report to the court and subsequently the judge cross-examined the expert. Attorneys did not pose questions to the expert. In the scientifically informed adversarial condition, the defense attorney examined the court-appointed expert during direct examination and the plaintiff attorney cross-examined the expert. The final third of jurors who heard testimony from the expert heard the plaintiff attorney cross-examine the expert, but the attorney did not ask scientifically informed questions (i.e., no follow-up questions that asked the expert to explain concepts of reliability and validity and to apply these concepts to the test administered by the expert). All participants heard the questions asked in the scientifically naïve condition. Finally, I included a no expert control condition to assess the impact of expert testimony on jurors' decisions. Participants rendered a verdict and, in the conditions containing an expert, rated their perceptions of the expert's testimony, scientific validity, and reliability.

Past research has demonstrated mixed effects regarding the ability of scientifically informed cross-examinations to educate jurors with evaluating scientific validity (Austin & Kovera, under review; Kovera et al, 1999); however, one reason Kovera and colleagues (1999) suggested that their cross-examination did not serve an educational purpose for jurors was because it did not specifically link the methodological aspect of the expert's research that had been violated to the study presented. Austin and Kovera (under review) addressed the relevant components of valid science and linked the presence of absence of those components to the experiment described by the expert. In the current study, I also linked the relevant validity and reliability concerns to the specific intelligence test and thus, I predicted that scientifically informed cross-examinations would provide jurors with the ability to evaluate scientific testimony. Thus, when participants heard a scientifically informed cross-examination, they would provide higher ratings of scientific quality to the valid and reliable intelligence test than of the invalid but reliable or the unreliable intelligence test. However, I also predicted that scientifically informed cross-examinations may not function equally in adversarial and inquisitorial courts. Thus, I also predicted that scientifically informed cross-examinations may be more likely to help jurors evaluate scientific evidence when the attorney performed the scientifically informed cross-examination than when it is conducted by a judge because that is when jurors may feel motivated and engaged to process the expert's testimony (Brekke et al., 1991). Thus, jurors may only provide higher ratings of scientific quality of the valid and reliable intelligence test than of the invalid but reliable or unreliable intelligence test when the attorney performed the cross-examination of the expert.

CHAPTER 8: STUDY 1: ATTORNEY ABILITY TO EVALUATE SCIENTIFIC QUALITY AND FORMUALTE SCIENTIFICALLY INFORMED CROSS-EXAMINATIONS

Method

Participants I contacted acquaintances (attorneys and non-attorneys) and used a chain-referral sampling technique to recruit attorneys to participate in the first study, resulting in a sample of ninety-five attorneys. Based on this recruitment technique, I am unable to compute a response rate because I cannot track how many times the survey was sent out. In addition, I capped the number of respondents on the online system and do not know how many more attorneys would have completed the survey if the survey had remained open. All attorneys received a letter explaining the purpose of the study and a link to the study website. To increase motivation to participate, we offered attorneys $30 in return for their participation. We asked participants to provide a mailing address to which we could send the check if they opted for compensation. I set up the online survey so that mailing addresses were filtered into a separate file, making it impossible to match data to the respondent who provided it.

Design

This study had a 3 (Scientific Quality: Valid and Reliable, Invalid but Reliable, Unreliable) group design.

Description of the Fact Pattern, Proposed Expert Evidence, and Manipulations

I asked attorneys to assume the role of the plaintiff attorney in a civil case based on the fact pattern of *Hoffman v. Board of Education of the City of New York* (1978). Attorneys read the following fact pattern: the plaintiff was placed in remedial classes based upon a determination by the defendant's certified psychologist that the child had an intelligent quotient (I.Q.) of 74. State laws permit placement in remedial classes when I.Q. scores fall below 75. The plaintiff remained in remedial classes until he was 17 years of age. At 17, the plaintiff was transferred to an occupational training center for the mentally challenged to learn a trade. He remained there for the period of one year before he was advised that he could no longer continue his education at the occupational training center because recent intelligence testing showed he was of average intelligence. The plaintiff made the claim that placement in remedial classes crippled his emotional and social development and motivation to achieve. Teachers and the principal at his school were expected to testify that the plaintiff suffered from a severe speech impediment that made him difficult to understand and influenced his ability to relate with other students. They would also testify that the plaintiff was not performing at a level consistent with students in normal classes and would not have excelled if not for remedial classes.

The attorneys also read a description of the expert testimony the defense wished to proffer. In this testimony, the expert planned to discuss standardized intelligence testing (e.g., the development of the intelligence test, the average score of such tests, what deviations from that score represent), intelligence quotients, and the results of the plaintiff's intelligence test from when he was in kindergarten. The expert used the fictional Bellevue Primary Scale of Intelligence for children to evaluate the plaintiff's intelligence. I created a fictional intelligence test name because I was concerned with whether it would be possible to manipulate perceptions of validity and reliability for a generally accepted intelligence test (e.g., the Wechsler Intelligence Test for Children) which is said to have excellent reliability and is one of the most widely used tests in clinical practice (Camara, Nathan, & Puente, 2000; Groth-Marnat, 2003; Watkins, Campbell, Nieberding, & Hallmark, 1995). Attorneys read that this test was standardized on 2,200 children between the ages of 6 and 16 years 11 months and that the clinicians responsible for standardizing this test examined children that were identified as normal, gifted, children with mild or moderate mental retardation, children with learning disorders, children with ADHD, and children with expressive and mixed receptive-expressive language disorders, autistic disorder, and children with motor impairment.

Within this summary description, I manipulated the validity and reliability of the intelligence test the expert administered to the plaintiff. In the valid version of the clinician's administration, the clinician was unaware of the plaintiff's predicted I.Q. prior to the administration of the exam. In the invalid version, the clinician/expert was aware of the plaintiff's predicted I.Q. prior to the administration of the exam. Therefore, in the invalid version, the expert may have biased ratings of the plaintiff's I.Q. by scoring the plaintiff's responses to items on the intelligence test so that it would align with an expectation of lower intelligence.

I manipulated reliability by varying the reliability indices of the internal consistency scores, the test/retest reliability score, and the inter-observer scores. The range for a reliability coefficient can fall between 0 and 1. The closer the reliability coefficient is to 1, the more confidently we can say that we are assessing one trait (e.g., intelligence). When the test was unreliable, the reliability coefficients fell between 0.64-0.73 indicating questionable reliability. When the test was reliable, the reliability coefficients fell between 0.89 - 0.96, suggesting excellent reliability.

Dependent Measures.

Motion to Exclude. After reading the case facts, attorneys rendered a dichotomous decision whether they would move to exclude the expert evidence and provided a justification for their decision. Attorneys also reported the likelihood that they would move to exclude the expert's testimony using a continuous scale that ranged from 0% to 100% in 10% increments.

Ratings of Scientific Evidence and Daubert Safeguards. I programmed my online survey (Qualtrics.com) to recode the data when necessary so that higher numbers reflected more positive evaluations of expert's testimony (i.e., more valid, more reliable, more helpful, more probative, and less prejudicial). Attorneys rated the proposed expert testimony on 7-point Likert-type scales by indicating their agreement with 12 statements (items recoded are indicated by [R]): (1) The intelligence test [the expert] administered to the plaintiff was reliable; (2) The intelligence test [the expert] administered to the plaintiff was scientifically valid; (3) If a child failed the intelligence evaluation performed by [the expert], he/she is more suitable for remedial classes; (4) [The expert's] intelligence assessment tool is based on good scientific principles; (5) [The expert] used an appropriate test of intelligence when testing [the plaintiff]; (6) Psychological intelligence tests are reliable indicators of someone's intelligence; (7) Psychological intelligence tests are created using valid scientific methods; (8) [The expert's] testimony would prejudice a jury [R]; (9) [The expert's] testimony is probative; (10) The findings from [the expert's] assessment can be used to understand whether [the plaintiff] should be in remedial classes; (11) Expert psychological testimony can improve juror decisions; (12) [The expert's] testimony would help jurors render a judgment. Attorneys also answered a series of three questions to assess their perceptions regarding different people's ability (i.e., jurors, attorneys, and judges) to identify flaws in scientific research. First, attorneys rated whether they believed that jurors, attorneys, and judges are able to recognize flaws in scientific research on individual 7-point Likert-type scales from 1 (strongly disagree) to 7 (strongly agree). Next, attorneys rated whether they believed that cross-examination and opposing experts are effective at educating jurors about flaws in scientific research and whether they help to reveal the reliability and validity of expert testimony on 7-point Likert-type scales from 1 (strongly disagree) to 7 (strongly agree).

Cross-examination questions. I provided open ended response boxes for attorneys to write two sets of five cross-examination questions they would pose to the expert if given the opportunity. For the first set, I asked attorneys to simply write five cross-examination questions that they would pose to the expert if given the opportunity. For the second set, I asked attorneys to craft five cross-examination questions that they would pose to the expert that might help them assess the validity and reliability of the intelligence test the expert administered to the plaintiff.

Manipulation Checks. Attorneys answered two multiple choice manipulation check items regarding the validity and reliability of the intelligence test employed by the expert. Both manipulation checks asked attorneys to choose from two possible responses. For the validity check, attorneys clicked either "yes" or "no" in response to a question of whether the expert was aware that the plaintiff was expected to be intellectually disabled before administering the exam. For the reliability check, I asked attorneys to check if the indices of reliably fell between 0.64 – 0.73 or 0.89 – 0.96.

Demographic Questionnaire. Attorneys completed a demographic questionnaire, providing information about their age, ethnicity, number of years in practice, types of expert testimony they have proffered in the past 12 months, types of law they practice, list the state(s) they are licensed to practice, and to answer questions designed to assess whether they received scientific training in undergraduate or graduate school, and to rate their familiarity with the Reference Manual on Scientific Evidence**.**

Procedure

All attorneys received a letter (either from me or from one of many referrals who may or may not have been another attorney) explaining the purpose of the study, basic information, and a link that would take them to the study (See Appendix A). When attorneys clicked on the link included in the letter, the online survey program opened and randomly assigned attorneys to one of the three experimental conditions. Attorneys provided informed consent by clicking a button at the end of the consent form indicating they wished to continue (See Appendix B). Attorneys assumed the role of the plaintiff attorney and read their randomly assigned fact pattern (See Appendix C). Before continuing, attorneys were instructed that they would not be allowed to return to the summary and to only click ‘continue’ when they were ready to proceed to the dependent measures.

Once they had indicated that they were ready to continue, attorneys proceeded to another page in the online survey. Attorneys first rendered a dichotomous judgment indicating whether they would file a motion to exclude the defense’s expert and provided a justification for their reason to file that decision. Based on these responses, I created a coding scheme to examine the justifications behind why attorneys would or would not file a motion to exclude the testimony (See Appendix D for full coding scheme). Two coders blind to condition and hypotheses examined the frequency with which attorneys provided legal (e.g., relevance, prejudicial, expert qualifications, common understanding of the jury, helpfulness, ultimate issue, hearsay), scientific (e.g., statements about internal or external validity or reliability), or trial strategy justifications for their decisions (e.g., likelihood that motion would succeed, cost of arguing motion, how they would handle expert through cross-examination or opposing experts if not excluded).

I calculated concordance estimates with the formula: $C = 2(C_{1,2})/(C_1 + C_2)$. In this formula, C = the agreement between coders regarding justifications for motions to exclude, $C_{1,2}$ represents the number of categories agreed upon by both coders and C_1 and C_2 represent the number of categories coded by each coder respectively. Codes did not match if one coder specified a reason to exclude and the second coder did not or if codes did not match at the subordinate category level. However, I counted agreement if one coder specified a reason to exclude at the superordinate category and the second coder used a more specific matching subordinate category. Overall, the concordance rate for the two coders on this measure was good (C = .83; Range 0.0 – 1.0); a third coder who was blind to condition resolved disagreements.

Attorneys then answered the 7-point Likert-type scales, rating the extent to which the proposed expert testimony was helpful, reliable, valid, of probative value, and prejudicial (see Appendix E). Attorneys judged the quality of the intelligence test employed by the expert by rating the reliability of the intelligence test administered, the validity of the intelligence test administered, the conclusions that could be drawn from the intelligence test administered, whether the intelligence test was created on good scientific principles, the appropriateness of the intelligence test used by the expert, the reliability of intelligence tests generally, and the validity of the methods used to create intelligence tests. I averaged across these seven items to create a single index of participants' perception of scientific quality, with higher numbers representing more positive evaluations of the expert's study (Cronbach's $\alpha = .82$). Attorneys then answered the questions to assess their perceptions regarding different people's ability (i.e., jurors, attorneys, and judges) to identify flaws in scientific research and answered questions to assess their perceptions regarding the utility of opposing experts and cross-examination to assist jurors with evaluating scientific validity and reliability (see Appendix E).

Following these Likert-type items, I asked attorneys to write two sets of five cross-examination questions. For the first set, I asked attorneys to simply write five cross-examination questions that they would pose to the expert if given the opportunity. For the second set, I asked attorneys to craft five cross-examination questions that they would pose to the expert that might help them assess the validity and reliability of the intelligence test the expert administered to the plaintiff. If attorneys had written questions in the previous section that accomplished this goal, I asked them to either note this for me in the space provided or to rewrite them in the space provided, but I did not request they create more questions. Two coders blind to condition and hypotheses coded the attorneys' questions to assess the purpose of the question (e.g., competence, case specific knowledge, expert qualifications, validity, reliability; see Appendix F for full coding scheme). Overall, the concordance rate on the first five cross-examination questions was good ($C = .77$; Range 0.0 – 1.0). The concordance rate for the second set, the scientifically informed cross-examinations questions, was very good ($C = .84$; Range 0.0 – 1.0). A third coder, blind to condition, resolved disagreements on both sets of questions. Within each coding category, an attorney could receive a code of 1 (they posed a question in this category within their 5 questions) or a code of 0 (they did not pose a question in this category within their 5 questions). Thus, if an attorney asked five questions regarding the expert's qualifications, they received a 1 under the category for expert qualifications.

Following the cross-examination question page, attorneys answered two manipulation check questions to assess sensitivity to my manipulations of reliability and validity. Finally, attorneys provided demographic information (See Appendix G). Two coders blind to the condition and hypotheses of the study classified the types of expert testimony proffered by attorneys in the past 12 months into one of the following categories of expert testimony: accident and traffic reconstruction, accounting, arson/fire, ballistics, drugs/toxicology, environmental, engineering, forensic, medical, psychological/psychiatric, safety, statistical, or other (see Appendix H for full coding scheme). I calculated concordance for types of experts proffered in court over the past 12 months using the same method described above and the concordance between the coders was good (C = 0.80, Range 0.0 – 1). A third coder who was blind to condition resolved disagreements.

Hypotheses

As discussed in the introduction, psychological research suggests that attorneys may not receive appropriate scientific training and may not recognize flawed scientific evidence (Lehman et al., 1988; Kovera & McAuliff, 2002). However, based on the Court's assumptions that judges are able to evaluate scientific evidence and the knowledge and judges and attorneys receive the same education, I predicted that attorneys would provide higher ratings of scientific quality for the valid and reliable intelligence test than they would for the invalid but reliable or unreliable intelligence test. Thus, I predicted that attorneys would provide more favorable ratings of scientific quality only when the intelligence test was both valid and reliable.

In Kovera and McAuliff (2002), 95% of attorneys stated they would move to exclude the expert's testimony irrespective of the study's scientific validity. Thus, I predicted that in the current study attorneys would move to exclude the expert's testimony regardless of the validity or reliability of the intelligence test (i.e., I predicted a ceiling effect). However, based on the Court's assumptions I predicted that attorneys' justifications for excluding would vary based on the validity and reliability of the intelligence test. When the intelligence test was valid and reliable, I predicted that attorneys would provide more legal and trial strategy justifications than when the intelligence test was either invalid but reliable or unreliable. When the intelligence test was unreliable, I predicted that attorneys would provide more reliability justifications than when the intelligence test was valid and reliable or invalid but reliable. Finally, when the intelligence test was invalid but reliable, I predicted that attorneys would provide more validity justifications in their reason to exclude than when the intelligence test was valid and reliable or unreliable. Thus, I predicted that attorneys would justify their decisions to move to exclude with the appropriate scientific justification when there was a validity or reliability threat present and that attorneys would justify their decisions to move to exclude with legal or trial strategy justifications when there was not a threat present.

I also predicted that attorneys who read that the expert administered an invalid but reliable intelligence test would pose more cross-examination questions that addressed administrator bias than would those attorneys that read the expert administered an unreliable or a valid and reliable intelligence test. I predicted that attorneys who read that the expert administered a valid but unreliable intelligence test would pose more cross-examination questions that addressed internal consistency, test-retest, and inter-observer reliability scores than would attorneys who read that the expert administered a valid and reliable or invalid but reliable intelligence test. Finally, I predicted that attorneys who read that the expert administered a valid and reliable test would ask more credibility damaging questions and qualification questions than would those who read that expert administered an unreliable or a invalid but reliable test.

Based on previous findings (Gatowski et al., 2001; Kovera & McAuliff, 2000), I predicted that attorneys would report high to moderate confidence in the ability of judges and attorneys to evaluate scientific evidence and low confidence in the ability of jurors to evaluate scientific evidence. I also predicted that attorneys would provide favorable ratings for the ability of opposing experts and cross-examination to serve as effective educational tools for jurors tasked with evaluating scientific quality.

CHAPTER 9: STUDY ONE RESULTS

Analytic Strategy

I subjected attorneys' dichotomous judgment of whether they would move to exclude the testimony to a three-group chi-square to examine whether the scientific quality of the intelligence test was associated with whether attorneys would move to exclude the testimony. Subsequently, I examined attorneys' stated justifications for why they would move to exclude or not exclude the testimony with chi-square analyses. I then examined whether the scientific quality manipulation affected attorney's justifications (e.g., did attorneys provide more validity arguments in invalid conditions etc.) with chi-square analyses.

I subjected the continuous measure of scientific quality to a one-way (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) analysis of variance (ANOVA) to determine whether ratings differed based on the type of intelligence test the expert administered to the plaintiff.

I subjected attorney ratings of how well judges, attorneys, and jurors can identify flaws in scientific research to a 3 (Decision Maker: Judge, Attorney, Juror) X 3 (Scientific Quality: Valid/Reliable, Valid/Unreliable, Invalid/Reliable) repeated measure ANOVA with decision maker as the within subject factor. I also subjected attorneys' ratings of the effectiveness of opposing experts and cross-examination for educating jurors about scientific validity and about scientific reliability to a 3 (Safeguard: Cross-examination vs. Opposing Experts) X 3 (Scientific Quality: Valid/Reliable, Valid/Unreliable, Invalid/Reliable) repeated measure ANOVA with safeguard as the within subject factor.

Manipulation Checks

Four attorneys did not complete the manipulation checks. There was a significant association between saying the clinician was blind to the plaintiff's I.Q. (i.e., the study was valid) and scientific quality condition, χ^2 (2, N = 91) = 25.86, p <.001, Φ = 0.53. Follow-up chi-squares showed that attorneys did not differ in their likelihood of saying the expert was blind to the plaintiff's I.Q. (i.e., study was valid) when the expert administered a valid/reliable (57%) or unreliable (78%) intelligence test, χ^2 (1, N = 62) = 3.26, p =.07, Φ = –0.23, [$CI_{.95}$ = -0.47, 0.05]. Attorneys stated that the clinician was blind to the plaintiff's predicted I.Q. more often when the clinician administered a valid/reliable intelligence test (57%) as compared to when the expert administered an invalid/reliable intelligence test (14%), χ^2 (1, N = 59) = 11.82, p =.001, Φ = 0.45, [$CI_{.95}$ = 0.16, 0.63]. Attorneys also stated that the clinician was blind to the plaintiff's predicted I.Q. more often when the clinician administered a unreliable intelligence test (78%) as compared to when the clinician administered an invalid/reliable intelligence test (14%), χ^2 (1, N = 61) = 25.25, p <.001, Φ = 0.64, [$CI_{.95}$ = 0.38, 0.81]. Thus, participants stated the expert did not know the plaintiff's I.Q. more often in conditions in which the science was valid suggesting that participants were sensitive to this manipulation check.

There was a significant association between saying the reliability indices fell between 0.64-0.73 and scientific quality condition, χ^2 (2, N = 95) = 95.00, p <.001, Φ = 1.00. Follow-up chi-squares showed that attorneys stated that the reliability indicies fell between 0.64 – 0.73 more often when then expert administered an unreliable intelligence test (96%) than when the expert administered a valid/reliable intelligence test (13%), χ^2 (1, N = 62) = 43.96, p <.001, Φ = –0.84, [$CI_{.95}$ = -0.90, -0.62]. Attorneys also stated that the indicies of reliability fell between 0.64 – 0.73 more often when the expert administered an unreliable intelligence test (96%) than when the expert administered an invalid/reliable intelligence test (10%), χ^2 (1, N = 61) = 46.17, p <.001, $\Phi = 0.87$, [$CI_{.95}$ = -0.65, 0.93]. Attorneys were equally likely to say the indicies of reliability fell between 0.64 – 0.73 when the expert administered a valid/reliable (13%) or invalid/reliable (10%) intelligence test, χ^2 (1, N = 59) = 0.13, p =.72, $\Phi = 0.05$, [$CI_{.95}$ = -0.22, 0.27].

I retained attorneys regardless of whether they passed or failed these manipulation checks because there is some reason to believe that attorneys will not pick up on these variations in real cases and I am interested in attorney behavior regardless of whether they noticed variations in validity and reliability. Moreover, if I had not retained these participants I would have reduced my statistical power and increased my chance of chance of committing a Type II error.

Attorney Demographics

Some participants opted not to provide demographic information, thus percentages may not sum to 100% because of missing data. Fifty percent (N = 48) of our sample was male and 45.3% was female (N= 43). Seventy-five percent of our sample identified as white/non-Hispanic, 9.5% identified as Hispanic, 4.2% identified as Asian, 2.1 % identified as black/non-Hispanic, and the remaining participants that responded to this question identified as Other (4.2%). Attorneys ranged in age from 26 to 75 (*M* = 34, *SD* = 8.43) and had been in practice between 1 and 32 years (*M* = 6.46 years, *SD* = 6.71). I asked attorneys to specify what type of law they practice. Two coders judged whether the law was a form of criminal law, civil law, both criminal and civil, or neither (e.g., family law), concordance = 1.00 (perfect agreement). Of the 91 attorneys who reported what type of law they practice, 45% reported that they practice criminal law and 70% reported practicing civil law (numbers do not sum to 100% because attorneys could be classified under both civil and criminal). Participants reported that on average 17% of their cases in the past 12 months had involved scientific testimony. When asked to consider all their cases over the past 12 months when they attempted to introduce scientific evidence, attorneys reported that the testimony was deemed admissible by the judge on average 51% of the time.

Attorneys described the types of expert testimony that they had proffered in the past 12 months. Thirty-two attorneys left this question blank, which either indicated that did not offer any expert testimony or that they chose not to respond. Of those who did respond (N=63), 71% (N=43) tendered some form of medial testimony (e.g., DNA, autopsy, or general medical), and 32% (N=20) tendered some form of psychological testimony (e.g., competency, risk assessment, abuse, counseling, insanity). Although medical testimony and psychological testimony were mentioned most often, attorneys also tendered testimony on drugs/toxicology, forensic evidence, engineering, and others with some frequency. The types of expert testimony tendered by attorneys are detailed in Table 1 below. There was a significant correlation between practicing criminal law and proffering a psychological expert, $r = .31$, [$CI_{.95} = .07, .52$], $p = .01$. Moreover, there was a small, albeit non-significant, negative correlation between practicing civil law and proffering a psychological expert, $r = -.19$, [$CI_{.95} = -.42, .60$], $p = .14$.

Table 1
Percentage of Attorneys Proffering Specified Types of Expert Testimony in the Past 12 Months

Expert Testimony Type	% Attorneys (n)
Medical	70 (44)
DNA	19 (12)
Autopsy	8 (5)
Psychological	21 (13)
Competency	17 (11)
Abuse	3 (2)
Risk Assessment	3 (2)
Domestic Violence	3 (2)
Counseling	2 (1)
Insanity	2 (1)
Drugs/Toxicology	16 (10)
Forensic	15 (9)
Fingerprints	8 (5)
Engineering	14 (9)
Accounting	11 (7)
Ballistics	10 (6)
Accident/Traffic Reconstruction	6 (3)
Environmental	6 (4)
Other	48 (30)

Note. Percentages are figured from attorneys who responded (N = 63).

During the demographic section, attorneys provided information about what types of educational training they have received regarding scientific issues. For example, attorneys told us their undergraduate major and whether they had any graduate training in the scientific method or any continuing legal education (CLE) classes focused on training them to evaluate scientific methods. Few attorneys (12%) reported pursuing an undergraduate major that would have included training in the scientific method (e.g., psychology, biology, chemistry, physics, mathematics, engineering, computer science). Only 7% reported scientific training during graduate school. Thirteen percent of the attorneys who responded reported that they had received training on the scientific method through continuing legal education (CLE) classes. In their open ended response to what the CLE covered, 10 of the 12 reported that the education covered the presentation and evaluation of DNA. Twelve attorneys reported that they had seen the Federal Judicial Center's reference manual, however, attorneys who said they had seen the Federal Judicial Center's reference manual did not report high familiarity with contents of the manual (M = 3.58 on a 1 (not at all familiar) to 7 (very familiar) scale).

Based on these responses, I categorized attorneys into two groups: those attorneys who majored in a scientific field during their undergraduate education or reported graduate training in the scientific method were labeled "scientifically trained" and those who lacked these criteria were labeled "scientifically untrained." I did not include whether attorneys had seen the Federal Judicial Center's reference manual as part of the training variable because even attorneys who had "seen" the manual did not rate high familiarity with the contents of the manual. Similarly, I did not include those attorneys who reported training from CLE because the focus of the CLE appeared to be almost entirely on evaluating and presenting DNA evidence and not about evaluating scientific methods broadly. Thus, 18% percent of attorney respondents (16 of 88) were classified as scientifically trained. I intended to create a natural group variable to examine whether scientific training improved attorneys' abilities to appropriately judge scientific quality; however the small sample of attorneys classified as scientifically trained did not allow for stable or meaningful comparisons and therefore I did not pursue these analyses.

Motions to Exclude

Overall, 55% of the attorneys reported that they would move to exclude the expert's testimony. The reliability and validity of the intelligence test did not impact whether attorneys would move to exclude the testimony, χ^2 (2) = 4.63, p =.09,

$\Phi = .22$. It is possible that I was unable to detect the effects of motions to exclude due to the dichotomous measure and nature of the chi-square comparison or because I lacked sufficient power and needed more attorneys to detect the difference. Indeed, 70 % of the attorneys reported that they would move to exclude the testimony when the intelligence test administered was unreliable, whereas only 48% of the attorneys would move to exclude the test that was valid and reliable and only 45% of the attorneys would exclude the test that was invalid but reliable. To examine whether reliability affected whether attorneys would move to exclude the testimony, I conducted a two group chi-square by creating a new variable (scientific reliability: reliable conditions v. unreliable condition). Here, scientific reliability was associated with whether attorneys would move to exclude the testimony, χ^2 (1) = 4.57, p =.03, Φ = –0.22, [$CI_{.95}$ = -0.41, 0.004]. Attorneys reported they would move to exclude the testimony 70% of the time when the intelligence test was unreliable and only 47% of the time when the intelligence test was reliable. To examine whether validity affects whether attorneys would move to exclude the testimony, I conducted a two group chi-square by creating a new variable (scientific validity: valid conditions v. invalid condition). Scientific validity was not associated with whether attorneys would move to exclude the testimony, χ^2 (1) = 1.70, p =.19, Φ = 0.13, [$CI_{.95}$ = -0.09, 0.34]. Attorneys moved to exclude 45% of the time when the intelligence test was invalid and 59% of the time when the intelligence test was valid. There was also a significant negative correlation between attorneys' perceptions of scientific quality and attorneys' dichtomous judgments of whether they would move to exclude the testimony, r = -.21, [$CI_{.95}$ = -0.32, 0.08], p = .04. Thus, as perceptions of scientific quality went up, the likelihood that attorneys would move to exclude the testimony decreased.

A full summary of attorneys' justifications for their decisions about moving to exclude the expert's testimony can be found in Table 5. I ran chi-squares between justifications and move to exclude (yes or no). Attorneys were more likely to move to exclude the expert's testimony when they felt the expert had vested interest in the outcome of the case, when they felt the testimony of the expert would be harmful to their side, and when they believed the intelligence test was generally not valid. Attorneys were less likely to move to exclude the expert's testimony when they felt the testimony was necessary for jurors to render an appropriate decision, when they were unable to think of a legal reason they could exclude the testimony, and if they felt they could do more damage by permitting the testimony and attacking the expert during cross-examination.

Although no one particular legal justification was signficantly associated with whether attorneys would move to exclude the expert's testimony or not, attorneys seemed to provide more legal justifications when opting to exclude testiomony. Thus, I created a new variable (Legal Justification), which indicated whether attorneys provided any of the legal justificatons decisions (1 = yes; 0 = no) in their justification or not. Attorneys provided more legal justifications when opting to move to exclude the testimony (N = 20) than when they choose not to move to exclude the testimony (N= 9), $\chi^2(1, N = 87) = 4.52$, $p = .03$, $\Phi = .23$, [$CI_{.95} = -0.01, 0.42$]. Scientific quality also affected how often attorneys provided legal justifications, $\chi^2(2, N = 87) = 6.47$, $p = .04$, $\Phi = .27$. Follow-up chi-squares revealed that attorneys were more likely to provide legal justifications when the testimony was valid. Specifically, attorneys provided more legal justifications when the testimony was valid/reliable (M = 44%) than when the testimony was invalid/reliable

(M = 16%), $\chi^2(1, N = 58) = 5.58$, $p = .02$, $\Phi = -.31$, [$CI_{.95} = -0.53, -0.02$]. Attorneys were also more likely to provide legal justifications when the testimony was unreliable (M = 41%) than when the testimony was invalid/reliable (M = 16%), $\chi^2(1, N = 60) = 4.71$, $p = .03$, $\Phi = -.28$, [$CI_{.95} = -0.50, 0.01$]. Attorneys were equally likely to provide legal justifications in the valid/reliable and unreliable conditions, $\chi^2(1, N = 56) = 0.05$, $p = .82$, $\Phi = -.03$, [$CI_{.95} = -0.32, 0.26$].

I also created a new variable that collapsed across all trial strategy subordinate categories. Attorneys provided more trial strategy justifications if they stated they would not move for exclusion (N = 26) than if they chose to move for exclusion (N = 15), $\chi^2(1, N = 87) = 9.81$, $p = .002$, $\Phi = -.34$, [$CI_{.95} = -0.54, -0.10$]. Scientific quality did not impact the likelihood that attorneys provided a trial strategy justification, $\chi^2(2, N = 85) = 0.50$, $p = .78$, $\Phi = .07$.

Of the 7 attorneys that mentioned the validity issue of administrator bias in their justification for exclusion or inclusion, all 7 read that the expert administered the invalid but reliable test. This association between condition and justification was statistically significant, $\chi^2(2, N = 87) = 13.75$, $p = .001$, $\Phi = .40$. Of the 14 attorneys who reported that the intelligence test was not reliable, three read that the expert had administered a valid and reliable intelligence test (n = 27), three read that the expert had administered the invalid but reliable intelligence test (n=31), and eight read that the expert had administered the unreliable test (n=29). Although more attorneys cited reliaiblity concerns in the appropriate condition, the association between condition and reliability concerns was not signficant, $\chi^2(2, N = 87) = 4.28$, $p = .12$, $\Phi = .22$. Finally, 10 attorneys argued generally that the intelligence test was not valid in their reason to move to exclude. Of these 10 attorneys, three read that the expert administered a valid and reliable intelligence test (n=27), five read that the expert administered an unreliable intelligence test (n=29), and two read that the expert administered an invalid but reliable intelligence test (n=31). This association between condition and justification was not significant , $\chi^2(2, N = 87) = 1.72$, $p = .42$, $\Phi = .14$.

Table 2
Frequency of Attorneys' Justifications for Moving or Not Moving to Exclude the Expert's Testimony

	Motion to Exclude					
Justification	Yes	No	χ^2	*p*	Φ	Confidence Interval
Legal						
Relevance (Rule 402)	2	2	0.01	.91	.01	-0.003, 0.03
Prejudicial v. Probative (Rule 403)	3	2	0.11	.74	.04	-0.18, 0.20
Qualifications (Rule 702)	9	3	2.74	.10	.17	-0.07, 0.32
Helpfulness (Rule 702)	1	1	0.01	.93	.01	-0.15, 0.14
Hearsay (Rule 802)	1	0	0.90	.34	.10	-0.10, 0.10
Frye/Daubert	3	0	2.77	.10	.18	-0.08, 0.18
General acceptance	4	3	0.06	.81	.03	-0.19, 0.21
Bias	6	0	5.74	.02*	.26	0.001, 0.25
Reliability						
Test is not reliable	9	5	0.87	.35	.10	-0.14, 0.29
Test is reliable	1	3	1.31	.25	-.12	-0.23, 0.11
Validity						
General Statement: Test is valid	0	2	2.30	.13	-.16	-0.16, 0.08
General Statement: Test not valid	10	0	10.07	.002*	.34	0.11, 0.34
External Validity	5	2	1.05	.31	.11	-0.13, 0.25
Internal validity/Admin bias	5	2	1.05	.31	.11	-0.13, 0.25
Peer review/Publication	3	0	2.77	.10	.18	-0.08, 0.18
Trial Strategy						
Likelihood motion would succeed	3	3	0.03	.86	-.02	-0.22, 0.18
Harmful to case	8	0	7.85	.005*	.30	0.06, 0.30
No regal reason to exclude	0	4	4.82	.03*	-.24	-0.24, 0.03
Too much to fight battle	0	2	2.30	.13	-.16	-0.16, 0.08
Attack expert on cross-examination	2	11	8.62	.003*	-.32	-0.42, -0.07
Counter with opposing expert	2	3	0.35	.55	-.06	-0.23, 0.15
Testimony necessary	1	6	4.55	.03*	-.23	-0.31, 0.02

Note. For all chi-squares, degrees of freedom = 1 and N = 87 (n exclude = 52, n not excluded = 43).

Evaluations of Scientific Evidence

I subjected data from attorneys who heard expert testimony to a one-way (Scientific Quality: Valid/Reliable, Valid/Unreliable, Invalid/Reliable) ANOVA on Scientific Quality. Attorneys average rating of scientific quality overall was 3.91 on a scale of 1-7. Attorneys rated scientific quality differently as a function of the validity and reliability of the intelligence test, F (1, 92) = 3.66, p = .03, η_p^2 = .07. Specifically, attorneys rated the unreliable intelligence test to be of lower scientific quality (M = 3.57, *SD* = 0.87) than either the valid and reliable test (*M* = 4.05, *SD* = 0.89), *p* = .03, *d* = 0.55 [$CI_{.95}$ = 0.34, 0.76] or the invalid but reliable test (*M* = 4.12, *SD* = 0.90), *p* = .02, *d* = 0.63 [$CI_{.95}$ = 0.42, 0.85]. Attorney ratings of scientific quality did not differ between the valid and reliable and invalid but reliable intelligence tests, *p* = .76, *d* = 0.08 [$CI_{.95}$ = 0.00, 0.30].

Attorneys' Ratings of Decision-Maker Ability to Identify Scientific Flaws

Attorneys rated how well they believe judges, attorneys, and jurors can identify flaws in scientific research. I subjected responses to these three questions to a 3 (Decision Maker: Judge, Attorney, Juror) X 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) repeated measures ANOVA with decision maker as the within-subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (2, N= 182) = 73.02, *p* < .001, therefore multivariate tests are reported (ε = .64), *F* (1.29, 116.99) = 58.86, *p* < .001, η_p^2 = .39. Attorneys reported significant differences in their perception of judges', attorneys', and jurors' ability to detect scientific flaws. Specifically, attorneys reported a belief that jurors are less able to identify flaws in scientific research (*M* = 3.18, *SD* = 1.68) than are judges (*M* = 4.34, *SD* = 1.47), *p* < .001, *d* = 0.74 [$CI_{.95}$ = 0.51, 0.96] or attorneys (*M* = 4.48, *SD* = 1.44), *p* < .001, *d* = 0.84 [$CI_{.95}$ = 0.61, 1.06]. Attorneys also rated their ability to identify threats as significantly higher than judges, *p* = .04, *d* = 0.10 [$CI_{.95}$ = 0.00, 0.30].

I also asked attorneys how well their legal training has prepared them to adequately assess issues of scientific reliability and validity on independent 7–point Likert-type scales. I subjected these data to a 2 (Threat: Validity, Reliability) X 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) repeated measure

ANOVA with threat as the within-subjects factor. There was a significant effect of threat, *F* (1, 88) = 42.88, *p* < .001, *d* = .64, [$CI_{.95}$ = 0.41, 0.87]. Attorneys stated their legal training had better prepared them to evaluate scientific reliability (*M* = 4.12, *SD* = 1.72) than scientific validity (*M* = 3.11, *SD* = 1.47).

Attorney Perceptions of Safeguard Effectiveness

Attorneys also provided their opinions on whether cross-examination and opposing experts are effective methods to educate jurors about problems with scientific evidence. I subjected these data to a 2 (Safeguard: Cross-examination vs. Opposing Expert) X 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) repeated measure ANOVA with safeguard as the within-subjects factor. Attorneys rated both cross-examination (*M* = 5.32, *SD* = 1.39) and opposing experts (*M* = 5.50, *SD* = 1.18) as equally effective safeguards for educating jurors about scientific quality, *F* (1, 91) = 1.56, *p* = .22, *d* = .14, [$CI_{.95}$ = -0.04, 0.33].

Attorneys also rated whether cross-examination and opposing experts help to reveal the reliability and validity of testimony. I subjected these data to a 2 (Safeguard: Cross-examination vs. Opposing Expert) X 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) repeated measure ANOVA with safeguard as the within-subjects factor. There was a significant effect of safeguard, *F* (1, 92) = 3.85, *p* = .05, *d* = .06, [$CI_{.95}$ = -0.10, 0.22]. Attorneys judged cross-examination to be more effective than opposing experts at helping to reveal the reliability and validity of testimony (cross-exam *M* = 5.62, *SD* = 1.11; opposing experts *M* = 5.36, *SD* = 1.23).

Cross-examination Questions

Attorneys generated two sets of five cross-examination questions. To assess whether attorneys generate questions addressing reliability and validity without prompting or whether they would simply attack the expert's qualifications, I first asked attorneys to craft five cross-examination questions without providing any more specific instructions. Recall that under each category an attorney could receive a 1 (they posed a question in this category within their 5 questions) or 0 (they did not pose a question in this category within their 5 questions). Table 3 presents the frequency of attorneys who self-generated cross-examination questions in each category by condition along with the sample size for each condition to determine frequencies. The table also includes chi-square values comparing each intelligence test with justification. For more meaningful comparisons, I collaped across subcategories of credibility, expert's case specific knowledge, and expert qualifications due to the low frequency with which these categories appeared in the attorneys cross-examination questions.

As predicted, attorneys who read that the expert administered an invalid but reliable intelligence test posed significantly more cross-examination questions addressing experimenter bias, the relevant validity flaw, than did those who read that the expert administered a valid intelligence test, irrespective of its reliability. As predicted, attorneys posed more questions to attack the credibility of the expert when the expert administered a valid and reliable intelligence test. Attorneys also asked more questions broadly concerning reliability when the intelligence test was valid and reliable. Further, attorneys asked more questions concerning the broad topic of validity when the intelligence test was valid and reliable than when the test was unreliable, $\chi^2(1, N = 53) = 5.57$, $p = .02$, $\Phi = -.32$, [$CI_{.95} = -0.42, -0.002$]. Attorneys also posed more general validity questions when the test was invalid but reliable as compared to unreliable, $\chi^2(1, N = 53) = 9.73$, $p = .002$, $\Phi = .43$, [$CI_{.95} = 0.12, 0.52$]. The number of general validity questions attorneys posed did not differ between the invalid but reliable or valid and reliable intelligence test conditions, $\chi^2(1, N = 52) = 0.79$, $p = .38$, $\Phi = .12$, [$CI_{.95} = -0.18, 0.40$].

Table 3

Frequency of Attorneys Self-Generated Cross-examination Questions by Condition

	Condition					
Cross-examination Goal	Valid/ Reliable n = 29	Valid/ Unreliable n = 32	Invalid/ Reliable n= 29	χ^2	*p*	Φ
Credibility	4	0	0	10.55	.03*	.34
Check Knowledge	0	4	2	3.82	.15	.15
Qualifications/ Background	7	3	3	3.27	.16	.19
Reliability	4	0	0	8.59	.01*	.33
Reliability I.Q. Scores	10	5	4	4.47	.11	.24
Reliability of this I.Q. test	7	6	5	0.50	.78	.08
Error rate	3	3	3	0.003	.998	.01
Test/Retest Reliability	0	0	2	4.18	.12	.23
Inter-observer Reliability	1	0	3	3.79	.15	.22
Validity	7	1	10	9.47	.009*	.35
Psychology not valid	1	1	1	0.001	1.00	.00
I.Q. testing not valid	0	4	0	8.12	.02*	.32
External Validity						
Participants generalize	6	9	4	2.36	.31	.17
Internal Validity	1	0	1	1.07	.59	.12
Experimenter Bias	2	2	10	11.43	.003*	.38
Subjective Rating	1	0	2	2.15	.34	.17
Unpublished/ peer review	0	5	0	10.28	.006*	.36

**Note*. For all chi-squares, degrees of freedom = 2, N = 90 (n of Valid/Reliable = 29, n of Unreliable = 32, n of Invalid/Reliable = 29).

**Validity effects are explained within text because not all comparisons are significant

I also asked attorneys to craft five cross-examination questions that might help them to assess the validity and reliability of the intelligence test administered to the plaintiff. I was most interested in whether attorneys would generate questions addressing reliability and validity and whether these questions would address the relevant validity and reliability threats when threats were present. Table 4 presents the frequency with which attorneys' cross-examination questions fell into different categories by condition. Attorney questions were not strictly limited to categories concerning the reliability and validity of the expert's testimony. Attorneys asked more questions inquiring about the publication and peer review status of the expert's work when the intelligence test administered by the expert was valid and unreliable, otherwise no comparisons were significant.

Table 4

Frequency of Attorneys Scientifically Informed Cross-examination Questions by Condition

	Condition					
Cross-examination Goal	Valid/ Reliable (n = 26)	Valid/ Unreliable (n = 24)	Invalid/ Reliable (n= 26)	χ^2	*p*	Φ
Credibility	7	8	4	2.54	.28	.19
Check Knowledge	3	1	2	0.95	.62	.11
Qualifications/ Background	0	0	0			
Reliability	2	1	2	0.31	.86	.06
Reliability I.Q. Scores	2	1	2	0.31	.86	.06
Reliability of this I.Q. test	8	6	7	0.25	.88	.06
Error rate	7	4	3	2.28	.32	.17
Test/Retest Reliability	1	0	2	1.86	.39	.16
Inter-observer	2	0	5	5.37	.07	.26
Validity	5	4	3	0.70	.71	.10
Psychology not valid	0	0	0			
I.Q. testing not valid	1	2	0	2.36	.31	.18
External Validity						
Participants generalize	4	4	5	0.09	.95	.04
Internal Validity	1	0	0	1.99	.37	.16
Experimenter Bias	2	1	3	0.85	.65	.11
Subjective Rating	2	1	1	0.50	.78	.08
Unpublished/ peer review	2	8	2	8.35	.02*	.33

Note. For all chi-squares, degrees of freedom = 2 and N = 76. (n of Valid/Reliable = 26, n of Unreliable = 24, n of Invalid/Reliable = 26).

CHAPTER 10: STUDY ONE DISCUSSION

When the *Daubert* Court made judges the gatekeepers of scientific evidence, judges' responsibility to be smart consumers of scientific methods increased but the ramifications of *Daubert* are not limited strictly to judges. If judges fail to recognize flawed science and admit it into evidence, attorneys are the next line of defense against the influence of invalid research on juror judgments. Attorneys may highlight scientific quality at two points during the trial process. First, attorneys can identify flawed science prior to trial and file a motion to exclude the testimony. Second, attorneys can identify the flawed science in trial and pose scientifically informed cross-examinations to reveal the reliability and validity of the scientific evidence for jurors. Only one study has examined attorney ability to recognize flawed science; in that study, attorneys were not sensitive to variations in scientific quality (Kovera & McAuliff, 2002). This is the first study to examine the ability of attorneys to recognize flaws in scientific reliability. Moreover, this is the first study to find that attorneys are likely to address validity concerns during cross-examination and the first to find that attorneys may be sensitive to issues of scientific reliability.

Evaluations of Scientific Quality

Attorneys' evaluations of scientific quality were negatively correlated with their decisions to file motions to exclude. Specifically, as attorneys' evaluations of scientific quality improved, they were less likely to move to exclude the expert's testimony, suggesting they were less likely to fight the admission of scientific evidence when they perceived the evidence to be of good scientific quality. Although attorneys' evaluations of perceived scientific quality negatively correlated with their decisions to file a motion to exclude the testimony, did actual scientific quality impact whether attorneys would move to exclude the expert's testimony? Nearly half of all attorneys who read the intelligence test was both valid and reliable stated that they would move to exclude the expert's testimony, which suggests that approximately half of attorneys will generally try to exclude evidence that is unfavorable to their case regardless of scientific quality. Scientific validity did not significantly increase the likelihood that attorneys would move to exclude the expert's testimony. Indeed, forty-five percent of attorneys that read that the expert administered an invalid test stated they would move to exclude the expert's testimony compared to the fifty-nine percent that read the expert administered a valid test. Attorneys, however, may be better at excluding unreliable expert testimony because seventy percent of attorneys who read the unreliable test stated they would move to exclude the evidence (as compared to the forty-seven percent who would move to exclude the reliable test). Attorneys also rated the unreliable intelligence test to be of lower scientific quality than either of the reliable tests; validity did not impact ratings of scientific quality. Additionally, attorneys perceived that their legal training had better prepared them to evaluate scientific reliability as compared to scientific validity. Taken together, these findings suggest that attorneys may recognize that research with low reliability should be excluded and thus attorneys may strive to keep unreliable testimony from entering the courtroom; however, these findings also suggest that attorneys may fail to keep invalid science from entering the courtroom.

What else drove attorneys' motions to exclude? Attorneys moved to exclude testimony more often when they believed that: the expert had vested interest in the outcome of the case or was biased in some manner, the expert's testimony would be harmful to their case, and the intelligence test lacked validity. However, general claims of invalidity were not associated with actual validity; attorneys used validity as a justification irrespective of whether the study was valid. When attorneys did express a specific concern about validity, they were most likely to mention concerns about administrator knowledge (bias). All attorneys who identified this specific validity threat read the invalid intelligence test. These data suggest that attorneys may argue against scientific validity as a tactic to keep opposing scientific evidence out of court, but also suggest that judges should pay heed when attorneys identify specific validity threats in their motions to exclude as a proportion of attorneys appear capable of identifying threats to validity, at least the threat of experimenter bias.

Attorneys were less likely to file a motion to exclude the testimony when they felt the testimony would be necessary for jurors to render an appropriate decision or when they were unable to think of a legal reason for why they could exclude the testimony. Attorneys who opted not to move to exclude the expert's testimony also cited various trial strategies for coping with expert testimony during trial. Most notable was attorneys' intent to handle the expert during cross-examination.

In this study, I also asked attorneys to write cross-examination questions that they would pose to the expert if given the opportunity. As predicted, attorneys posed cross-examination questions directed at exposing the specific threat to validity more often when the threat was present than not present. In partial support of my hypothesis, attorneys also posed more questions attacking the expert's credibility when the intelligence test was both valid and reliable, but these attorneys also posed more questions about the broad concepts of reliability and validity. Contrary to my predictions, attorneys who read about the unreliable intelligence test did not ask more questions directed at exposing the particular threats to reliability (internal consistency, test/retest, or inter-observer). Furthermore, attorneys who read about the invalid intelligence test did not pose more questions regarding the overall validity of the test. It is also interesting to note the content of attorneys' questions regarding the error rate and margin of error associated with the test because it appears that many attorneys interpreted the reliability values provided for the internal consistency, test/retest, and inter-observer scores as indicative of values representing the margin of error or error rates. For example, one attorney in the unreliable condition in which reliability indices fell between 0.64-0.73 asked: "This test had nearly a 30 percent error rate, correct?" Similarly, an attorney in a condition in which indices of reliability fell between 0.89 - 0.96 asked: "Do you believe that a test administered to a child that contains an error rate of over 10% is a reliable assessment?" These questions provide insight into attorneys' understanding of reliability. Thus, although attorneys provided lower ratings of scientific quality for unreliable scientific evidence, these questions suggest that attorneys may not fully understand the concept of reliability. In essence, at least in regards to reliability, attorneys may be able to get to the right answer (it is not good because it is unreliable), but they lack understanding of what the indices of reliability mean. This presents the question of whether their motion to exclude unreliable testimony would be successful if they are unable to correctly interpret the issue. Moreover, these questions also suggest that the scale designed to measure attorneys' perceptions of scientific quality may be inadequate for understanding these data because although attorneys provided lower ratings of scientific quality for the invalid test on the measure designed to assess scientific quality, the lower ratings did not measure whether attorneys understood and operationalized scientific reliability accurately. Further research is needed to determine how attorneys operationalize scientific reliability.

In regards to validity, it is important to note that some attorneys appear to possess the skills necessary to pose cross-examination questions that address the appropriate validity concerns. However, the percentage of attorneys who accomplished this goal was small. Specifically, only 34% of the sample that read the invalid intelligence test posed questions that addressed the particular validity concern. Although significantly more attorneys self-generated cross-examination questions to address the threat of administrator bias in the invalid condition than in either valid condition, the overall proportion of attorneys who identified this flaw was low. Thus, it may be a relatively small percentage of attorneys who are equipped with the skills necessary to evaluate scientific validity and craft scientifically informed cross-examination questions.

Attorneys' Perceptions of Judges, Attorneys, Jurors and Safeguards

As predicted, attorneys rated judges and attorneys as better able to identify flaws in scientific testimony than jurors. Moreover, attorneys rated their abilities to identify flaws in scientific testimony as slightly better than judges. Are attorneys better than jurors and judges at evaluating scientific testimony? Jurors seem to struggle with identifying most threats to scientific validity (Kovera et al., 1999; Levett & Kovera, 2008; McAuliff & Duckworth, 2010; McAuliff & Kovera, 2008), with the exception that jurors do appear to recognize missing controls groups (Austin & Kovera, under review, McAuliff & Duckworth, 2010; McAuliff et al., 2009). Similarly judges, at least those without scientific training, struggle to accurately evaluate scientific quality (Kovera & McAuliff,

2000) as do attorneys (Kovera & McAuliff, 2002). Data from the current study may suggest that attorneys are more confident in their ability to reason about scientific quality than is warranted. Although attorneys rated an unreliable intelligence test to be less reliable than a more reliable test, it is unclear from their cross-examination questions if they have a conceptual understanding of reliability issues. Further, the attorneys who read about the unreliable test did not pose cross-examination questions to highlight reliability concerns for jurors. With regards to validity, attorneys did not rate valid science as more valid than invalid science and only 34% of attorneys posed cross-examination questions that identified the relevant validity issue. Before strong conclusions can be made about the ability of attorneys to carry out the intended function of the Court, more research is needed on the ability of attorneys to operationalize reliability. From this study, it seems that attorneys may recognize that we seek indices of reliability that approach the value of one, but they may not understand what is meant by inter-observer reliability, test-retest reliability, or internal consistency. Moreover, this study only examined whether attorneys could identify one threat to validity: experimenter bias. Although a small proportion of attorneys did demonstrate sensitivity to experimenter bias, more research is needed on attorney sensitivity to other factors that contribute to valid scientific methods (e.g., the absence of confounds, issues of statistical conclusion validity, the necessity of random assignment to condition) and their ability to apply these concepts to experimental studies.

Attorneys indicated that they had faith in the ability of cross-examination and opposing experts to educate jurors about scientific quality. Attorneys also provided favorable ratings for their abilities in cross-examination and opposing experts to help reveal the reliability and validity of expert testimony, but attorneys provided higher ratings of cross-examination on this item. Their faith in the effectiveness of these safeguards has mixed empirical support. In a test of opposing experts, Levett and Kovera (2008) found that opposing experts did not sensitize jurors to the absence of a control group or issues relevant to counterbalancing of interview items. In another study, opposing experts sensitized jurors to the absence of a control group; however, missing control groups may be an easier validity threat for jurors to recognize (Austin & Kovera, under review, McAuliff & Duckworth, 2010). In general, studies on the presentation of opposing experts suggest that opposing experts increase skepticism but generally fail to educate jurors about scientific evidence or fail to influence trial decisions enough to influence key decisions. Cross-examination also has mixed support. In one study, cross-examination did not sensitize jurors to issues of construct validity (Kovera et al., 1999). However, like opposing experts, cross-examination was effective as educating jurors about a missing control group (Austin & Kovera, under review) and again, cross-examination failed to influence trial decisions enough to influence verdicts or perceptions of evidence strength. In general, it appears that at least two of *Daubert's* safeguards may only work to educate jurors about scientific quality when the threat is easily understood by those who lack methodological training. If cross-examination and opposing experts are to serve as effective safeguards against the untoward influence of unreliable or invalid science then they must also be able to educate jurors about threats less easily identified. Research should examine techniques that could be incorporated into cross-examination that could serve to educate jurors about more sophisticated threats to scientific validity and reliability. Furthermore, educating jurors about methodological concerns is not a complete remedy against the admission of invalid and unreliable evidence. Jurors also need to adjust trial decisions (e.g., verdict) to reflect the value and weight that should be assigned to an expert's testimony. Thus, we need a better understanding of how jurors assign value to the testimony of an expert witness.

Conclusions

These findings suggest that attorneys will move to exclude scientific testimony when they perceive the scientific evidence to be flawed. Attorneys may work to keep unreliable scientific evidence from entering the courtroom because attorneys recognize the relationship between reliability and scientific quality. However, because attorneys do not seem to draw the link between scientific quality and validity, invalid science may be presented to jurors. In sum, there may be a greater likelihood that jurors will be exposed to invalid (as compared to unreliable) science.

Perhaps this issue is not concerning because from the data it seems that attorneys may be better able to pose scientifically informed cross-examination questions that assess relevant validity issues. If attorneys move to exclude unreliable testimony, perhaps their ability to pose scientifically informed cross-examination questions that assess reliability is unnecessary. Although attorneys may not move to exclude invalid testimony, it appears that at least some attorneys can address issues of validity with cross-examination.

Attorneys expressed faith in their ability during cross-examination to help educate jurors and reveal the reliability and validity of testimony. Because research suggests that jurors struggle with validity evaluations on their own, it is important to understand whether cross-examination is effective for educating jurors about scientific validity. It seems that some attorneys may be capable of formulating scientifically informed cross-examination questions, but will cross-examination function as the Court intends and help jurors distinguish between valid and invalid evidence?

CHAPTER 11: STUDY TWO: THE EDUCATIONAL ABILITY OF CROSS-EXAMINATION OF VALIDITY AND RELIABILITY

Method

Participants

Two hundred fifty-nine community members participated in this study in exchange for $20. Of those participants, one participant did not complete dependent measures because he was called into work during the study, 12 were excluded from the analyses because they were not jury eligible, and another 14 participants were excluded from the analyses because they failed to recall whether the judge or attorney performed the cross-examination of the expert (final N = 232). Fifty-two percent of the participants were women (N = 120) and the sample was racially diverse (39.2% white/non-Hispanic, 29.7% black/non-Hispanic, 10.8% Hispanic, 9.5% Asian, and 10.3% identified as 'Other'). Participants ranged in age from 18 to 67, with a mean age of 37.

Design

This study had a 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) X 3 (Cross-examination: Naïve, Informed by Judge, Informed by Attorney) + 1 (No Expert) factorial design.

Videotaped Simulation of a Civil Trial

I recruited actors and actresses from the Actors Fund Work Program and filmed a videotaped simulation of a civil trial in a laboratory space designed to resemble a courtroom, complete with a judge's bench, a witness box, and a podium for attorneys. When the simulation included expert testimony, the duration of the video was approximately one hour and fifteen minutes, which included opening and closing instructions from the judge, opening statements and closing arguments from the attorneys, and 6 witnesses. Without expert testimony, the video was approximately fifty-five minutes.

The trial testimony was based on an actual case: *Hoffman v. Board of Education* (1978). In the simulation, the plaintiff charged the Board of Education with negligence because the Board placed the plaintiff, in remedial classes based upon a determination by the defendant's certified psychologist that the he had an I.Q. of 74 and state law permitted placement in remedial classes when I.Q. scores were 75 or below. The plaintiff remained in remedial classes until he was 17 years of age. At 17, the plaintiff was transferred an occupational training center for mentally challenged to learn a trade. The plaintiff remained at the occupational training center for a period of one year before he was advised that he could no longer continue with his training because recent intelligence testing suggested he was of average intelligence. During trial, mock jurors heard testimony from the plaintiff's mother as well as the plaintiff who both testified about the effects that his ungraded status has had on his inability to secure employment, make friends, and gain self-confidence. The plaintiff also presented testimony from one of his teachers at the occupational training center who testified about the plaintiff's success in class during the one year that he was a student.

On behalf of the defense, the plaintiff's high school teacher testified about the plaintiff's achievement in school and stated that he was not reading at a level consistent with those students in average classes. The principal at the plaintiff's school discussed the laws that permit placement in remedial classes and opined that his intellectual improvement could be due, in part, to his placement in classes designed to suit his needs and abilities. Further, both defense witnesses discussed the plaintiff's speech impediment and how this affected his ability to relate with other students and make friends. Forty participants heard only these witnesses testify.

The remaining participants also viewed a version of the trial containing expert testimony from the clinician who performed the plaintiff's I.Q. testing in kindergarten. This expert explained that she recommended placement in remedial classes because the plaintiff's score qualified him for remedial classes and because she struggled to understand the plaintiff. Furthermore, the expert discussed the creation of generalized intelligence tests, the validation of the intelligence tests, and the descriptive statistical properties of the test she administered (i.e., mean and standard deviation). The expert also discussed the properties of a "good" psychological test (i.e., tests are good when they are valid and reliable). Subsequently, the expert briefly explained variables that might invalidate a test, explained that reliability deals with consistency of scores, and explained the statistical cut-off scores for evaluating reliability.

Validity and reliability manipulations. I varied the validity and reliability of the intelligence test administered by the expert. During the testimony, the expert stated that she administered the Bellevue Primary Scale of Intelligence for Children to the plaintiff. The test was described by a fictional name because I was concerned with whether I could manipulate perceptions of a widely used I.Q. tests such as the Wechsler Intelligence Scales for Children (WISC), which is said to have excellent reliability and is one of the most widely used tests in clinical practice (Camara, et al., 2000; Groth-Marnat, 2003; Watkins, et al., 1995).

To manipulate validity, I varied whether the expert was aware of the plaintiff's predicted I.Q. prior to the administration of the exam or not. Thus, in the valid version, the clinician tested the intelligence quotients of all the students in kindergarten regardless of their predicted I.Q. as part of standard practice. In the invalid version, the clinician only tested the intelligence quotients of the students whom the school district believed may need remedial classes. Knowledge of the plaintiff's predicted I.Q. score poses a problem with experimenter bias because this knowledge may have interfered with the clinician's administration or scoring of the exam.

To manipulate reliability, I manipulated the reliability of the internal consistency score, the test/retest score, and the inter-observer reliability score to be either of high or acceptable standards. When reliability was high, the indices of reliability fell between the range of 0.90 - 0.96. When reliability was low, the indices of reliability fell between the range of 0.70 – 0.76.

Cross-examination Manipulation. To test the motivation hypothesis, I designed the expert's testimony to emulate either adversarial or inquisitorial courts. In the adversarial conditions, the defense attorney presented the expert as a court appointed expert and proceeded to conduct a direct examination of the witness and the plaintiff's attorney conducted a cross-examination. In the inquisitorial condition, the judge instructed jurors that they would now hear the testimony of an expert in clinical psychology on intelligence testing. Following the judge's introduction of the expert, the judge requested that the expert present an oral report for the court. The information relayed by the expert during her oral report emulated the information elicited from the expert by the defense attorney in the adversarial conditions. Following the expert's report, the judge posed cross-examination questions to the expert. The content of the expert's testimony and the cross-examination questions posed to the expert were constant between the inquisitorial and adversarial conditions. Besides the change in who posed the cross-examination questions (judge vs. attorney), the only remaining difference between these court systems was that in the inquisitorial condition, the judge posed fewer questions to the expert during her oral report because experts provide a report in place of direct examination in the inquisitorial system.

In addition to the adversarial vs. inquisitorial manipulation, the cross-examination of the expert either highlighted the reliability and validity of the I.Q. test administered by the expert (scientifically-informed cross-examination) or it did not (scientifically naïve cross-examination). Specifically, naïve cross-examinations addressed scientific debates about whether we can really capture intelligence with a test, the stability of intelligence as a trait, and the expert's credibility (e.g., questioned whether the expert had ever made a mistake on an assessment, whether other psychologists often disagree with her, the expert's fee for testifying, and whether the expert had lied before).

In addition to these questions, scientifically informed cross-examinations also had the expert elaborate on topics alluded to during direct examination/the expert's report. Importantly, the cross-examiner questioned the expert about the validity and reliability of the intelligence test employed by the expert. Specifically, scientifically informed cross-examinations explained the concepts of internal consistency (i.e., how well scores on a test correlate with one another), test-retest (i.e., the ability to get the same score over multiple administrations of a test), and inter-observer reliability (i.e., the probability that different administrators would achieve the same score on the same person if the test was given by different administrators). Further, scientifically informed cross-examinations addressed how scientific methods can affect validity and connected the concepts of validity and administrator knowledge (presence or absence) to the present administration of the exam. Either the judge or the attorney conducted the scientifically informed cross-examination, however, only attorneys conducted scientifically naïve cross-examinations because it is unlikely that a judge would strictly attack the credibility of an expert appointed by the court without seeking additional information. Attorneys, however, are instructed by trial technique manuals to attack witness credibility during cross-examination (Mauet, 2007). Because we did not fully cross this design, there are three cross-examination conditions: scientifically informed by judge, scientifically informed by attorney, and scientifically naïve attorney.

Dependent Measures

***Voir Dire*.** On this questionnaire, participants provided general demographic information, which assessed juror eligibility and solicited demographic information from the participant (e.g., age, gender, citizenship status, racial/ethnic background, voter registration, felony convictions, and possession of a license/state sponsored identification.

Recognition test. Participants who heard expert testimony completed a true/false memory test questionnaire composed of 16 items. Participants who did not hear expert testimony completed a true/false memory test questionnaire composed of 11 items. These questions assessed participants' ability to recognize information such as whether the plaintiff brought a charge of negligence against the Board of Education, whether the plaintiff's father died as a teenager, what the plaintiff is doing now and what he wishes to do, what I.Q. score he received in kindergarten, how much the expert got paid, which side carries the burden of proof, and the level of proof necessary for the plaintiff to prove their case.

Post-trial questionnaire. Participants rendered a dichotomous liability judgment indicating whether they believed the plaintiff had proven by a preponderance of the evidence that the Board was negligent in placing and keeping the plaintiff in remedial classes. Participants rated their confidence in this liability decision and judged the probability that the Board was negligent in placing the plaintiff in remedial classes using scales that ranged from 0 (not at all confident/probable) to 100% (completely confident/probable). If the participant rendered a verdict of liable, they could award compensatory and punitive damages.

Participants rated the mother and the plaintiff on a series of 7-point bipolar adjectives scales: immoral-moral, respectable-non respectable, intelligent-unintelligent, good-bad, unlikable-likable, trustworthy-untrustworthy, honest-dishonest, sincere-insincere, not believable-believable, convincing-unconvincing, certain-uncertain, not credible-credible, and competent-incompetent. Those participants who heard expert testimony also rated the expert on the series of 7-point bipolar adjectives including one additional scale (qualified-unqualified).

Following the bipolar adjective scales, participants indicated their agreement with 22 statements using seven-point Likert-type scales that ranged from strongly disagree to strongly agree, with higher numbers indicating greater agreement (items recoded indicated with [R]: (1) Other psychologists respect [the expert]; (2) [The expert's] intelligence tool is based on good scientific principles; (3) [The expert's] testimony seemed biased [R]; (4) I did not learn anything about intelligence testing from [the expert's] testimony [R]; (5) [The expert] is not qualified to serve as an expert in this case [R]; (6) [The expert's] testimony is fair; (7) [The expert's] testimony was helpful in reaching my verdict; (8) [The expert's] testimony was difficult to understand [R]; (9) The intelligence test [the expert] administered to the plaintiff was scientifically valid; (10) If a child failed the intelligence evaluation performed by [the expert], he/she is more suitable for remedial classes; (11) [The expert] used an appropriate test of intelligence when testing [the plaintiff]; (12) Psychological intelligence tests are reliable indicators of someone's intelligence; (13) Psychological intelligence tests are created using valid scientific methods; (14) The findings from [the expert's] assessment can be used to understand whether [the plaintiff] should be in remedial classes; (15) [The expert] did not use a valid measure of intelligence in [the plaintiff's] assessment [R]; (16) The intelligence test [the expert] administered to the plaintiff was reliable; (17) Psychologists know enough about intelligence testing to justify allowing them to testify as experts; (18) Expert psychologists should be allowed to testify about intelligence even if psychologists do not entirely agree about what the tests have demonstrated; (19) Psychology is a subjective, unscientific discipline [R]; (20) Expert psychological testimony can improve juror decisions; (21) The cross-examination of the expert damaged my perceptions of the expert's credibility; (22) The cross-examination of [the expert] helped me to evaluate the expert's testimony critically. Participants also answered a series of opinion questions regarding the ability of attorneys and decision makers (i.e., jurors and judges) to identify flaws in scientific research: (1) Jurors are able to identify flaws in scientific research; (2) Attorneys are able to able to recognize flaws in scientific research; (3) Judges are able to recognize flaws in scientific research.

Manipulation checks. Participants answered three multiple choice manipulation check questions to assess whether they noticed who cross-examined the expert, whether the administration suffered from administrator bias, and the reliability indices of the intelligence test administered to the plaintiff by the expert. All manipulation checks asked participants to choose from two possible responses. For the validity check, attorneys clicked either 'yes' or 'no' in response to a question of whether the clinician who tested the plaintiff during kindergarten was aware that plaintiff was expected to be intellectually disabled before administering the exam. For the reliability check, jurors chose between two indices of reliability: 0.70 – 0.76 or 0.90 – 0.96. For the manipulation check to determine whether the participant could recall who questioned the expert, participants chose between the attorney and the judge.

Procedure

Participants responded to an ad posted on Craigslist to participant in a study on jury decision making (see Appendix I). I scheduled participants to participate in groups that ranged in size from 2 – 15 participants and I randomly assigned these sessions to condition. Upon arrival, participants were given a consent form to read and sign (see Appendix J). All participants agreed to participate. Prior to the video, participants completed the *voir dire* questionnaire (see Appendix K). Next, participants watched approximately one hour and fifteen minutes of videotaped trial testimony (fifty-five minutes when the expert did not testify).

After viewing the trial, participants filled out a true/false recognition test of trial testimony. If participants heard the expert testify, they received a recognition test with 16 questions (see Appendix L). If participants did not hear the expert testify, they received a recognition test with 11 questions (see Appendix M). I recoded items when necessary on the test so that a correct answer received a value of 1 and an incorrect answer received a value of 0. I then created a recognition test score by summing the response values and dividing by the number of items on the test (i.e., 16 with expert testimony, 11 without). Scores on this memory test ranged from 45 - 100% with an average score of 87%.

Following the memory test, participants completed the post-trial questionnaire (see Appendix N). In this questionnaire, I asked participants to rate the plaintiff's mother and the plaintiff on a series of 7-point bipolar adjectives. If participants heard the expert testify, they also rated the expert on a series of 7-point bipolar adjectives. I recoded items so that higher numbers always represented more positive evaluations of the witness. Participants judged the credibility of the plaintiff and his mom on a series of 7 bipolar adjective pairs: immoral—moral, respectable—not respectable, intelligent—unintelligent, good—bad, unlikable—likable, certain—uncertain, not credible—credible. For both the plaintiff and his mom, I created scales representing credibility by averaging across these 7 items (Cronbach's α = .83 and .87 for the plaintiff and plaintiff's mom, respectively). Participants judged the expert's credibility on the same 7 bipolar adjective pairs as the plaintiff and the plaintiff's mom and 1 other adjective pair (not qualified—qualified; Cronbach's α = .88). Participants judged the trustworthiness of the plaintiff, the plaintiff's mom, and the expert by rating them on 5 bipolar adjective pairs: honest—dishonest, trustworthy—untrustworthy, sincere—insincere, not believable—believable, convincing—unconvincing. I again created scales by averaging across these five items (Cronbach's α = .88, .90, and .92 for the plaintiff, the plaintiff's mom, and the expert's trustworthiness, respectively).

On this questionnaire, participants also provided perceptions of the expert's testimony by responding to a series of 7-point Likert-type items. Participants provided ratings on three items: (1) Other psychologists respect [the expert]; (2) The expert's testimony seemed biased [R]; and (3) [The expert's] testimony is fair. I averaged these items to create a single index of participants' perceptions of expert bias, with higher numbers representing less bias/more positive evaluations of the expert (Cronbach's α = .72). In addition, participants judged the quality of the intelligence test used by the expert by rating the scientific principles of the intelligence tool administered, the scientific validity of intelligence tests generally, the reliability of psychological intelligence tests generally, the validity of the methods used to create psychological intelligence tests, the appropriateness of the intelligence test used to examine the plaintiff, the validity of the measurement tool used to assess plaintiff, and the reliability of the intelligence test used to assess the plaintiff. I averaged across these 7 items to create a single index of participants' perception of scientific quality, with higher numbers representing more positive evaluations of the expert's study (Cronbach's α = .90). Finally, participants answered questions to assess their perceptions regarding different people's ability (i.e., jurors, attorneys, and judges) to identify flaws in scientific research and three multiple choice manipulation checks to assess whether participants noticed who cross-examined the expert, whether the administration suffered from administrator bias, and the reliability indices of the intelligence test administered to the plaintiff by the expert. At the conclusion of the questionnaire, I paid and debriefed the participants.

Hypotheses

Because expert testimony is influential on jurors' decisions, I predicted that participants who heard expert testimony would render more pro-defense judgments compared to those who did not hear expert testimony. I also predicted that when jurors heard a scientifically naïve cross-examination they would perceive the scientific quality of the valid and reliable, invalid but reliable, and unreliable intelligence tests to be of equal scientific quality. In addition, I predicted that ratings of expert credibility, trustworthiness, and bias would not differ when participants heard a scientifically naïve cross-examination.

If the scientifically informed cross-examination increased jurors' ability to evaluate scientific quality and jurors were equally motivated to process the expert's testimony systematically, I predicted a two-way interaction between scientific quality and cross-examination type. Specifically, I predicted that when jurors heard a scientifically informed cross-examination, they would provide higher ratings of scientific quality when they heard the expert discuss an intelligence test that was valid and reliable than when they heard the expert discuss a test that was invalid but reliable or an unreliable intelligence test. Further, I predicted that jurors would find the Board liable less often when the expert administered a valid and reliable test than when the expert administered an invalid but reliable or an unreliable intelligence test. I also predicted that jurors would provide lower ratings of expert bias and provide higher ratings of expert credibility and trustworthiness when the expert administered a valid and reliable intelligence test than an invalid but reliable or unreliable intelligence test.

Alternatively, scientifically informed cross-examinations may not uniformly motivate jurors to evaluate scientific evidence systematically. Specifically, jurors may feel unmotivated to systematically process testimony elicited by judges (Brekke et al., 1991). Thus, scientifically informed cross-examinations may be more likely to help jurors evaluate scientific quality when the attorney cross-examines the expert. Thus, I predicted an interaction between Scientific Quality and Cross-examination, and post-hoc tests would demonstrate that jurors who heard the plaintiff attorney perform a scientifically informed cross-examination would provide higher ratings of scientific quality for the valid and reliable test than for the invalid but reliable or the unreliable intelligence tests. Moreover, those jurors who heard scientifically informed cross-examinations by the judge or scientifically naïve cross-examinations would provide similar ratings of scientific quality regardless of the validity and reliability of the intelligence test. Jurors who heard a scientifically informed cross-examination by the attorney would also be less likely to find the Board liable in valid and reliable intelligence test conditions than in invalid but reliable or unreliable intelligence test conditions. Finally, I predicted that jurors who heard that the expert administered a valid and reliable intelligence test and received a scientifically informed cross-examination by the attorney would also rate the expert as less biased, more credible, and more trustworthy than if they heard about an invalid but reliable or unreliable intelligence test. Those jurors who heard a scientifically naïve cross-examination or scientifically informed cross-examination by a judge would rate the scientific quality of the intelligence test similarly because they would either lack the motivation or ability to process the information systematically.

CHAPTER 12: STUDY TWO RESULTS

Analytic Strategy

To examine the influence of expert testimony on liability decisions, I conducted a chi-square analysis on the dichotomous liability variable. I subjected the verdicts of those who heard expert testimony to a 3 (Scientific Quality) X 3 (Cross-Examination) logistic regression. Continuous scales and questions were subjected to 3 (Scientific Quality) X 3 (Cross-examination) ANOVAs. Finally, I conducted a 3 (Decision Maker: Judge, Attorney, Juror) X 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) X 3 Cross-examination (Informed Judge, Informed Attorney, Naïve) repeated measures ANOVA with decision maker as the within-subjects factor to assess jurors' perceptions of how well judges, attorneys, and jurors can identify flaws in scientific research.

Manipulation Checks

Manipulation checks were relevant only to participants who heard expert testimony; thus, of the 246 jury-eligible participants only 206 participants completed manipulation checks. Ninety-three percent (192 of the 206 jury-eligible participants viewing expert testimony) passed the manipulation check regarding who questioned the expert during the trial. Participants who failed this manipulation check were not included in the analyses because participants, if alert, should have seen and heard who questioned the expert. Missing this manipulation check suggested that participants were either not paying attention during the video, did not understand basic trial processes (e.g., cross-examination), or could not distinguish between the judge and attorney. Scientific quality, cross-examination, nor the interaction of these variables impacted whether participants passed this manipulation, all *F*s < 1. Participants were significantly more likely to say that the judge performed the cross-examination when the judge performed the cross-examination (84%) than if the attorney performed the cross-examination (2%), χ^2 (1, N = 206) = 150.21, p <.001, Φ = 0.85, [$CI_{.95}$ = 0.76, 0.89].

Scientific quality affected whether participants stated that the expert was aware of the plaintiff's predicted I.Q., F (2, 208) = 41.26, p < .001, η_p^2 = .28. Specifically, participants who heard that the expert administered an invalid intelligence test stated that the expert knew of the plaintiff's predicted I.Q. more (M = .86, SD = .43) than those who heard about the unreliable I.Q. test (M = .27, SD = .43), p < .001, d = 1.38, [$CI_{.95}$ = 1.31, 1.45] and more than those who heard about the valid/reliable I.Q. test (M = .31, SD = .41), p < .001, d = 1.32, [$CI_{.95}$ = 1.25, 1.39]. Participants who heard that the expert administered an unreliable or valid/reliable intelligence test did not differ from one another, p = .57, d = 0.10, [$CI_{.95}$ = 0.03, 0.16]. Thus, eighty percent of participants (153 of 192) passed the validity manipulation check that asked participants to recall whether the expert was aware of the plaintiff's predicted I.Q. prior to administering the exam. I predicted that jurors would not be sensitive to variations in validity unless they received procedural help (i.e., a scientifically informed cross-examination) and predicted that they may be less sensitive to variations in validity when the judge performed the examination. Neither scientific validity, F (2, 182) = 1.92, p = .15, η_p^2 = .02, cross-examination, F (2, 183) = 2.12, p = .11, η_p^2 = .02, nor the interaction of these variables, F < 1 affected participants' ability to pass this manipulation check. I retained data from all participants regardless of whether they passed this manipulation because I would expect that a proportion of actual jurors would also not be sensitive to this manipulation and jurors who missed the manipulation may be representative of those jurors engaging in heuristic processing and thus removing them may reduce my ability to test my hypotheses.

Scientific quality again affected whether participants stated that the reliability indices fell between 0.70 - 0.76 or 0.90 – 0.96, F (2, 209) = 38.44, p < .001, η_p^2 = .27. Specifically, participants who heard that the expert administered an unreliable intelligence test stated that the reliability indices fell between 0.90 - 0.96 less often (M = .16, SD = .43) than those who heard about the valid/reliable I.Q. test (M = .76, SD = .44), p < .001, d = 1.39, [$CI_{.95}$ = 1.32, 1.46] and less often than those who heard about the invalid I.Q. test (M = .64, SD = .44), p < .001, d = 1.11, [$CI_{.95}$ = 1.04, 1.18]. Participants who heard the expert administered a valid/reliable or invalid intelligence test did not differ from one another, p = .12, d = 0.28, [$CI_{.95}$ = 0.20, 0.35]. Thus, 77% of participants (147 of 192) passed the reliability manipulation check, which asked participants to indicate the indices of reliability discussed by the expert. I found a significant main effect of scientific quality, F (2, 183) = 5.60, p = .004, η_p^2 = .06. Participants passed this manipulation less often when the intelligence test was invalid but reliable (M = .64, SD = .41) than when it was unreliable (M = .88, SD = .41), p = .001, d = 0.59 [$CI_{.95}$ = 0.52, .66] or valid and reliable (M = .79, SD = .41), p = .001, d = 0.37 [$CI_{.95}$ = 0.30, .44]. No differences were observed between the valid and reliable and unreliable conditions, p = .26, d = 0.22 [$CI_{.95}$ = 0.15, .29]. Cross-examination did not affect performance on this manipulation check, F (2, 183) = 1.04, p = .36, η_p^2 = .01, nor did the interaction of these variables, F (4, 183) = 1.39, p = .24, η_p^2 = .03. Again, I retained participants regardless of their performance on this manipulation check.

Recognition Test

Participants averaged 87% accuracy on the recognition test. Scores on the recognition test were not affected by scientific quality, cross-examination type, *F*s <1, nor the interaction of these variables, F (4, 180) = 1.51, p = .20, η_p^2 = .03. Furthermore, accuracy on the recognition test was not affected by whether participants heard the expert testify (M = .87, SD = .09) or not (M=.86, SD = .13), t (226) = -0.783, p = .43, d = .10, [$CI_{.95}$ = 0.09, 0.12].

Liability

Experts versus no expert. I conducted a chi-square analysis on verdicts to examine the impact of expert testimony for all 232 participants. Fifty-five percent of the participants held the Board liable for negligence; however, whether the expert testified did not affect liability decisions, χ^2(1, N = 2232) = 1.05, p = .31, Φ = .07.

Expert manipulations. To examine whether scientific quality or cross-examination impacted verdicts, I subjected data from the participants who viewed expert testimony to logistic regressions. No main effects or interactions were significant. Additionally, confidence in verdict did not differ as a function of scientific quality, F < 1, cross-examination, F (2, 182) = 1.20, p = .32, η_p^2 = .01, nor the interaction of these variables, F (4, 182) = 1.37, p = .25, η_p^2 = .03. There was a marginally significant effect of scientific quality on participants' probability of negligence assessments, F (2, 181) = 2.57, p = .08, η_p^2 = .03. Participants provided higher ratings of negligence when the intelligence test was invalid but reliable (M = 63.43, SD = 31.69) than when the intelligence test was unreliable (M = 52.31, SD = 31.71), p =.05, d = 0.35 [$CI_{.95}$ = 0.00, 5.80] or when the test was valid and reliable (M = 52.58, SD = 31.69), p =.06, d = 0.35 [$CI_{.95}$ = 0.00, 5.81]. There was no difference between valid and reliable and unreliable conditions, p = .96, d = 0.01 [$CI_{.95}$ = 0.00, 5.50]. Probability of negligence assessments were not affected by cross-examination, nor its interaction with scientific quality, Fs < 1. Neither scientific quality, cross-examination, nor their interaction impacted participants' monetary awards (i.e., compensatory and punitive damages) Fs < 1.

Perceptions of the Expert and Evaluations of Scientific Quality

Only data from participants who viewed expert testimony (N=192) were included in the analyses relevant to examining participants' perceptions of expert evidence. I subjected data from participants who heard expert testimony to 3 (Scientific Quality: Valid/Reliable, Valid/Unreliable, Invalid/Reliable) X 3 (Cross-examination: Informed by Judge, Informed by Attorney, Naïve) ANOVAs on scientific quality, ratings of expert biasness, rating of expert credibility, and ratings of expert trustworthiness.

Overall, jurors assigned an average score of 4.45 (possible range 1-7) on ratings of scientific quality. There were no main effects or interactions for scientific quality based on my manipulations: scientific quality, F (2, 183) = 2.09, p = .13, η_p^2 = .02, cross-examination, nor the interactions of these variables, both Fs < 1. As a more sensitive test, pairwise comparisons revealed that participants provided lower ratings of scientific quality for the invalid but reliable intelligence test (M = 4.20, SD 1.26) than the unreliable intelligence test (M = 4.64, SD = 1.26), p = .05, d = 0.35 [$CI_{.95}$ = 0.14, 0.57]. Ratings of scientific quality did not differ between the valid and reliable (M = 4.52, SD 1.23) or invalid but reliable (M = 4.20, SD 1.26) intelligence tests, p = .15, d = 0.30 [$CI_{.95}$ = 0.05, 0.47]. Ratings of scientific quality also did not differ between the valid and reliable (M = 4.52, SD 1.23) and the unreliable intelligence test (M = 4.64, SD = 1.26), p = .61, d = 0.30 [$CI_{.95}$ = 0.05, 0.47].

In addition, participants ratings of expert bias did not differ as a function of scientific quality, F (2, 183) = 1.69, p = .19, η_p^2 = .02, cross-examination, or the interaction of these variables, Fs < 1. Additionally, participants did not differentially rate the expert's competence based on cross-examination, F < 1, scientific quality, F (2, 183) = 1.88, p = .16, η_p^2 = .02, or the interaction of these variables, F (4, 183) = 1.42, p = .23, η_p^2 = .03. Participants did rate expert trustworthiness differently as a function of scientific quality, F (2, 183) = 3.55, p = .03, η_p^2 = .04. Follow-up tests revealed that participants rated the expert as more trustworthy when the expert administered an unreliable test (M = 5.18, SD = 1.44) as compared to an invalid but reliable intelligence test (M = 4.51, SD = 1.44), p = .009, d = 0.47 [$CI_{.95}$ = 0.22, 0.72]. Trustworthiness ratings did not differ significantly between the unreliable and valid/reliable conditions (M = 4.93, SD = 1.45), p = .34, d = 0.17 [$CI_{.95}$ = 0.00, 0.43]. They were also the same for invalid but reliable and valid and reliable conditions, p = .34, d = 0.11 [$CI_{.95}$ = 0.00, 0.77]. Neither cross-examination nor the interaction of scientific quality and cross-examination impacted participant ratings of expert trustworthiness, Fs < 1.

Ratings of Decision-Maker Ability and Perceptions of Cross-examination

I asked participants to rate the extent to which they believe that judges, attorneys, and jurors can identify flaws in scientific research. I subjected these responses to a 3 (Decision Maker: Judge, Attorney, Juror) X 3 (Scientific Quality: Valid/Reliable, Invalid/Reliable, Valid/Unreliable) X 3 Cross-examination (Informed Judge, Informed Attorney, Naïve) repeated measures ANOVA with decision maker as the within-subjects factor. Jurors reported significant differences in their perception of judges', attorneys', and jurors' ability to detect scientific flaws. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (2) = 23.34, p < .001, therefore multivariate tests are reported (ε = .89), *F* (1.78, 326.68) = 26.16, *p* < .001, η_p^2 = .13. Specifically, participants reported that judges are better able (*M* = 4.55, *SD* = 1.61) to identify flaws in scientific research than are jurors (*M* = 3.92, *SD* = 1.73), *p* < .001, *d* = 0.38 [$CI_{.95}$ = 0.21, 0.55]. Moreover, participants reported that attorneys are better at identifying flaws (*M* = 4.43, *SD* = 1.57) than are jurors, *p* < .001, *d* = 0.31 [$CI_{.95}$ = 0.15, 0.47]. However, participants did not report differences between judges' and attorneys' abilities to identify flaws, *p* = .12, *d* = 0.07 [$CI_{.95}$ = 0.0, 0.23].

I also asked participants whether the cross-examination of the expert damaged their perceptions of the expert's credibility on a 7-point Likert-type scale from 1 (strongly disagree) to 7 (strongly agree). The scientific quality of the intelligence test administered by the expert impacted whether participants reported that cross-examination damaged their perceptions of the expert's credibility, *F* (2, 182) = 8.08, *p* < .001, η_p^2 = .08. Pairwise comparisons demonstrated that jurors believed the cross-examination was damaging more when the expert presented an intelligence test that was invalid but reliable (*M* = 4.23, *SD* = 1.82) than when the expert presented an intelligence test that was unreliable (*M* = 3.00, *SD* = 1.82), *p* < .001, *d* = 0.68 [$CI_{.95}$ = 0.37, 0.99] or a test that was valid and reliable (M = 3.48, SD = 1.82), *p* = .01, *d* = 0.42 [$CI_{.95}$ = 0.10, 0.73]. Ratings of the damaging effects of the cross-examination did not differ between valid and reliable and unreliable intelligence tests, *p* < .001, *d* = 0.26 [$CI_{.95}$ = 0.00, 0.58]. Cross-examination did not influence participants' ratings, *F* (2, 182) = 1.87, *p* = .18, η_p^2 = .02, nor did the interaction of cross-examination and scientific quality, *F* < 1.

Finally, I asked participants whether the cross-examination helped them to evaluate the expert's testimony on a 7-point Likert-type scale from 1 (strongly disagree) to 7 (strongly agree). Scientific quality, cross-examination, *F*s < 1, nor the interaction of these variables, $F = (4, 183) = 1.92$, $p = .11$, $\eta_p^2 = .04$, affected helpfulness of cross-examination ratings.

Ratings of the Plaintiff and the Plaintiff's mother

I subjected ratings of competence and trustworthiness for both the plaintiff and plaintiff's mother to 3 (Scientific Quality: Valid/Reliable, Valid/Unreliable, Invalid/Reliable) X 3 (Cross-examination: Informed by judge, Informed by attorney, Naïve) ANOVAs. Participants' ratings of the plaintiff's competence, plaintiff's trustworthiness, mother's competence, mother's trustworthiness were not affected by scientific quality, cross-examination, or the interaction of these variables, all *F*s < 1.

CHAPTER 13: STUDY TWO DISCUSSION

This study examined the Court's assumption in *Daubert* that attorneys can educate jurors about scientific quality through the legal proceeding of cross-examination. I applied tenets of dual process models of persuasion (Chaiken et al., 1989; Petty et al., 1984) to examine whether scientifically informed cross-examinations equip jurors with the ability to recognize scientific flaws and adjust important trial decisions (e.g., verdict) to reflect the quality of the scientific evidence. Further, I examined one variable hypothesized to impact motivation to process information systematically (i.e., cross-examiner). I predicted that jurors would be unable to identify the scientific flaws without the assistance of a scientifically informed cross-examination. Further, I also predicted that jurors may be less likely to use information learned during a scientifically informed cross-examination if the testimony was elicited by a judge rather than attorney because jurors may not be as motivated to process testimony from judges (Brekke et al., 1991). Thus, I predicted that jurors may only be able to identify flaws in scientific testimony during attorney conducted scientifically informed cross-examinations.

Past research has demonstrated mixed effects regarding the ability of scientifically informed cross-examinations to educate jurors with evaluating scientific validity (Austin & Kovera, under review; Kovera et al, 1999). Why might scientifically informed cross-examinations only work sometimes? One reason may be the sophistication of the validity threat present in the trial stimulus. Specifically, jurors may be sensitive to missing control groups, which may make this particular validity threat easier for jurors to recognize (McAuliff & Duckworth, 2010; McAuliff et al., 2009) and may also, in turn, increase the ability of a scientifically informed cross-examination to assist with quality evaluation.

I manipulated scientific quality by varying whether the clinician knew the plaintiff's predicted I.Q. prior to evaluating his intelligence (i.e., experimenter bias) and the reliability of the intelligence test on three measures: the test/retest, internal consistency, and the inter-observed reliability. Past research has examined the ability of jurors to recognize issues of internal validity, but no research had yet examined juror' sensitivity to values of reliability.

In my study, participants scored reasonably high on all measures designed to assess whether participants were attending to information within the trial (i.e., recognition memory test and manipulation checks). Equivalent performance on the recognition memory test between manipulations suggested that most jurors attended to information in all conditions equally well. Although our manipulation checks suggested that jurors were generally able to recognize our manipulations of validity and reliability, scientifically informed cross-examinations did not help jurors to evaluate the validity and reliability of the intelligence test administered by the expert regardless of whether the judge or attorney performed the cross-examination. This finding aligns with results from Kovera et al. (1999) and suggests that cross-examination may only be effective against unreliable scientific evidence when the threat is easily recognizable and understood by jurors (e.g., a missing control group). If we consider this finding along with jurors' perceptions regarding their inability to evaluate scientific quality (jurors provided low ratings in their ability to identify flaws in scientific methods) juror insensitivity to scientific validity and reliability may come as no surprise. Jurors do, however, have faith that attorneys and judges are better able to identify flaws, despite the fact that judges and attorneys do not typically receive scientific training.

This research did not support my predictions regarding the effectiveness of scientifically informed cross-examinations. Jurors remained insensitive to variations in experimenter bias and reliability, irrespective of whether the judge or the attorney highlighted these concepts and their relationship to scientific quality. This research provides further evidence that cross-examinations may not provide jurors with the ability to evaluate complex threats to scientific validity and reliability. If scientifically informed cross-examinations are only able to educate jurors about simple threats to validity, the safeguard fails to function effectively and does not meet the court's intent as a safeguard against the admission of unreliable expert testimony.

It is, of course, possible that my scientifically informed cross-examination was simply ineffective at educating jurors about scientific validity. Potentially, jurors can be trained to reason about science with cross-examination, but my particular operationalization of a scientifically informed cross-examination may have been ineffective at helping jurors to evaluate evidence scientifically. During direct examination, the expert explained that "good" psychological tests are both valid and reliable and simply introduced the topics of reliability and validity to the jury. On cross-examination, the expert elaborated on these issues by responding to questions about the validity and reliability of the intelligence test employed by the expert. Furthermore, the expert made a direct link between the concepts of validity and reliability and the intelligence test during this cross-examination. Specifically, the expert stated whether the intelligence test was of excellent or average reliability and explicitly admitted that the test could be invalid due to experimenter bias because she only tested the I.Q. of children the school officials thought might be in need of remedial classes or she stated that the test could not be invalid due to experimenter bias because she did not know the students' predicted intelligence before she administered the test and examined all students regardless of their predicted intelligence. It is unclear what else could be incorporated during cross-examination to provide jurors with the tools necessary to evaluate methodological concepts within the confines of a cross-examination.

In addition to my manipulations not influencing evaluations of scientific quality, the validity and reliability of the intelligence test did not influence verdicts. This simulation generated nearly 50/50 verdicts split (55% held the Board liable; 45% did not). Accordingly, there was room for movement in verdicts and yet neither of my manipulations, nor the presence of the expert affected participants' verdicts.

Caveats

Certain limitations must, of course, be addressed before generalizing these findings to an actual courtroom experience. To start, we examined individual juror judgments rather than post-deliberation jury judgments. Although research suggests that jurors rarely discuss expert testimony during deliberations (Brekke & Borgida, 1988; Kovera, Gresham, Borgida, Gray, & Regan, 1997) and pre- and post-deliberation judgments tend to be similar (Hastie, Penrod, & Pennington, 1983), it is always possible that jurors would have judged scientific quality differently as a function of deliberation.

Additionally, I did not fully cross my experimental design. I manipulated the validity and reliability of the intelligence test employed by the expert; however, I did not include a condition in which the expert discussed an unreliable and invalid test. Perhaps if the intelligence test had suffered from both a validity threat and low reliability then jurors may have provided lower ratings of scientific quality. It is possible that jurors did not rate the invalid or unreliable intelligence test of lower scientific quality than the valid/reliable intelligence test because our participants did not understand the relationship between validity and reliability. Specifically, it is possible that they provided relatively high ratings of scientific quality because each test had at least one good trait and they did not know how to discount the weakness appropriately. Although it is unlikely that jurors would be exposed to testimony that lacked both validity and reliability because a clinician would not likely use a test of questionable reliability and validity, an important question is raised of whether jurors are capable of evaluating scientific quality on multiple dimensions.

Conclusions

In spite of these limitations, the results of this study contribute to our understanding regarding the effectiveness of *Daubert's* safeguards. Although the Court intended for cross-examination to act as a form of scientific training for jurors, it appears that cross-examination may be limited to educating jurors about simple threats to scientific validity. Unfortunately, scientific studies do not only suffer from simple, easily identifiable threats. Because most jurors will lack the scientific training necessary to identify threats without assistance, the utility of the safeguards is a necessity.

However, if the safeguards continue to fall short of the Court's expectation and legal researchers cannot identify an implementable method for educating jurors about scientific quality in Court, then the judge's role as the gatekeeper for admissible evidence (i.e., to keep unreliable evidence from being presented at trial) is crucial. That is, judges may need to be vigilant in evaluating the reliability of scientific evidence before ruling on its admissibility to ensure that only quality science is admitted.

Future research should continue to assess viable methods of training jurors to reason scientifically. To this end, a deeper understanding is needed to understand how jurors evaluate scientific quality and to identify the factors that may be effective in educating jurors. Rarely is science ever "good" or "bad". How do jurors weigh evidence when there are both good and bad components present (e.g., the study is reliable, but lacks validity)? No empirical study is ever perfect, but when scientists choose whether to accept an article for publication or not, they weigh the strengths and limitations of a study's methodology. Can we expect untrained laypeople to do the same? Thus, we need a better understanding of how untrained people evaluate scientific quality. Further, we also need to understand whether jurors can appropriately apply judgments about scientific quality in their evaluations of trial evidence and strength. If we can sensitize jurors to scientific quality, can they then translate those skills into a verdict? In other words, can jurors use validity assessments to improve decision making or can they only be trained to differentiate between valid and invalid science? Again, if jurors are unable to adjust critical trial judgments after identifying substandard expert evidence then additional effort should be concentrated on equipping judges with the skills necessary to keep inferior science out of court.

CHAPTER 14: GENERAL DISCUSSION

In two studies, I examined the utility of the cross-examination safeguard established in the 1993 Supreme Court case of *Daubert v. Merrell Dow Pharmaceuticals, Inc*. In Daubert, the justices expressed faith in the ability of cross-examination to reveal the reliability of testimony for jurors. For cross-examination to function as an effective safeguard, attorneys must recognize flaws in expert testimony and craft cross-examination questions that expose the relevant methodological flaws. However, it is not enough for attorneys to simply recognize and expose flaws during cross-examination. Cross-examinations must also function as a form of scientific training for jurors. To examine these issues, I conducted two studies to test the Court's assumptions regarding the utility of cross-examination.

In Study One, 95 attorneys read a summary of the fact pattern in a civil case that contained expert testimony regarding the administration and validation of an intelligence test. I varied the validity and reliability of the test. Attorneys provided lower ratings of scientific quality when the expert administered an unreliable version of the intelligence test. However, attorneys did not craft cross-examination questions designed to expose the unreliability of the scientific test. Thus, attorneys could recognize the issues associated with low reliability, but could not formulate questions to help expose this issue during cross-examination. Attorneys' inability to formulate questions that help expose reliability issues may not be a problem because attorneys can move to exclude expert testimony prior to trial if they recognize the flaw. Accordingly, it follows that attorneys may move to exclude unreliable scientific evidence, protecting jurors from the untoward influence of flawed scientific testimony, and reducing their need to formulate scientifically informed cross-examination questions addressing reliability. Future research should examine the success of attorneys' motions to exclude expert testimony. Although attorneys in my sample recognized the reliably indices as low and rated the scientific quality accordingly, their subsequent cross-examination questions suggest that they have a false understanding of what is meant by reliability.

Attorneys did not provide lower ratings of scientific quality when the intelligence test suffered from a threat to validity; however, a proportion of attorneys did craft cross-examination questions to expose the particular validity threat. Thus, there may be a greater likelihood that jurors will be exposed to invalid (as compared to unreliable) science, but a greater probability that attorneys will expose the threat during cross-examination. Attorneys expressed strong faith in the ability of cross-examination to reveal the reliability and validity of testimony. Do scientifically informed cross-examinations help jurors evaluate scientific quality?

Whether scientifically informed cross-examinations help jurors evaluate scientific quality was the focus of Study Two. New York City community members watched a videotaped trial of a civil case that contained expert testimony. I again varied the reliability and validity of the intelligence test. In addition, I varied whether the cross-examination educated jurors about the study's flaws (scientifically informed) or did not (scientifically naïve). I applied tenants of dual process of models of persuasion (Chaiken, 1980; Petty et al., 1986) to understand the conditions under which jurors may be both able and motivated to process the expert's testimony systematically.

Scientifically informed cross-examinations did not assist jurors with evaluating scientific reliability or validity, irrespective of whether the judge or attorney performed the cross-examination. Previous research demonstrated mixed effects regarding the ability of cross-examination to educate jurors (Austin & Kovera, under review; Kovera et al., 1999). From the available data, it seems that cross-examination may only improve jurors' ability to evaluate scientific evidence when the threat is easily understood (Austin & Kovera, under review; Levett & Kovera, 2009; McAuliff & Duckworth, 2010; McAuliff et al., 2009).

Attorneys' ability to expose methodological issues relevant to validity and reliability may not matter if jurors cannot be trained to evaluate scientific quality through cross-examination. Thus, if scientifically informed cross-examinations do not act as a form of methodological training for jurors, then the judge's role as the gatekeeper for admissible evidence (i.e., to keep unreliable evidence from being presented at trial) becomes significantly more important. However, little empirical research suggests that judges will be able to detect methodological flaws (Gatowski et al., 2001; Kovera & McAuliff, 2000; Manuto & O'Rourke, 1991). Perhaps judges are knowledgeable about what questions to ask of experts that would help to educate them about threats to validity and reliability and thus concerns about the gatekeeping function are irrelevant. Future research should examine judge's ability to identify validity threats and elicit information about reliability and validity during questioning of the expert. If judges cannot elicit this information from experts then the legal field should begin to examine ways for judges and attorneys to receive continuing legal education about scientific evidence. Perhaps law schools could include classes intended to improve judge and attorney ability to reason scientifically or judges could gain scientific training in their continuing legal education courses or manuals (Faigman, Kaye, Saks, & Sanders, 1997; Federal Judicial Center, 1995; Kovera & McAuliff, 2000).

More understanding is also needed in regards to attorneys' abilities to understand concepts of reliability and validity. Are attorneys' motions to exclude scientific evidence because of arguments against validity or reliability successful? Can judges recognize when attorneys make a valid motion that should result in exclusion or do they dismiss motions to exclude because they view the motion as a trial strategy? These questions would better help us understand what role attorneys play in the admission of scientific evidence.

Finally, more research is needed to understand how jurors evaluate scientific quality. Perhaps participants in Study Two provided relatively high ratings of scientific quality because each test was either valid or reliable and jurors did not understand how to appropriately weigh the weakness. Although we know that jurors lack training in scientific methods, we have little understanding how they approach the task of evaluating scientific quality. Future research should also continue to assess viable methods that could train jurors to reason scientifically. This again requires a deeper understanding of how jurors evaluate scientific quality. Answering these research questions will increase our understanding regarding the Court's assumptions in *Daubert* and will help inform the legal field about the need for judges and attorneys to receive scientific training.

Appendix A

John Jay College of Criminal Justice
The City University of New York

I am a doctoral student in Forensic Psychology at John Jay College of Criminal Justice and I, along with my dissertation adviser Dr. Margaret Bull Kovera, am conducting dissertation research on the admission of scientific testimony in courts. This dissertation research is being funded by the National Science Foundation. As part of my research, I am examining the questions attorney and judges may ask to experts to help themselves and jurors evaluate the expert's testimony. Results of this study will serve to inform the psychological and legal communities as to whether attorneys and judges can ask questions that assess scientific quality.

This is an online survey which will take approximately 30 minutes to complete. If you choose to take part in the study, you will be paid $30 for your participation. In addition, we will share the results of our study with you upon its completion. Results of this survey will be distributed as aggregate data; your name will never be associated with the project.

Your survey responses will be entirely anonymous. For your convenience, the study can be accessed online at https://gccunyep.qualtrics.com/SE/?SID=SV_4NEPg4XryThymLH

There is a space at the end of the survey for you to enter your name and the mailing address to which you would like your payment sent. Your name will not be associated with the data – this information is set up to be sent to another database to ensure confidentiality of responses. You must complete the study in one session (so best not to click on the link until you are ready to complete the survey). If you encounter any problems, please send an email to lawlab@jjay.cuny.edu and we will email you a hard copy of the questionnaire to complete.

If you would be willing to provide us with names and email addresses of other civil attorneys or judges who may be interested in participating, please let us know. Any assistance with our recruitment efforts would be greatly appreciated.

If you would like more information about the researchers conducting this study, please email my dissertation adviser, Dr. Margaret Bull Kovera at mkovera@jjay.cuny.edu. If you have any questions about this research, please feel free to contact me at lawlab@jjay.cuny.edu. I appreciate your assistance and would like to thank you in advance for your participation.

Sincerely,

Jacqueline Austin
Doctoral Candidate, Psychology
John Jay College – CUNY
New York, NY 10019

Appendix B

You are invited to participate in a study about courtroom testimony. We ask that you read this form before agreeing to participate in this study.

If you agree to be in this study, you will read a short trial summary that includes expert testimony. After the summary, you will be asked about your perceptions and opinions of the testimony you read. You will answer both closed and open ended questions. The total time required to complete the experiment should be about 30 minutes. You will be one of 180 participants in this study (90 judges, 90 attorneys).

Participation in this study will provide you with an opportunity to learn more about the legal system and research methods in social psychology. You will also be offered $30 for your time. Risks associated with participation are minimal.

Any information obtained in connection with this study will be strictly confidential. In any written reports or publications, no one will be identified or identifiable and only aggregate data will be presented. Research records will be kept in a locked location; only the researcher will have access to the records.

Your decision whether or not to participate is voluntary and will not affect your future relations with John Jay College of Criminal Justice or the researcher involved. If you decide to participate, you are free to discontinue participation at any time without penalty. Moreover, you may choose not to answer any question that is asked of you.

The researcher conducting this study is Jacqueline Austin. You may ask her any questions via email jaustin@jjay.cuny.edu. If you have questions later, you may contact the project director, Dr. Margaret Bull Kovera, at (212) 484-1112 or mkovera@jjay.cuny.edu. If you have any questions about human research subjects' rights, please contact the Institutional Review Board at John Jay College of Criminal Justice at (212) 237-8961.

You may print off a copy of this form if you wish.

You are making a decision whether or not to participate. Please click I consent below to indicate that you have read the information provided above and have decided to participate. You may withdraw at any time without prejudice after signing this form should you choose to discontinue participation in this study.

Appendix C

EVALUATING SCIENTIFIC EVIDENCE: A SURVEY OF JUDGES AND ATTORNEYS

SUMMARY OF THE CASE THE SCIENTIFIC EVIDENCE

The plaintiff, Sam Shay, is charging the Board of Education of the City of New York with negligence for improper intelligence testing and placement in remedial classes. Mr. Shay was classified by a clinical psychologist in kindergarten as borderline retarded with an I.Q. of 74. State laws permit placement in remedial classes when I.Q. scores fall below 75. Recent I.Q. testing of Sam as an adult demonstrates Sam is of average intelligence (an I.Q. of 100). The plaintiff claims that placement in remedial classes significantly crippled his emotional and social development as well as his motivation to achieve. Teachers and the principal at Sam's school report that Sam suffered from a severe speech impediment which made him difficult to understand in class and reported that his speech impacted his ability to relate with other students. Further, they report that Sam was not performing at a level consistent with students in normal classes and argue that he would not have excelled if not for remedial classes.

The defense intends to tender testimony from Dr. Lawrence, a clinical psychologist, who tested Sam's I.Q. at the age of 6. The defense will ask the expert to discuss standardized intelligence testing (e.g., the development of the intelligence test, the average score, what deviations from that score represent). Dr. Lawrence will explain that I.Q. is considered a stable trait is not expected to change over time. The expert will also relay the results of Mr. Shay's test indicating borderline mental retardation.

Dr. Lawrence used to the Bellevue Primary Scale of Intelligence for children to evaluate Sam's intelligence. This test was standardized on 2,200 children between the ages of 6 and 16 years 11 months. The clinicians responsible for standardizing this test examined children that were identified as normal, gifted, children with mild or moderate mental retardation, children with learning disorders, children with ADHD, and children with expressive and mixed receptive-expressive language disorders, autistic disorder, and children with motor impairment. The clinicians who standardized test did not know before administering the exam whether the children they were testing were expected to be gifted, normal, or intellectually disabled. [However/In addition], Dr. Lawrence [was/was not] aware that Sam Shay was expected to be intellectually disabled before administering the exam. The internal consistency score for the Full Scale I.Q. is [.96, .71]. The test-retest reliability score is [.98, .73] for the Full Scale I.Q. and the inter-observer reliability score is estimated at [.89, .64].

Appendix D

Reasons to File (or not file) Motions Coding Scheme

L Law

- L1 Rules of Evidence
  - L1.1 Rule 402—Relevance
  - L1.2 Rule 403 — Prejudicial v. Probative
  - L1.3 Rule 702
    - L1.3.1 Qualifications
    - L1.3.2 Common knowledge/province of jury
    - L1.3.3 Helpfulness
  - L1.4 Rule 704 – Ultimate Issue
  - L1.5 Rule 802 –Hearsay
- L2 Case Law (Frye, Daubert)

GA General Acceptance

B Expert is biased/has interest in results supporting his assessment

NR Reliability — Test is not reliable

R Reliability — Test is reliable

V Validity — Test is valid

NV Validity — Test is not valid

- V1 External (Ecological) Validity
  - V1.1 I.Q. test created on (number, type, etc.) of people
- V2 Internal Validity
  - V2.1 Dr. Lawrence/Doctor not blind to predicted I.Q.
- V3 Unpublished, no peer review

TS Trial Strategy

- TS1 Likelihood that motion will succeed (likely or unlikely)
- TS2 The testimony is harmful
- TS3 Allow so I could
  - TS3.1 Use an opposing expert to attack expert
  - TS3.2 Attack exert during cross-examination
- TS4 No legal reason to exclude
- TS5 Testimony necessary
- TS6 No b/c cost of fighting battle (time, resources, etc).

O Need more information/Other

Appendix E

Instructions: In this section of the questionnaire, please respond to the following questions about the case summary and the expert testimony.

Please pretend or assume the role of plaintiff attorney. Based on what you have read about the case, would you move to exclude the testimony of Dr. Lawrence?

YES NO

What are the most important reasons for your decision?

____
____
____
____
____

What is the probability that you would move to exclude the testimony of Dr. Lawrence?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
not at all probable — completely probable

Instructions: For the following questions, please indicate the number that best reflects your opinion.

The intelligence test Dr. Lawrence administered to the plaintiff was reliable.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

The intelligence test Dr. Lawrence administered to the plaintiff was scientifically valid.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

If a child failed the intelligence evaluation performed by Dr. Lawrence, he/she is more suitable for remedial classes.

1 2 3 4 5 6 7
STRONGLY — STRONGLY

DISAGREE AGREE

Dr. Lawrence's intelligence assessment tool is based on good scientific principles.

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

Dr. Lawrence used an appropriate test of intelligence when testing Sam.

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

Psychological intelligence tests are reliable indicators of someone's intelligence

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

Psychological intelligence tests are created using valid scientific methods.

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

Dr. Lawrence's testimony would prejudice a jury.

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

Dr. Lawrence's testimony is probative.

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

The findings from Dr. Lawrence's assessment can be used to understand whether Sam should be in remedial classes.

1 2 3 4 5 6 7
STRONGLY STRONGLY
DISAGREE AGREE

Expert psychological testimony can improve juror decisions.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Dr. Lawrence's testimony would help jurors render a judgment.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

<u>Jurors</u> are able to recognize flaws in scientific research.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

<u>Attorneys</u> are able to recognize flaws in scientific research.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

<u>Judges</u> are able to recognize flaws in scientific research.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Cross-examination is an effective method for educating jurors about problems with scientific evidence.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Cross-examinations help to reveal the reliability and validity of testimony.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Proffering an opposing expert is an effective method for educating jurors about problems with scientific evidence.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Opposing experts help to reveal the reliability and validity of testimony.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Appendix F

Cross-Examination Topics Coding Scheme

CR Credibility
- CR1 Bias
  - CR1.1 You did the test – you have invested interest
  - CR1.2 How much expert is being paid
  - CR 1.3 Number of times expert testified for defense/plaintiff
- CR2 Competence: How good are you at your job

CK Expert's case-specific knowledge
- CK1 No interview of plaintiff/Sam; how many times did you interview
- CK2 Did you review other records of Sam's (medical, psychological)

GA General acceptance of findings

Q Expert qualifications/Background
- Q1 Education
- Q2 Was not part of initial creation of original I.Q. test
- Q3 Research record
- Q4 Number of times performed test/tested I.Q.

R Reliability
- R1 Questioning reliability of intelligence scores/stability of I.Q.
- R2 Questioning reliability of this I.Q. test
  - R2.1 Error rate/margin of error/confidence intervals
  - R2.2 Reliability of test discussed
    - R2.2.1 Internal Consistency
    - R2.2.2 Test/Retest
    - R2.2.3 Inter-observer

V Validity — Questioning validity of the I.Q. test
- V1 Psychology is not valid/scientific
- V2 Intelligence testing is not valid/scientific
- V3 External (Ecological) Validity/Generalize?
  - V3.1 Participants used to create test (number, type, etc.) (children w/speech impediment generalize)
  - V3.2 Setting (laboratory, interview, etc.) *only if specifically about creation of test
- V4 Internal Validity
  - V4.1 Missing Control
  - V4.2 Dr. Lawrence not blind to Sam's predicted I.Q.
  - V4.3 Subjective ratings of intelligence
- V5 Unpublished, no peer review

O Other

Appendix G

DEMOGRAPHIC CHARACTERISTICS AND LEGAL EXPERIENCE

We ask the following information so we can describe the sample of respondents when we report the results of the research. Under no circumstances will responses of individual judges be reported. Naturally, you are not obligated to respond, but we hope you will do so.

Gender: Male, Female, Undisclosed

Age:

Race/Ethnicity

- Hispanic, regardless of race
- Asian or Pacific Islander
- Black/African American (not of Hispanic origin)
- Native American or Alaskan
- White (not of Hispanic origin)
- Other

In what year did you graduate from law school?

As a practicing attorney, what types of law does your practice emphasize?

Please indicate the number of years you have been in practice __________

In which state do you practice?

Of the cases you've handled in the past 12 months, approximately what percentage involved scientific testimony? (please provide a number) __________

Please consider all of your cases in the past 12 months in which you have attempted to introduce expert scientific evidence. In what percentage of these cases was that expert testimony deemed admissible (please provide a number) __________

Expert testimony can address a variety of different topics. In the space below, please describe the types of expert testimony (e.g., medical evidence, competency evaluations) that have been proffered by you in the past 12 months.

____
____
____

Have you received any additional education designed to help you evaluate scientific methods?
Circle One: YES NO

If yes, please describe the additional education you have received to help you evaluate scientific methods:

____

____

____

Have you seen the Federal Judicial Center's Reference Manual on Scientific Evidence?

Yes No

I'm familiar with the contents of the Reference Manual.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

In college, what was your undergraduate major? _______________

Did you receive any graduate training in the scientific method? Yes No

If yes, please describe this training _________________________________

Have you received continuing legal education on the scientific method? Yes No

If yes, please describe the course__________________________________

My legal training has prepared me with the skills I need to adequately assess issues of scientific validity.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

My legal training has prepared me with the skills I need to adequately assess issues of reliability.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

If you have any additional comments for us, please leave them in the space provided below

Appendix H

Types of Expert Testimony in Past Year Coding Scheme

AC	Accident/traffic reconstruction
ACT	Accounting
AF	Arson/Fire
BA	Ballistics
DR	Drugs/Toxicology
ENV	Environmental
ENG	Engineering (Product Liability, Standards of Construction)
FR	Forensic
	FR1 Fingerprints
MD	Medical
	MD1 DNA
	MD2 Autopsy
PY	Psychological/Psychiatric
	PY1 Counseling
	PY2 Abuse
	PY3 Competency
	PY4 Risk Assessment
	PY5 Insanity
	PY6 Domestic Violence
SF	Safety
STA	Statistical
O	Other

Appendix I

Earn $20 CASH for participation in study on Jury Decision Making! (Midtown West)

We are seeking jury eligible community members to participate in a study on jury decision making. To be eligible for this study you must be jury eligible, which means (1) you are at least 18 years old, AND (2) You are a U.S. Citizen, AND (3) you are not a convicted felon, AND (4) you fall into at least of the following 3 categories: (A) you have a valid Driver's License or (B) you have a valid State Sponsored Identification Card or (C) you are registered to vote. If you do not meet these qualifications, you are not eligible for this study.

The study approximately 2 hours to complete and pays $20 cash. The study requires that you visit John Jay College of Criminal Justice located at 860 11th Avenue. During the study, you will watch a simulated civil trial in groups and be asked to render a verdict. If interested, please send an email to psychlab3@gmail.com and we'll contact you to schedule you for a session!

Appendix J

Jury Decision Making Study

You are invited to participate in a study of jury decision making. I ask that you read this form and ask any questions you have before agreeing to participate in this study.

If you agree to be in this study, you will watch a videotape of trial. After the video, you will be asked to complete a questionnaire about the trial participants and about your attitudes/perceptions regarding the trial. The total time required to complete the experiment should be about 2 hours. You will be one of 260 participants in this study.

Participation in this study will provide you with an opportunity to learn more about the legal system and research methods in social psychology. You will also receive $20 for your time. Risks associated with participation are minimal.

Any information obtained in connection with this study will be strictly confidential. In any written reports or publications, no one will be identified or identifiable and only aggregate data will be presented. Research records will be kept in a locked location; only the researcher will have access to the records.

Your decision whether or not to participate is voluntary and will not affect your future relations with John Jay College of Criminal Justice or the researcher involved. If you decide to participate, you are free to discontinue participation at any time without penalty. Moreover, you may choose not to answer any question that is asked of you.

The researcher conducting this study is Jacqueline Austin. You may ask any questions you have now. If you have questions later, you may contact the project director, Dr. Margaret Bull Kovera, at (212) 484-1112 or mkovera@jjay.cuny.edu. If you have any questions about human research subject's rights, please contact the Institutional Review Board at John Jay College of Criminal Justice at (212) 237-8961.

If you so request, you will be given a copy of this form to keep.

You are making a decision whether or not to participate. Your signature indicates that you have read the information provided above and have decided to participate. You may withdraw at any time without prejudice after signing this form should you choose to discontinue participation in this study.

______________________________ ______________________

Participant Signature Date

Appendix K

Voir Dire Questionnaire

Instructions: In this section of the questionnaire, I would like to ask you a few general background questions about yourself. Recall that all answers to this questionnaire are confidential and anonymous; your name will not be linked to any of the information that you provide today.

1. Gender (Circle One) FEMALE MALE

2. How old are you? ______

3. Are you a U. S. citizen? YES NO

4. Are you registered to vote? YES NO

5. Which of the following characterized your background?
   1. White, non-Hispanic
   2. Hispanic
   3. Black, non-Hispanic
   4. Asian
   5. Pacific Islander
   6. Native American
   7. Other

6. Do you have a Driver's license or other state sponsored identification card?
   1. No
   2. Yes

7. How would you evaluate your political views:
   1. Liberal
   2. Slightly Liberal
   3. Slightly Conservative
   4. Conservative

8. Have you ever served on a jury in a **criminal** trial?
   1. No
   2. Yes

9. Have you ever served on a jury in a **civil** case?
   1. No
   2. Yes

10. Have you ever been convicted of a felony?
    1. No
    2. Yes

Appendix L

True/Fall Recognition Test. Think back to the information presented during the trial. Please answer the questions to the best of your ability.

1) The plaintiff, Sam Shay, is charging the Board of Education with negligence.
TRUE FALSE

2) Rachel Shay's husband, Sam's father, died when Sam was a teenager.
TRUE FALSE

3) Sam Shay works as a messenger now.
TRUE FALSE

4) Sam is interested in making art for living
TRUE FALSE

5) Sam Shay has been applying to art schools since he found out he is not classified as mentally retarded.
TRUE FALSE

6) Sam was assigned an I.Q. of 74 in Kindergarten.
TRUE FALSE

7) Mr. Kuntz, Sam's teacher the Occupational Training Center, immediately thought that Sam should have been in regular classes.
TRUE FALSE

8) Sam's primary school teacher, Mr. Duncan, sent Sam to the school psychologist to have his intelligence retested when Sam was in middle school.
TRUE FALSE

9) Sam's reading score improved while he was in Junior High.
TRUE FALSE

10) Dr. Lawrence is a clinical psychologist.
TRUE FALSE

11) Dr. Lawrence administered the Bellevue Primary Scale for Children to Sam when he was in kindergarten.
TRUE FALSE

12) Sam's score on the intelligence test was 100 when he was in kindergarten.
TRUE FALSE

13) When experimenters have knowledge of experimental conditions, this knowledge may bias any results.

TRUE FALSE

14) The expert got paid $30/hour for her testimony.
TRUE FALSE

15) The burden of proof rests on the plaintiff.
TRUE FALSE

16) The plaintiff must prove his case beyond a reasonable doubt.
TRUE FALSE

Appendix M

True/Fall Recognition Test. Think back to the information presented during the trial. Please answer the questions to the best of your ability.

1) The plaintiff, Sam Shay, is charging the Board of Education with negligence.
TRUE FALSE

2) Rachel Shay's husband, Sam's father, died when Sam was a teenager.
TRUE FALSE

3) Sam Shay works as a messenger now.
TRUE FALSE

4) Sam is interested in making art for living
TRUE FALSE

5) Sam Shay has been applying to art schools since he found out he is not classified as mentally retarded.
TRUE FALSE

6) Sam was assigned an I.Q. of 74 in Kindergarten.
TRUE FALSE

7) Mr. Kuntz, Sam's teacher the Occupational Training Center, immediately thought that Sam should have been in regular classes.
TRUE FALSE

8) Sam's primary school teacher, Mr. Duncan, sent Sam to the school psychologist to have his intelligence retested when Sam was in middle school.
TRUE FALSE

9) Sam's reading score improved while he was in Junior High.
TRUE FALSE

10) The burden of proof rests on the plaintiff.

TRUE FALSE

11) The plaintiff must prove his case beyond a reasonable doubt.

TRUE FALSE

Appendix N

Based on what you heard about the case so far, do you think the plaintiff (Sam Shay) proved that the defendant (Board of Education of the City of New York) was negligent in placing Sam into remedial classes with a preponderance of the evidence (see below). (CIRCLE ONE)

> *A preponderance of the evidence means evidence that persuades you that the plaintiff's (Sam Shay's) claim is more likely to be true than not to be true. If the proof fails to establish any essential part of the plaintiff's claim by a preponderance of the evidence you should find for the defendant (render Board of Education as not liable) as to that claim.*

YES	NO
(Defendant (Board of Education) is liable)	(Defendant (Board of Education) is not liable)

How confident are you about this decision?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

not at all confident — completely confident

What is the probability that the Board of education was negligent in placing Sam Shay into remedial classes?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

not at all probable — completely probable

> ***If you believe that the defendant is not liable (i.e., you answered NO to the first question), turn to the next page. If you believe the defendant is liable (i.e., you answered YES to the first question), answer the following questions before completing the rest of the questionnaire.***

If you found for the plaintiff (Sam Shay) in this case, you may (but you are not required to) award him **compensatory** damages. The purpose of compensatory damages is to make the plaintiff whole—that is, to compensate the plaintiff for the damage that the plaintiff has suffered. Please indicate the amount of money (if any) you would award the plaintiff to compensate him.

$_________________

If you found for the plaintiff (Sam Shay) in this case, you may (but you are not required to) award her **punitive** damages. The purpose of punitive damages is to punish a defendant for shocking conduct, and to deter the defendant and others from engaging in similar conduct in the future. Please indicate the amount of money (if any) you would award the plaintiff to punish the defendant.

$_________________

Next, I would like you to describe your impressions the witnesses. In particular, I would like you to rate some of the potential witnesses using a series of adjective pairs. The scales are designed so that you can express the degree to which the person that you are rating seems to fit one end of the scale or the other. Which space you check should depend on the degree to which the word describes the person you are rating. For example, if you thought that Sam was *slightly* tall, you would mark the item as follows:

Tall :______:______:__X__:______:______:______:______: Short

However, if you thought that Sam was *extremely* short, you should place the "X" next to short:

Tall :______:______:______:______:______:______:__X__: Short

If you have any questions about this task, you may ask them at this time.

Based on the testimony you heard, carefully rate your impressions of ***Rachel Shay*** (the plaintiff's mother) as best you can on each of the following dimensions:

immoral :______:______:______:______:______:______:______: moral

respectable :______:______:______:______:______:______:______ :not respectable

intelligent :______:______:______:______:______:______:______: unintelligent

good :______:______:______:______:______:______:______: bad

unlikable :______:______:______:______:______:______:______: likable

trustworthy :______:______:______:______:______:______:______: untrustworthy

honest :______:______:______:______:______:______:______: dishonest

sincere :______:______:______:______:______:______:______: insincere

not believable :______:______:______:______:______:______:______: believable

convincing :______:______:______:______:______:______:______: unconvincing

certain :______:______:______:______:______:______:______: uncertain

not credible :______:______:______:______:______:______:______: credible

competent :______:______:______:______:______:______:______: not competent

Based on the testimony you heard, carefully rate your impressions of ***Sam Shay*** (the plaintiff) as best you can on each of the following dimensions:

immoral :______:______:______:______:______:______:______: moral respectable :_______:_______:_______:_______:_______:_______:_______: not respectable

intelligent :_______:_______:_______:_______:_______:_______:_______: unintelligent

good :_______:_______:_______:_______:_______:_______:_______: bad

unlikable :_______:_______:_______:_______:_______:_______:_______: likable

trustworthy :_______:_______:_______:_______:_______:_______:_______: untrustworthy

honest :_______:_______:_______:_______:_______:_______:_______: dishonest

sincere :_______:_______:_______:_______:_______:_______:_______: insincere

not believable :_______:_______:_______:_______:_______:_______:_______: believable

convincing :_______:_______:_______:_______:_______:_______:_______: unconvincing

certain :_______:_______:_______:_______:_______:_______:_______: uncertain

not credible :_______:_______:_______:_______:_______:_______:_______: credible

competent :_______:_______:_______:_______:_______:_______:_______: not competent

Based on the testimony that you heard, carefully rate your impressions of ***Dr. Lawrence*** (the expert who gave Sam the intelligence evaluation to assess whether he was suitable for general education) as best you can on each of the following dimensions:

immoral :_______:_______:_______:_______:_______:_______:_______: moral

respectable :_______:_______:_______:_______:_______:_______:_______: not respectable

intelligent :_______:_______:_______:_______:_______:_______:_______: unintelligent

good :_______:_______:_______:_______:_______:_______:_______: bad

unlikable :_______:_______:_______:_______:_______:_______:_______: likable

not qualified :_______:_______:_______:_______:_______:_______:_______: qualified

trustworthy :_______:_______:_______:_______:_______:_______:_______: untrustworthy

honest :_______:_______:_______:_______:_______:_______:_______: dishonest

sincere :_______:_______:_______:_______:_______:_______:_______: insincere

not believable :_______:_______:_______:_______:_______:_______:_______: believable

convincing :_______:_______:_______:_______:_______:_______:_______: unconvincing

certain :_______:_______:_______:_______:_______:_______:_______: uncertain

not credible :_______:_______:_______:_______:_______:_______:_______: credible

competent :_______:_______:_______:_______:_______:_______:_______: not competent

In this section of the questionnaire, I am interested in your reactions to the expert testimony (Dr. Lawrence). Please read each item carefully and circle the NUMBER that best reflects your opinion.

Other psychologists respect Dr. Lawrence.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Dr. Lawrence's intelligence assessment tool is based on good scientific principles.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Dr. Lawrence's testimony seemed biased.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

I did not learn anything about intelligence testing from Dr. Lawrence's testimony.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Dr. Lawrence is not qualified to serve as an expert in this case.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Dr. Lawrence's testimony is fair.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Dr. Lawrence's testimony was helpful in reaching my verdict.

1	2	3	4	5	6	7
STRONGLY DISAGREE						STRONGLY AGREE

Dr. Lawrence's testimony was difficult to understand.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

The intelligence test Dr. Lawrence administered to the plaintiff was scientifically valid.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

If a child failed the intelligence evaluation performed by Dr. Lawrence, he/she is more suitable for remedial classes.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Dr. Lawrence used an appropriate test of intelligence when testing Sam.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Psychological intelligence tests are reliable indicators of someone's intelligence.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Psychological intelligence tests are created using valid scientific methods.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

The findings from Dr. Lawrence's assessment can be used to understand whether Sam should be in remedial classes.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Dr. Lawrence did not use a valid measure of intelligence in Sam's assessment.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

The intelligence test Dr. Lawrence administered to the plaintiff was reliable.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

Psychologists know enough about intelligence testing to justify allowing them to testify as experts.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

Expert psychologists should be allowed to testify about intelligence even if psychologists do not entirely agree about what the tests have demonstrated.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

Psychology is a subjective, unscientific discipline.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

Expert psychological testimony can improve juror decisions.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

The cross-examination of the expert damaged my perceptions of the expert's credibility.

1 2 3 4 5 6 7
STRONGLY DISAGREE — STRONGLY AGREE

The cross-examination of Dr. Lawrence helped me to evaluate the expert's testimony critically.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

<u>Jurors</u> are able to recognize flaws in scientific research.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

<u>Attorneys</u> are able to recognize flaws in scientific research.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

<u>Judges</u> are able to recognize flaws in scientific research.

1 STRONGLY DISAGREE	2	3	4	5	6	7 STRONGLY AGREE

Was Dr. Lawrence aware that Sam Shay was expected to be intellectually disabled before administering the intelligence test?

YES NO

The different indices of reliability fell within which of the following ranges:

0.70 – 0.76 0.90 - 0.96

Who questioned the expert? (Circle One)

Attorney Judge

References

Austin, J. L, & Kovera, M. B. (under review). The effect of inquisitorial versus adversarial cross-examination on juror evaluations of expert evidence validity.

Anderson, R.A., & Otto, A.L. (2003). Perceptions of fairness in the justice system: A cross-cultural comparison. *Social Behavior and Personality, 31*, 557-564. doi:10.2224/sbp.2003.31.6.557

Axsom, D., Yates, S., & Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. *Journal of Personality and Social Psychology, 53,* 30-40. doi:10.1037/0022-3514.53.1.30

Baddeley, A. D., Scott, D., Drynan, R., & Smith, J. C. (1969). Short-term memory and the limited capacity hypothesis. *British Journal of Psychology, 60,* 51-55. doi:10.1111/j.2044-8295.1969.tb01175.x

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior, 19*, 79-88. doi:10.1007/BF01499074

Blandón-Gitlin, I., Sperry, K., & Leo, R. (2011). Jurors believe interrogation tactics are not likely to elicit false confessions: Will expert witness testimony inform them otherwise?. *Psychology, Crime & Law*, *17*(3), 239-260. doi:10.1080/10683160903113699

Brekke, N., & Borgida, E. (1988). Expert psychological testimony in rape trials: A social-cognitive analysis. *Journal of Personality and Social Psychology, 55,* 372–386. doi:10.1037/0022-3514.55.3.372

Brekke, N., Enko, P., Clavet, G., & Seelau, E. (1991). Of juries and court-appointed experts: the impact of nonadversarial versus adversarial expert testimony. *Law and Human Behavior, 15,* 451-475.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116-131. doi:10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in the need for cognition. *Psychological Bulletin, 199*, 197-253. doi: 10.1037/0033-2909.119.2.197

Cacioppo, J. T., Petty, R. E., Kao, C., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, *51*, 1032-1043. doi:10.1037/0022-3514.51.5.1032

California v. Green, 399 U.S. 149, 90 S. Ct. 1930 (1970).

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*, 141-154. doi:10.1037/0735-7028.31.2.141

Cecil, J. S., Hans, V. P., & Wiggins, E. C. (1991). Citizen comprehension of difficult issues: Lessons from civil jury trials. *American University Law Review, 40,* 727-774.

Cecil, J. S., Lind, E. A., & Bermant, G. (1987). Jury service in lengthy civil trials. Washington, DC: Federal judicial center. http://www.fjc.gov/public/pdf.nsf/lookup/jurylnth.pdf/$file/jurylnth.pdf

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

Chaiken, S., Lieberman, A., & Eagly, A.H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J.S. Uleman & J.A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford Press.

Chaiken S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66,* 460-473. doi: 10.1037/0022-3514.66.3.460

Charrow, R. P, & Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review, 79*, 1306-1374.

Collett, M. E., & Kovera, M. B. (2003). The effects of British and American trial procedures on the quality of juror decision-making. *Law and Human Behavior, 27,* 403-422. doi: 10.1023/A:1024037100696

Conway, A. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4,* 577-590. doi:10.1080/741940997

Cooke, D. J., & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior*, *34*, 259-274. doi:10.1007/s10979-009-9176-x

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

Cooper, J., & Neuhaus, I.M. (2000). The "hired gun" effect: Assessing the effect of pay, frequency of testifying, and credentials on the perception of expert testimony. *Law and Human Behavior, 24*, 149-171. doi: 10.1023/A:1005476618435

Cutler, B. L., Dexter, H. R., & Penrod, S. D., (1989). Expert testimony and jury decision making: How do jurors make decisions? *Law and Human Behavior, 20,* 379-394. doi: 10.1002/bsl.2370070206

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

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Cutler, B.L., Penrod, S.D., & Dexter, H.R. (1989). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior, 13*, 311-322. doi:10.1007/BF01067032

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579, 113 S.Ct. 2795 (1993).

Devenport, J. L., & Cutler, B. L. (2004). Impact of defense-only and opposing eyewitness experts on juror judgments. *Law and Human Behavior*, *28*, 569-76. doi:10.1023/B:LAHU.0000046434.39181.07

Dixon, L., & Gill, B. (2002). Changes in the standards for admitting expert evidence in federal civil cases since the *Daubert* decision. *Psychology, Public Policy, and Law*, *8*, 251-308. doi:10.1037/1076-8971.8.3.251

Elwork, A., Sales, B. D., & Alfini., J. J. (1977). Juridic decisions: In ignorance of the law or in light of it? *Law and Human Behavior, 1,* 163-89.

Elwork, A., Sales, B. D., & Alfini, J. J. (1982). *Making jury instructions understandable.* Charlottesville, VA: Michie.

English, P. W., & Sales, B. D. (1997). A ceiling or consistency effect for the comprehension of jury instructions. *Psychology, Public Policy, and Law*, *3*, 381-401. doi:10.1037/1076-8971.3.2-3.381

Faigman, D. L., & Monahan, J. (2005). Psychological Evidence at the Dawn of the Law's Scientific Age. *Annual Review of Psychology*, *56*631-659. doi:10.1146/annurev.psych.56.091103.070316

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*, 31-35. doi:10.1126/science.3291114

Federal Judicial Center. (2000). *Reference manual on scientific evidence* (2nd ed.). Washington, DC: Federal Judicial Center.

Federal Rules of Evidence, 28 U.S.C. (West 1975).

Federal Rules of Evidence Handbook 2000-2001. (2000). Cincinnati, OH: Anderson.

Finkelman, J. M. (2005). Live from the firing line: The practice of management: Juror perception of employment litigation. *Psychologist-Manager Journal, 8*, 45-54. doi: 10.1207/s15503461tpmj0801_5

Flowe, H. D., Finklea, K. M., & Ebbesen, E. B. (2009). Limitations of expert psychology testimony on eyewitness identification. In B. L. Cutler, B. L. Cutler (Eds.) , *Expert testimony on the psychology of eyewitness identification* (pp. 201-221). New York, NY US: Oxford University Press. doi:10.1093/acprof:oso/9780195331974.003.009

Fong, G. T., Krantz, D. H, & Nisbett R. E. (1986). The effects of complainant age and expert psychological testimony in a simulated child sexual abuse trial. *Law and Human Behavior, 10*, 317-32. doi:10.1007/BF01044540

Fox, S. G., & Walters, H. A. (1986). The impact of general vs. specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10*, 215-28. doi: 10.1007/BF01046211

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

Fulero, S. M., & Turner, D. (1997). Using British trial procedures in American cases: A more civil trial? *Law and Human Behavior, 21*, 439-448. doi: 10.1023/A:1024863320673

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior, 25*, 433-458. doi: 10.1023/A:1012899030937

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

Greene, E. (1988). Judge's instruction on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology*, *18*, 252-276. doi:10.1111/j.1559-1816.1988.tb00016.x

Greene, E., Downey, C., & Goodman-Delahunty, J. (1999). Juror decisions about damages in employment discrimination cases. *Behavioral Sciences & the Law*, *17*, 107-121. doi:10.1002/(SICI)1099-0798

Griffith, J. D., Libkuman, T. M., & Poole, D. A. (1998). Repressed memories: The effects of expert testimony on mock jurors' decision making. *American Journal of Forensic Psychology*, *16*, 5-23.

Groscup, J. L., Penrod, S. D., Studebaker, C. A., Huss, M. T., & O'Neil, K. M. (2002). The effects of *Daubert* on the admissibility of expert testimony in state and federal criminal cases. *Psychology, Public Policy, and Law*, *8*, 339-372. doi:10.1037/1076-8971.8.4.339

Groth-Marnat, G. (2003). Handbook of psychological assessment (4th ed.). Hoboken, NJ US: John Wiley & Sons Inc.

Hagen, M. A. (1997). *Whores of the court: The fraud of psychiatric testimony and the rape of American justice.* New York: Harper Collins.

Hamill, R., Wilson, T. D., & Nisbett, R. E. (1980). Insensitivity to sample bias: Generalizing from atypical cases. *Journal of Personality and Social Psychology*, *39*, 578-589. doi:10.1037/0022

Harkins, S. G., & Petty, R. E. (1981). The multiple source effect in persuasion: The effects of distraction. *Personality and Social Psychology Bulletin*, *7*, 627-635. doi:10.1177/014616728174019

Hastie, R., Penrod, S. D., & Pennington, N. (1983) *Inside the Jury.* New Jersey: The Lawbook Exchange, LTD.

Hoffman v. Board of Education, 64 App. Div. 2d 369, 410 N. Y.S.2d 99 (1978).

Hosch, H. M., Beck, E., & McIntyre, P. (1980). Influence of expert testimony regarding eyewitness accuracy on jury decisions. *Law and Human Behavior*, *4*, 287-296. doi:10.1007/BF01040620

Holly, J. L. (1999). Why the *Daubert* standard should apply to both expert opinions based upon "technical" or "other specialized knowledge" and to expert opinions based upon "science" or a "scientific method." *Cumberland Law Review*, *30*, 247-275.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237-251. doi: 10.1037/h0034747

Kassin, S. M., Williams, L. N., & Saunders, C. L. (1990). Dirty tricks of cross-examination: The influence of conjectural evidence on the jury. *Law and Human Behavior*, *14*, 373-384. doi:10.1007/BF01068162

Kovera, M. B., & Cutler, B. L. (2013). *Jury selection*. New York: Oxford University Press.

Kovera, M., Gresham, A., Borgida, E., Gray, E., & Regan, P. (1997). Does expert psychological testimony inform or influence juror decision making? A social cognitive analysis. *Journal of Applied Psychology, 82*, 178-191. doi:10.1037/0021-9010.82.1.178

Kovera, M. B., Levy, R. J., Borgida, E., & Penrod, S. D. (1994). Expert testimony in child sexual abuse cases: Effects of expert evidence type and cross-examination. *Law and Human Behavior, 18*, 653-674. doi: 10.1007/BF01499330

Kovera, M. B., & McAuliff, B. D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: Are judges effective gatekeepers? Journal *of Applied Social Psychology*, 85, 574-586. doi: 10.1037/0021-9010.85.4.574

Kovera, M. B., & McAuliff, B. D. (2002). *Attorneys' evaluations of psychological science: Does evidence quality matter?* Unpublished manuscript.

Kovera, M. B., McAuliff, B. D., & Hebert, K. S. (1999). Reasoning about scientific evidence: Effects of juror gender and evidence quality on juror decision in a hostile work environment case. *Journal of Applied Psychology, 84*, 362-375. doi: 10.1037/0021-9010.84.3.362

Kovera, M. B., McAuliff, B. D., & Russano, M. B. (2002). Assessment of the commonsense psychology underlying *Daubert*: Legal decision makers' abilities to evaluate expert evidence in hostile work environment cases. *Psychology, Public Policy, and Law, 8*, 180-200. doi:10.1037/1076-8971.8.2.180

Krafka, C., Dunn, M. A., Johnson, M., Cecil, J. S., & Miletich, D. (2002). Judge and attorney experiences, practices, and concerns regarding expert testimony in federal civil trials. *Psychology, Public Policy, and Law*, 309-332. doi:10.1037/1076-8971.8.3.309

Kramer, G. P., Kerr, N. L., & Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior*, *14*, 409-438. doi:10.1007/BF01044220

Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, *7*, 267-310. doi:10.1037/1076-8971.7.2.267

Kumho Tire Co. v. Carmichael, 526 U.S. 137,119 S. Ct. 1167 (1999).

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43*, 431-42. doi:10.1037/0003-066X.43.6.431

Leippe, M. R., & Elkin, R. A. (1987). When motives clash: Issue involvement and response involvement as determinants of persuasion. *Journal of Personality and Social Psychology, 52*, 269-278. doi: 10.1037/0022-3514.52.2.269

Levett, L. M., & Kovera, M. (2008). The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence. *Law and Human Behavior*, 32, 363-374. doi:10.1007/s10979-007-9113-9

Levett, L. M., & Kovera, M. B. (2009). Psychological mediators of the effects of opposing expert testimony on juror decisions. *Psychology, Public Policy, and Law, 15*, 124-48. doi:10.1037/a0016309

Lieberman, J. D., & Sales, B. D. (1997). What social science teaches us about the jury instruction process. *Psychology, Public Policy, and Law, 3*, 589-644. doi:10.1037/1076-8971.3.4.589

Lilly v. Virginia, 527 U.S. 116, 119 S. Ct. 1887 (1999).

Mackie, D. M., & Worth, L. T. (1989). Processing deficits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology, 57*, 27-40. doi:10.1037/0022-3514.57.1.27

Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology, 61*, 13-25. doi:10.1037/0022-3514.61.1.13

Manuto, R., & O'Rourke, S.P. (1991). Federal judges' perceptions of social research in judicial decision making. *Communications Reports, 4*, 103-106. http://www.westcomm.org/publications/comm_reports.asp

Martin, R., Hewstone, M., & Martin, P. Y. (2007). Systematic and heuristic processing of majority- and minority-endorsed messages: The effects of varying outcome relevance and levels of orientation on attitude and message processing. *Personality and Social Psychology Bulletin, 33*(1), 43-56. doi:10.1177/0146167206294251

Maryland v. Craig, 497 U.S. 836 (1990).

Mauet, T.A. (2007). *Trial Techniques, 7th ed.* New York: Aspen Publishers, Inc.

McAuliff, B. D. (2009). Judging the validity of psychological science from the bench: A case in point. *Journal of Forensic Psychology Practice, 9*, 310-320. doi:10.1080/15228930902936121

McAuliff, B. D., & Duckworth, T. D. (2010). I spy with my little eye: Jurors' detection of internal validity threats in expert evidence. *Law and Human Behavior, 34*, 489-500. doi: 10.1007/s10979-010-9219-3

McAuliff, B. D., & Kovera, M. (2008). Juror need for cognition and sensitivity to methodological flaws in expert evidence. *Journal of Applied Social Psychology, 38*, 385-408.

McAuliff, B. D., Kovera, M. B., & Nunez, G. (2009). Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? *Law and Human Behavior, 33*, 247-257. doi: 10.1007/s10979-008-9133-0

Melton, G. B. & Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers.* New York: Guildford Press.

Monahan, J., & Walker, L. (1988). Social science research in law: A new paradigm. *American Psychologist, 43*, 465-472. doi:10.1037/0003-066X.43.6.465

Moore v. Ashland Chemical, Inc., 126 F.3d 679 (5th Cir. 1997)

Moore v. Ashland Chemical, Inc., 151 F.3d. 269 (5th Cir. 1998)

Murdock, B. B. (1965). Test of the 'limited capacity' hypothesis. *Journal of Experimental Psychology, 69*, 237-240. doi:10.1037/h0021796

Petty, R. E., & Cacioppo, J. T. (1984). The effect of involvement on response to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology, 46*, 69-81. doi: 10.1037/0022-3514.46.1.69

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-203). New York: Academic Press.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument based persuasion. *Journal of Personality and Social Psychology, 41*, 847-855. doi: 10.1037/0022-3514.41.5.847

Raitz, A., Greene, E., Goodman, J., & Loftus, E. F. (1990). Determining damages: The influence of expert testimony on jurors' decision making. *Law and Human Behavior, 14*, 385-395. doi:10.1007/BF01068163

Ramirez, G., Zemba, D., & Geiselman, R. (1996). Judges' cautionary instructions on eyewitness testimony. *American Journal of Forensic Psychology*, 14, 31-66.

Reifman, A., Gusick, S. M., & Ellsworth, P. C. (1992). Real jurors' understanding of the law in real cases. *Law and Human Behavior, 16*, 539-554. doi:10.1007/BF01044622

Saks, M.J. & Kidd, R.F. (1980). Human information processing and adjudication: Trial by heuristics. *Law and Society Review, 15*, 124-160.

Salerno, J. M., & McCauley, M. R. (2009). Mock jurors' judgments about opposing scientific experts: Do cross-examination, deliberation and need for cognition matter?. *American Journal Of Forensic Psychology*, *27*(3), 37-60.

Sanders, J., Diamond, S. S., & Vidmar, N. (2002). Legal perceptions of science and expert knowledge. *Psychology, Public Policy, and Law*, *8*, 139-153. doi:10.1037/1076-8971.8.2.139

Severance, L. J., Greene, E., & Loftus, E. F. (1984). Toward criminal jury instructions that jurors can understand. *Journal of Criminal Law and Criminology*, *75*, 198-233. doi:10.2307/1143210

Severance, L. J., & Loftus, E. F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. *Law & Society Review, 17*, 153-197. doi:10.2307/3053535

Schuller, R. A. (1992). The impact of battered woman syndrome evidence on jury decision processes. *Law and Human Behavior, 16,* 597-620. doi:10.1007/BF01884018

Schuller, R. A., McKimmie, B. M., & Janz, T. (2004). The impact of expert testimony in trials of battered women who kill. *Psychiatry, Psychology and Law, 11*, 1-12. doi: 10.1375/1321871041336091

Schweitzer, N. J., & Saks, M. J. (2009). The gatekeeper effect: The impact of judges' admissibility decisions on the persuasiveness of expert testimony. *Psychology, Public Policy, and Law, 15*, 1-18. doi:10.1037/a0015290

Smith, S. M., & Petty, R. E. (1996). Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin*, *22*, 257-268. doi:10.1177/0146167296223004

Sue, S., Smith, R. E., & Caldwell, C. (1973). Effects of inadmissible evidence on the decisions of simulated jurors: A moral dilemma. *Journal of Applied Social Psychology*, *3*, 345-353.

Tanford, S., & Cox, M. (1988). The effects of impeachment evidence and limiting instructions on individual and group decision making. *Law and Human Behavior*, *12*, 477-497. doi:10.1007/BF01044629

Thompson, W. C., Fong, G. T., & Rosenhan, D. L. (1981). Inadmissible evidence and juror verdicts. *Journal of Personality and Social Psychology, 40*, 453-463. doi:10.1037/0022-3514.40.3.453

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131. doi:10.1126/science.185.4157.1124

United States v. Addison, 498 F.2d 741, 743 (DC Cir. 1974).

Watkins, C., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*, 54-60. doi:10.1037/0735-7028.26.1.54

Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64*, 440-448. doi:10.1037/0021-9010.64.4.440

5 J. Wigmore, Evidence §1367 (J. Chadbourn rev. 1974).

Wigmore, J. H. (1978). *Evidence in trials at common law* (Vol. 7, Chadbourn rev.). Boston: Little, Brown.

Wissler, R. L., & Saks, M. J. (1985). On the inefficacy of limiting instructions: When jurors use prior conviction evidence to decide on guilt. *Law and Human Behavior, 9*, 37-48. doi:10.1007/BF01044288

Wood, W., & Kallgren, C. A. (1988). Communicator attributes and persuasion: Recipients' access to attitude-relevant information in memory. *Personality and Social Psychology Bulletin*, *14*, 172-182. doi:10.1177/0146167288141017