The impact of institutional authority on forensic evidence evaluation by criminal justice professionals

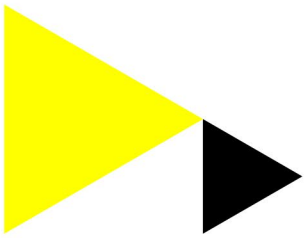
Author(s)
de Roo, Rosanne H. D.; Stevens, Lonneke; de Poot, Christianne J.

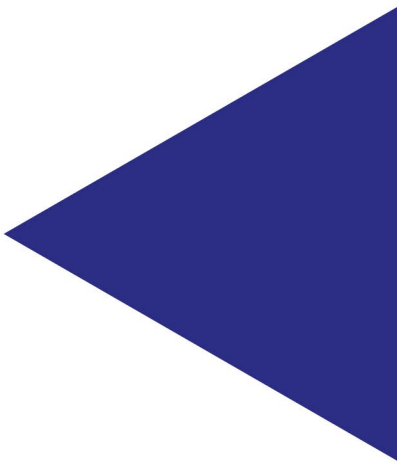
DOI
10.1111/1556-4029.70190

Publication date
2025

Document Version
Final published version

Published in
Journal of Forensic Sciences





Received: 20 May 2025 | Revised: 4 August 2025 | Accepted: 11 September 2025

DOI: 10.1111/1556-4029.70190

JOURNAL OF FORENSIC SCIENCES AAFS

ORIGINAL PAPER

General

The impact of institutional authority on forensic evidence evaluation by criminal justice professionals

Rosanne H. D. de Roo MSc[1,2] | Lonneke Stevens PhD[2,3] | Christianne J. de Poot PhD[1,2,3]

[1]Amsterdam University of Applied Sciences, Forensic Sciences, Amsterdam, The Netherlands

[2]Criminology Department, VU University Amsterdam, Amsterdam, The Netherlands

[3]Police Academy of the Netherlands, Apeldoorn, The Netherlands

Correspondence
Rosanne H. D. de Roo, Amsterdam University of Applied Sciences, Forensic Sciences, Tafelbergweg 51, 1105 BD Amsterdam, The Netherlands.
Email: r.h.d.de.roo@hva.nl

Abstract

Accurate and reliable decision-making in the criminal justice system depends on accurate expert reporting and on the correct interpretation of evidence by the judges, prosecutors, and defense lawyers. The present study aims to gain insight into the judiciary's capability to assess the accuracy and reliability of forensic expert reports by first examining the extent to which criminal justice professionals are able to differentiate between an accurate (or sound) expert report and an inaccurate (or unsound) expert report. In an online questionnaire, 133 participants assessed both a sound and an unsound expert report. The findings show that, on average, participants were unable to significantly distinguish between sound and unsound forensic expert reports. Second, the study explored the influence of institutional authority on the evaluation of forensic expert reports. Reports that were not recognized as flawed—particularly those originating from well-known and reputable institutions—were subjected to less critical examination, increasing the risk of evaluation errors. These results suggest that the perceived institutional authority influences the assessment of forensic evidence. The study highlights the need for tools to support criminal justice professionals in evaluating forensic evidence, particularly when experts are unregistered. Recommendations include adhering to established quality standards, consulting counter-expert evaluations, improving courtroom communication, and enhancing forensic knowledge through training. Overall, the findings underscore the importance of critical evidence evaluation to reduce the risk of misinterpretation and wrongful convictions in the judicial process.

KEYWORDS
admissibility, cognitive bias, criminal justice professionals, forensic evidence evaluation, institutional authority, validity

Highlights

- Criminal justice professionals struggle to distinguish sound from unsound forensic expert reports.
- Reports from reputable institutes face less scrutiny, risking flawed legal decisions.
- Tools and training are needed to improve forensic evidence evaluation by justice professionals.

1 | INTRODUCTION

The use of forensic technological innovations in the process of criminal investigation and prosecution has increased exponentially in recent decades [1–4]. These technologies help reduce the possibility of wrongful convictions [5] and have exonerated individuals who were wrongly convicted [6]. On the other hand, the increasing complexity of forensic research may hamper the interpretation of forensic evidence, potentially leading to judicial errors and even wrongful convictions [5, 7, 8].

Many wrongful convictions are based on valid and reliable forensic evidence reports that were misused or misunderstood by court officers. A wrongful conviction may occur when the limitations of forensic conclusions are poorly understood, evidence is either overvalued or undervalued, or alternative interpretations are not considered [5, 9, 10]. In addition to judicial errors, wrongful convictions are also associated with incompetent or fraudulent examiners and the use of experts who do not practice within accredited forensic science organizations [5, 11]. In summary, accurate and reliable decision-making in the criminal justice system depends first on accurate expert reporting and second on the correct interpretation of evidence by the judges, prosecutors, and defense lawyers (collectively referred to as criminal justice professionals) [5].

In one of the most recent wrongful convictions in the Netherlands (the "Rosmalense flatmoord"), it became evident that the scientists, embedded in reputable institutes, who initially acted as forensic experts in the case, made overly assertive and unsupported claims. These claims played a significant role in shaping the judges' verdicts. During the case review, new reports highlighted the experts' unsubstantiated claims, leading to a re-evaluation (Dutch Supreme Court, ECLI:NL:HR:2018:2095; ECLI:NL:HR:2020:1604; ECLI:NL:PHR:2020:684 [12–14]) that ultimately resulted in the defendant's acquittal [15]. This case underscores the risks associated with the assumed credibility of scientific sources in the evaluation of evidence, as well as the presumed judiciary's capability to assess the accuracy and reliability of expert reports. But is this assumption justified? In recent literature [16–18], the question is raised to what extent criminal justice professionals are equipped to critically evaluate forensic reports and their reliability, posing the question of which guidelines can assist criminal justice professionals in assessing the quality of such reports.

Dutch law holds that the (selection and) assessment of evidence is entrusted to the judgment of the court [16]. In practice, this means that the court determines the weight and credibility of a piece of evidence, guided by criteria for evaluating the expert's expertise, such as reliability of their methods, their competence in applying them, and their expertise regarding the specific question at hand, see Article 51l *Wetboek van Strafvordering (Sv) [Dutch Code of Criminal Procedure]*. Additionally, registries such as the NRGD and LRGD can be consulted as a tool to examine the reliability of the expert. These registries establish specific quality standards that an expert must meet to become and remain registered. The objective of these standards is to provide reasonable assurance that the information provided by the expert is of a certain quality (art. 12(2) *Besluit register deskundige in strafzaken [Decree on the Register of Experts in Criminal Cases]*). However, there are no strict requirements as to what makes an expert. As a result, the assessment is largely based on the judge's own interpretation and weighting, which potentially leads to arbitrary and difficult-to-verify considerations, making it challenging to understand how criminal justice professionals determine the admissibility of evidence [18, 19].

Research has shown that the credibility of a statement is strongly influenced by the perceived credibility of the source [20–24]. Although scientific information is often complex and/or counterintuitive [25], scientists are generally considered competent sources [26]. When complex or incomprehensible information is supplied by a credible source or authority, their audience is more likely to accept unclear scientific information [27, 28]. When the source is perceived as a trustworthy expert, people are willing to believe claims from that source without fully comprehending them, to the point where even conclusions based on "gobbledygook" or incomprehensible content are accepted as actionable [24]. However, to date, no studies have examined how the perceived credibility of a forensic source or authority impacts the evaluation of forensic evidence by criminal justice professionals.

The source of a forensic report can refer both to the writer of the report and the publishing institution. In the field of forensics, authority is generally divided into two types: (1) Epistemic authority, which is granted to individuals with expert knowledge in a specific field (e.g. DNA experts, pathologists, blood pattern analysis experts, etc.), and (2) Administrative authority, referring to those in positions of power due to their official roles such as the police [22]. Previously discussed papers focus on the impact of an expert as a credible source, leaving open the question of how the publishing institution affects the perception of credibility.

The present study aims to gain insight into the judiciary's capability to assess the accuracy and reliability of expert reports by first examining the extent to which criminal justice professionals are able to differentiate between accurate (or sound) expert reports and inaccurate (or unsound) expert reports. Second, it investigates the effect of the perceived institutional authority on criminal justice professionals' assessment of the reliability and evidential strength of both sound and unsound expert evidence reports. We hypothesize that criminal justice professionals have limited ability to differentiate between sound and unsound expert reports, and that institutional authority is likely to influence their evaluation of expert evidence. This may, on the one hand, lead to inadequacies in the forensic report not being identified, resulting in qualitatively inadequate reports being accepted as adequate. On the other hand, it may also lead to insufficient value being attached to adequate reports. Both potentially increasing the risk of misinterpretation and wrongful convictions. Therefore, it is crucial that criminal justice professionals are equipped with the right tools to critically assess forensic expert reports and (to some extent) understand its content in order to support accurate and reliable decision-making.

2 | METHOD

2.1 | Sound and unsound forensic reports

This study uses both "sound" and "unsound" forensic expert reports. A sound report refers to one with supported claims and logically correct reasoning, while an unsound report contains (overly) assertive, unsupported claims and errors in logical reasoning. To make the reports as realistic as possible, they were developed in collaboration with two forensic experts from the Netherlands Forensic Institute. With their input, realistic sound and unsound reports were created, based on real case elements, in the fields of pathology and blood pattern analysis (BPA). A sound report included three supported claims and two logically correct conclusions, whereas an unsound report contained three assertive and unsupported claims and two instances of the prosecutor's fallacy. A model of the reports can be found in Appendix 1.

2.2 | Institutional authorities

To explore the influence of potential differences in the perceived institutional authority by criminal justice professionals on their assessment of the reliability and evidential strength of both sound and unsound expert evidence reports, we distinguish between institutional authority at epistemic and administrative levels. Reporting experts in forensic institutes are regularly examined on their expertise, registered as experts, and recognized in their field. We therefore hypothesize that reports from such institutes are highly reliable due to their epistemic authority. In the Netherlands, the Netherlands Forensic Institute is the largest distributor of forensic reports and is well-known within the Dutch judiciary. Therefore, this institute will be included in this study. In addition to the NFI, the Netherlands Police (hereafter referred to as the Police) is a significant distributor of forensic reports. While experts from the Police are considered knowledgeable in their field, they are not officially registered as such, which may impact the perceived reliability of the expert evidence reports they produce. Furthermore, the Police hold a position of power, partly due to the assumption that the Police are more credible than other individuals and the fact that a Police report carries more weight than other pieces of evidence (art. 344(2), Sv). As a result, they receive both epistemic and administrative authority. To study potential differences in the perceived authority of institutions and its influences on the reliability and assessment of the forensic expert reports they produce, we included a third nonexistent (and thus unknown) forensic institute "ForensX" in our study, to serve as a baseline. We are aware that this nonexistent forensic institution will have a certain degree of institutional epistemic authority, due to the fact that it is a forensic institute producing forensic reports. However, the experts are not registered, and the institution is not recognized in the field. Given the varying types of institutional authority attributed to these three institutes, we hypothesize that the perceived authority of the NFI, Police, and ForensX will differ, leading to variations in the perceived reliability of the reports they produce and the assessment of these forensic reports.

2.3 | Design

In an online questionnaire, participants were asked to read two forensic reports: a BPA report and a pathology analysis report. Each report topic was selected for specific reasons: BPA, as reports are written by both the police and the NFI (with the NFI potentially holding greater epistemic authority than the police, while the police maintain administrative authority over the NFI), and pathology, because the police hold administrative authority but lack epistemic authority (as police officers are not pathologists), whereas the NFI possesses epistemic authority.

Each report had two versions, namely (1) a sound logically correct version, and (2) an unsound logically incorrect version. All reports were designed specifically for this study in collaboration with forensic experts and reported an evidential strength ranging from "likely" to "more likely" (Likelihood Ratio (LR) = 100). The reports were randomly allocated to the participants such that each participant received one sound and one unsound report.

Additionally, the two reports featured the institutional logos of one of the three institutions—the Police, the NFI, or ForensX—that were hypothesized to differ in the institutional authority. The results of this manipulation are described in Section 3.2.1. Manipulation check. See Figure 1 for a visualization and Figure 2 for an example of the research

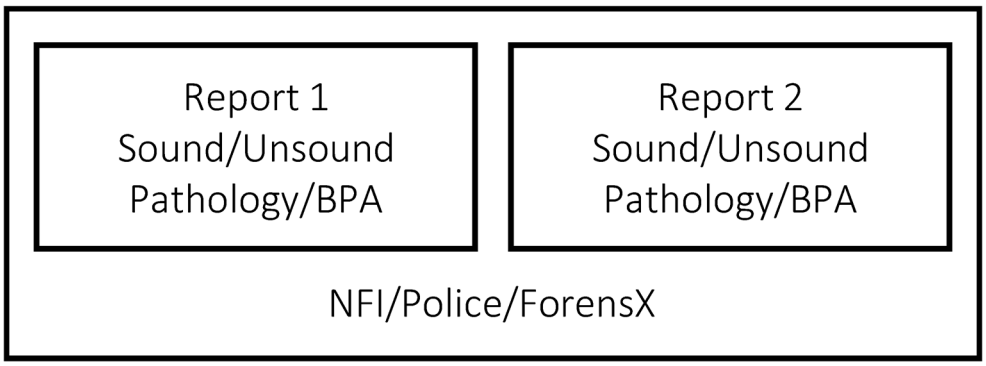


FIGURE 1 Visualization of the research design.

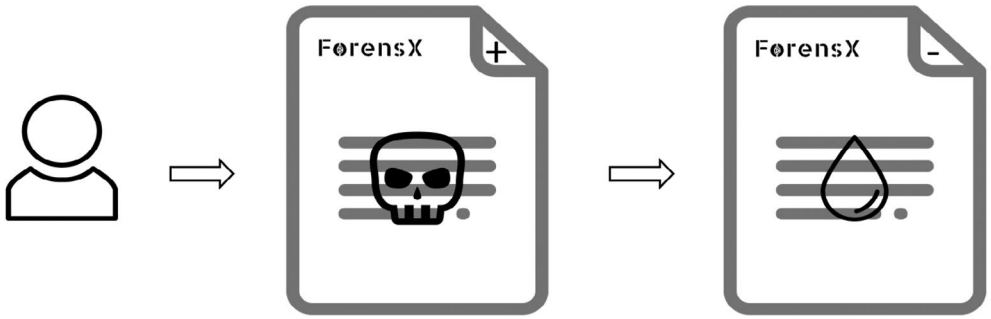


FIGURE 2 Example: Participant receives two reports. In this case, a sound (+) pathology and an unsound (−) BPA report, both from the institution ForensX.

design. The survey tool (www.qualtrics.com) was used to randomly allocate participants over the conditions. To control for an order-effect, the order in which the reports were presented was randomized.

2.4 | Procedure

Participants voluntarily participated after a request was sent out via email within their organization. In the invitation email, the goal of the study—improving the understanding and valuation of forensic evidence—was briefly explained to the participants. Via a URL in this email, participants were directed to the questionnaire on the website of the survey tool (www.qualtrics.com). A welcome message explained that, as part of a study on interpreting forensic evidence, participants were invited to read forensic reports and answer questions. On the next page, participants were asked general questions about their profession, age, gender, years of working experience in criminal law, frequency of reading forensic reports, and their self-rated understanding of Bayesian statistics.

After these questions, the first report was displayed, accompanied by a brief case description. Both cases portrayed with the BPA reports and pathology reports described an ambiguous scenario that could be attributed to multiple explanations. All reports (BPA report, both sound and unsound; and pathology report, both sound and unsound) had the same text style and layout, except for the institutional logo. After each version of the report, the exact same set of questions appeared. First, the participants were asked about their perceived reliability of the report. Next, they were asked to assess and rate the reported evidential strength in the report. Finally, they were asked about the defendant's presumed guilt and were required to elaborate on their answer.

Once the participants completed the questions on the first report, the second report was displayed. To prevent them from altering their responses after reading and answering the questions of the second report, participants were unable to modify their answers to the first report. Finally, participants were asked from which institute they had just read two reports and how credible they generally perceived all three organizations (the Police, NFI, and ForensX) in their daily practice, aside from the reports just shown in the survey. Completing the questionnaire required 15–20 min.

2.5 | Participants

Participants were criminal justice professionals (i.e., criminal judges, criminal lawyers, and public prosecutors) who could be tasked with assessing forensic evidence reports in their work practice.

A total of 234 participants started the online questionnaire, of which 94 participants discontinued during the general questions and 7 participants stopped before completing both reports. The remaining 133 participants completed a total of 266 reports, of which 133 were sound and 133 were unsound. The sound reports consisted of 62 BPA and 71 pathology reports, and (logically) the unsound reports consisted of 71 BPA and 62 pathology reports. The participants consisted of 71 criminal judges, 48 criminal lawyers, and 14 public prosecutors. Of the participants, 45% were female, 53% male, and 2% preferred not to answer the question regarding gender. The majority of participants fell into the age category of 51–60 years, and most participants reported having 16–20 years of working experience in criminal law.

2.6 | Assessing test scores

The data generated by the participants were scored in four sections: (1) perceived reliability of the report, (2) evaluation of the reported evidential strength, (3) presumed guilt of the defendant, and (4) general credibility of the institutions.

1. The perceived reliability was assessed using a 5-point Likert scale with the scores 1=very unreliable, 2=unreliable, 3=neither reliable nor unreliable, 4=reliable, and 5=very reliable.
2. The perceived evidential strength of the reports ("more likely" to "much more likely," LR=100) was rated using a 5-point Likert scale, with the scores 1=very low, 2=low, 3=neither low nor high, 4=high, and 5=very high.
3. The presumed guilt of the defendant was rated using a score: 1=not incriminating, 2=somewhat incriminating, 3=incriminating, and 99=I don't know. The answers to the open-ended question were scored binary (yes=1 and no=0) based on whether participants mentioned (1) assertive and unsupported claims and (2) logical reasoning errors.

4. The general credibility of the forensic institutes, including NFI, Police, and ForensX, was assessed using a 5-point Likert scale with the scores 1 = very unreliable, 2 = unreliable, 3 = neither reliable nor unreliable, 4 = reliable, and 5 = very reliable. Additionally, participants had the option to indicate that they could not assess the institute if they were unfamiliar with it.

The responses to the open-ended question concerning the content of the reports were categorized into statements about the presence of a prosecutor's fallacy (1 = mentioned, 0 = not mentioned) and/or indications that the report contained (a form of) inimitable inadequate reasoning (1 = mentioned, 0 = not mentioned).

2.7 | Data analysis

All data were exported from the survey tool to IBM SPSS Statistics (version 27). The research design combined within-person factors (sound/unsound and BPA/pathology) with a between-persons factor (type of forensic institution). We employed a repeated measures ANOVA for the within-person comparisons and a multivariate ANOVA for the between-persons comparison. The binary scores for the open-ended questions were analyzed with a Chi-square test.

3 | RESULTS

The main question of this study is whether criminal justice professionals are able to differentiate between sound and unsound expert evidence reports. After discussing these results, we will discuss the impact of the perceived institutional authority on criminal justice professionals' assessment of the reliability and evidential strength of both reports.

3.1 | Sound vs. unsound expert reports

3.1.1 | Perceived reliability

On average, participants ($n = 133$) did not statistically differentiate between the rated reliability of sound vs. unsound reports (sound, M = 3.3; unsound, M = 3.1; $F(1) = 2.839$, $p = 0.094$). Fifty percent of the participants correctly rated a sound report as sound, by rating the report as reliable or very reliable ($n = 67$). However, only 32% of the participants correctly rated the unsound report as unreliable, rating it as unreliable or very unreliable ($n = 42$). Twenty-two percent of the participants ($n = 29$) rated a sound report as unreliable or very unreliable, and 44% of the participants ($n = 58$) rated an unsound report as reliable or very reliable (see Figure 3).

The majority of the participants (41%, $n = 55$) rated the reliability of the sound and the unsound reports as equal. Thirty-five percent of the participants ($n = 46$) rated unsound reports as less reliable than sound reports, and 24% ($n = 32$) rated unsound reports as more reliable than the sound reports. See Figure 3 for a visualization of these results.

Report topic (BPA and pathology)

On average, the participants rated the pathology reports significantly more reliable than the BPA reports (respectively M = 3.3, M = 3.0; $F(1) = 8.382$, $p = 0.004$), independent of the "soundness" of the reports (sound vs. unsound). Noticeable in the data is that the pathology reports seem to be assessed less critically than the BPA reports. On average, only 12% of the participants rated the

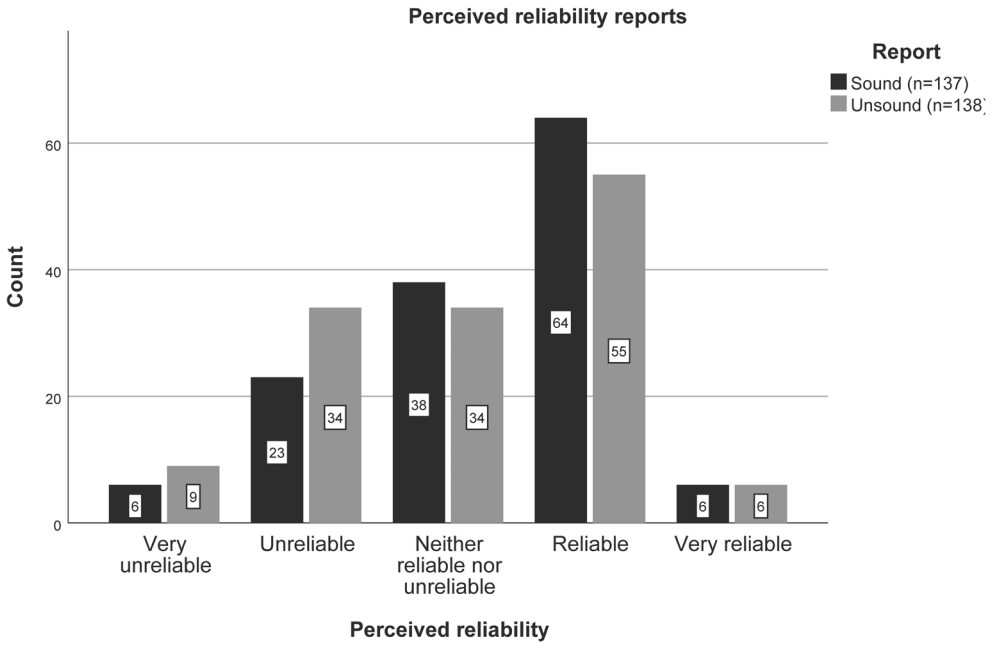


FIGURE 3 The assessed reliability of sound reports and unsound reports by criminal justice professionals.

pathology reports ($n=133$) as unreliable to very unreliable (11% sound (8 out of 71) and 13% unsound (8 out of 62)), while on average 30% of the participants rated the BPA reports as unreliable to very unreliable (24% sound (15 out of 62) and 35% unsound (25 out of 71)). Despite this difference in rated reliability of the two report types, on average the assessed reliability of the sound vs. unsound reports was not significantly different (pathology sound, M=3.4, unsound, M=3.3, $p=0.68$; BPA sound, M=3.1, unsound, M=3.0, $p=0.17$), see Table 1. Next to that, a form of inadequate reasoning in the reports was mentioned by the participants in 31% (41 of 133) of the BPA reports and in 17% (23 of 133) of the pathology reports, indicating a significant impact of report type on the number of comments given by the criminal justice professionals ($\chi^2(1)=14.548$, $p=0.001$). The prosecutors' fallacy was mentioned only seven times in total, six times in a pathology report and once in a BPA report.

3.1.2 | Evidential strength

Participants ($n=133$) valued the evidential strength of the sound and unsound reports the same on average (sound, M=2.8; unsound, M=2.9; $F(1)=0.211$, $p=0.647$), where a score of "3" indicates a neutral assessment ("neither low nor high"). Figure 4 shows that participants' responses to the question regarding the assessment of the evidential value of either the sound and unsound reports were distributed.

The reports had an evidential value or LR of 100. This could arguably be interpreted as neither high nor low. As Figure 4 shows, it is striking how divided the participants are in their interpretation of this evidential value. A few responses of participants who rated the sound (pathology) report, discussed below, show the discrepancy in how individuals interpret the same numerical value. Two participants rated the report as having a "high evidential value," stating "The results are likely to more likely, not very much more likely, which is incriminating, but not the highest degree" and "Given the indicated probability, the results are incriminating." While two other participants rated the same sound pathology report as having a "low evidential value," stating "The incriminating conclusion is only likely to more likely, not much more likely" and "More likely is insufficient for legal and convincing evidence."

3.1.3 | Presumed guilt defendant

On average, 126 of the 133 participants rated the sound reports as "somewhat incriminating" (M=2.0) for the defendant. Similarly, the

TABLE 1 Average perceived reliability per report topic (BPA and pathology) for the sound and unsound reports.

	Sound			Unsound		
	Total	BPA	Pathology	Total	BPA	Pathology
Perceived reliability	3.3	3.1	3.4	3.1	3.0	3.3

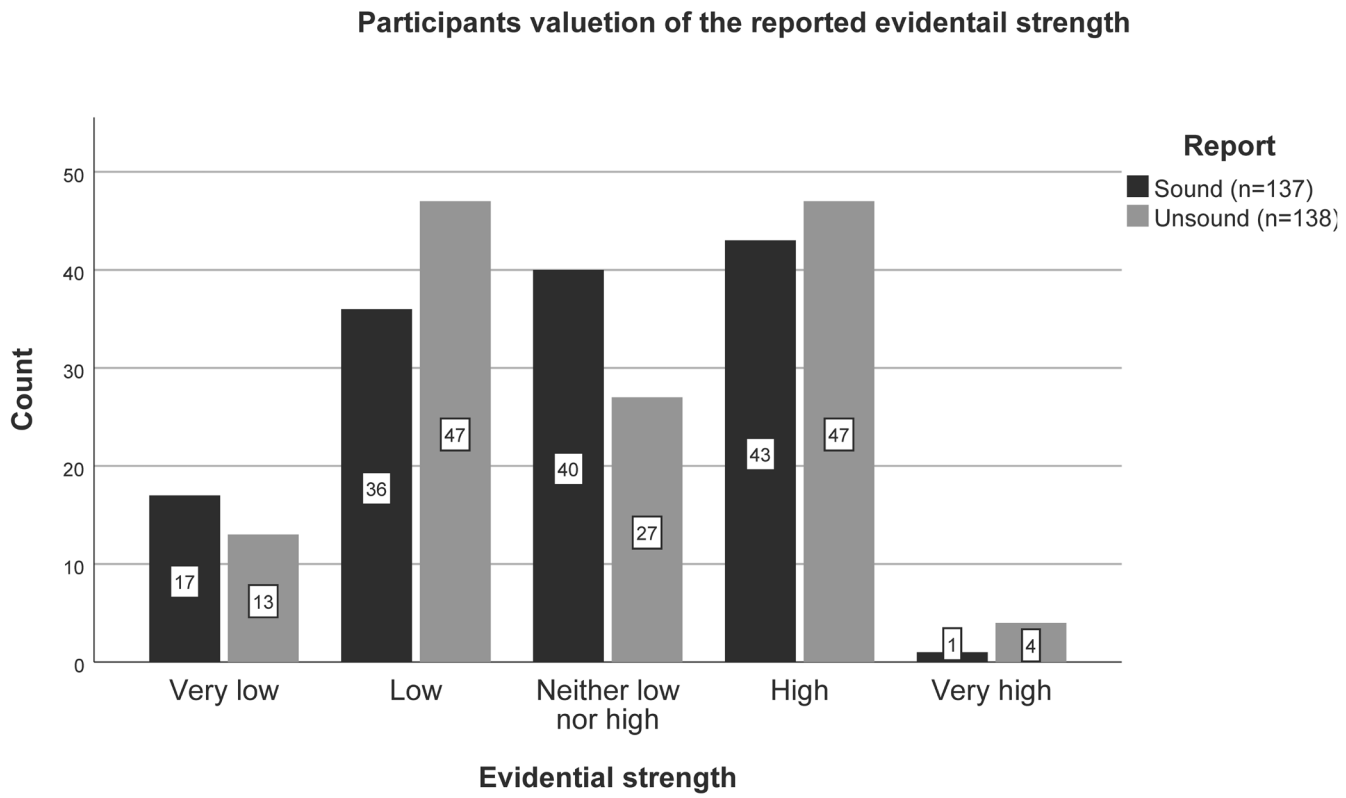


FIGURE 4 The assessed evidential value of sound and unsound reports by criminal justice professionals.

unsound reports were rated by 108 of the 133 participants (M = 2.1) reports as "somewhat incriminating" for the defendant ($p = 0.342$). Despite these average means, as shown in Figure 5, the participants were highly divided in their responses regarding the presumed guilt in both the sound and unsound reports.

The majority of the participants (33%, $n = 44$) rated the presumed guilt of the defendant equally after reading either the sound or unsound report. Seventeen percent ($n = 23$) rated the sound report as more incriminating than the unsound report, and 26% ($n = 34$) rated unsound reports as more incriminating than the sound reports. In 19% of the cases ($n = 25$), the participant did rate the sound reports' incriminating value, but not the unsound reports – by rating it as I don't know – and 5% ($n = 7$) of the participants did rate the unsound reports' incriminating value, but not the sound reports.

When looking further into the results, it shows that in all cases a participant stated that a report was "not incriminating" ($n = 32$), they indicated not to not have sufficient (case) information to assess the defendant's guilt. Notable is the number of participants who did evaluate the sound report, but not the unsound report ($n = 25$) against the participants who did rate the unsound but not the sound report ($n = 7$), potentially indicating that the participants did notice "something" was off about the unsound reports.

When reviewing each participant's explanation of why they did or did not find the report incriminating for the accused, the results revealed that in about 25% of the assessed reports (64 of the 266), a comment on the adequacy of the report was given that potentially influenced their assessment. In 32% of the unsound reports (43 of 133), participants mentioned the report contained (a form of) inimitable inadequate reasoning (40 times) or the presence of a prosecutor's fallacy (7 times); four participants mentioned both errors. For the sound reports, 15% of the participants (20 out of 133) mentioned some form of inadequate reasoning in their response. A Chi-square analysis showed a significant effect of report type (sound and unsound) on the number of scored comments on the adequacy of the report, $\chi^2(1) = 6.666$, $p = 0.01$.

For the unsound reports, 32 (27 inimitable inadequate reasoning and 5 prosecutor's fallacy) of the 47 comments were given by judges ($n = 71$). Prosecutors ($n = 14$) made 8 comments (6 inimitable inadequate reasoning and 2 prosecutor's fallacy), and lawyers ($n = 48$) mentioned inimitable inadequate reasoning seven times. Also, with the sound reports, most of the 20 comments were given by judges, mentioning 14 times inimitable inadequate reasoning 14 times next to lawyers commenting this six times.

3.2 | Impact perceived institutional authority

To analyze the impact of the perceived institutional authority on criminal justice professionals' assessment of the reliability and

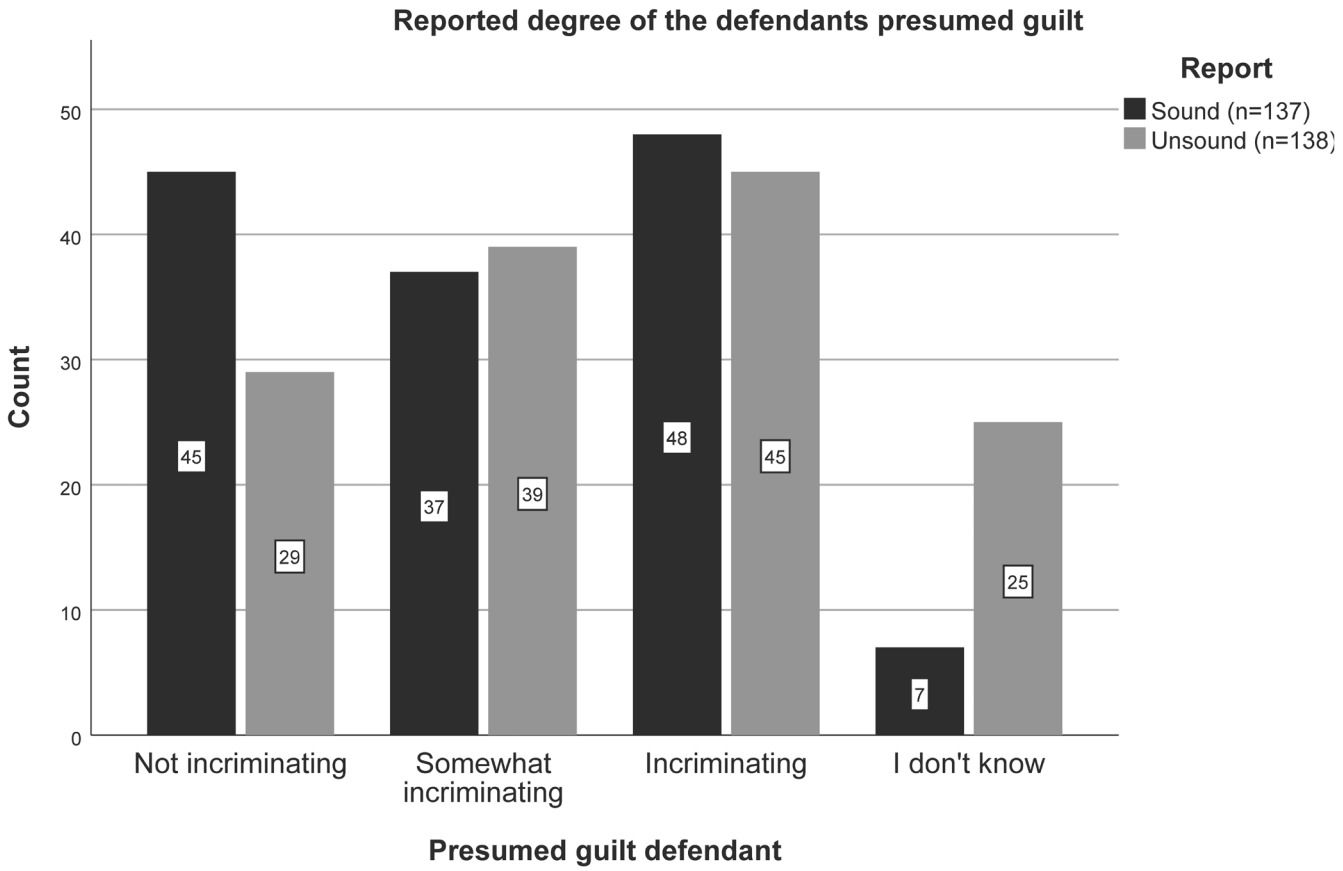


FIGURE 5 Count presumed guilt of the defendant by criminal justice professionals for the sound reports and unsound reports.

evidential strength of both reports, a manipulation check was conducted on the 133 participants.

3.2.1 | Manipulation check

While assessing the sound and unsound reports, the participants were primed by viewing the institution names and logos of either the NFI, the Police, or ForensX. We expected that these institutions would differ in the institutional authority attributed to them. To assess the effectiveness of this manipulation in our design, the following manipulation check was conducted.

The general perceived credibility was significantly different for the three institutes: NFI, Police, and ForensX ($F(2, 130) = 561.48$, $p < 0.001$). The post hoc test showed that participants rated the NFI more credible (M = 4.5, $n = 133$) than the Police and ForensX (respectively M = 3.6, $n = 126$; M = 3.8, $n = 17$). No significant difference was found between the rated perceived credibility of the Police and ForensX. Although the majority of the participants ($n = 116$) stated they could not evaluate the credibility of ForensX because they were not familiar with it. Next to that, seven participants responded that they could not evaluate the credibility of the Police.

The 133 participants responded to the question about whether they remembered which institute had produced the two forensic reports they had just seen. Ninety-seven participants (73%) remembered this correctly, seven (5%) mistakenly believed they had seen NFI reports while they had actually seen reports from the Police ($n = 3$) or ForensX ($n = 4$), and 29 participants (22%) stated that they did not remember which institution had produced the reports.

Only the 97 participants for whom the manipulation was successful were included in this analysis, comprising 52 criminal judges, 35 criminal lawyers, and 10 public prosecutors. Of the participants, 38 were female, 58 were male, and 1 preferred not to answer the question regarding gender. The average age category was 51–60 years, and participants had, on average, 16–20 years of working experience in criminal law. The participants assessed 194 forensic reports, of which 97 were 97 sound and 97 were unsound.

3.2.2 | Impact perceived institutional authority on the sound and unsound reports

The participants assessed 194 reports, containing 68 NFI reports, 64 Police reports, and 62 ForensX reports. The perceived reliability of the reports produced by the three institutes differed significantly depending on the institute, regardless of whether the reports were sound or unsound ($F(2,191) = 3.985$, $p = 0.02$). The post hoc test showed that the NFI reports (M = 3.4) were valued significantly more reliable than the ForensX reports (M = 2.9) ($p = 0.017$), and that there were no significant differences between Police reports (M = 3.2) and NFI reports, and between Police reports and reports from ForensX ($p = 0.13$); see Table 2.

Further analyses of the data revealed that the unsound reports ($n = 97$) were assessed differently between the institutes ($F(2, 94) = 3.209$, $p = 0.045$). The post hoc test showed that reports with an NFI logo (M = 3.4) were considered more reliable ($p = 0.05$) than the same reports when they came from the ForensX institute (M = 3.1). No significant differences were found between the reliability of the unsound police reports (M = 3.3) and the unsound reports of the other two institutes. However, for the sound reports ($n = 97$), no significant difference in the perceived reliability was found between the three institutes (NFI, M = 3.4; Police, M = 3.3; ForensX, M = 3.1; $F(2, 94) = 1.152$, $p = 0.320$).

Within each institute, no significant difference was found between the sound and unsound reports. Notably, the reliability of the NFI and police reports was rated almost equally, while a trend was observed for the ForensX reports, where the unsound reports were rated lower than sound reports. However, this difference is not significant ($p = 0.21$).

To examine whether the manipulation also works in participants who did not pass the manipulation check, for example due to implicit priming, the data from these participants ($n = 36$, 72 reports) were analyzed. No effect of the reporting institute on the rated reliability was found for either the sound and unsound reports ($F(2, 70) = 0.787$, $p = 0.378$).

It is notable that the reporting institution influenced the number of comments provided on the adequacy of the reports. In the 194 reports, 50 comments were made by the criminal justice professionals regarding the adequacy of the report. Forty-four instances of inadequate reasoning were mentioned, including 14 times after reading an NFI report (21% out of 68 NFI reports), 9 times after a Police report (14% out of 64 Police reports), and 21 times after a ForensX report (34% out of 62 ForensX reports). A Chi-square analysis showed a significant effect of reporting institution (NFI, Police, and ForensX) on the number of scored comments χ^2 (2) = 7.308, $p = 0.026$, indicating that the reporting institution affected the criminal justice professionals' assessment of the report. Next to this, 6 prosecutors' fallacies were commented on, 4 times after reading an NFI report, once after a Police report, and once after a ForensX report. These numbers were too small to perform a Chi-square analysis.

TABLE 2 Average perceived reliability per institute for the sound and unsound reports.

	Sound				Unsound			
	Total	NFI	Police	ForensX	Total	NFI	Police	ForensX
Perceived reliability	3.3	3.4	3.2	3.1	3.1*	3.3^{a}	$3.2^{a,b}$	2.7^{b}

Note: The means that do not share the same superscript letters differ at $p < 0.05$.

*Effect of reporting institute significant at $p < 0.05$ level.

No impact of the perceived institutional authority of the reporting institutions was found on the rated evidential strength ($p=0.472$) of the reports, nor on the assessed presumed guilt of the defendant ($p=0.634$).

3.3 | Correlation analysis

The correlation between various factors—including years of working experience in criminal law, number of Bayesian statistics courses taken, number of forensic reports encountered per month, and participants' professions—was examined in relation to their ability to differentiate between sound ($n=133$) and unsound reports ($n=133$). The perceived reliability of the reports was found to correlate with participants' years of experience in criminal law. Specifically, the longer participants had worked in criminal law, the more reliable they rated both sound and unsound forensic reports (χ^2 (20)$=40.891$, $p=0.004$). No significant correlation was found with the number of Bayesian statistics courses taken (χ^2 (16)$=12.958$, $p=0.676$), nor with the number of forensic reports read per month (χ^2 (16)$=11.036$, $p=0.807$). The profession of the criminal justice professionals did not influence their assessment of the reliability of the forensic reports ($p=0.168$), see Table 3. Notably, both judges and prosecutors rated the sound reports on average slightly more reliable, yet not significantly, than the unsound reports. In contrast, lawyers rated both report types similarly, on average as "neither reliable nor unreliable."

Another noteworthy result is that all participants, regardless of their profession, rated both the sound and unsound reports on average neutral to reliable ("neither reliable nor unreliable" to "reliable"), suggesting that criminal justice professionals have difficulties distinguishing between accurate (sound) and inaccurate (unsound) forensic expert reports. However, due to the small sample size (71 criminal judges, 48 criminal lawyers, and 14 public prosecutors), especially in the latter group, no firm conclusions can be drawn regarding the influence of participants' professions on their ability to differentiate between sound and unsound reports.

4 | DISCUSSION

The study focused on the question of whether criminal justice professionals (criminal judges, prosecutors, and lawyers) were able to differentiate between an accurate (or sound) expert report and an inaccurate (or unsound) expert report, and to what extent this assessment was influenced by the reporting institute (NFI, Police, or ForensX). The results on the first part of our question have shown that the participants were, on average, not able to significantly differentiate between a sound and an unsound forensic expert report. Moreover, the majority of the participants, 41%, valued the sound and unsound reports as equally reliable. Thirty-five percent valued the unsound reports as less reliable than the sound reports, and 24% rated the unsound reports as more reliable than the sound reports. Although no significant difference was found between the perceived reliability of the sound and unsound reports, a significant correlation ($p=0.01$) was found between the number of remarks made on the adequacy of the report and the type of report (sound and unsound). In 32% of the unsound reports (43 of 133) and 15% of the sound reports (20 of 133), a form of inimitable inadequate reasoning and/or the presence of a prosecutor's fallacy was mentioned. The findings of this study, which show that legal professionals struggle to evaluate the adequacy of forensic reports and have difficulty identifying logical errors such as the prosecutor's fallacy (only identified in 7 out of 133 unsound reports), are consistent with previous research indicating that criminal justice professionals are often unable to properly assess the scientific justification of methods used and (Bayesian) conclusions reached in forensic reports (among others; [8, 29–31]). Nonetheless, criminal justice professionals are expected to determine the weight and credibility of forensic evidence by evaluating forensic evidence reports. Therefore, it is of the utmost importance that (logically) incorrect forensic reports are recognized as such, as these can negatively influence judgment in the criminal trial process. As mentioned previously, some studies raise the question of to what extent this task should be performed by legal professionals [16–18] and argue that it might be more appropriate for experts to undertake this role through peer review, where the reports are evaluated by other professionals in the field [18]. It could potentially benefit the Dutch judiciary to consult registries such as the NRGD, where experts are assessed through peer review based on quality standards, more regularly as a tool to evaluate the reliability of experts. A notable challenge is that not all fields of expertise are currently represented in these registries, and some may never be included due to their rarity. Furthermore, such expert registries are relatively unique in the Dutch inquisitorial legal system, meaning that other countries with different legal frameworks may continue to encounter challenges in evaluating forensic expert evidence. This raises an important question: what tools can be provided to criminal justice professionals, particularly in cases where the expert is not formally registered? Two potential paths to consider are (1) the enhancement of the judiciary's capacity to comprehend forensic reports, and (2) stricter quality assurance of forensic evidence reports.

The first point presents a degree of contradiction, as prior research has shown that even with additional training criminal justice

TABLE 3 Average perceived reliability per profession for the sound and unsound reports.

	Sound				Unsound			
	Total	Judge	Prosecutor	Lawyer	Total	Judge	Prosecutor	Lawyer
Presumed guilt defendant	3.3	3.4	3.6	3.0	3.1	3.1	3.4	3.0

professionals show limited capacity for understanding forensic (Bayesian) reasoning [8, 30, 32] and this research shows that criminal justice professionals are limited in their ability to distinguish between accurate and inaccurate forensic expert evidence reports. But, if criminal justice professionals fail to recognize (logical) inaccuracies in forensic reports, how can they enhance their evaluative skills? In other words, how can one address inaccuracies if one is unaware of their existence. Potentially it could be considered to standardize counter-expert evaluations from qualified professionals when the (first) reporting expert is not registered as such, or, if the judiciary does not feel adequately equipped to evaluate the expert report, or has doubts about the expert's expertise. Another potential consideration to enhance the judiciary's capacity to comprehend forensic reports is the improvement of the forensic experts' communication. A trend that has become increasingly apparent over the past years is the growing emphasis on improving the communication of complex scientific information in layman's terms that are accessible to legal professionals [33, 34]. For example, by adding information leaflets and subject appendices to the scientific report, or—especially in the inquisitorial system—making it more common practice to invite experts to criminal court hearings. Next to that, the concept of "courtroom communicators" is a relatively recent development and is currently under investigation and discussion within legal and academic circles (extensively discussed during the workshop "Courtroom Communication" at IAFS2023 (23rd Triennial Meeting of the International Association of Forensic Sciences) in Sydney). The primary objective of these communicators is to present complex scientific information in a manner that is comprehensible to laypersons, including jurors and judges, thereby contributing to the enhancement of both the fairness and efficacy of the legal system. In line with this concept, the Dutch (inquisitorial) judiciary has already introduced scientific "forensic advisers" to support judges in all matters related to forensic science [35, 36] and Belgium introduced similar "forensic advisers" to assist the prosecution [37]. Another way to enhance the judiciary, we argue, is to raise awareness on the limited capability of the criminal justice professionals to evaluate forensic expert reports, emphasizing the importance of a continued investment in the forensic education of criminal justice professionals. Especially as our study showed that the longer participants have worked in criminal law, the more reliable they rated both sound and unsound forensic reports. It is essential to equip the judiciary with the knowledge to assess the quality of forensic reports to potentially improve their critical assessment, for example, by learning to recognize "red flags" such as the persecutor's fallacy, incorporating Bayesian reasoning into the educational program of (criminal) law students, and further educating lawyers to make forensic reasoning more ingrained within the (criminal) judiciary.

The second path to consider is stricter quality assurance of forensic evidence reports. If the quality of forensic reports is sufficiently high, the potential limitations in legal professionals' understanding may be of lesser concern. In this regard, professional organizations, such as the IAFS (International Association of Forensic Sciences), AAFS (American Academy of Forensic Sciences), and EAFS (European Academy of Forensic Science) could play a significant role in establishing standards and promoting the professionalization of their respective disciplines. One potential solution could involve making the assessment of quality standards established for nonregistered experts (semi-) mandatory. Further elaboration on this topic will be discussed later on.

As a second objective, this study investigated to what extent criminal justice professionals' assessments of both sound and unsound expert evidence reports were influenced by perceived institutional authority (from the NFI, the Police, or ForensX). Our study shows that the reporting institution influenced the assessment of the criminal justice professionals. Independent of whether they were sound or unsound, the NFI reports were valued significantly more reliable than the ForensX reports ($p < 0.05$). While the reliability of Police reports was not rated significantly different from either the NFI reports or the ForensX reports. Especially in the evaluation of the inaccurate unsound reports, the NFI reports were assessed as significantly more reliable than those from the ForensX institute. Notably, the reporting institution significantly influenced the number of comments given on the adequacy of the reports ($p = 0.026$). Especially, participants (34%) who read reports from ForensX commented on (a form of) inimitable inadequate reasoning in the reports, compared with 21% after reading NFI and 14% after reading Police reports. These results show that the authority attributed to the reporting institutes affected the criminal justice professionals' assessment of the reports, especially for the unsound reports. Our findings are in line with the scored general perceived credibility of the participants, showing that the NFI was generally considered more reliable than both the Police and ForensX. This leads us to conclude that criminal justice professionals who read a forensic expert report tend to rate it as more reliable, especially when it is inaccurate, if it is written by an institution they consider trustworthy, like the NFI, compared with when it is written by an unknown (nonexistent) institution like ForensX. Next to that, criminal justice professionals are more critical toward a report from an unknown institution, like ForensX, and therefore tend to make more critical notes on its adequacy compared with participants reading the same report from an institution they are familiar with, like the Police or NFI. Moreover, participants reading reports from the unknown (nonexisting) institute (ForensX) were more likely to indicate they needed more information before making a decision on the presumed guilt of the defendant or wanted to ask additional questions to the expert about their reasoning steps in the reports.

The fact that authority influences the assessment of forensic reports by legal professionals is, in itself, logical and not inherently negative, as long as authority is equated with quality. The existence of expert registers and guidelines for assessing the qualifications of experts underlines their importance, especially now that it has become clear that the majority of legal professionals are limited in their ability to distinguish between sound and unsound reports. However, even with these safeguards in place, registered and/or qualified experts may still make (logical) reasoning errors that may influence their reports and conclusions. Therefore, the judiciary should never

blindly rely on the adequacy of a report simply because it has been written by an institute that enjoys a high level of perceived epistemic authority and/or administrative authority, and should always stay critical when evaluating forensic expert evidence.

This also brings us to the evaluation of the police reports by the criminal justice professionals. The police reports in this study were not valued significantly different than either the NFI reports or the ForensX reports. On the other hand, the police reports had the least number of comments given on their adequacy (14% against, 21% NFI, and 34% ForensX). This may be in line with our earlier statement that the police enjoy both administrative and epistemic authority, being considered knowledgeable in their field, yet not officially registered as such, which may have an impact on participants' perceived reliability. In Dutch law (art. 150 Sv), there is a distinction between the technical investigators of the police and "experts." This means that investigators of the police who specialized, for example, in arson or BPA, are not formally considered "experts," but do produce reports with specialized knowledge that are used as evidence in the criminal trial process. We argue that, due to the important position of the police in the investigative process and the administrative and epistemic authority they hold, a tool to document and establish the capabilities of these specialized technical investigators should be implemented based on quality standards to evaluate their professionalism, which can serve as a foundation to justify their expertise. Overall, this would potentially improve the perceived reliability of police reports and provide the judiciary with more tools to assess police expertise.

Another noteworthy finding of this study is the responses to the question regarding the perceived evidential value of the reports. Participants were divided in their evaluation of how high or low they assessed the evidential value of both the sound and unsound reports. This division was particularly interesting to us because all the reports had an evidential value or likelihood ratio (LR) of 100. Certainly, the assessment of the evidential value largely depends on the context of the case; however, in this study, the context was consistent within the pathology reports and within the BPA reports. Our findings show that criminal justice professionals interpreted the same numerical value in strikingly different ways, ranging from describing it as insufficient to use as evidence to considering it incriminating.

Another point worth discussing is the significant difference in the rated reliability of the reporting topic: BPA and pathology. Criminal justice professionals rated the pathology reports significantly more reliable than the BPA reports ($p = 0.015$) and were significantly more critical of the BPA reports ($p = 0.001$), regardless of the "soundness" of the reports (sound vs. unsound). These findings may be explained by several factors. First, the expertise of a pathologist may be perceived as higher than that of a BPA expert. Pathology is a specialized field of expertise within the medical field, which takes years of training, whereas BPA is potentially seen as a more generalized field of expertise. Furthermore, these findings are consistent with the study by Hoogeveen et al. 2022, which argues that the more complex a topic is and the more the source is perceived as a trustworthy expert, the more people are willing to believe claims made by that source without fully understanding them [24]. Furthermore, Howes' study discusses that the more familiar judges are with a particular topic, the more they are able to critically assess the adequacy of the report and conclusions [38, 39]. Emphasizing our previous statement that increasing forensic knowledge among criminal justice professionals could potentially lead to improving their ability to critically assess forensic expert reports.

Although this study only briefly addressed the impact of a person's profession on how they value and interpret forensic evidence, this topic warrants further exploration in future research. Legal professionals may have varying approaches to assessing the probative value of forensic evidence, depending on their role within the legal process. For instance, lawyers, driven by the imperative to represent their client's best interests, often focus on identifying weaknesses or errors. Successfully challenging flawed evidence can significantly strengthen their defense strategy. Judges, on the other hand, must assess the evidence objectively to reach a fair judgment, independent of the interests of the parties involved. Given these differences, it would be valuable to investigate the influence of profession on the evaluation of forensic evidence. Unfortunately, the sample sizes in this current study were too small to draw definitive conclusions on this subject.

This study design has its limitations. The design involved the use of shortened (1 page) forensic reports based on real reports (developed in collaboration with forensic experts from the Netherlands Forensic Institute). In addition, apart from a brief description of an ambiguous case, participants were not provided with further case and/or contextual information. Typically, criminal justice professionals have access to a full case file. However, due to the limited time capacity of the participants, it was not possible to present a full case file or a full-length forensic report to the participants.

5 | CONCLUSION

Criminal justice professionals face difficulties distinguishing between accurate (sound) and inaccurate (unsound) forensic expert reports. Reports that are not recognized as flawed, specifically reports that originate from well-known and reputable institutions, are subjected to less critical examination, increasing the risk of errors in evaluation. The fact that institutional authority influences the assessment of forensic reports by legal professionals is, in itself, logical and not inherently negative, as long as authority is equated with quality. Safeguards to ensure this quality may include the use of expert registers with clear guidelines for evaluating expert qualifications, as well as making the assessment for nonregistered experts (semi-) mandatory. In addition, criminal justice professionals should be aware of the influence of institutional authority on their judgment of forensic expert evidence. By making them aware, we hope that the judiciary may become less susceptible to biases stemming from the perceived credibility of an expert's institutional affiliation and focus more critically on the actual quality and reliability of the evidence presented. To address these challenges, we strongly recommend that forensic experts focus on

improving the communication of complex scientific information in layman's terms, and criminal justice professionals should assess expert evidence with the aid of available quality standards, obtain counter-expert evaluations—especially if the expert is not registered as such, and improve their forensic knowledge to equip themselves to critically evaluate forensic evidence and potentially decrease the risk of misinterpretation and wrongful convictions in the judicial process.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Dutch Council for the Judiciary (*Raad voor de Rechtspraak*), the Board of Procurators General (*college van procureurs-generaal*), and the Dutch Association of Criminal Defense Lawyers (*Nederlandse Vereniging van Strafrechtadvocaten*) for granting permission to conduct this research. We would like to thank the scientists who contributed to the design of the study through their valuable input and reflections, and thank the experts of the Netherlands Forensic Institute (NFI) for their collaboration in drafting the forensic reports.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are restricted to the public. Restrictions apply to the availability of these data, which were used under license for this study. Data are obtained by the author(s) with the permission of the Dutch Council for the Judiciary and the Board of Procurators General.

REFERENCES

1. Butler J. The future of forensic DNA analysis. Philos Trans R Soc Lond Ser B Biol Sci. 2015;370(1674):20140252. https://doi.org/10.1098/rstb.2014.0252
2. Jakovski Z, Ajanovska R, Stankov A, Poposka V, Bitoljanu N, Belakaposka V. The power of forensic DNA data bases in solving crime cases. Forensic Sci Int Genet Suppl Ser. 2017;6:e275–e276. https://doi.org/10.1016/j.fsigss.2017.09.085
3. Mapes A. Rapid DNA technologies at the crime scene: "CSI" fiction matching reality. Amsterdam, Netherlands: University of Amsterdam; 2017.
4. de Poot C. Het gebruik van DNA in het opsporingsproces [The use of DNA in the investigation process]. Justitiële Verkenn. 2021;47(1):20–43. https://doi.org/10.5553/jv/016758502021047001003
5. Morgan J. Wrongful convictions and claims of false or misleading forensic evidence. J Forensic Sci. 2023;68(3):908–61. https://doi.org/10.1111/1556-4029.15233
6. Cole SA, Meterko V, Chu S, Cooper G, Paredes J, Possley M, et al. The contribution of forensic and expert evidence to DNA exoneration cases: an interim report. National Registry of Exonerations and Innocent Project. 2022. https://n2t.net/ark:/88112/x2g890
7. O'Brien É, Daeid N, Black S. Science in the court: pitfalls, challenges and solutions. Philos Trans R Soc Lond Ser B Biol Sci. 2015;370(1674):20150062. https://doi.org/10.1098/rstb.2015.0062
8. van Straalen E, de Poot C, Malsch M, Elffers H. The interpretation of forensic conclusions by criminal justice professionals: the same evidence interpreted differently. Forensic Sci Int. 2020;313:110331. https://doi.org/10.1016/j.forsciint.2020.110331
9. Mnookin J, Cole S, Dror I, Fisher B, Houck M, Inman K, et al. The need for a research culture in the forensic sciences. UCLA Law Rev. 2011;58(3):725–80.
10. Laporte G. Wrongful convictions and DNA exonerations: understanding the role of forensic science. Natl Inst Justice J. 2018;279:10–25.
11. Bonventre C. Wrongful convictions and forensic science. WIREs Forensic Sci. 2021;3:e1406. https://doi.org/10.1002/wfs2.1406
12. Dutch Supreme Court. ECLI:NL:HR:2018:2095. 2018.
13. Dutch Supreme Court. ECLI:NL:HR:2020:1604. 2020.
14. Dutch Supreme Court. ECLI:NL:PHR:2020:684. 2020.
15. Court of Appeal Arnhem-Leeuwarden. ECLI:NL:GHARL:2022:7605. 2022.
16. Corstens G, Borgers M, Kooijmans J. Het Nederlandse strafprocesrecht [The Dutch criminal procedure law]. Alphen aan den Rijn, Netherlands: Wolters Kluwer; 2018.
17. Cheng E. The consensus rule: a new approach to scientific evidence. Vand L Rev. 2022;75:407–74. https://doi.org/10.2139/ssrn.3826782
18. Visser M. De deskundige in het recht: een interne rechtsvergelijking van de normering van deskundigeninbreng in het straf-, privaat-en bestuursrecht bestuursrecht [The expert in law: an internal law comparison of the standardisation of expert input in criminal, private and administrative law]. Dissertation. The Hague, Netherlands: Boom Juridische Uitgevers; 2023.
19. Stevens L. Bewijs waarderen. Hoe doen strafrechters dat? [Valuing evidence. How do criminal judges do that?]. Nederlands Juristenblad. 2014;40:2842–50.
20. McGinnies E, Ward C. Better liked than right: trustworthiness and expertise as factors in credibility. Personal Soc Psychol Bull. 1980;6:467–72. https://doi.org/10.1177/014616728063023
21. Pornpitakpan C. The persuasiveness of source credibility: a critical review of five decades' evidence. J Appl Soc Psychol. 2004;34:243–81. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x
22. Harris A, Hahn U, Madsen J, Hsu A. The appeal to expert opinion: quantitative support for a Bayesian network approach. Cogn Sci. 2016;40:1496–533. https://doi.org/10.1111/cogs.12276
23. Costa S. DNA as 'ready-made evidence': an analysis of Portuguese judges' views. Int J Evid Proof. 2022;26:121–35. https://doi.org/10.1177/13657127211070331
24. Hoogeveen S, Haaf JM, Bulbulia JA, Ross RM, McKay R, Altay S, et al. The einstein effect provides global evidence for scientific source credibility effects and the influence of religiosity. Nat Hum Behav. 2022;6:523–35. https://doi.org/10.1038/s41562-021-01273-8
25. McCloskey M, Washburn A, Felch L. Intuitive physics: the straight-down belief and its origin. J Exp Psychol Learn Mem Cogn. 1983;9:636–49. https://doi.org/10.1037/0278-7393.9.4.636
26. Krause N, Brossard D, Scheufele D, Xenos M, Franke K. Americans' trust in science and scientists. Publ Opin Q. 2019;83:817–36. https://doi.org/10.2307/26952238
27. Chaiken S, Maheswaran D. Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity and task importance on attitude judgment. J Pers Soc Psychol. 1994;66:460–73. https://doi.org/10.1037/0022-3514.66.3.460
28. Mercier H. The argumentative theory: predictions and empirical evidence. Trends Cogn Sci. 2016;20:689–700. https://doi.org/10.1016/j.tics.2016.07.001
29. de Keijser J, Elffers H, Kok R, Sjerps M. Bijkans begrepen? Feitelijk en vermeend begrip van forensische deskundigenrapportages onder rechters, advocaten en deskundigen [Understood at chance? Actual and perceived understanding of forensic expert reports among judges, lawyers and experts]. The Hague, Netherlands: Boom Juridische Uitgevers; 2009.

30. de Keijser J, Elffers H. Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. Psychol Crime Law. 2012;18:191–207. https://doi.org/10.1080/10683161003736744
31. Malsch M, Taverne M, Kranendonk R, Elffers H, de Keijser J. DNA-rapporten: makkelijker kunnen we het niet maken, begrijpelijker wel [DNA reports: we can't make it easier, but we can make it more understandable]. Amsterdam, Netherlands: Boom–Lemma; 2013.
32. Cashman K, Henning T. Lawyers and DNA: issues in understanding and challenging the evidence. Curr Issues Crim Just. 2012;24:69–83. https://doi.org/10.1080/10345329.2012.12035945
33. Twisk K, Dubelaar M, Berger C. De kenniskloof verkend. Een onderzoek naar de waardering van (complex) DNA-bewijs in strafzaken [Exploring the knowledge gap. An investigation into the valuation of (complex) DNA evidence in criminal cases]. Expertise En Recht. 2019;3:105–11.
34. Hackman L. Communication, forensic science, and the law. Forensic Sci. 2021;3:e1396. https://doi.org/10.1002/wfs2.1396
35. De Roo R, Remijn W, Meeuwissen J, Kruithof-Van Esch J, Van Der Heijden S, Claushuis M, et al. Forensisch adviseurs actief bij alle gerechten in Nederland: tijd voor een terugblik [Forensic advisers active at all courts in The Netherlands: time for a review]. Expertise En Recht. 2021;6:234–9.
36. Meeuwissen J, de Roo R, Kruithof-van Esch J, van der Heijden S, Claushuis M, van Blijswijk-Kieftenbeld L, et al. Forensic advisers working for all district courts and courts of appeal in The Netherlands: an overview and discussion. J Forensic Sci. 2024;69(1):182–8. https://doi.org/10.1111/1556-4029.15385
37. Federale Overheidsdienst (FOD) Justitie Belgium [Justice Belgium]. Jaarverslag 2010 [Annual report 2010]. 2010. Available from: https://justitie.belgium.be/nl/publicaties/jaarverslag_2010. Accessed 9 Sep 2025.
38. Howes L, Kirkbride KP, Kelty S, Julian R, Kemp N. The readability of expert reports for non-scientist report-users: reports of forensic comparison of glass. Forensic Sci Int. 2014;236:54–66. https://doi.org/10.1016/j.forsciint.2013.12.031
39. Howes L, Julian R, Kelty S, Kemp N, Kirkbride KP. The readability of expert reports for non-scientist report-users: reports of DNA analysis. Forensic Sci Int. 2014;237:7–18. https://doi.org/10.1016/j.forsciint.2014.01.007

How to cite this article: de Roo RHD, Stevens L, de Poot CJ. The impact of institutional authority on forensic evidence evaluation by criminal justice professionals. J Forensic Sci. 2025;00:1–14. https://doi.org/10.1111/1556-4029.70190

APPENDIX 1

MODEL SOUND AND UNSOUND REPORTS

The content of the reports used in this study is derived from real cases. Therefore, the exact text cannot be made publicly available due to confidentiality restrictions. The sound variant remains strictly within the limits of what the findings support, avoiding speculation beyond the available data. The unsound variant goes a step too far by claiming excessive certainty or by implicitly suggesting that a hypothesis is more likely without direct justification from the data. For example:

Our analysis indicates that the bloodstain pattern observed at the crime scene exhibits characteristics consistent with both impact patterns and expirated patterns.

Sound report

An impact pattern arises from a forceful action on liquid blood, such as stepping in blood or blunt force trauma, whereas an expirated pattern occurs when blood is expirated through breathing or coughing. Based on the evidence, it is more likely that the observed pattern resulted from a forceful action impacting liquid blood than from the expiration of blood by the victim.

Unsound report

We analyzed the size and shape of the bloodstains and concluded definitively that the bloodstain pattern resulted from a forceful act, such as kicking or punching the victim. The bloodstain pattern did not result from blood expiration by the victim (e.g., breathing or coughing blood) and is therefore inconsistent with the claim that the injury was self-inflicted.

Remarks unsound report:

- The statement that the pattern resulted from kicking or punching is overly definitive.
- The conclusion that the absence of an expirated pattern excludes self-infliction is logically flawed; the lack of expirated bloodstains alone does not rule out self-inflicted injury.

To provide readers with a clear understanding of the structure of the used reports, the following model has been constructed.

[Title of the Report] Institutional logo [NFI/Police/ForensX] [DD-MM-YYYY]

Dear Commissioner,

Please find enclosed a concise report presenting the conclusions of the forensic [pathological/bloodstain pattern analysis] examination written by the [NFI/Police/ForensX].

[Pathology: Previous investigations established that the victim died as a result of [type of injury and cause of death]]/[BPA: The forensic bloodstain pattern analysis conducted is based on the bloodstains [description blood patterns] present at the crime scene].

Based on the forensic [pathological examination/bloodstain pattern analysis], the following conclusions can be drawn:

- Description of [Pathology: fatal wound]/[BPA: bloodstain pattern] and interpretation
  - Sound (supported claim)
  - Unsound (overly assertive and unsupported claim)
- Description of [Pathology: condition of the skin around the wound]/[BPA: blood expiration pattern] and interpretation
  - Sound (supported claim)
  - Unsound (overly assertive and unsupported claim)
- Description of [Pathology: additional found injuries]/[BPA: luminol testing] and interpretation
  - Sound (supported claim)
  - Unsound (overly assertive and unsupported claim)
- Literature-based likelihood of [Pathology: wound location]/[BPA: observed blood pattern]
  - Sound (logically correct conclusion)
  - Unsound (instance of prosecutors' fallacy)

The findings have been evaluated in light of the scenarios put forward by the parties involved under the following mutually exclusive hypotheses:

- Hypothesis 1: [Scenario 1]
- Hypothesis 2: [Scenario 2]

Concluding

- Sound (logically correct conclusion): The [pathological/bloodstain pattern analysis] findings are “more likely” to “much more likely” under Hypothesis 1 than under Hypothesis 2.
- Unsound (instance of prosecutors' fallacy): It is “more likely” to “much more likely” that Hypothesis 1 is true than Hypothesis 2.

For any additional research questions, a new request must be submitted.

Kind regards,

[Name], Forensic [Pathologist/bloodstain pattern expert] [NFI/Police/ForensX]

[Signature]