

I have a question about evaluative reporting: a comprehensive collection of FAQs

R. Brent Ostrum

Canada Border Services Agency, Ottawa ON, Canada

Corresponding author. Email: Brent.Ostrum@cbsa-asfc.gc.ca

ORCID: 0000-0001-9214-1940

Abstract

With various calls in the forensic scientific literature for pattern-matching disciplines to adopt evaluative reporting (ER), which is also known as the likelihood ratio or logical approach, practitioners used to reporting evidence under a more traditional paradigm may have a level of apprehension, and face a steep learning curve. There may be doubts about the benefits of ER, when it applies and when it does not, what it actually entails, its validity, and how to explain it in court. These questions and more are answered in this comprehensive collection of frequently asked questions on ER. While the focus is on providing context for forensic handwriting examiners, many of the questions and answers are general in nature and the concepts can be applied to other disciplines.

Introduction

In recent years, numerous publications have discussed approaches to forensic science reporting, with calls for disciplines to abandon conclusions expressed as a degree of confidence about a categorical source attribution in favor of an approach whereby conclusions are offered on the relative support that the evidence provides for one proposition over an alternative proposition (see, for example, Association of Forensic Science Providers, 2009; Willis et al., 2015). This is variously known as evaluative reporting (ER), the likelihood ratio (LR), or logical approach. This paradigm shift may be difficult or daunting for some practitioners to contemplate, which is reflected in the number and type of queries received and discussions had during various educational sessions on the topic.

The frequently asked questions (FAQs) included in this paper have been drawn from questions asked of the author and others during workshops and training events on ER in forensic document examination (FDE) or forensic handwriting examination (FHE). The answers that are provided reflect the personal understanding and beliefs of the author, with a few key references provided where relevant. Others may see things differently and, as a result, may prefer a different approach. In that regard, it must be noted that the ER approach is very flexible, and can be implemented in many different ways.

For simplicity, the FAQs have been grouped by topic. Section 1 covers basic thoughts and key ideas, Section 2 delves into the theory and basis of the approach, and Section 3 discusses error rates and validation. In Section 4, details of the process are addressed, including setting propositions, defining the relevant population (RP), determining expectations, and the effect of limitations on the evaluation. Section 5 deals with assigning numeric values, and Section 6 answers questions related to explaining the approach, particularly in testimony.

Section 1: basic thoughts and key ideas

FAQ #1: What are the key benefits of doing things this way?

ER helps to ensure balance, transparency, logical consistency, and robustness in the formation of the final opinion.

"Balance" comes from considering at least two competing and mutually exclusive propositions.

"Transparency" comes from clear disclosure of what was, and was not, considered in the reasoning.

"Logical consistency" comes mainly from the examiner focusing on the probability of the evidence, and not the probability of the propositions (i.e., avoiding transposition). "Robustness" relates to the ability of the examiner to explain their opinion and address any variation in relevant parameters (e.g., different propositions or framework information).

The ER approach provides structure to the process and to our reporting, makes it easier for all parties to understand the process and the outcome, ensures that the evaluation is complete and fair, and lets the examiner adjust their belief, as appropriate and without fault, when being cross-examined.

FAQ #2: What do I gain using this approach? How does it help me?

Learning how to apply this approach is well worth the effort.

It helps to ensure that the examiner is undertaking a thorough and balanced evaluation, it is much easier to be transparent about what was, and was not, carried out to reach any opinion, it makes it easier to be logically consistent in your reasoning, and, finally, it makes your opinion easier to explain and to defend.

The latter point should not be overlooked or undervalued. Using this approach, it is remarkably easy to handle ANY cross-examination – assuming your process was undertaken properly in the first place.

Why and how? Any legitimate cross-examination question is just some variation on the conditioning factors at play in the evaluation. That variation is equivalent to a change in either a specific sub-proposition under one of the main propositions, or some variant on the framework information, or both. When asked a question, you need only restate it using proper terminology, then explain the expectations that would apply with that new information in the mix, and finally explain the effect that has on the resulting opinion, whatever that might be. For some examiners, this will be basically what they have always done, even though they may not have been fully aware of it. For others, it is a different way of doing things.

Examiners have historically been taught to "defend" their position or opinion strongly and without truly considering the possibility of a change in their opinion. Very often, that has led to responses to legitimate challenges that involve misdirection, wordplay, or even deception, rather than addressing the matter at hand. Understanding that any opinion may change with a change in the conditioning elements is critical, and fairly obvious once you think about it.

Of course, silly or inconsequential questions are sometimes asked, but they can readily be dismissed by explaining how they do not apply to the evaluation (i.e., they are not part of the findings/evidence, propositions/hypotheses, or framework/information elements), or they relate to something outside the examiner's domain of expertise.

Finally, it is important to understand that there is nothing in the ER approach that should not also be in the traditional approach.¹ The traditional approach is literally an extension of the ER approach wherein the latter ensures that elements of reasoning have been made clear, and limitations in the process have been observed.²

Thus, if an examiner should encounter a situation where they are having difficulty conducting the evaluation (i.e., assigning a probability) whereas they felt that they never had any trouble when

using the traditional approach, they will have discovered a weakness or issue in the traditional approach. At the very least, it is an area they need to consider carefully because something may have been missed or overlooked in the past.

FAQ #3: What changes when using the ER approach, relative to the “traditional” approach?

Some examiners find it difficult to grasp the key change in reasoning that is required when using the ER approach. Historically, examiner training has been predicated on the idea that it is possible to determine (with varying degrees of certainty) the cause of some observed traditional outcome. In other words, that an expert can tell what actually happened based on the features observed in the resulting evidence. In many instances, examiners are taught that evidence pointing one way effectively rules out the possibility of it pointing in any other direction.

This idea makes it difficult to accept that evidence can be, and generally is, more or less expected in a variety of situations. The bottom line is that the observed evidence may arise from more than one cause – but it is more likely the result of one cause, rather than some other cause.

Expressing an opinion based on the likelihood ratio (LR) or some similar construct reflects that reality. The LR is the ratio of two conditional probabilities that is ultimately a statement about the relative support for possible potential causes, taking into account relevant information. In other words, when considering a source determination, the observed feature set (i.e., the features seen in the questioned item together with the observed “match” between the questioned item and a proposed source) exhibits non-zero probability when considering either the proposed source or some other source.³ In that sense, a proper evaluation will always be at least bidimensional and often multidimensional in nature, especially if you break things down into sub-propositions, or try to address multi-writer scenarios.

This can be a challenge for forensic examiners who are trained to think that the better the match between the questioned and specimen samples, the more likely it is that the suspect is the actual source. However, that is not necessarily the case. The problem with this belief is that even a perfect match cannot equate to the same source on its own. Such an assignment means that the evidence conforms perfectly with what is expected **if** the suspect was the perpetrator who deposited the evidence.

Thus, the numerator of the LR may be close to 1, or even equal to 1, but that is only part of what must be considered. The real question, and the harder thing to assess, is the probability of observing that same evidence **if someone else** was the perpetrator. In some situations, that value conforms to a type of random match probability (RMP), but it may also be something quite different.⁴ If that probability is high (and, in some situations, it may potentially be equal to 1), then the overall value of the match will be considerably less than definitive.

Assessing evidence in the above manner is very straightforward in some disciplines, particularly those where quantified data are available to inform the assignment of probabilities (e.g., forensic

DNA analysis). However, it is more challenging in disciplines such as FHE where the probability reflects the personal belief of the examiner in accordance with their training and experience. Expressing subjective probability for this purpose is fine, but it is a very difficult concept to embrace for examiners, who very often reject the concept of uncertainty or the need for a probabilistic expression in their conclusion.

Transposition is another aspect of the reasoning process. Traditional opinion scales are statements of the posterior probability (or odds) of the propositions. An LR, or any equivalent construct, is a statement about the evidence, not about the propositions, and the relationship between the two is clearly expressed by Bayes' theorem. When lacking knowledge of prior odds, there is no logical way to use a LR to determine posterior odds for the propositions.

As a result, the two forms of opinion are not at all equivalent, comparable, or transferable. However, when a new way of reasoning and reporting is presented, examiners seek some sort of concordance between the two forms. It is possible to achieve this, but it is a non-trivial issue because of the issue of prior odds, which relate to other elements of the case that fall outside the scope of the forensic document examiner. The main benefit of discussing such a concordance would be to disclose the complex interrelationship that exists between the LR and our traditional conclusions, thereby exposing the nature of traditional opinions, which are much more complicated and involved than any LR.

See also **FAQs #5, 7, and 8**.

FAQ #4: How is this approach different from what we have always done?

For some examiners, it is not very different at all.

For others, it may help to consider the following points outlining how the ER approach differs from the traditional approach:

- 1) In ER, the opinion speaks to the evidence, and not the propositions. There is no transposition of the conditional.
- 2) In ER, conditionality and uncertainty (which are omnipresent in any evaluation regardless of the approach taken) are expressly acknowledged and addressed.
- 3) In ER, all elements of the evaluation process are clearly stated and permit more detailed reporting and disclosure.
- 4) In ER, the examiner never risks overstepping the bounds of their expertise, and cannot possibly be charged with usurping the authority of the court. This ties back to point 1 above.
- 5) In ER, clarity and thoroughness of reasoning are more easily achieved. This relates back to points 2 and 3 above.

Section 2: theory and basis

FAQ #5: What is the basis for the opinion?

Ultimately, **the examiner** has to explain the process that they used, whatever that might be. The process is **unique to the examiner**, being based on their knowledge, skills, abilities, and whatever information they brought to the process.

The basis for a given examiner's opinion is whatever they used to assign probabilities to the findings, and thus that is ultimately what matters in terms of the answer to this question. Those probabilities, the numerator and denominator of the LR, represent the examiner's belief about the findings in terms of each applicable proposition. Hence, **those probabilities** are the basis for the final opinion.

The examiner may choose to discuss probability assignment as being either data-driven or otherwise. In addition, the examiner will likely want to expand their response to discuss the opinion wording/scale being used, but that should be a straightforward process.

See also **FAQ #27**.

FAQ #6: When is ER applicable, and when is it not applicable?

ER is necessary whenever the report involves some level of evaluation of the findings aimed at providing an assessment of the strength of those findings in the context of the alleged circumstances. Hence, most reports issued by forensic document examiners will be evaluative in nature. At the same time, the European Network of Forensic Science Institutes (ENFSI) guideline (Willis et al., 2015) briefly discusses four general types of reports that may be issued, one of which is evaluative in nature. The other three are intelligence-based, investigative, and technical. If the work conducted on the matter does not involve evaluation of the findings, then one of the other report types will suffice.

FAQ #7: Why can't we say that someone wrote something when it is obvious that they did?

This question refers to a categorical or definitive conclusion, which is a transposed statement about the probability that the proposition is true (i.e., "The specimen writer wrote the questioned signature") **and**, further, is expressed in absolute terms (i.e., the examiner believes that **is** what actually happened and nothing else is possible). There are examiners who, in at least some situations, will feel that such an opinion is obvious and perfectly justified.

In general, it is safe to say that many examiners understand that categorical conclusions cannot, or should not, be given. This may be an acceptance of the argument that we can never be 100% sure about anything, or it may relate to some form of policy directive to that effect from their agency/laboratory.

This type of question generally reflects a belief derived from the training that most examiners (or at least those using the traditional approach) have received that teaches them to automatically

transpose the conditional while ignoring any logical inconsistencies this may involve. Specifically, examiners tend to ignore (or are unaware of) inherent conditionality that affects such an opinion. In addition, issues relating to prior odds (i.e., beliefs about the propositions) are ignored.

Basically, if an examiner wants their opinion to be logically justified and fully representative of the evidence, it is critical to use procedures that **ensure** that 1) proper logic has been followed, 2) all conditioning elements are included and considered, and 3) any uncertainty is properly acknowledged in the opinion. The latter, in particular, means that any statement about actual authorship, regardless of whether it is conclusive or less definitive in nature, cannot be logically justified.⁵

Examiners must remain focused on the probability of the evidence and be careful not to extend the discussion to any statement about the probability/certainty of the propositions. Fortunately, it is possible to achieve all of this using the ER approach. It simply requires a slightly different approach to our work.

The topic of reaching formal decisions via decision theory is beyond the scope of this article, but some significant issues are involved. In that regard, there is a large body of interesting literature worth reading (e.g., Biedermann et al., 2016; Biedermann et al., 2017, Biedermann et al., 2018; Taroni et al., 2020; Taroni et al., 2021).

FAQ #8: I understand that categorical conclusions are not scientifically acceptable, but why can't we say that it seems very likely that someone wrote something?

This issue is very closely related to the point discussed in **FAQ #7**. **Any** statement about the probability (or truth) of the proposition(s) extends beyond the scope of our expertise because it requires information that we do not have and **should not** have. This applies equally to a categorical opinion and to one that is expressed in probabilistic terms.

As noted in **FAQ #7**, examiners must remain focused on the probability of the evidence and be careful not to extend the discussion to any statement about the probability/certainty of the propositions.

See also **FAQ #3**.

FAQ #9: What are the "two faces of probability"?

It is not so much a matter of "two faces", which implies that the two are somehow in opposition, but that there are simply different views about the nature of probability, both of which are equally valid and appropriate.

Most people are familiar, if not completely comfortable, with empirical probability (i.e., frequentist or aleatory probability). Empirical probability is data-driven, being based on observed or historical

data, and is often generated by formal research and study that involves data collection. That is, it applies to situations where events can be repeated.

People are generally less comfortable with subjective probability (i.e., epistemic probability), which relates to a person's judgment about the probability of occurrence of a single (generally non-repeatable) event. This probability is based on that individual's past training, experience, and/or analysis of a situation. It is **very** important to note that subjective probability is not arbitrary in any sense, so long as the examiner can provide justification for the probability assignment (i.e., they can explain why they assigned a particular probability).

The latter is key to FDE work, and thus a good understanding of it is essential. If you have difficulty with this, please consider consulting the following references:⁶

- Lindley DV. The philosophy of statistics. J Royal Statistical Society. Series D (The Statistician). 2000;49:293–337.
- Lindley DV. Letter to the editor. Teaching statistics. 2002;24:22–23.
- Lindley DV. Understanding uncertainty. Hoboken (NJ): John Wiley & Sons; 2006.

As an introduction to the latter text, the author recommends Biedermann's 2013 review which states, in part, that:

The book is not primarily about the *calculus* of probability (indeed, mathematics are kept to a strict minimum); it is about the very *meaning* of probability — how probability ought to be understood in order to deal with uncertainty. On this latter point, Professor Lindley is uncompromisingly clear and, at the same time, draws yet another parallel to de Finetti's two-volume work on probability...: probability is the measure for strength of belief, but probability *does not exist* in the sense of being a property of the outside world.

In the absence of empirical data, and even when such data exist, the view of probability that applies to our work will necessarily always be subjective. This is key to understanding how the ER approach works.

FAQ #10: Why do I have to work with probability? I hate probability.

First, it may help to realize that examiners have **always** worked with probability. Certainly, if one has been doing FDE work, they have been assessing uncertainty in one way or another, and it is important to understand that *probability is simply the language of uncertainty*. An examiner may not have been using that language in any formal way, but they will have been working with probability all along.

Second, some people are very concerned about the mathematics that may be involved. It is true that some of the approaches can involve complex calculations. However, that is not strictly necessary. In reality, very little math is needed to apply the basic approach, and what math is involved is quite

straightforward. In that regard, consider this quote from Lindley's book Understanding Uncertainty (2006, p. 23):

Anyone taking this volume off the shelves of a bookstore and flipping through the pages will see some mathematical symbols, which may frighten them and lead to their replacing it on the shelves unread and unpurchased. Perhaps there should be a health warning attached—"Danger, this book contains mathematics".

This would be a pity for it would better read—"The germ of mathematics contained herein is harmless." There is some mathematics here, so let me try to explain why and of what it consists, for it really is harmless, indeed it is positively therapeutic.

Third, another issue for many people is the idea of quantifying their personal beliefs. However, that is ultimately an exercise in precision—specifically, being precise about the probability (uncertainty) that applies to a situation. We may not be used to doing this, but it is not all that difficult. It just takes practice, not unlike the way in which examiners learn to express opinions using the traditional approach.

As Lindley puts it, probability is a personal/subjective expression of a belief held by a person. The trick to all of this is simply learning how to express those numeric values in a meaningful way. That is a uniquely personal process.

Note: epistemic and aleatory probabilities may relate to one another. For example, when empirical data exist, they can be used to inform one's personal (subjective) belief and, in fact, that **should** be done. Note also that a frequency is not a probability, although frequencies can be useful in helping to elicit or assign probabilities.

FAQ #11: What is the hierarchy of propositions, and how does it apply to FDE/FHE work?

Cook et al. presented a hierarchy of propositions as part of the case assessment process in two articles (Cook et al., 1998a, 1998b).

The concept is fairly straightforward when you consider the overall process. We assess evidence in terms of competing propositions, but those propositions will always be somewhat removed from those that the jury must address. To aid the discussion of these issues, it is helpful to consider propositions as falling into three broad levels: source, activity, or offence.⁷

The first, and usually simplest, level of propositions relates to the potential **source** of evidentiary material. At this level, the assessment focuses on elements inherent in the evidence that serve to distinguish it from other samples of the same type of evidence, as well as factors that cause those elements to vary. The next level of propositions relates to **activities** that may have led to the evidentiary material. Proper interpretation at this level is more complicated because it requires consideration of additional sources of variation beyond those inherent to the trace, essentially shifting in most situations toward human behavior and related issues. The highest level of

propositions relates to the actual **offence**, and thus is clearly beyond the realm of the examiner. This level effectively corresponds with the propositions that a judge/jury must consider in the course of a trial.

Sets of source-level propositions generally require little in the way of circumstantial information. However, activity-level propositions cannot be addressed without a framework of circumstances because they relate to such things as evidential transfer and persistence, among others.

The issue, from our point of view, is that handwriting is not easily classified. Does our work belong to the source level or the activity level? Many examiners think of the problem as one related to source, but a strong argument can be made that it falls within the realm of activity.

First, writing is an activity, with handwriting samples forming a written record of dynamic human behavior. Second, it follows that handwriting can be affected by a huge number of sources of variation, meaning that relevant contextual information is critical to proper interpretation.

Thus, it is always best to function at the higher level because this helps to ensure that consideration is given to all relevant sources of variation.

FAQ #12: What are appropriate prior probabilities to use?

Prior odds reflect the probability that the propositions are true, before or without consideration of the evidence the examiner is considering. The overall process is one of updating prior odds to **posterior odds** through the provision of new, domain-specific evidence from an expert.

A key question is: whose prior odds require updating? Clearly, the belief that matters most is that of the **trier-of-fact**, and **not** the examiner, even in those situations in which the examiner may hold some belief about the matter that could form the basis for setting prior odds.

To be clear, both prior and posterior odds reflect the belief of the trier-of-fact about the propositions. Testimony by the expert provides new information to the trier-of-fact, information that they cannot obtain for themselves because it requires our expertise. In that regard, we will never know and, arguably, **should** never know the information required to set such priors because that must relate to other non-forensic aspects of the case.

This is very much a matter of who *should do* what, and not really who *knows* what. It is likely that most examiners would not feel comfortable setting prior odds even if given all the information relevant to the case because that is not the role of the expert

Equal priors have been suggested as a default by some authors. However, that approach raises significant issues because equal priors are inappropriate in almost all instances, and thus the examiner is required to make numerous unjustifiable assumptions regarding the situation (Biedermann et al., 2007; Taroni & Biedermann, 2005). Most importantly, prior probabilities **are not required** to enable us to do our work, which is to assess the findings under each proposition (i.e., to determine an LR).

Another option is to outline various possible prior odds in an effort to demonstrate the end effect when a given LR/opinion is used. This approach is fine because the examiner is still focused on the LR. Discussion of possible priors occurs solely with the aim of helping the trier-of-fact to understand the impact of including the expert's evidence and opinion. For more information on this, see Appendix C in Marquis et al.'s (2016) article "Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings."

Section 3: Error Rates and Validation

FAQ #13: How do laypeople perceive ER, and how might it be validated?

This question touches on two issues.

Layperson perception

In the author's experience, laypersons have no difficulty grasping the essence of the ER approach, nor what it involves in application. Indeed, most of them do not think it is unusual or different, in fact it is often precisely what they *expect* us to be doing.

In places where the ER approach has been used, often for many years, there are rarely questions from clients. Feedback on the process and the results is generally positive.⁸

However, there is a legitimate issue of whether or not a layperson actually understands what is going on. In that regard, clear explanation is key. In terms of explaining the approach to a layperson, it is important to explain that the ER approach is, first and foremost, a system of logic and reasoning, a process that, by design, helps to ensure complete and proper reasoning. Our aim is to fully assess the evidence in terms of all issues of interest to the trier-of-fact, as far as we understand them to be. The opinion that we present provides the decision-maker with a proper understanding of where the evidence points and the degree of strength that should be allocated to that evidence in the estimation of the examiner.

It is a system of logical reasoning that helps ensure that the four key requirements are met (see **FAQ #1**). Its aim is to explain the value or weight of evidence in a specified context, and it is based on the concept of the LR, or the Bayes factor.

While the author believes that laypersons can easily understand the fundamentals of the ER approach (see **FAQ #28**), it is acknowledged that further research is needed to determine 1) how laypersons actually perceive or apply this form of opinion (or any other opinion for that matter) for their own reasoning, and 2) the most appropriate and effective wording for communication purposes.

The goal of that research would be to gain a better understanding of how best to communicate this information effectively. In other words, it is the author's belief that the core concepts underlying

the ER approach are entirely sound, in that they convey the information that we should be giving to the trier-of fact.

There remains the issue of how best to convey this information to laypersons.

Validation

When the issue of validation arises, it can be helpful to begin by considering the present state of affairs. Very often, issues regarding validation are raised by people who want to continue to use traditional methods and wording. Thus, a first question is, “How well validated are our present methods?”

Why start there? Mainly because any existing validation should, for the most part, still apply. This derives from the fact that the basic method used for analysis, examination, and comparison of material does not change when using the ER approach. If we assume that examiners have been following the standard procedures all along, then it should not change. What changes in the ER approach is that the key elements relating to evaluation are clearly defined and easier to explain, with much better transparency and clarity (see **FAQs #2 and 4**).

The evaluation process **is** different, being focused solely on the findings and not the proposition(s), but the analysis and comparison remain the same. Thus, a solid argument can be made that whatever validation has been conducted using the traditional approach should also apply, for the most part, to key aspects of the ER approach.

It is important to distinguish between a logical construct (i.e., a mathematical truth) and the application of a practical method. In this regard, the ER approach to reasoning and reporting cannot be considered invalid in any conceptual or overall sense because it simply involves the use of well-established logic and reasoning.⁹ Thus, the structure and approach are inherently valid.

The conclusion scale that one chooses to use might also be considered. In that regard, it is difficult to see how a descriptive scale can be considered invalid. A descriptive scale is aimed at accurately describing or conveying the outcome of any forensic evaluation. Thus, the scale is designed to span all possible outcomes in a sensible and logical manner. Most importantly, the meaning of the scale is clearly defined. In that sense, the scale is also inherently valid.

Therefore, where does validation come into play? It applies to how examiners do their work and express their opinions using that scale. In other words, when an examiner **uses** the scale to express their belief, is it being done in a reliable and accurate manner? Another factor is how well the end result accurately reflects the underlying situation it is intended to portray. Those are both important issues that remain to be addressed.

In fairness, though, the same can be said for our traditional opinion scale. The research that has been undertaken thus far to validate our method and scale suggests that the methods are reasonably accurate and reliable, certainly better than any alternative. However, there is more that the discipline

can and should do in that regard. In the author's opinion, if further research is going to be undertaken, it should focus on the ER approach.

Does the lack of complete validation preclude the use of the ER approach? The answer is clearly "No," otherwise we should not be using the traditional approach either.

First, validation is a process that is never complete, but rather is more an issue of determining how well a process works, and applying some (often vague) criteria in an attempt to validate it. Second, if using any method, one that is at least logically sound and proper is preferred.

See also **FAQ #15**.

FAQ #14: What is the error rate associated with the ER approach?

This is not an easy question to answer, but in fairness, it is not easy to answer in relation to the traditional approach either. Insofar as there is no significant change in the basic method being used, the existing error rate continues to apply.

Why? Existing proficiency studies that give some idea of error rates do so by collapsing answers into binary outcomes (i.e., strength is usually discarded). Thus, opinions expressed using the ER approach can be considered equivalent so long as the assumption is made that traditional opinions of association/non-association require LR values >1 and <1 , respectively, and would be expressed in that manner. This is an obvious and reasonable relationship, which means that whatever error rate applies based on existing studies should also be reasonably accurate in relation to the ER approach.

Remember, the key difference between the two approaches is related to how the evaluation is conducted, by ensuring that the process is balanced and thorough. If anything, proper and complete evaluation **should** improve the accuracy of the results, although inter-examiner reliability may remain the same. However, this appears to be a legitimate approach, it is difficult to know the precise outcome until appropriate formal testing has been conducted using the ER approach.

FAQ #15: How can error be incorporated into the LR and the opinion?

Once again, it helps to consider how error is currently addressed in our work and how it is reflected in traditional opinions. There are two aspects to the subject of error:

- 1) The error rate, which can be useful in determining whether a discipline should be accepted by the court. That information does not need to be integrated into an LR value.
- 2) Potential error in the opinion, or what some would call the error risk in a case, that is, the chance of providing a wrong opinion. Here, the term "wrong" may be interpreted in various ways, but we can say that an opinion is wrong if it points toward a proposition that does not fit the ground truth. Of course, in casework, the ground truth is unknown.

First, it should be acknowledged that we do not have a good handle on a specific error rate that applies to our work. There have been several studies related to different tasks, mainly handwriting

and signature assessments, the results of which indicate that there is a reasonably low error rate for those tasks, certainly much lower than that of laypersons.¹⁰ However, it is doubtful that most examiners attempt to formally incorporate an error rate into any current method or procedure.

Second, notwithstanding the above point, traditional opinion scales have always related to potential error in our work, although how they do so is unclear to most people. This is the basic purpose for the scale. On the basis of the terminology used to explain most traditional scales, they relate to an examiner's degree of confidence. On reflection, it seems clear that the degree of confidence should be inversely related to the potential for error in some way. That is, the strongest opinion should be expressed when the potential for error is lowest, while the weakest opinion should be expressed when the potential for error is highest. However, how many examiners express things in that form when discussing their opinions, or the scale they have used? It should also be noted that this approach has been used for a long time, even though we do not have a good handle on the error rates that apply to our work.

Little changes when using the ER approach, although the relationship between the expressed opinion and potential for error may ultimately be clearer. As noted in **FAQ #13**, the analysis and comparison steps are the same in the ER approach, and only the evaluation changes, with a focus on the probability of the findings rather than the propositions. Thus, there is little reason to think that the potential error rate associated with the ER approach would differ significantly from that associated with the traditional approach.

A key point to remember is that the ER approach expressly incorporates uncertainty into the evaluation. After all, the main elements of the LR are conditional probabilities. Ultimately, the opinion must reflect the level of uncertainty present in the evaluation, captured in either probability, with the basic relationship being that as uncertainty increases, the strength of the opinion should decrease (i.e., move toward LR=1). This is what has been done using traditional scales, although it is true that those scales do it in an obscure way.¹¹

The ER approach clarifies the relationship between the level of uncertainty and the opinion. That is particularly true when using a scale that conforms to the ENFSI guideline (Willis et al., 2015) because that guideline provides a clearly defined relationship between LR values and the wording of the opinion.

In practical terms, and as noted in **FAQ #14**, the presence of limitations, which are directly related to the potential for error, tends to shift the LR numerator and/or denominator values toward the most indeterminate value. While the specific effects on the numerator and denominator may differ, the overall end result is that the LR will tend to move toward a value of 1, that is, a state of roughly equal support for both propositions, which is also the highest level of uncertainty.

Finally, it is even possible to base the strength of an opinion on an empirically derived LR for a given task, which is currently impossible because of the lack of appropriate research and data, but is definitely a feasible approach.¹²

Section 4: process details

FAQ #16: Is there a standard set of propositions we should use?

Not really. In theory, and ideally, we want to use whatever propositions reflect the arguments to be made by the parties in court. If those positions are known, we should set the propositions accordingly. However, in practice, we rarely know the arguments that are going to be made by both parties.

Examiners should have a good idea of at least one position (that of the party engaging them), and thus one proposition for the set is known.

It is the alternative position that is more difficult to determine. To set that, we need to consider the information that we are given regarding the scenario and framework.

The bottom line is that there is no standard set of propositions that we should use.

FAQ #17: Why aren't alternative propositions formulated as a negation of the main proposition? Why are we told that "Someone other than X wrote the questioned sample", and not "X did not write the questioned sample"?

This is not a hard and fast rule.

To understand this issue, it helps to differentiate between forming an alternative based upon the negation of the main proposition and the use of negative terminology for the alternative proposition.

The former is an acceptable option. When no alternative is known or provided, an examiner can consider the negation of the main proposition to form the alternative. On the plus side, this provides an alternative proposition that they can use to conduct a viable evaluation. It should be noted that such a choice is often less than ideal, and would likely result in extensive cross-examination or debate. The simple negation may be poorly defined and/or likely to be inappropriate because it may not be representative of the real alternative.

However, that does not mean that the wording used in the proposition itself needs to be stated in the negative. The main reasons for using positive terminology are as follows:

- Negative wording has an implicit/inherent bias insofar as people tend to consider such descriptions in a negative light. If both propositions are formulated in the positive sense, this avoids a potential, albeit subtle, bias.
- It is important to emphasize that the specimen writer is NOT the key to the alternative (in most cases), and framing the proposition in the positive sense helps to make that

completely clear. Under the alternative proposition, the writer of the questioned item is not the specimen writer, but rather someone other than the specimen writer.

As for the wording that might be used, there are a few more or less equivalent options. For example:

- Someone other than the specimen writer wrote the questioned sample.
- Some unknown person, drawn from the RP, wrote the questioned sample.
- Some randomly-selected writer, other than the specimen writer, wrote the questioned sample.

It should be noted that any bias deriving from this wording will be subtle. As an aside, similar biases may be present in the choice of which proposition goes first, or the use of colors such as green and red to signify similarities and differences on charts (e.g., green = similar = good, red = different = bad). Examiners can introduce implicit/latent signals of a very subtle nature into the work, often unintentionally. However, we should try to avoid this as much as possible.

FAQ #18: What if I have no propositions?

In a situation where there are no propositions at all, a proper, balanced evaluation is impossible. However, that is a rare situation. Submissions generally include a question from the contributor, and as a rule, that question provides one proposition — usually the main proposition of interest.

It may also be possible to evaluate the findings under a single proposition, but that will result in some form of non-evaluative report type (see **FAQ #6**).

During an investigation, if there is no set of competing propositions, the client may simply ask about potential causes that might be considered. Forensic document examiners have done this sort of thing for a long time, but it is important to understand a key distinction between an explanation and a proposition. As the ENFSI guideline (Willis et al., 2015) indicates:

In the context of a forensic science evaluation, an explanation has been recognised as an intermediate consideration for use when exploring less formal alternatives. A key characteristic of explanations is that they are generated after the forensic findings have been obtained. While an explanation has the potential to account for particular observations, it does not qualify as a formal proposition because - often - it may be a statement of the obvious, speculative or fanciful. Moreover, an explanation can be offered provided that parties have presented no exclusive alternatives.

Further:

Provision of such a range of explanations is not an evaluation of the probative force of the findings.

[author's emphasis]

In contrast:

Propositions are statements that are either true or false, and that can be affirmed or denied... Propositions should be formulated in pairs (e.g., views put forward by the parties to the cases) and against a background of information and assumptions. Moreover, they should be amenable to a reasoned assignment of credibility by a judicial body and be useable for rational inference. Propositions should be distinguished from explanations that do not have the aforementioned properties.

The key point is that propositions are based on competing theories regarding the situation, and thus the set of competing propositions should be formulated and presented before any forensic findings have been determined.

To conduct a complete and proper evaluation, the examiner needs at least two competing, mutually exclusive propositions. In practical terms, the issue often comes down to setting an appropriate alternative/competing proposition because, as noted above, the question asked in the submission will likely provide a reasonable main proposition of interest.

Of course, there is always one alternative, even when none is either given to you or apparent from the scenario: the negation of the main proposition. However, as noted in **FAQ #17**, that is rarely a perfect choice. Nonetheless, in the absence of any other options, it can permit a balanced evaluation.

Strictly speaking, the best, or even merely an appropriate alternative proposition may not be known to the examiner, at least, not until they are testifying in court. When considering propositions to be used, the examiner should:

- 1) Consider all facets of the scenario/framework and remain as general as possible, and
- 2) Present the proposed set of propositions to the submitter to determine whether they agree that these are acceptable and appropriate, given their understanding of the matter.

In any event, whatever set of propositions is used **must** be disclosed. Doing so permits the set to be either accepted or challenged, as appropriate.

To ensure that the trier-of-fact understands the significance of the propositions to the opinion, it is important to include a disclaimer to the effect that the opinion depends upon the set of propositions used, as well as other information. Such a disclaimer helps to ensure that the significance of the set used is more fully understood.

FAQ #19: Why is it necessary or desirable to use a particular set of propositions?

To conduct a balanced evaluation, the examiner needs a set of at least two competing and mutually exclusive propositions. Ideally, we should be using whatever is desired by the parties/court. However, unfortunately, these are often not known to us.

Basically, the issue is whether there are any particular propositions of interest. If so, we should use them. If not, we need to address the issue differently.

On a more fundamental (and critical) level, evidence really has **no** strength in itself. It **must** be contextualized in relation to the matter at hand, and it gains significance only when considered in the light of specific propositions.

The evidence may be stronger if a given proposition is true, rather than if the alternative proposition is true. However, if the alternative proposition changes, the strength of the evidence can change.¹³

That is why it is necessary to specify the propositions clearly when providing any opinion relating to them.

In that regard, see **FAQ #3**.

FAQ #20: What background information (framework) was considered?

The answer to this is completely case-dependent, which is not surprising because every evaluation is undertaken in a specific case context or framework.

The examiner should respond to this question by explaining any effect/use of the information they were given (or to which they were exposed). In theory, critical elements of any framework should be outlined in the report, and thus this should not be difficult to explain.

Framework information of value is anything that might affect the opinion, or is needed for proper evaluation. That information 1) may be used to formulate a specific alternative proposition, or 2) may serve to moderate the opinion in some other way.

A key factor here is relevance. Basically, any information that influences the value of the evidence should be considered relevant (e.g., writing conditions, medical condition of the writer, and whether the writer had an opportunity to practice beforehand). This information can impact the probability of the evidence either when a given proposition is true or when the alternative proposition is true. Relevant information **SHOULD** always be considered in the evaluation. Conversely, irrelevant background information refers to any information that is not useful for the evaluation (e.g., other expert opinions, testimonies, and suspect confessions), and should not be made available to the examiner because it may bias the evaluation.

FAQ #21: What is the RP?

The RP can be conceptualized and explained as follows:

1) When considering the findings under a particular proposition (H_1 or H_2), evaluate the findings in terms of a specific RP (sometimes well-defined but often not). That RP depends on the proposition **AND** any particular framework information known to apply under the proposition.

2) It may seem strange to consider an RP for H_1 , but the concept applies to both propositions. Under H_1 , the RP is generally determined, and delimited, by the specimen writer as a special and specific case within the more general population of writers. Therefore, the evaluation under H_1 focuses on demonstrated abilities shown in the provided samples, but also takes into account some

broader understanding of writing in general, particularly for selected sub-propositions under H_1 . Here, the framework would include information such as the age and medical condition of the specimen writer, and any environmental issues applicable to the specimen samples.

3) For H_2 and a standard open set, the RP is usually determined in terms of an average writer drawn randomly from the population at large. This should be refined if meaningful, relevant information is received **about the perpetrator (not the suspect)**, or the conditions applicable to the production of the questioned material,¹⁴ that is, information that permits the RP to be delimited in a more meaningful manner. In general, the assessment is made in terms of the expert's belief about what such an average writer would display in their writing, and whether the questioned sample conforms to those expectations. This would be done by default, unless there is some reason to consider a non-average writer for some reason (e.g., the "skilled forger" argument as a sub-proposition).

4) Note that under any alternative proposition, H_2 , where a specified specimen writer is not involved, details regarding a given specimen writer's health and physical condition are **irrelevant**. The relevant framework under H_2 would include only information known to apply to the production of the questioned material that might affect the writing, for example, the environment, circumstances, or the claimed behavior of the perpetrator. Thus, consideration is only given to conditions known to apply to the production of the questioned material.

5) In contrast, where H_2 is a closed set, the RP would be determined by a second suspect and based on specimen samples from that person. In that situation, the evaluation for both the numerator and denominator becomes a process based on an RP that is well defined (at least, in relation to the samples provided), similar to the situation regarding H_1 and the initial suspect. In that constrained case, relevant framework information extends to things like the age and medical condition of the second suspect and environmental issues applicable to the specimen samples.

From a more statistical perspective, the RP would be a set of writers that needs to be sampled to give us an appropriate probability distribution function that applies under each of the relevant propositions/sub-propositions.

See also **FAQ #20**.

FAQ #22: What expectations are appropriate for FHE work?

The answer to this question depends entirely upon the wording of the proposition under consideration. Every proposition (or sub-proposition) invokes a set of expectations for the features that *should* be seen in the evidence, if/when that proposition is true.

It should be noted that the ability to express expectations is a key attribute of any expert, and in many ways, is the true **hallmark** of expertise. The examiner must ask, for each of the propositions (and any sub-propositions), "What do I expect to observe if this particular proposition is true?"¹⁵

If they can answer that question, they are ready to proceed with the evaluation. If not, the examiner must 1) conduct a review of literature related to the problem, 2) undertake research aimed at providing appropriate expectations, or 3) NOT undertake the work at all (because they lack the requisite knowledge to conduct the evaluation). If the examiner is unable to describe any expectations, the evaluation cannot, and should not, be done.

Determination of expectations really comes down to 1) training, 2) literature, and 3) research. A good starting point for handwriting examinations is Huber and Headrick's 1999 textbook, as well as more recent textbooks.

Note that the set of expectations is necessarily somewhat unique to a given examiner. This is a key reason why any opinion offered by an examiner must be considered to be specific to that examiner. That factor also adds a degree of uncertainty to the situation, which should ultimately be reflected in the opinion.

FAQ #23: How should we handle superimposable signatures?

An observation that two signatures were superimposable or identical in form would be made in the comparison phase. Thus, it is part of the evidence to be assessed. What that means in a given evaluation depends on the specifics of the matter.

When it comes to evaluation of the evidence under a proposition involving natural writing, congruence is not expected because habitually written signatures do not superimpose, given sufficient complexity. Thus, there is already some conditionality. Hence, in this case, there is very little support for any such proposition. Note that how little support there is depends on the degree of congruence, among other things. Therefore, the probability to be assigned would likely be effectively zero.

Conversely, superimposable signatures are expected for tracings, simulations, or reproductions. Thus, such evidence provides support for any proposition involving various forms of unnatural writing (i.e., tracings or careful simulations), or some types of reproduction. The probability to be assigned in an evaluation of the evidence under such a proposition would likely be close to, or effectively one. Of course, other observations regarding the signatures might permit better differentiation of any relevant propositions in play.

FAQ #24: How do limitations affect the evaluation and resulting LR?

The FDE literature includes numerous extensive discussions of limitations that may affect our work.

In simple terms, limitations are factors that make it difficult for the examiner 1) to identify the details of the writing action, or 2) to evaluate those features fully and properly. Different types of limitations affect different aspects of the process, but all limitations must be addressed or considered at some point in the process. Common limitations in relation to materials include:

- Reproductions – type/quality/distortion.

- Small amount of material – questioned or specimen.
- Limited comparability (sometimes related to the above point).
- Having only collected specimen samples.
- Having only requested specimen samples.
- Self-serving specimen samples.

Proper procedures should help to ensure that all of these potential limitations are adequately considered and, ultimately, reflected in the final opinion.

In any given case, the examiner must consider how a particular limitation affects their evaluation. Whether or not a limitation has a significant effect on the outcome depends on the specific situation.

In practical terms, limitations tend to moderate the opinion, although of course, they do not always preclude something from being said. In probabilistic terms, limitations tend to shift the numerator/denominator in the LR toward the most indeterminate value (although the effects on the numerator and the denominator may differ).

The term "most indeterminate value" may be difficult to understand. Both the numerator and denominator are allocated values on a scale from 0 to 1 (exclusive).

Therefore, what value from 0 to 1 is most indeterminate? First, let us consider the endpoints. If either probability is set to ~ 0 , there is essentially no chance of observing the findings if that particular proposition is true. If either probability is set to ~ 1 , the findings will almost certainly be observed if that particular proposition is true. It should be clear that a value at either end of the range is going to be the most informative or determinate. It also follows that the mid-point of the range (i.e., 0.5) is the least informative, or the most indeterminate.

Now consider the LR, which is the ratio of two such probabilities. Any combination involving a value around 0.5 will tend to produce inconclusive or limited value regarding the evidence (i.e., the resulting LR will tend to have a value close to 1). This is particularly true because most limitations affect both the numerator and denominator, often moving *both* of them towards a value of 0.5.

Finally, when an opinion is expressed in the presence of limitations, there will be an increased likelihood of potential error in that opinion (see also **FAQ #15**).

FAQ #25: I keep hearing about assumptions and the need to disclose them in our reports. What assumptions apply to our work?

It is important to understand that the concept of an assumption is NOT something new or unique to the ER approach. Examiners have always made certain assumptions, regardless of whether or not they were aware of doing so. However, because the logical approach helps to clarify all of the elements in our evaluation process, it also helps to clarify some of these assumptions. An assumption

is simply some premise or belief about the situation, the materials we have been given, or any other aspect of the matter, which may or may not be true, but applies to and affects the evaluation.

This topic is fairly case-specific, but it can be discussed in general terms.

First, it is important to recognize that most assumptions can be tested to some degree.

For example, we generally work on the assumption that the specimen samples all come from a particular person, even though we do not know whether that is the case. We simply accept the word of the submitter, who is ultimately required to provide proof for their belief to the trier-of-fact. At the same time, while we cannot address the question of authorship, we do "test" a key part of the assumption, namely, the idea that the samples are all written by one writer, whoever that may be. That is why we conduct internal comparisons as part of our method. Hence, this is a type of tested assumption.

Most of the basic concepts we apply involve some degree of assumptions. When we decide that the samples that are provided are adequate for comparison purposes, we are assuming that those samples are sufficiently representative of the writer's habits that nothing will be misinterpreted. However, we do not make this assumption blindly. Again, there are rules that we apply to the situation.

If the samples are limited, we should moderate our opinion to some degree. However, in the event that an examiner decides that no limitations exist, they are assuming that nothing important is missing, or has been missed. Again, that assumption is never simply accepted without question. Instead, the examiner applies guidelines developed by the discipline relating to the adequacy, comparability, and sufficiency of the material, all of which are aimed at reducing the level of uncertainty regarding that assumption. However, as a rule, those guidelines are quite flexible, and the final decision is that of the individual examiner.

In applying accepted discipline guidelines, there is yet another assumption involved. Everything we do is predicated on the belief/assumption that the examiner is adequately trained and competent in performing a given task. This assumption can also be tested, usually through proficiency testing and the like, but a key underlying assumption is that the examiner's knowledge is applicable to and sufficient for the evaluation.

We might also make assumptions about aspects such as the propositions. Whenever we set an alternative proposition based on the framework and question asked, we make an assumption that this is a good choice. Again, we do not just blindly accept that assumption: we test it by asking our client to either confirm or reject that choice of proposition.

We make similar assumptions about the appropriate RP regarding the alternative proposition (see **FAQ #21**). We may also make assumptions about the size or nature of that population, although these are often somewhat vaguely defined.

In some situations, we might also make assumptions regarding whether particular sub-propositions apply to a given evaluation, for example, whether disguise by the specimen writer is a worthwhile scenario to consider (e.g., in relation to a contested will). Various examiners might view this sub-proposition in very different ways, and whatever an examiner decides in that regard is an assumption regarding the need to consider that sub-proposition. Whatever an examiner opts to do is fine, **so long as it is fully disclosed**, and thus the parameters of the evaluation are clear to all parties.

In any event, our work involves a number of elements that can be seen as assumptions or premises upon which the final opinion rests. The examiner should consider their own reasoning process as much as possible, and disclose these elements as they apply, particularly when they have a significant effect on their opinion.

Finally, something that often arises when discussing assumptions is the concept of disclosure, and how thorough it should be. The idea of disclosure (or transparency) is absolutely key to the ER approach.

We **want** to be clear about what was done, how it was done, and why it was done. In addition, we **need** to be clear about information and conditioning factors that may have affected our evaluation. For many aspects of our work, that should be fairly easy to do. However, assumptions will likely prove difficult in terms of transparency, mainly because examiners do not tend to think about their assumptions overtly or consciously.

In general, it's important for examiners to pay close attention to both their overall and specific thought processes. Experts tend to see things, and then quickly evaluate and interpret what those things mean, while rarely thinking about how they think or do things. That is fine, but examiners should slow down and ensure that everything in the process has been considered, including whatever pieces of information they have heard, read, or otherwise encountered relating to the file. Information comes to us in many ways, some of which are subtle or incidental, and thus we may not fully realize that we have taken something in. To assist this process, examiners should undertake the evaluation in a procedurally rigorous manner, making notes about key elements as they go. This is likely very familiar to most people, even though they may not adopt the ER approach. It should be noted that a written dissertation is not the aim, merely ensuring that the process and outcomes are documented.

Regarding assumptions, there are a few basic guidelines that may help:

- First, assumptions, like all other aspects of the process, should be disclosed. This includes information that you choose to include in or exclude from your evaluation process. As noted, this may relate to the framework and background information you have received, or to elements of propositional decomposition.
- Second, if you are aware that a particular factor is in play, then disclose it.

- Third, if you are unsure whether a particular factor matters to the trier-of-fact, then disclose it. You may choose to explain what it would mean to have it included or excluded from the evaluation.
- Fourth, if you are unaware of a particular factor, then obviously you cannot disclose it, and thus there is no need to be worried. You can only disclose what you know. Conversely, you are the expert, and thus all aspects related to handwriting and, for example, how propositions may break down should be known to you. In addition, you should always be open to the presentation of new information or an assumption requiring re-evaluation and the expression of a new opinion. Openness to change when the conditioning information changes is another key facet of the ER approach.

Section 5: numbers

FAQ #26: How can we allocate values to specific features?

As a starting point in answering this question, consider how things are done using the traditional approach to evaluation.

The difference between the traditional approach and the ER approach is much less than one might think. The main thing examiners should now do is to try to put a number on their belief in an effort to provide a more precise and meaningful opinion. Of course, examiners do not **have** to assign a number to their belief, but if they can do so, the process becomes clearer and easier to explain. Furthermore, even if they do not assign a numeric value, they should be aware that such values still exist, and *could* be elicited and assigned.

The key concern for many examiners is the need to quantify a subjectively held belief (**their** belief), and how that might be done. Of course, in that process, we must focus on the evidence, and not the propositions. At present, examiners clearly hold some non-numeric belief about the weight to be allocated to a set of features (i.e., the overall set of observed features: similar/different/not-accounted-for elements). The term "clearly" has been used because if that was not the case, it would be impossible to express any opinion.

We know from our training, experience, and the literature that some features are more or less indicative of specific types of writing. **How much** more or less is the probabilistic aspect, the area where uncertainty exists. This approach requires the examiner to try to learn how to assign a value to a probability that **they** feel is most applicable, whether that be a specific value (ideal) or a category/range (less than ideal).

Second, we might consider the process down to the level of specific features, as cited in the question, but keep in mind that in general, we have never done that. Most examiners consider the entirety of the writing, with all of its elements/features taken together as a whole. At the same time, there may

be instances where a single feature might stand out and have more significance than the rest of the features combined. If an examiner wants to assign a value to a specific feature, they may do so. However, that may not be either necessary or beneficial; probability values can be assigned based on the overall set of features.

In terms of the process of numeric assignment, see **FAQ #27**.

FAQ #27: How can we assign numbers to the probability?

It is worth noting that the assignment of numbers is nothing new for document examiners, having been discussed by various authors in the past.¹⁶ However, it becomes more important than those authors have suggested if/when we want to adopt the ER approach.

First, we need to understand that probability has two equally valid forms. It can be considered as either aleatory (empirical) or epistemic (subjective) in nature. See **FAQ #9** in that regard. In general, we are assigning subjective probabilities that reflect the examiner's belief about a matter.¹⁷

Second, and most importantly, how each examiner assigns probabilities will differ, just as is presently the case, and thus they need to be prepared to explain how they do this, if/when asked. This is a variation on "What is the basis for your opinion?" Currently, there is no right or wrong way to do this. See **FAQ #5** for more on this point.

The bottom line is that there is no set formula or procedure for this purpose. Thus, we need to consider several elements when assigning numeric values to probabilities.

The assignment of probability is generally based on how well the findings conform to the expectations held for each of the propositions and sub-propositions.¹⁸ Having said that, the assignment process has a different focus for each proposition.

Under a common source proposition, the focus is on how well the specimen and questioned samples match. That is, we must consider the evidence (i.e., the degree of match) assuming that the suspect writer actually wrote the material (i.e., given H_1). To do that, we consider our predetermined expectations under that proposition.

We expect that in the case of natural writing, the two samples will display the same writing habits with no divergences beyond the displayed range of variations, assuming that there are no limitations in the material (e.g., inadequacy). For unnatural writing (under H_1 , focused on the specimen writer), our expectations will differ depending on the type of unnatural writing. Under the common source proposition involving outright disguise, the expectation would be almost complete divergence between the specimen and the questioned sample. In contrast, expectations when considering disguise for later denial (where the writer needs to produce something that will pass initial scrutiny while still being disavowed in the future) might include divergences in obvious, superficial features, as well as similarities in relation to inconspicuous and subtle features in the writing.¹⁹

Under a different source proposition (i.e., given H_2), our focus shifts to the question of how likely it is that we will encounter the same or a better degree of matching if the questioned sample was written by another person and **not** the specimen writer. In this case, the evidence is considered with a focus on how likely it is that someone else would produce as good a match or better.

For natural writing under H_2 (i.e., a coincidental or random match), we expect that another person's writing will display fundamental differences of a repeated nature relative to the specimen writing. We know that the probability of seeing a *coincidental/chance match* in the natural writing of two different writers is extremely low so long as the samples are adequate in all respects (i.e., the writing is mature and developed, complex, and comparable). In this case, the other person is an individual selected at random from the RP of possible perpetrators (see **FAQ #21**). Obviously, the **complexity** of the writing is key here, but the **manner of production** should also be considered.

The assessment method for unnatural writing under H_2 is different. Here, we mainly consider simulation or tracing.²⁰ Under those sub-propositions, the assignment is based on **the degree to which the specimen writing is susceptible** to such actions, and how closely the questioned writing conforms to what we would expect if they occurred. For example, under a proposition of simulation, we generally expect to observe good graphic conformity between the questioned and known samples (especially at a superficial graphic level), but with divergences in the dynamic elements of the writing. The **complexity** of the writing remains important, but the **manner of production** becomes even more crucial.

In general, probabilities can be assigned in terms of what may be called threshold values: for example, 1/2, 1/10, 1/100, 1/1 000, 1/10 000, 1/100 000,²¹ and so on. An examiner can develop an understanding of how common/rare a given event is when it occurs with one of those frequencies. The author strongly recommends following this concept, although there is no easy or straightforward way to do so. A study of the FHE literature relating to frequency data is good of course,²² but it is equally or even more important to study probabilities for events of other types in an effort to gain a more general appreciation of such frequencies. Values between the thresholds can be assigned of course, but the assignment usually comes down to the question of whether a given assignment exceeds one of those thresholds.

Expectations expressed at the level of the main propositions will necessarily be probabilistic, uncertain, and personal because they vary based in accordance with any specific sub-proposition under consideration (see **FAQ #22**).

A valuable exercise would be the production of a master list of expectations in relation to the various propositions that examiners encounter, or at least those most often encountered. This would help to minimize inter-examiner variations in those expectations, and thus would be invaluable for both training purposes and casework.

In assigning probabilities, it is also important to consider limitations and how they affect our ability to assess features. In general, limitations introduce uncertainty into the process, which leads to a less definitive result. Those limitations might include reproduction quality, the amount of comparable material, or even the examiner's basis for their belief in relation to the matter, among other things.²³

Finally, the ER approach **can**, of course, be applied without assigning numbers because it is, after all, a system of logical reasoning. It is completely flexible, and thus it can be used for any application we might have. However, doing so **without** numbers makes the process less clear, more imprecise, and more challenging to explain (e.g., the precise relationship between the elements of the LR are difficult to explain).

Section 6: explanation and testimony

FAQ #28: This is difficult to explain to a layperson – how can I do it so that they understand?

At its most basic, the ER approach is little more than an expression that the evidence favors one belief over another, supplemented with a statement about the strength of that evidence. For example, the opinion may take the form “The findings strongly support the belief that the specimen writer wrote the questioned entry rather than it having been written by some other writer.” While that may seem somewhat wordy, it is not difficult to understand or explain.

It is difficult to see how this is any more complicated than the opaque explanations that we have historically used to describe and explain both our processes and our opinions.

See also **FAQ #13**.

FAQ #29: Judges want us to tell them what happened. How do I address a directive to that effect given that I could be held in contempt if I refuse to do what they want?

To some degree, the judge's direction is understandable. After all, **they** need to answer that question, and they want/need our help in that regard. However, similar to many other people, judges do not always understand the limitations of our science, and thus, as the expert in our domain, we need to explain these to them.

Even if we are not in a position to answer the question of what happened directly, we can definitely help the trier-of-fact to answer the question for themselves. Frankly, that **is** the role of the expert – not to take on the role of the judge, but to help them incorporate our information appropriately into their own decision-making process.

As for how to address this issue with the court, something along the following lines is suggested.

First, explain why, as an expert, you are uncomfortable giving that type of opinion. You should point out that what they are asking for is **not** an expert opinion, *per se*, because it requires

information extending beyond your domain. That will undoubtedly lead to additional questions and discussion relating to why it goes beyond your domain of expertise, during which you could outline the type of information you are referring to, including detailed case circumstances and/or witness statements.

Second, explain to the court that to answer their question, you would need to make several assumptions that may or may not be valid, and clearly extend into their domain as the ultimate decision-maker. Those assumptions involve the setting of prior odds of some sort, which you could transform into posterior odds via multiplication using the LR, and the application of some personal threshold regarding the posterior odds that warrant a particular decision about what actually happened. This might be done using some statement like the following. See the sample wording below (see also **FAQ #12**).

“I would be happy to address that question, your honor, but I must note that doing so requires information 1) outside my area of expertise, and 2) that is unknown to me. I am referring to information about the matter to which I am not privy, but which may affect any such decision (for example, the prior odds of the propositions). There is also the related issue of the decision criteria I must apply to decide what posterior odds warrant a specific decision. I can apply my personal beliefs about this but, in reality, anyone could do that, and I believe that it is generally best to leave that sort of thing to the court. I do not want to inadvertently cross over into the court’s domain or usurp the court’s authority in relation to any such decision.”

If the court requires further explanation, you could point to the various published discussions in this area including those in the ENFSI guideline (Willis et al., 2015) and the Human Factors in Handwriting Examination report (Taylor et al., 2021), among many others.

Finally, if the court still wants an opinion, you could provide whatever you felt was the most appropriate response given the particular situation. At this point, the examiner can do whatever they want because they have done their best to avoid any potential problems.

This suggested approach should be done with the knowledge that it would almost certainly elicit an appeal if a competent examiner or lawyer was engaged by the other party.

FAQ #30: Couldn't these findings be the result of a particular cause, event, or activity suggested by some party, either the defense, the prosecution, a barrister, or the judge?

This is an example of basic cross-examination. Usually, such questions come from opposing counsel looking for an alternative that might shift the opinion in favor of their client. In brief, the answer to this question is almost always "Yes," because it is really a question regarding the *potential* possibility that the findings might arise from the specified cause, and such potential almost always exists.

However, a big problem with this question is that it can easily be misconstrued as a transposed statement about the proposition. When answering such a question, the examiner needs to be careful

to always focus on the **findings** and the effect, **if any**, on the LR. After all, the critical issue is the degree to which the findings are expected to occur when X is true, but equally importantly, such information only becomes valuable (in a probative sense) when contrasted with the probability of occurrence should some alternative proposition be true, that is, when X is not true.

To respond fully to the question requires a simple process of 1) determining what expectations exist under the newly specified proposition, and 2) determining if/how well the findings from the matter at hand conform to those expectations. Then, the result of that process must be compared with the results of any alternative proposition that is under consideration. It should always be kept in mind that a proper opinion speaks to the **relative** support provided by the findings under at least two propositions. Therefore, considering only a single proposition in isolation would be both improper and incomplete, at least in terms of undertaking a comparative evaluation process.

It is unlikely that the examiner will not have already considered the possibility that is presented, although that may not be obvious to the person asking the question. Hence, the answer might be easy to provide. However, an examiner should never feel compelled to provide an immediate answer, in fact doing so can sometimes appear to be dismissive. Often, it is strategically wise to suggest that the matter requires re-evaluation, and then to ask for the necessary time to do so.

It is always important to remember that **whenever** the propositions change in some way, a new evaluation is necessary. The original statement does not necessarily become invalid, but the new situation must be assessed on its merits.

See also **FAQs #2** and **22**.

Acknowledgments

The FAQs included in this article were compiled for the workshop "Evaluative Reporting: Bridging Theory and Practice" held as part of the European Network of Forensic Handwriting Experts (ENFHEX) conference in Zagreb, Croatia in September 2022. The author acknowledges the input of workshop co-facilitators Nicole Crown, Erich Kupferschmid, Raymond Marquis, and Carolyne Bird in reviewing and making suggestions regarding the content of the FAQs.

Compliance with ethical standards

Not applicable.

Disclosure statement

The author has no competing interests to declare.

References

Arkes, H. R. and Koehler, J. J. (2022). "Inconclusives and error rates in forensic science: a signal detection theory approach." *Law, Probability and Risk* **20**(3): 153-168
<https://doi.org/10.1093/lpr/mgac005>

Association of Forensic Science Providers. (2009). Standards for the formulation of evaluative forensic science expert opinion, *Science & Justice*, **49**(3): 161-164.
<https://doi.org/10.1016/j.scijus.2009.07.004>

Biedermann, A. (2013). Your uncertainty, your probability, your decision. *Frontiers in Genetics*, **4**: 148. <https://doi.org/10.3389%2Ffgene.2013.00148>

Biedermann, A., Bozza, S. and Taroni, F. (2016). The decisionalization of individualization. *Forensic Sci Int*, **266**: 29-38. <https://doi.org/10.1016/j.forsciint.2016.04.029>

Biedermann, A., Bozza, S. and Taroni, F. (2008). Decision theoretic properties of forensic identification: underlying logic and argumentative implications. *Forensic Sci Int*, **177**(2-3): 120-132. <https://doi.org/10.1016/j.forsciint.2007.11.008>

Biedermann, A., Taroni, F. and Garbolino, P. (2007). Equal prior probabilities: can one do any better? *Forensic Sci Int*, **172**(2-3): 85-93. <https://doi.org/10.1016/j.forsciint.2006.12.008>

Biedermann, A., Bozza, S., Taroni, F. and Aitken, C. (2017). The consequences of understanding expert probability reporting as a decision. *Science & Justice*, **57**(1): 80-85.
<https://doi.org/10.1016/j.scijus.2016.10.005>

Biedermann, A., Taroni, F., Bozz, S., Augsburg, M. and Aitken, C.G.G. (2018). Critical analysis of forensic cut-offs and legal thresholds: A coherent approach to inference and decision. *Forensic Sci Int*, **288**: 72-80. <https://doi.org/10.1016/j.forsciint.2018.04.030>

Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. (1998a). A model for case assessment and interpretation. *Science & Justice*, **38**(3): 151-156. [https://doi.org/10.1016/S1355-0306\(98\)72099-4](https://doi.org/10.1016/S1355-0306(98)72099-4)

Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. (1998b). A hierarchy of propositions: deciding which level to address in casework. *Science & Justice*, **38**(4): 231-239.
[https://doi.org/10.1016/S1355-0306\(98\)72117-3](https://doi.org/10.1016/S1355-0306(98)72117-3)

European Network of Forensic Science Institutes. (2022). *Best practice manual for the forensic handwriting examination*, ENFSI-FHX-BPM-01, Issue 04. ENFHFX. <https://enfsi.eu/wp-content/uploads/2022/12/4.-BPM-Edition-4-Draft-20220822-3-1.pdf>

Found, B. and Bird, C. (2016). The Modular Forensic Handwriting Method—2016 Version. *Journal of Forensic Document Examination*, **26**: 7-83. <https://doi.org/10.31974/jfde26-7-83>

Found, B., Sita, J. and Rogers, D.K., (1999). The development of a program for characterising forensic handwriting examiners' expertise: Signature examination pilot study. *Journal of Forensic Document Examination*, **12**: 69-80.

Huber, R.A. and Headrick A.M. (1999). *Handwriting Identification: Facts and Fundamentals*. (1st ed.). CRC Press. <https://doi.org/10.1201/9781420048773>

Huber, R. A. and Headrick A. M. T. (1990). Let's do it by the numbers. *Forensic Science International* **46**(3): 209-218. [https://doi.org/10.1016/0379-0738\(90\)90307-K](https://doi.org/10.1016/0379-0738(90)90307-K)

Kam, M., Fielding, G. and Conn, R. (1997). Writer identification by professional document examiners. *Journal of Forensic Sciences*, **42**(5): 778-786. <https://doi.org/10.1520/JFS142071>

Kam, M., Gummadidala, K., Fielding, G. and Conn, R. (2001). Signature authentication by forensic document examiners. *Journal of Forensic Sciences*, **46**(4): 884-888. <https://doi.org/10.1520/JFS15062J>

Kam, M., Wetstein, J. and Conn, R. (1994). Proficiency of professional document examiners in writer identification. *Journal of Forensic Sciences*, **39**(1): 5-14. <https://doi.org/10.1520/JFS13565J>

Köller, N., et al. (2004). *Probability Conclusions in Expert Opinions on Handwriting: Substantiation and Standardization of Probability Statements in Expert Opinions*. München, Wolters Kluwer Deutschland GmbH

Lindley, D. V. (2000). The Philosophy of Statistics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **49**(3): 293-337. <https://doi.org/10.1111/1467-9884.00238>

Lindley, D. V. (2002). Letter to Editor. *Teaching Statistics*, **24**(1): 22-23. <https://doi.org/10.1111/1467-9639.00075>

Lindley, D. V. (2006). *Understanding Uncertainty*. Hoboken, New Jersey, John Wiley & Sons. <https://doi.org/10.1002/0470055480>

Marquis, R., et al. (2016). Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings. *Science & Justice*, **56**(5): 364-370. <https://doi.org/10.1016/j.scijus.2016.05.009>

Petraco, N. and Mohammed, L. A. (2011). *Basic Statistical Concepts for Questioned Document Examiners*. Annual Meeting of the American Society of Questioned Document Examiners. Philadelphia, Pennsylvania.

Taroni, F. and Biedermann, A. (2005). Inadequacies of posterior probabilities for the assessment of scientific evidence. *Law, Probability and Risk* **4**(1-2): 89-114. <https://doi.org/10.1093/lpr/mgi008>

Taroni, F., Bozza, S. and Biedermann, A. (2020). Decision theory. In: *Handbook of Forensic Statistics*. Banks, D. L., Kafadar, K., Kaye, D.H. and Tackett, M. (Eds.). Chapman & Hall/CRC: 103–130.

Taroni, F., Bozza, S. and Biedermann, A. (2021). The logic of inference and decision for scientific evidence. In: Philosophical Foundations of Evidence Law. Dahlman, C., Stein, A. and Tuzet, G. (Eds.). Oxford, Oxford University Press: 251-266. <https://ssrn.com/abstract=3955309>

Taylor, M., et al. (2021). Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, [online]. <https://doi.org/10.6028/NIST.IR.8282r1>

Willis, S., et al. (2015). ENFSI guideline for evaluative reporting in forensic science. European Network of Forensic Science Institutes, [online]. <https://enfsi.eu/docfile/enfsi-guideline-for-evaluative-reporting-in-forensic-science/>

Note

1. Using the term "traditional" here is not ideal, but have been included because most examiners think of things in this way. In reality, the ER/logical approach goes back to the beginnings of forensic science and, in a few laboratories, has been used all along.
2. See Biedermann et al. (2008).
3. This discussion ignores the legitimate possibility that our propositions might not invoke multiple potential sources. Activity-level evaluations, for example, may relate to a single suspect such that the identity of some unknown perpetrator is not the issue at hand. That is, the concern becomes one of "what happened," rather than "who was responsible." Of course, these matters are not always clear-cut (see also **FAQ #11**).
4. Particularly, when considering matters related to, among others, the ease of simulation and knowledge of the graphic abilities of the suspect.
5. Note that conclusive opinions may be possible for some types of evidence, but not for handwriting. Note also that such reports would be one of the non-evaluative forms. See also **FAQ #3**.
6. Several authors have discussed this topic, but Lindley is one of the easiest to understand.
7. Note an additional level, "sub-source," was later added to the basic scheme to address evidence such as touch DNA.
8. Or, at least, any questions raised are no more common or problematic than those encountered when using the 'traditional' approach.

9. The logic that is invoked is often called Bayesian, but in reality, concepts such as the LR (which is the basis of the opinion) exist independently of Bayes' theorem. Thus, while this can be viewed as a Bayesian philosophy or approach, it is not constrained to that approach.
10. See, for example, Found et al. (1999), Kam et al. (1994), Kam et al. (1997), Kam et al. (2001), and Sita et al. (2002).
11. Technically speaking, another key distinction is that traditional scales are transposed to speak about the proposition, meaning that the error is expressed in terms of the proposition rather than the evidence alone. However, that is the result of our training, as discussed in **FAQ #3**.
12. See, for example, Petraco and Mohammed (2011) and Arkes and Koehler (2022).
13. The observant reader will note that, technically speaking, this statement is incomplete. The strength of the evidence may also change with some change in the relevant framework information.
14. Thus, for example, our suspect under H_1 might be an artist. However, under H_2 , there may be no reason to assume that the perpetrator would be a skilled writer or an artist like the suspect. Similarly, our suspect under H_1 might be an elderly or infirm person, but under H_2 , there may be no reason to assume that the perpetrator would be elderly or infirm. Alternatively, there may be times when the situation warrants some belief about the perpetrator under the alternative (i.e., someone other than the specimen writer). However, that belief comes not from any knowledge of the specimen writer but rather is based on the framework. For example, the suspect writer under H_1 may be of Albanian nationality. In general, we should not assume, based on the suspect, that the perpetrator under H_2 would also be Albanian. However, if the questioned text were written in fluent Albanian or if the scenario otherwise suggests that the perpetrator is Albanian, then it might be reasonable to apply that information to our RP.
15. Note that this concept applies to both what we expect to see and what we do not expect to see, when any specified proposition is true. In addition, such expectations will be conditional in nature, that is, they will be more or less likely to occur (i.e., they will be assigned a different probability of occurrence) under a given scenario.
16. See, among many others, Huber and Headrick (1990) and Köller et al. (2004).
17. This approach will undoubtedly change over time as more quantified data become available to examiners. However, quantified data should serve to complement or improve subjective beliefs based on training, rather than replace them.
18. The topic of expectations has not been fully explored here because this is an area where the discipline should be developing specific guidelines, although there are numerous worthwhile references available (see, for example, the ENFSI Best Practice Manual for the Forensic Handwriting Examination (2023) and Found and Bird's 2016 Modular Approach document). See also **FAQ #22**.

19. Note that some examiners place matters such as writing under the influence of an intoxicant under a proposition of unnatural writing, while others place them under natural writing. It really depends on how you view such things (as well as on your training and the method being used). It does not matter where it is placed, so long as the examiner makes it clear how they view the situation and where/how this fits into their evaluation, either in their work notes or in their report.
20. Disguise might also be considered under H_2 , but this is no longer disguise by the specimen writer, but rather disguise by the perpetrator. Therefore, expectations would be more aligned with a chance match because we have no samples for the perpetrator.
21. For clarity, these may be read as “1 in 2”, “1 in ten”, “1 in one hundred”, etc.
22. The issues with previous studies on the frequency of features are 1) they only focus on natural writing (i.e., they only apply to H_1 , natural), and 2) they are dependent on some classification system specific to the study that may or may not translate very well to the real world.
23. It is also important to note that this type of belief adjustment has always been undertaken by examiners. For example, it is normal to adjust our opinion to reflect any limitations in the material. The use of quantified statements of belief enables greater precision and understanding of the relative strength of the opinion being expressed.

UNCORRECTED MANUSCRIPT