

# **Understanding forensic expert evaluative evidence: A study of the perception of verbal expressions of the strength of evidence**

## **Abstract**

Verbal expressions of evidential strength are routinely used when presenting forensic expert evaluative evidence. The degree to which these verbal expressions are interpreted uniformly among different individuals requires further empirical study. This study focussed on groups of individuals with different roles within the criminal justice system and individuals with varying degrees of expertise and knowledge. Three groups of individuals were identified: laypeople, legal professionals and those with some forensic or investigative knowledge. The participants in the study (n=230) were provided with a case summary to which a verbal expression of the strength of evidence was randomly assigned. Participants were subsequently invited to indicate their perception of the strength of the evidence on a scale that was provided. Generally, across the study groups, the trend was one of increased perceived strength of evidence as the intended strength of the verbal expression was increased, with some notable exceptions. In general, there was good concordance between the groups in the way the different expressions were perceived. It was found that participants performed poorly when it came to differentiating between expressions at the 'strong' end of the scale ('strong', 'very strong' and extremely strong'). The findings resonate with calls for validated and robust communication frameworks for evaluative opinions. Further empirical research in this area is warranted and that such research can represent an important contribution towards improving the communication and presentation of forensic evidence.

*Keywords: Verbal scale, Likelihood ratio, Perception, Strength of evidence*

## **1. Introduction**

The calculation of the likelihood ratio is a central component in the interpretation of forensic evidence. While there is consensus that this can provide a basis for achieving a balanced interpretation of evidence, the manner in which the result of this calculation is reported and presented is the subject of debate and discussion. There are challenges associated with communicating uncertainty and the concordance between intention and understanding when using different methods to expressing this uncertainty has been the subject of much psychological research [1]. The need for frameworks for expressing conclusions in forensic science has been identified [2, 3, 4] and a scale of verbal equivalents is routinely employed as a means of conveying the strength of evidence. The interpretation of these expressions by different audiences and decision-makers is of great importance. This is the subject of a number of empirical investigations, but there is a need for further work to explore variations in the way expressions are perceived and understood.

Previous empirical studies that have explored this issue have reported some issues with regard to the use of verbal expressions. Mullen et al. [5] examined the perception of verbal expressions among volunteers and while there was some evidence that median perceptions ascended with each level of the scales, there were significant inconsistencies.

The study concluded that terms may be misunderstood by lay people and that, generally, the majority of descriptors were found not to convey the intended level of support. The authors question the degree to which the verbal scale fulfils its purpose of assisting the court in understanding the strength of evidence. Martire and Watkins [6], in a re-examination of the data – conclude that the correspondence between expert intentions and lay perceptions is low, meaning that the potential for miscommunication is high. They argue that the verbal scales ‘do not appear to fulfil purpose of assisting court or facilitating effective and accurate communication’ [6 p.272]. Accordingly, the need for research into an alternative means for expressing likelihood ratios and the requirement for empirical validation when it comes to scales of expressions are both highlighted.

In an investigation into the expression and interpretation of the verbal scale, Martire et al. [7] undertook experiments that revealed evidence of a “weak evidence effect” whereby some participants inverted the direction of support when presented with evidence that provided weak support. Martire et al. [8] explored different methods of communicating support. As well as observing the undervaluing of expert testimony, a “weak evidence effect” was also identified when participants were presented with expressions conveying low strength. The utility of low strength verbal expressions was therefore questioned, given that the potential for miscommunication was found to be high.

Therefore, in order to develop the empirical evidence base from which it is possible to further understand the effectiveness of verbal expresses in conveying the intended strength of evidence, this present study explored the perceptions of these verbal expressions among different groups. The research sought to assess the extent to which verbal expressions of the strength of evidence are perceived differently by individuals with different roles within the criminal justice system and individuals with varying degrees of expertise or knowledge.

## **2. Methods**

### **2.1 The experiment**

Participants took part in a questionnaire exercise that required them to read a brief synopsis of the details of a fictional case in which a piece of footwear mark evidence was presented, along with a verbal expression of the strength of support. Participants belonged to one of three sample groups.

The case précis provided to each participant concerned the same piece of footwear mark evidence. Footwear mark evidence was chosen for this study as it was one of the evidence types employed by Mullen et al [5]. A volume crime – a burglary - was selected as it was hypothesised that a more “serious” offence may have had an impact on responses.

Having read the case detail and the presentation of the footwear evidence, participants were instructed to indicate the perceived strength of this evidence on a 20 point line which ran from ‘no support’ at one end, to ‘conclusive support’ at the other. The questionnaire materials were identical in all respects other than the accompanying expression from the

verbal scale. Versions of the questionnaire were prepared with each of the six verbal expressions (see table 1) and these were randomly allocated to participants.

The scale chosen for this study was the six point verbal scale recommended by the Association of Forensic Science Providers (Table 1). Participants were not provided with a reference scale as it was deemed that this would have confounded any findings regarding perception accuracy.

Table 1. Recommended likelihood ratio terminology [4]

<b>Numerical expression</b>	<b>Verbal expression (support)</b>
>1-10	Weak
10-100	Moderate
100-1,000	Moderately strong
1,000-10,000	Strong
10,000-1,000,000	Very strong
>1,000,000	Extremely strong

Numerical labels were not included on the line that participants used to indicate the strength of evidence. This was because it was deemed that respondents may have been inclined to rely on some form of numeric reasoning when forming their perceptions and this would have had a confounding effect the findings of the experiment. Numbers were added to the 20 point scale during analysis in order to measure the 'perception of strength'. The first point on the line, 'no support', was assigned a value of zero (0), while the 20<sup>th</sup> point, 'conclusive support' was assigned the number 19.

## 2.2 Sampling

Three sample groups were recruited from populations with different roles in the Criminal Justice System and different degrees of expertise; lay jurors, legal professionals and a group of participants with knowledge of forensic science and criminal investigations.

1. Potential lay jurors were recruited through a convenience sampling strategy via several multimedia platforms. Eligibility to vote in the U.K. was use as a qualification criterion. A total of 88 respondents were recruited in this manner. This group was labelled as the 'lay' group.

2. Legal professionals were recruited through a snowball-convenience sampling strategy. Legal practitioners were identified and requested to distribute the questionnaire on behalf of the researchers. The Criminal Bar Association advertised the research to readers of its weekly newsletter. This dual strategy enabled access to legal networks. Responses were filtered after collection to ensure that only responses from legal practitioners were included in the analysis. A total of 84 respondents were recruited in this manner. This group was labelled as the 'legal' group.

3. The final group was composed of forensic practitioners, students of forensic science and crime/security professionals. These participants were recruited through targeted snowball-convenience sampling. Questionnaires were distributed through contacts at police forces and it was also distributed to attendees at a parallel session on forensic science at the 8<sup>th</sup> International Crime Science Conference. A total of 58 respondents were recruited in this manner. This group was labelled as the 'forensic' group.

### **3. Results & Analysis**

#### **3.1 Perception of verbal expression**

##### *All responses*

Data for the three sample groups were combined and figure 1 displays the perceptions of each respondent across the three groups (n=230). The general trend is one of increased perceived strength of evidence as the intended strength of the verbal expression is increased, albeit with some exceptions. Boxplots for the same data, combined across the three sample groups, are displayed in figure 2. A general trend of increased perceived strength of evidence with increased intended strength can be observed, while the greatest variability in perceived strength was evident when the evidence presented provided 'weak', 'moderately strong' or 'extremely strong' support, although the influence of extreme responses should be acknowledged here. It is notable that, across the three groups, 'strong', 'very strong' and 'extremely strong' were perceived similarly by respondents. Descriptive statistics for the combined data are provided in table 2. The mean perception rating generally increases with each gradation. However, the mean rating for 'extremely strong' is lower than that of 'strong' and 'very strong'. It is evident that the differences in mean values between the expressions at the upper end of the scale are small ('strong' = 12.4, 'very strong' = 12.8 and 'extremely strong' = 12.3). However, as indicated in figure 2, the mean value for 'extremely strong' is affected by a small number of extreme responses between 0 and 2 on the perception scale. The median perception ratings, which are less affected by these responses, do reflect the trend of increased perceived strength as the intended strength is increased, with smaller increases at the upper end of the scale.

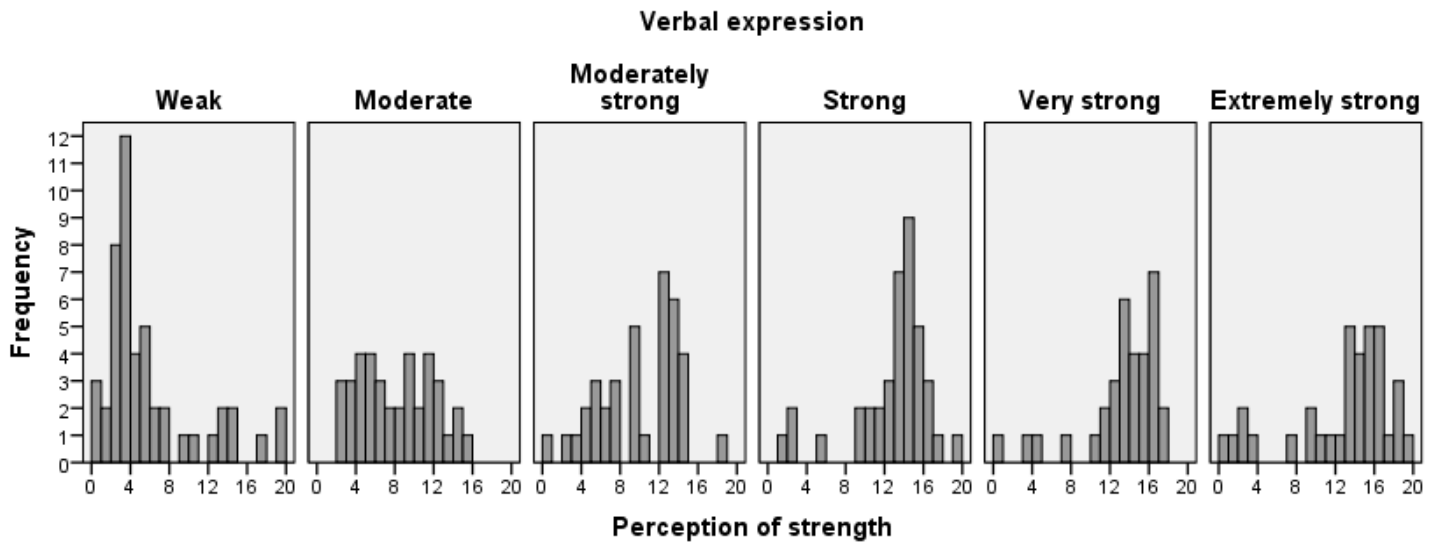


Figure 1. Perception of evidential strength for each verbal expression – all participants (three groups combined).

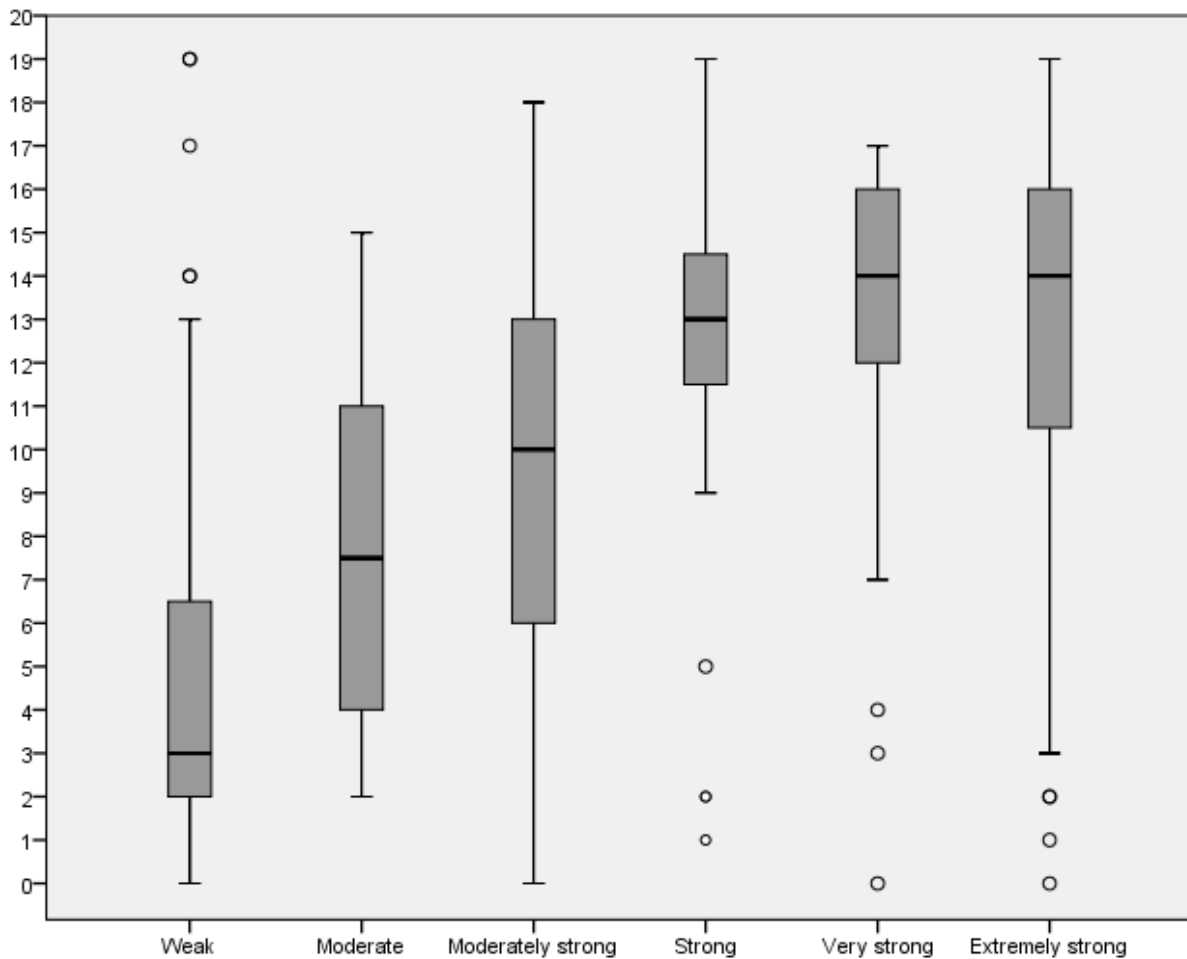


Figure 2. Boxplots displaying the perception of evidential strength for each verbal expression – all participants (three groups combined). Y axis indicates perception of strength (0-20).

Table 2. Descriptive statistics. The perception of evidential strength for each verbal expression – all participants (three groups combined).

	Weak	Moderate	Moderately strong	Strong	Very strong	Extremely strong
<b>Mean</b>	5.4	7.7	9.5	12.4	12.8	12.3
<b>Median</b>	3	7.5	10	13	14	14
<b>Standard deviation</b>	4.9	3.8	4.2	4	4	5.2

### *The three sample groups*

Table 3 provides descriptive statistics for the response data for the three sample groups. The perceptions of the respondents in each sample group are displayed in boxplots (figures 3-5). In all three groups, we observe the general trend of increased perceived strength as the intended strength is increased, particularly if we examine the median values which are affected by the extreme responses to a lesser degree. Again, the differences in perception at the upper end of the scale are generally small. Mean responses reveal that in the lay group, 'very strong' was perceived to convey more strength than 'extremely strong' (mean values = 13.4 and 13.1, respectively), although the median value was the same for both groups (14). Meanwhile, from the mean values, it appears that the legal group perceived 'strong' to convey greater strength than both 'very strong' and 'extremely strong'. However, this is likely to result from a number of extreme responses to 'very strong' and 'extremely strong' given that the median values are the same for these three expressions (13). It is again clear that the perceptions of expressions at the upper end of the scale were very similar.

Upon an examination of the data, it appears that verbal expressions were generally considered to be indicative of greater evidential strength by the forensic practitioners than by the other two groups. The highest degrees of variation were observed when the legal and forensic groups were presented with 'weak' (standard deviations = 5.9 and 5.8) and when these groups were presented with 'extremely strong' (standard deviations = 5.9 and 5.8), although in all cases this does appear to result from a small number of extreme responses which will be considered in the discussion. Overlapping of perceptions was observed among all three groups (figures 3-5).

Table 3. Descriptive statistics. The perception of evidential strength for each verbal expression for the three sample groups.

		Verbal expression					
		Weak	Moderate	Moderately strong	Strong	Very strong	Extremely strong
<b>Lay</b>	Mean	5.0	7.6	8.5	11.8	13.4	13.1
	Median	4	7	9	13	14	14
	Standard deviation	3.5	3.3	4.9	4.2	3.1	2.7
<b>Legal</b>	Mean	5.2	5.6	10.0	12.0	11.5	10.5
	Median	3	4.5	9.5	13	13	13
	Standard deviation	5.9	3.5	3.5	3.6	5.1	5.9
<b>Forensic</b>	Mean	6.6	10.4	10.9	13.5	13.9	14.1
	Median	3.5	11	12	14	16	16
	Standard deviation	5.8	3.3	3.4	4.4	3.6	5.8

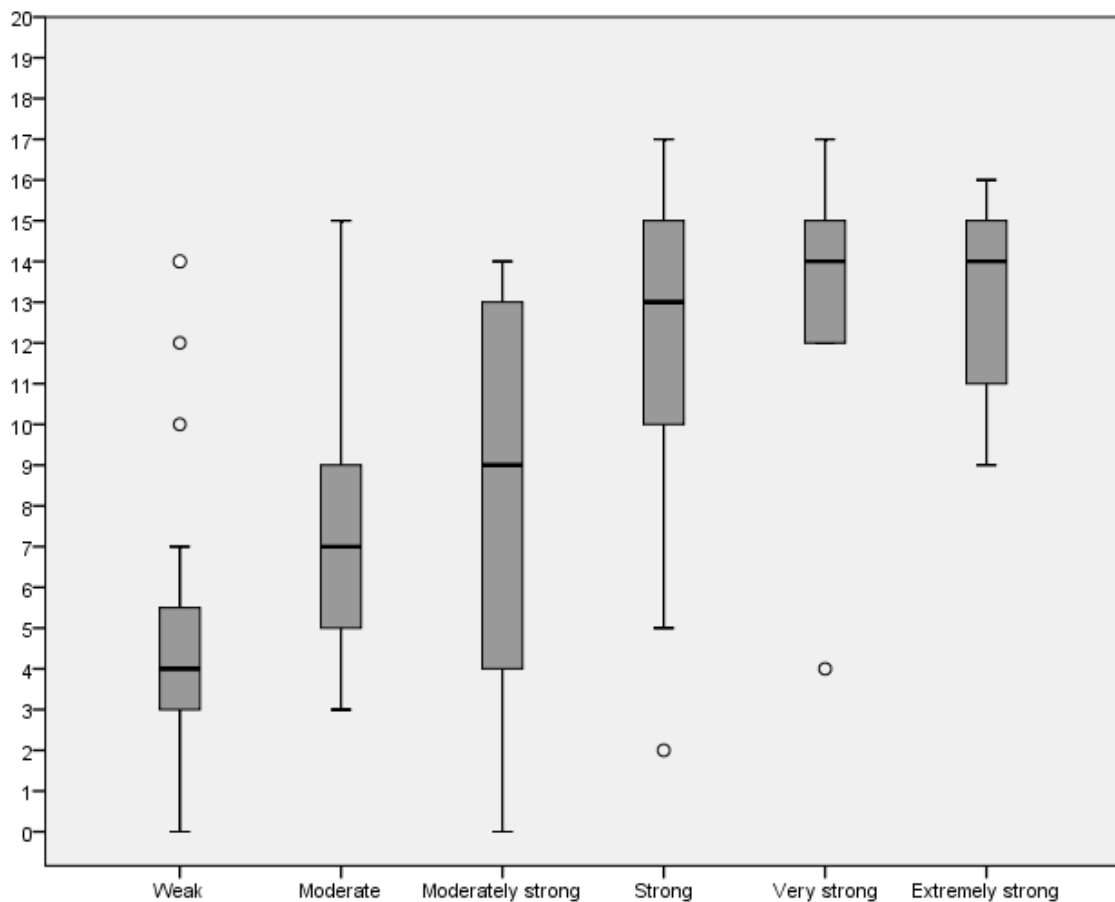


Figure 3. Perception of evidential strength for each verbal expression – lay group. Y axis indicates perception of strength (0-20).

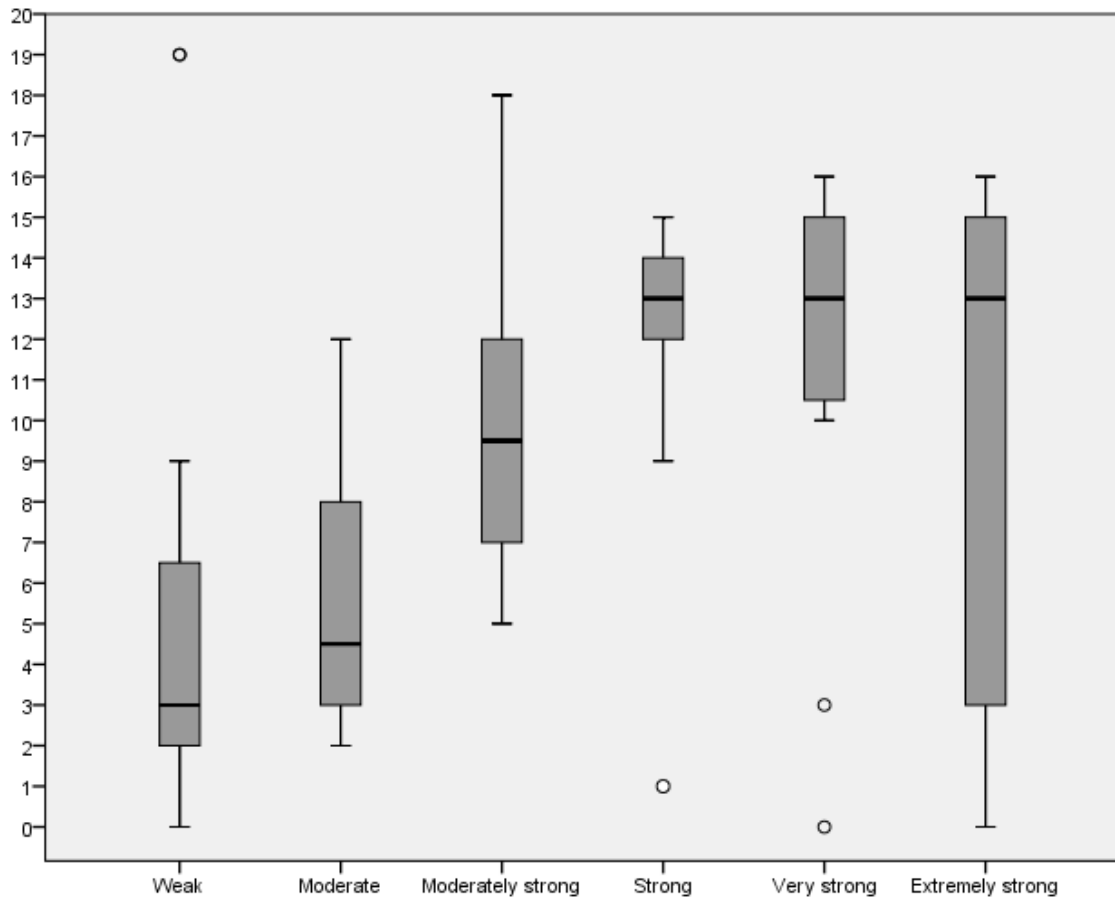


Figure 4. Perception of evidential strength for each verbal expression – legal group. Y axis indicates perception of strength (0-20).

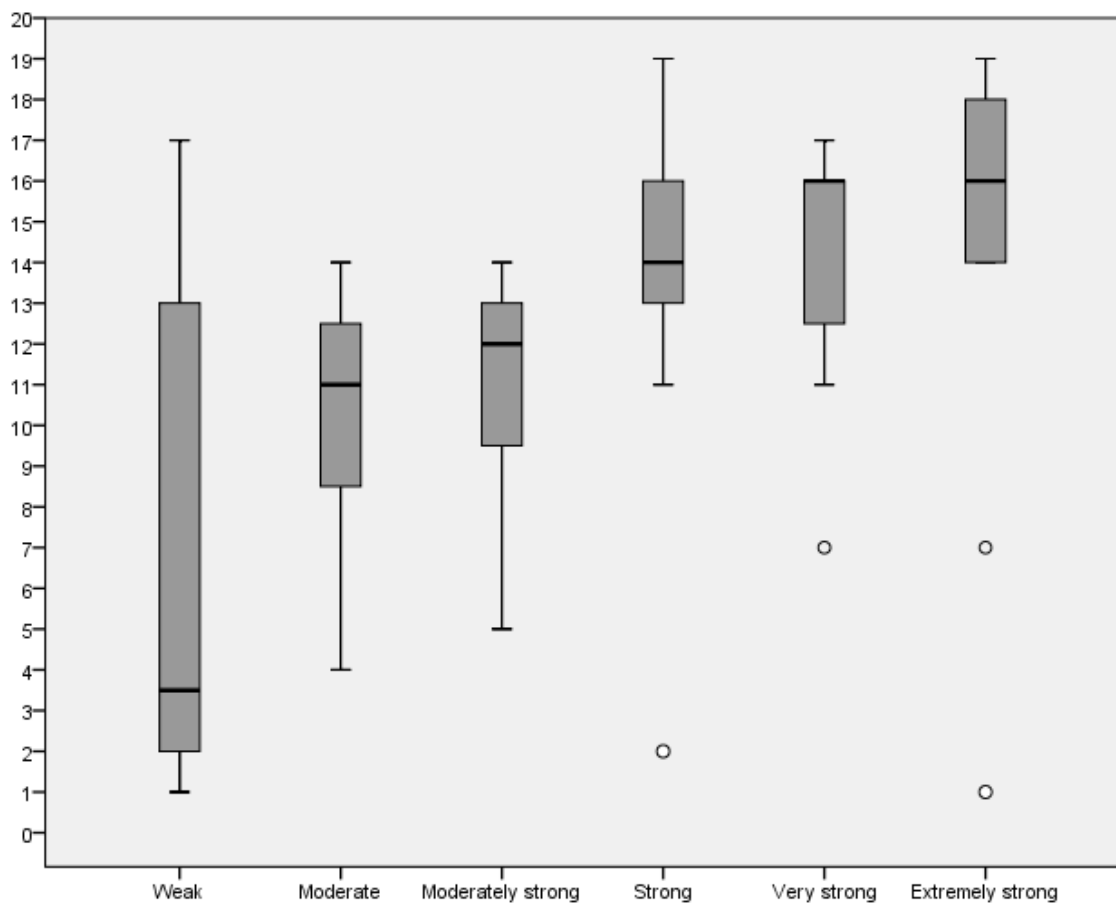


Figure 5. Perception of evidential strength for each verbal expression – forensic group. Y axis indicates perception of strength (0-20).



The significance of observed differences between the strength assigned to each expression was examined for each of the three sample groups. Non-parametric Mann-Whitney *U* tests were employed. Table 4 displays the results of the tests for significance. Those instances in each sample group where the distinction between two gradations was not statistically significant ( $p > 0.05$ ) and therefore, where expressions were not differentiated effectively are indicated by “No”. The remaining differences, indicated by “Yes”, were significant ( $p < 0.05$ ). The results demonstrate that, in each group, the distinction between the expressions at the top of the scale (‘strong’, ‘very strong’ and ‘extremely strong’) was not clear. Meanwhile, both the lay and forensic groups did not significantly distinguish between ‘moderate’ and ‘moderately strong’. The lay group did not distinguish between ‘moderately strong’ and ‘strong’, while the forensic group did not distinguish between ‘moderately strong’ and ‘very strong’. There were issues at the lower end of the scale in the forensic group, as this group did not significantly distinguish between ‘weak’ and ‘moderate’, or between ‘weak’ and ‘moderately strong’. The influence of an extreme response of 17 should, however, be acknowledged. Finally, while the legal group did not distinguish between ‘weak’ and ‘moderate’, the non-significant differences between ‘weak’ and ‘extremely strong’, between ‘moderately strong’ and ‘very strong’, and between ‘moderately strong’ and ‘extremely strong’ are likely to be attributable so the aforementioned small number of extreme responses to ‘very strong’ and ‘extremely strong’ in this group.

Table 4. The significance of observed differences in perceived strength between expressions for each of the three sample groups ( $p < 0.05$ ).

	Weak	Moderate	Moderately strong	Strong	Very strong	Extremely strong
Weak		Yes- Lay No - Legal No - Forensic	Yes - Lay Yes - Legal No - Forensic	Yes - Lay Yes - Legal Yes - Forensic	Yes - Lay Yes - Legal Yes - Forensic	Yes - Lay No - Legal Yes - Forensic
Moderate			No - Lay Yes - Legal No - Forensic	Yes - Lay Yes - Legal Yes - Forensic	Yes - Lay Yes - Legal Yes - Forensic	Yes - Lay Yes - Legal Yes - Forensic
Moderately strong				No - Lay Yes - Legal Yes - Forensic	Yes - Lay No - Legal No - Forensic	Yes - Lay No - Legal Yes - Forensic
Strong					No - Lay No - Legal No - Forensic	No - Lay No - Legal No - Forensic
Very strong						No - Lay No - Legal No - Forensic
Extremely strong						

Next, the significance of observed differences in the perception of the same expression by the different sample groups was examined. Non-parametric Mann-Whitney *U* tests were employed. In table 5 those differences which were deemed to be significant, i.e. those instances where the difference in perception of the same expression across two of the sample groups was deemed statistically significant at  $p < 0.05$  and therefore, where there was significant discordance between the groups, are indicated by “Yes”. The remaining differences were not significant ( $p > 0.05$ ) and are indicated by “No”. Overall, there were very few instances where a significant difference was observed in the perception of the same expression by two of the sample groups; there appeared to be concordance across the sample groups. The two exceptions involved the forensic and legal groups – the differences observed in the perception of both ‘moderate’ and ‘extremely strong’ between these groups were deemed to be significant ( $p < 0.05$ ) (table 5).

Table 5. The significance of observed differences in perceived strength of verbal expressions between the sample groups: lay/legal, lay/forensic and legal/forensic ( $p < 0.05$ ).

	Weak	Moderate	Moderately strong	Strong	Very strong	Extremely strong
Weak	No - Lay/Legal No - Lay/Forensic No - Legal/Forensic					
Moderate		No - Lay/Legal No- Lay/Forensic Yes - Legal/Forensic				
Moderately strong			No - Lay/Legal No- Lay/Forensic No - Legal/Forensic			
Strong				No - Lay/Legal No- Lay/Forensic No - Legal/Forensic		
Very strong					No - Lay/Legal No- Lay/Forensic No - Legal/Forensic	
Extremely strong						No - Lay/Legal No- Lay/Forensic Yes - Legal/Forensic

#### 4. Discussion

In general, the observed trend in this study was one of increased perceived strength as the intended strength was increased. There was concordance between the sample groups in this respect, albeit with some instances of divergence from this trend. The participants in this study tended to perceive differences in the strength of evidence indicated by different verbal expressions and between the upper and lower ends of the scale. However, upon testing the significance of the differences in perception of the verbal expressions, it is clear that participants performed less well when it came to distinguishing between the highest gradations on the scale (i.e. 'strong', 'very strong' and 'extremely strong'). In addition, when all responses were combined, the measures of central tendency for the perceptions of evidential strength for these three expressions – 'strong', 'very strong' and 'extremely strong' - were similar ( means = 12.4, 12.8 and 12.3 and medians = 13, 14 and 14, respectively). It is possible, therefore, to question the effectiveness of the scale of verbal expressions in communicating the intended evidential strength at the higher end of the scale.

Relatively high levels of variability and overlap were observed in the way in which expressions were perceived by participants and there is evidence to suggest that participants also performed poorly when it came to distinguishing between 'weak and 'moderate' and between 'moderate' and 'moderately strong'. It should be acknowledged, particularly in the case of the legal group, that a small number of extreme responses at either end of the scale rendered the differences between 'weak' and 'extremely strong', for example, non-significant. It is possible that a small number of respondents misinterpreted the orientation and meaning of either end of the scale. However, without the opportunity to interview these respondents, this cannot be assumed. When the perceptions of the verbal expressions between the different sample groups were compared, very few significant differences were observed; only two of the 18 pairs that were compared (in table 5) revealed significant differences. Therefore, it is possible to conclude that the correspondence between the groups was good; large divergences were not generally observed.

When considering the results of this study, it is important to acknowledge that the scenario was fictional and the scenario was simplified. It is not clear whether differences would exist between decisions made in real and hypothetical settings. However, the extension of this finding to the validity of jury decisions is not straightforward and requires extensive further exploration. In addition, it is acknowledged that the lay responses were provided by individuals, rather than as a result of consensus between jury members.

While sample sizes in each group were relatively modest (n=88, 84, 58 for the lay, legal and forensic participant groups, respectively), the data did reveal noteworthy trends and observations, particularly with regard to the 'clustering' of perceptions at the top end of the scale and the relative lack of divergence between sample groups. Nonetheless, it would be worthwhile to extend the exercise to larger and more diverse sample groups. For example, it would be productive to gauge the perceptions of a group that consisted exclusively of

casework forensic scientists who are routinely engaged in presenting evidence in this manner. This would enable a comparison between those engaged in presenting evidence and other actors in the criminal justice system. Further studies might also consider the impact of presenting a reference scale and whether this might be a means of improving the correspondence between intentions and perceptions in practice. The study did not compare responses to those from an alternative scale, but instead measured the extent of any divergence from a model in which intentions and perceptions align. Following previous work [5, 6, 7, 8], this study provides additional evidence that the potential for miscommunication when verbal expressions are used and the perception problems observed suggest a number of divergences between the intended meaning of the expressions and the way in which they are perceived. The results and findings generated by this study highlight the need for the undertaking of further validation studies in alternative contexts. They also provide support for calls to pursue the development of validated and reliable scales to aid the communication of evaluative opinions [6].

## 5. Conclusion

In conclusion, this study represents a contribution to the body of work concerned with the perception of verbal expressions of evidential strength in a forensic context. Specifically, the data generated by this study indicate that, while perceived strength was found to increase with the intended strength of the expression, there were notable exceptions. The similarity of perceptions at the higher end of the scale calls into question the effectiveness of these expressions in conveying large differences in the strength of evidence, while the same may be argued for the distinctions between 'weak and 'moderate' and between 'moderate' and 'moderately strong'. These results suggest we may not be able to assume that decision-makers will be able to discern between these expressions. However, the results of this study indicate that these trends were not confined to a particular sample group and that there was relatively good concordance between the groups of decision-makers in the way that expressions were perceived. Findings such as these should represent a contribution to continued efforts to understand the way that uncertainty in forensic science is conveyed and understood and they highlight further the need for empirical validation of presentation and communication methods in the forensic sciences. The manner of expressing and reporting the strength of evidence should be high on the research agenda. Indeed, the need for the empirical validation of frameworks for communicating evaluative opinions [6] should resonate with calls for a "research culture" in the forensic sciences [9]. As well as striving to improve the validity and robustness of the interpretation of evidence at the source and activity levels, research should also be focused on ensuring clear, effective and unambiguous communication when it comes to conveying the strength of evidence.

## References

- [1] Martire, K., Kemp, R., and Newell, B. 2013a. The psychology of interpreting expert evaluative opinions. *Australian Journal of Forensic Sciences* 45 (3) 305-314
- [2] Evett, I. 1998. Towards a uniform framework for reporting opinions in forensic science casework. *Science and Justice* 38 (3) 198–202.
- [3] Aitken, C., et al. 2011. Expressing evaluative opinions: a position statement. *Science and Justice* 51 (1) 1–2.
- [4] Association of Forensic Science Providers. 2009. Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice*, 49 (3) 161–164
- [5] Mullen, C., Spence, D., Moxey, L., and Jamieson, A. 2014. Perception problems of the verbal scale. *Science and Justice* 54 (2) 154-158
- [6] Martire, K., and Watkins, I. 2015. Perception problems of the verbal scale: A reanalysis and application of a membership function approach. *Science and Justice* 55 (4) 264-273
- [7] Martire, K., et al. 2013b, The expression and interpretation of uncertain forensic science evidence: verbal equivalence, evidence strength, and the weak evidence effect. *Law and Human Behaviour* 37 (3) 197-207
- [8] Martire, K., Kemp, R., Sayle, M., and Newell, B. 2014. On the interpretation of likelihood ratios in forensic science evidence: presentation formats and the weak evidence effect. *Forensic Science International*, 240 61-68
- [9] Mnookin, J., et al. 2011. The need for a research culture in the forensic sciences. *UCLA Law Review* 58 (3) 725-780