Dean&Francis

ISSN 2959-6149

https://doi.org/10.61173/e99e1276

The Impact of Retention Interval on Recall Confidence and Eyewitness Testimony Accuracy

Yixuan Zhang

Abstract:

It is widely acknowledged that eyewitness testimony plays a crucial role in the criminal justice system. There has been a significant public concern regarding the number of people being wrongly accused of certain crimes. An essential factor contributing to this issue is the high malleability and fallibility of memory due to various elements. This study focuses closely on the retention interval and its impact on recall confidence and eyewitness accuracy. Twenty participants were invited to take part in an experiment. They were required to answer open-ended questions at different retention intervals (immediate, 5-10 minutes, a day, a week) after watching a video about a crime. Our overall analysis indicates that the length of retention interval significantly affects both overall accuracy and confidence level; i.e., the accuracy and confidence level tend to decrease as the retention interval increases. The overall Confidence-Accuracy (CA) relationship shows a significant and positive correlation. However, when considering the retention interval, a non-significant relationship is observed for all CA relationships. Meanwhile, both eyewitness accuracy and confidence level decrease as item difficulties increase. Among all categories, "scene" items have the highest accuracy, while "action" items have the highest rate of "don't remember".

Keywords: Eyewitness testimony, retention interval, eyewitness accuracy, recall confidence, confidence-accuracy relationship

1. Introduction

It is widely acknowledged that eyewitness testimony plays a crucial role in the criminal justice system (Wells & Olson, 2003). There has been a significant public concern regarding the number of people being wrongly accused of certain crimes. Evidence has shown the estimation that 6 out of every 100 prisoners in a general state prison are innocent even after the advent of DNA tests (Loeffler, Hyatt, &

Ridgeway, 2019). The Innocence Project (2024) states that eyewitness misidentification serves as the primary leading cause of wrongful convictions, being involved in 69% of DNA exonerations.

A crucial factor contributing to this issue is the high malleability and fallibility of memory (Loftus, 2005; Wade, Rowthorn, & Sukumar, 2017). Various factors such as emotions (Glomb, 2022), lineup models (Wells & Turtle, 1986), race difference (Brigham, Bennett, Meissner, & Mitchell, 2007), ways of questioning (Loftus & Zanni, 1975), and time delay (Wheatcroft, Wagstaff, & Manarin, 2015) can pose threats to the accuracy of eyewitness testimony, resulting in misidentification. According to the research by Wheatcroft et al. (2015), after justice associations became aware of the issue created by time delay, relevant legislation was introduced to minimize time delays (Manarin, 2009; Riddle, 2012), and numerous benefits can be achieved thereby (Hanna, Davies, Henderson, Crothers, & Rotherham, 2010; Manari, 2009).

When misidentification occurs, witnesses do not feel like lying. On the contrary, they possess great confidence in the accuracy of their recalls (Garrett, 2011). Taking the well-known Ronald Cotton case as an example, victim Jennifer Thomson identified Cotton as the sex rapist with extremely high level of confidence. Despite her determination at the time, she was wrong (Loftus, 1996).

As relatively few studies have investigated the relationship among time delay, recall confidence and accuracy of eyewitness testimony. Meanwhile, yet, the studies regarding the relationships between these factors are inconclusive and highly inconsistent. This present research aims to investigate the impact of retention interval on recall confidence and memory accuracy, along with the relationship between recall confidence and eyewitness accuracy on a relatively more comprehensive consideration of controlled factors.

2. Literature Review

2.1 Confidence, an unreliable factor

Recall confidence refers to the degree of certainty of retrieving memory. Confidence level plays an important role in convincing the juries as a lack of confidence would lead to doubt in the reliability of one's testimony (Sauer, Brewer, Zweck, & Weber, 2010; Wells, 1985). However, despite the significance of recall confidence in making people believe in their words, recall confidence during identification tests is highly unreliable (Wixted & Wells, 2017; Berkowitz, Garrett, Fenn, & Loftus, 2020) and even if the witnesses hold higher confidence because they have experienced a longer exposure to culprit's face, it does not invariably guarantee that witnesses will be able to identify the person on a subsequent occasion (Shepherd, Ellis, & Davies, 1982; Devlin, 1976). The experiment conducted by Wells et al.. (2000) also provided evidence by finding eyewitnesses who were assertive but incorrectly identified the defendants. Therefore, the curiosity about the relationship between recall confidence and eyewitness accuracy was further triggered among many researchers, who laid their focus on the examination of this relationship. However, many have not reported a significant CA relationship (Cutler, Penrod, & Martens, 1987; Sporer, Penrod, Read, & Cutler, 1995), unless under specific conditions.

2.2 Factors affecting CA relationships

Studies have provided information that confidence could vary due to different factors. According to Wixted & Wells (2017), identification confidence increases after the confirmatory feedback from the police. To be more specific, when the identification officer shows a positive attitude in confirming the decisions made by witnesses who have just completed identifying one of the suspects, the confidence level of witnesses will increase when they are asked about it again afterward. It is also illustrated that the confidence of eyewitnesses was higher in correct identification than incorrect witnesses in short exposure condition, while the long exposure condition shows no difference in confidence ratings (Memom, Hope, & Bull, 2003). Moreover, age functions greatly in affecting the confidence level. Children below 11 years old tend to exhibit excessive confidence in their incorrect identifications (Brewer & Day, 2005), and the significance of an eyewitness's confidence diminishes when the individual reaches 40 years old or above (Martschuk, Sporer, & Saunderland, 2019), as mentioned by Berkowitz et al. (2020). Meanwhile, Poole and White (1993) have stressed that children perform poorly to open-ended questions (less accurate) and yes-no questions (less consistent), they also tend to fabricate certain responses and confuse the action of research assistants.

2.3 Past research on examination of CA relationships

Several studies have demonstrated the correlations between recall confidence and accuracy of eyewitness testimony. These correlations concluded by previous studies will be illustrated in the following paragraphs. However, the outcomes were highly inconsistent and controversial as some of them were mutually contradictory due to different research variables and methodologies, such as retention interval (e.g., Sauer et al., 2010; Odinot & Wolters, 2006), exposure duration (e.g., Lindsay, Read, & Sharma, 1998; Wells & Murray, 1983), and divided attention (e.g., Reinitz, Morrissey, & Demb, 1994), according to Palmer,

ISSN 2959-6149

Brewer, Weber and Nagesh (2013).

Since previous studies have provided sufficient evidence that CA relations can be influenced by multiple factors, each factor may have a different impact on the correlation of confidence and accuracy (Palmer et al., 2013). Therefore, in order to conclude a result with better precision and validity, a certain factor for the examination of the CA relationship is required.

2.3.1 Impact of retention interval on confidence level: a matter of repetition

When we exclusively concentrate on the factor of retention interval, which refers to the time gap between the event and the identification test, the associated research has yielded conflicting results as well.

Some research has illustrated that eyewitnesses' confidence level increases or decreases as the retention interval extends. For instance, confidence increased after 2 days with repeated questioning of the equivalent material (Hastie, Landsman, & Loftus, 1978), and after 3 weeks (Turtle & Yuille, 1994). While, confidence decreased after 1 week with the same condition and even dropped further for incorrect responses (Ryan & Geiselman, 1991). However, Granhag (1997) emphasizes the important side-effect of repetition, suggesting that the reason for confidence level to rise after delay may be due to the fact that memory was strengthened by repeated questioning but not delay itself. His results also provided support to the hypothesis, which showed improvement in CA relationships after a week's delay with repeated test and deteriorated in CA relationships if being tested for the first time. It is also worth mentioning that repeated recall can have both facilitating and detrimental effects on later retention since retrieving memory is not a neutral process, which contaminates memory (Roediger, McDermott, & Goff, 1977).

2.3.2 Research examples: "Significant and positive" and "insignificant" CA correlations

Research conducted by Odinot and Wolters (2006) has shown the result of relatively high and positive confidence-accuracy (CA) correlations in all conditions (delay and repetition) and the CA relations are very likely to predict accuracy. By referring back to their experiment data calculated by gamma correlations, they have concluded that "once information is retrieved it 'survives' and the content (accuracy) and confidence ratings remain stable". While Odinot and Wolters have indicated an opposite conclusion from Granhag (1997) that there are no clear indications of memory enhancement with repeated recall attempts as repeated questioning seems to only solidify retrieved information, but not significantly influence the confidence of accuracy. The same positive relationship was also obtained by Ryan and Geiselman, (1991) via their data, which two factors (confidence & accuracy) found to be significant. However, it is noted that accuracy remained statistically stable following a control passage and decreased following a misleading passage.

Meanwhile, some results have shown that the CA correlations are not significant. For instance, the study operated by Shaw and McClure (1996) produced such an outcome that the correlation between accuracy and confidence was low and insignificant in the initial test before declining even more when the test was repeated. Noticeably, this study stated similarly to Granhag (1997) about the great impact of repeated questioning and suggested participants under all conditions have expressed overall overconfidence in their responses. This insignificant relationship could also be shown by Wheatcroft et al. (2015) in their within-subject CA relationships with repeated testing after a delay. However, such relations do not apply to difficult items (item difficulty) as the data has shown a significant main effect on them.

2.3.3 Previous research concerns and limitations

It is expressed by Odinot et al., (2006) that incorrect items with high confidence ratings are always inevitable if even the proportion is small. Therefore, although confidence plays an important role in accuracy prediction, it can not be identified alone as it could potentially cause harm to court situations and police investigations. Furthermore, even though a positive CA relationship was stated, it is only applicable when no misleading information is involved (Ryan et al., 1991).

Wheatcroft et al. (2015) illustrated that although the negative impact of using leading questions in identification tests is suggested by several researchers (e.g., Loftus, 1979; Clifford & Scott, 1978), it is allowed as this study attempted to simulate a courtroom condition where layers may ask such questions to witnesses and stating that "a firm rationale has developed in legal culture whereby leading questions may be permitted during cross-examination (Keane and Fortson, 2011)". Moreover, it is stressed that the CA calibration might improve if witnesses were allowed to review and alternate their previous responses, and the effect of reviewing previous responses on CA relationships for items of varying difficulty has yet to be investigated. Eventually, this study is not generalizable enough due to the limited sample size.

3. Methodology

3.1 Participants

A total of 25 students (19 females), aged 15-21 (M = 17.75, SD = 1.178), were recruited as volunteers. 5 participants were involved in pilot study (informal/test experi-

ment). All participants were either high school students or undergraduates.

3.2 Experiment Design

The independent group experiment with only initial questioning (no repetition questioning group) is designed to investigate the impact of retention interval on the confidence, accuracy, and CA relationship by recording and analyzing data from participants. Participants were randomly assigned to one of the four conditions. Condition 1 (n=5) had no retention interval (immediate), Condition 2 (n=5) had a retention interval of 5-10 minutes, Condition 3 (n=5) had a retention interval of a day, and Condition 4 (n=5) had a retention interval of a week.

3.3 Materials

3.3.1 Consent form

Participants were asked to read and sign a consent form before they participated in the experiment. The consent form includes basic information, cautions, and requirements of the experiment. After reading the consent form, participants shall understand their rights during the experiment that any information recorded in the investigation will remain confidential and no information that identifies them will be made publicly available.

3.3.2 Video clip

A 56 second long colour video clip (1080HD pixels) without audio, selected from YouTube, was shown individually to the participants online, using a high-quality 14-inch laptop screen. The video presents a view from the surveillance camera of a real crime, which consists of an appropriate level of violence about a man trying to shoot another man in a parking lot at night with cars driving by while a bystander is walking closely next to the crime scene.

3.3.3 Questionnaires

After a certain retention interval, a questionnaire was constructed of 22 open-ended questions (Loftus & Goodman, 1985) concerning several aspects of the video and 22 questions towards participants' confidence level for each open-ended question given on a Likert scale ranging from 1 to 6, where (1) represent 'pure guess' and (6) represent 'absolutely certain' (Wheatcroft et al., 2015). The questions are classified into difficulty groups (e.g., "easy", "moderate", "hard") and category groups (e.g. "action", "appearance", "scene").

3.4 Procedure

The participants were informed that the experiments will be held online via an online meeting software (Tencent Meeting), their devices should be either iPad or laptop (computer) with a screen size above 10-inch. During the first session, participants were told to observe a non-audio video clip individually of an event for a period of around 1 minute on their devices with instructions from the investigator. They were then instructed to watch the video in a quiet environment and try to treat the video as a personal experience. All participants were then shown the video clip.

In the following recall session, they were asked to complete a questionnaire after a certain retention interval. Before filling in the questionnaires, participants were told to answer in their mother tongue even though they should answer according to the English version questions, fill in "Don't remember" for questions which they have no clue, read carefully and fully understand each question, and reach out when they fail to do so. All participants have no repetition condition, answering the questions for the first time (Roediger, McDermott, & Goff, 1977).

After completing the questionnaire, the investigator will ask follow-up questions upon participants' ambiguous and incomprehensible answers with open-ended questions when necessary, participants were asked if they could elaborate more about the answer and re-rate their confidence level if they are able to answer more specifically.

All answers from the participants were recorded as correct, incorrect, and don't remember. Correct information consists of information that compiles with the actual information in the video clip. Incorrect information consists of information that contradicts the actual information or is not present in the video clip. Answers include both correct and incorrect information are considered to be incorrect. For example, in the experiment, the question "What is the type of the transport? Please describe it.". The correct answer is "A white SUV". The answer "A white van" or "A black SUV" is regarded to be wrong. For questions participants no longer remember and responded as don't remember were neither considered correct nor incorrect.

3.5 Statistical analysis

In this research, the Pearson correlation coefficient and average mean calculation were utilized to determine the CA relationship considering questions with different difficulties and different categories. These analyses were carried out via SPSSAU.

4. Result

4.1 Overall accuracy and confidence level with retention intervals

Total percentages of accuracy (Percentage of the correct

ISSN 2959-6149

responses in total question number), and mean confidence level (Average the sum of average confidence level for each participant) were calculated for both correct and incorrect responses in the within-subject method (Wheatcroft et al., 2015). The overall results are shown in Table 1.

Preliminary analysis showed that the overall accuracy of correct responses is significantly larger than incorrect responses. Meanwhile, as the retention interval increases from the "immediate" condition to the "a week" condition, the accuracy for correct responses decreases, and accuracy for incorrect responses increases. However, noticeably the 5-10 minutes condition shows differently against the above trend, the contribution factor to this result will be illustrated in the discussion section.

Furthermore, participants tend to have higher confidence level in their correct responses rather than incorrect responses, and the confidence levels for correct responses drop consistently as retention interval increases (except 5-10 minutes condition). Surprisingly, the confidence level for incorrect responses in the "a week" group dramatically increased (M = 3.484) and even exceeded the confidence level for incorrect in the "immediate group" (M = 3.222).

Table 1: Overall Results for Within-Subjects' Mean Total Accuracy and Mean Confidence Level, at 4 different retention intervals

	Retention interval							
	Immediate		5-10 minutes		A day		A week	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Accuracy%	65.45	16.36	74.55	15.45	59.09	24.55	57.27	28.18
Confidence	4.694	3.222	4.854	3.875	4.477	2.667	4.318	3.484

4.2 Accuracy and confidence level with item difficulties among different retention intervals

A further analysis on the accuracy and confidence level for correct responses and incorrect responses on the within-subject was conducted in Table 2. The analysis separated four retention intervals further into three item difficulties, including "easy", "moderate", and "hard" considering the difficulty of spotting answers for each question. Mean accuracy and confidence level for each participant among different item difficulties are calculated, summed up, and taken average. DR in the chart stands for don't remember, which was neither regarded as correct nor incorrect.

Apparently, correct accuracy for easy items is much higher than for moderate and hard items, and moderate items have better accuracy than hard items no matter the length of retention intervals. As expected, participants' confidence level for easy items was the highest, which decreases as the item difficulty increases (no matter the retention interval). In addition to the results, the rate of "don't remember" for hard items was significantly higher than other item difficulties.

Table 2: Results on the accuracy and confidence level at different item difficulties

		Retention interval											
		Immediate			5-10 minutes			A day			A week		
		Easy	Mod	Hard	Easy	Mod	Hard	Easy	Mod	Hard	Easy	Mod	Hard
Cor	Rate%	86.7	67.3	25.0	90.0	76.4	40.0	93.3	52.7	20.0	93.3	54.6	10.0
	Confi	5.31	4.65	2.00	5.44	4.81	2.75	5.11	3.93	3.75	5.00	3.70	2.00
Incor	Rate%	3.33	23.6	20.0	10.0	16.4	25.0	6.67	29.1	40.0	6.67	32.7	50.0
	Confi	1.00	3.31	2.75	5.33	3.88	3.00	2.50	2.75	2.50	3.00	3.83	3.00
DR	Rate%	10.0	9.09	55.0	0.00	7.27	35.0	0.00	18.2	40.0	0.00	12.7	40.0
	Confi	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

4.3 CA relationship analysis

The statistical analysis for the overall CA relationship was carried out (see Table 3). The total CA relationship shows a positive and significant proportion ($r=.553$, $p<0.05$), showing a tendency that the higher the confidence level the higher the accuracy for correctly responding to the question.

Table 3: Pearson correlation analysis on the overall CA relationship

		Confidence level
Accuracy%	Pearson correlation coefficient	.533*
	Significance	.016
	Case number	20

Since the Pearson correlation coefficient in the overall CA relationship shows significance, a further analysis based on different retention intervals is investigated (see Table 4). It is shown that the retention intervals weakened the CA relationship. As retention interval increases, the CA relationship remains positive but decline, which then suddenly decreases to negative when retention interval gets to "a week". However, the results appeared very insignificant ($p>0.05$).

Table 4: Pearson correlation on CA relationship among different retention intervals

Pearson correlation coefficient (significance)		Confidence level			
		Immediate (n=5)	5-10 minutes (n=5)	A day (n=5)	A week (n=5)
Accuracy%	Immediate (n=5)	.792 (.775)			
	5-10 minutes (n=5)		.611 (.379)		
	A day (n=5)			.533 (.110)	
	A week (n=5)				-.617 (.268)

4.4 Addition analysis on item categories with retention interval and accuracy

An additional analysis on the accuracy and confidence level of the within-subject was conducted in Table 5. The analysis separated four retention intervals further into three item categories, including "action", "scene", and "appearance" considering the classification of each question.

The table provides clear information that the "scene" category has the highest accuracy, while "action" category has nearly the highest rate of not remembering the responses.

Table 5: Results on the accuracy and confidence level at different item categories

		Retention interval											
		Immediate			5-10 minutes			A day			A week		
		Act	Sce	App	Act	Sce	App	Act	Sce	App	Act	Sce	App
Correct	Rate%	45.7	91.4	60.0	65.7	91.4	67.5	48.6	94.3	37.5	51.4	80.0	42.5
DR	Rate%	40.0	2.86	12.5	14.3	5.71	10.0	31.4	0.00	17.5	20.0	0.00	22.5

5. Discussion and evaluation

5.1 Discussion

5.1.1 Overall impacts on accuracy and confidence

In sum, the overall analysis demonstrated that the length of retention interval had a significant impact on both overall accuracy and confidence level. Specifically, as the time delay lengthens, the accuracy and confidence level tend to declined. Additionally, incorrect responses generally exhibit lower accuracy and confidence level than correct responses. The above result is consistent with data obtained

ISSN 2959-6149

by Odinot and Wolters (2006), which showed that longer retention intervals before first questioning led to fewer correct units of information being recalled and lower confidence scores; higher accuracy for correct responses than incorrect responses is also shown in their study. Furthermore, as the retention interval increases, the accuracy of giving correct responses decrease while the accuracy giving incorrect responses increase. This finding stands in stark contrast to that of Odinot and Wolters (2006) who illustrated opposite results (accuracy of correct responses and incorrect responses increase and decrease respectively as time delay increase) in their table, however, it is in line with the study conducted by Palmer et al., (2013). Moreover, surprisingly, the confidence level for incorrect responses in the "a week" was unexpectedly high, indicating that when time delay increases to a certain point, the confidence level is no longer matches with accuracy.

The above results underscore the significance of the retention interval in eyewitness identification. This implies that identification officers are obligated to minimize the time delay to the greatest extent possible in order to guarantee the freshness of memory and thereby enhance the accuracy of testimony.

5.1.2 CA relationship

The subsequent discussion (5.1.2.1 & 5.1.2.2) regarding the CA relationship demonstrates that the CA relationship undergoes variation (ranging from significant and positive in the overall case to insignificant for different retention intervals) when retention intervals are taken into consideration.

5.1.2 .1 Overall CA relationship

Referring to the Pearson correlation coefficient test for CA relationship, the overall CA relationship shows both significant and positive, indicating a direct proportionality between confidence level and accuracy that it is very possible for witnesses to obtain higher accuracy responses when given higher confidence level. This outcome is supported by Sauer et al. (2010) that CA relationship was shown to be positive and significant for choosers before and after delay.

5.1.2 .2 CA relationship regarding retention interval

Additionally, a non-significant relationship is observed for all CA relationships when considering the retention interval. Shaw and McClure (1996) also yielded the same insignificant CA correlation outcome. They stressed that "the correlation between accuracy and confidence was low and insignificant in the initial test before declining even more when the test was repeated".

The CA relationship, in specific, remains positive but continues to decrease as retention intervals increase (from "immediate", "5-10 minutes", to "a day"). This result is in line with the study by Odinot and Wolters (2006), which found a high and positive relationship between confidence and accuracy, especially with the shortest recall interval.

5.1.3 Whether confidence level is usable for predicting accuracy

However, noticeably, a deterioration of negative correlation is shown in the "a week" group, suggesting that for longer retention intervals, the confidence level is very likely to be skewed and imprecise to predict accuracy. Thus, this study demonstrates that confidence level is a crucial and valuable instrument for predicting accuracy, however, serves mainly for shorter retention interval (within 24 hours) since the outcomes referred.

It is indeed intriguing to note that the highest confidence ratings are accorded to a high level of accuracy, especially after a brief delay. Although the overall correlation between confidence level and eyewitness accuracy appears significant and positive, confidence level can never, by itself alone, become a definitive judgement of a statement's correctness. This is because there are cases where low-accuracy testimony is expressed with high confidence, even though such occurrences are only a small proportion.

5.1.4 Item difficulty and item category impact on recall confidence and accuracy

Furthermore, in accordance with varying item difficulties, both eyewitness accuracy and confidence level decline as item difficulties increase. Among different item difficulties, the "easy" items possess the highest average accuracy and mean confidence level among other item difficulties. This finding was depicted in the table reported by Wheatcroft et al., (2015) and Pulford et al., (1997). Meanwhile, "hard" items have the highest rate of "don't remember", referring that they are the most prone to be forgotten. To explain why such outcome occurs, studies have found support for the reason behind the tendency of "easy items" to show better recall performance; they have indicated that "easy items" are typically remembered in an "all or none" fashion (Wheatcroft et al., 2015; Lindsay, Read, and Sharma, 1998), meaning that if participants pay attention to it then they remember it, otherwise, they do not. Consequently, the CA correlations for "easy items" tend to be higher than other item difficulties.

It is also worthy of mention that item categories may likewise exert an influence on accuracy. "Scene" items exhibit the highest accuracy among all categories ("action"," scene" and "appearance"), meaning that participants are comparatively more likely to provide correct responses to "scene" questions. Conversely, "action" items have the highest rate of "don't remember", referring that participants tend to forget their answers to "action" questions. The research on the "weapon focus" (E.F. Loftus, G.R.

Loftus, & Messo, 1987) might offer a possible explanation for why "action" and "appearance" questions hold lower accuracy than "scene" questions. They positioned that subjects who witnessed an event containing a weapon were not only less likely than controls to accurately identify the perpetrator, but they were less accurate when answering specific questions about him.

Therefore, during eyewitness identification, the occurrence of weapon needs to be in careful consideration. In a real-life situation, "weapon focus" could emphasize the existence of a weapon and intensify people's anxiety level, leading to higher stress and causing the perceptual focus to be narrowed in range (Easterbrook, 1959), which eventually lower the accuracy of the memory about the event.

5.1.5 Anxiety, a potential effect on eyewitness performance

As Table 1 and Table 2 have revealed, an unexpected result emerged in the "5-10 minute" retention interval did not follow the decreasing trend as the time delay increases. It is known that the level of anxiety and stress can impact memory accuracy and confidence, leading to inferior eyewitness performance (Alho, Rodrigues, & Fidalgo, 2019). The "inverted U" shape is obtained in the result (Yerkes & Dodson, 1908). In Yerkes and Dodson's law, it is stated that the reason for the "inverted U" shape is that optimal performance often occurs when the anxiety level is at moderate state. Hence, a possible explanation for the "5-10 minutes" retention interval group having better accuracy for their responses and standing out against the trend might be due to the relatively moderate anxiety level the "5-10 minutes" participants were experiencing. Another potential reason for this phenomenon is the opportunity for memory rehearsal. The "immediate" group was not provided with time for rehearsal, while any retention interval other than "immediate" could allow participants to retrieve their memory of the event. However, this hypothesis is not supported by Craik and Watkins (1973), who have demonstrated that maintenance rehearsal does not lead to an improvement in memory performance.

This outcome serves as a reminder that anxiety is a crucial matter during identification task. Great attention should be paid to the questioning method. It should neither overly arouse witnesses' stress nor render them overly relaxed. Ensuring that their anxiety level is at moderate could lead to optimal performance in retrieving the memory.

5.2 Evaluation

Cognitive Interview (CI) is a valuable investigative tool as it is an effective, efficient, and legally acceptable memory-enhancement technique (Geiselman, Fisher, & MacKinnon, & Holland, 1986). However, this study deviated from the Cognitive Interview Technique due to limited training and knowledge. Since CI is designed to enhance the eyewitness recall accuracy (it assists to elicited significantly more correct information from the subjects than the standard police interview (SI) without an increase in incorrect information or confabulation), employing CI in experiments could potentially increase the rigor and representativeness of the results.

Moreover, in light of the "inverted U" shape phenomenon, this unexpected outcome might be due to the insufficient sample size (n=5 per group, total 20 participants). As a potential solution, enlarging the sample size could significantly reduce the proportion of outliers, leading to a more accurate result and eventually yield a different outcome.

In addition, the experiment was conducted online under less rigid requirement for devices. For example, the experiment required participants to use device which screen size should be above 10-inch instead of using the same device with a fixed screen size. Studies have stated the significant influence of screen size on memory (Reeves, Lang, Kim, & Tatar, 1999) If the experiment could be carried out offline though one laptop, this extraneous variable could be eliminated by employing the same device.

Finally, the age and gender factor should be considered. Multiple studies have illustrated the crucial impact of age groups on memory accuracy and recall confidence (e.g., Brewer & Day, 2005; Martschuk et al., 2019; Poole & White, 1993) and the effect of gender on memory performance; neither sex has superior memory ability, but that they do differ in terms of what is remembered (Loftus, Banaji, Schooler, & Foster, 1987).

6. Conclusion

In this investigation, the objective was to appraise the influence of retention interval on recall confidence and eyewitness accuracy, as well as the confidence-accuracy relationship. Concurrently, assess the potential schemes for enhancing the accuracy, precision, and rigor of eyewitness testimony.

Taken as a whole, this study emphasizes that time delay has indeed exerted a significant influence on recall confidence and eyewitness accuracy, highlighting the importance and necessity of minimizing the length of retention interval in the justice system to ensure the freshness of memory and improve the quality and quantity of eyewitness testimony. The result also underscores that the confidence level can only serve as an auxiliary reference for accuracy prediction since cases of high-confidence but low-accuracy testimony occur. Thus, over-reliance on recall confidence is inappropriate. However, this study has several limitations (as shown in 5.2 Evaluation). Future research could improve the experiment by increasing sam-

ISSN 2959-6149

ple size and exploring different factors (such as age, gender, the use of CI, etc) with more rigorous standardization. In summary, the present results lend support to the public's concern regarding the negative impact of time delay on eyewitness testimony and emphasize the crucial importance of minimizing the length of retention interval for identification tests.

References

Alho L, Rodrigues PF, Fidalgo C. Do stress and anxiety impact memory? an exploratory Portuguese study. Journal of Psychological Research. 2019 Sep 30;1(3):31-8.

Berkowitz, S. R., Garrett, B. L., Fenn, K. M., & Loftus, E. F. (2022). Convicting with confidence? Why we should not overrely on eyewitness confidence. Memory, 30(1), 10-15.

Brewer, N., & Day, K. (2005). The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. Psychiatry, Psychology and Law, 12, 119–128. https://doi. org/10.1375/pplt.2005.12.1.119

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In The Handbook of Eyewitness Psychology: Volume II (pp. 271-296). Psychology Press.

Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. Journal of Applied Psychology, 63(3), 352.

Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. Journal of verbal learning and verbal behavior, 12(6), 599-607.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. Law and Human Behavior, 11, 233–258. doi:10.1007/BF01044644

Devlin, Hon. Lord P. (1976). Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal Cases. London: Her Majesty's Stationery Office.

Easterbrook, J. A. (1959). The effect of emotion on the utilization and orfanization of behaviour. *Psychological Review*, 66, 183-201.

Garrett, B. L. (2011). Convicting the innocent: Where criminal prosecu- tions go wrong. Harvard University Press.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. The American journal of psychology, 385-401.

Granhag, P.A. (1997). "Realism in eyewitness confidence as a function of type of event witnessed and repeated recall." Journal of Applied Psychology, 82(4): 599-613. http://dx.doi.org/10.1037/0021-9010.82.4.599

Głomb, K. (2022). How to improve eyewitness testimony research: Theoretical and methodological concerns about experiments on the impact of emotions on memory performance. Psychological research, 86(1), 1-11.

Hanna, K., Davies, E., Henderson, E., Crothers and Rotherham, C. (2010). "Child witnesses in the Criminal Courts: A Review of Practice and Implications for Policy." Auckland: The New Zealand Law Foundation and Institute of Public Policy, Auckland University of Technology.

Hastie, R., Landsman, R. and Loftus, E.F. (1978). "Eyewitness testimony: The dangers of guessing." Jurimetrics Journal, 19: 1-8. http://dx.doi.org/10.2307/29761641

Innocence project (2024). Innocence Project. Available at: https://www.innocenceproject.org/

Keane, A. and Fortson, R. (2011). "Leading Questions: A critical analysis." Criminal Law Review, 4: 280-295.

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. Psychological Science, 9, 215–218. doi: 10.1111/1467-9280.00041

Loeffler, C. E., Hyatt, J., & Ridgeway, G. (2019). Measuring self-reported wrongful convictions among prisoners. Journal of Quantitative Criminology, 35, 259-286.

Loftus, E. F. (1996). Eyewitness testimony. Harvard University Press.

Loftus, E. F. (2005). Memories of things unseen. Curr. Dir. Psychol. Sci. 13, 145–147.

Loftus, E. F., & Goodman, J. A. N. E. (1985). Questioning witnesses. The psychology of evidence and trial procedure, 253-279.

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. Bulletin of the Psychonomic Society, 5(1), 86-88.

Loftus, E. F., Banaji, M. R., Schooler, J. W., & Foster, R. (1987). Who remembers what?: Gender differences in memory (pp. 64-85). Ann Arbor,: University of Michigan.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and human behavior*, 11(1), 55-62.

Manarin, B. (2009). "Bedeviled by delay: Straight talk about memory loss, procedural manipulation, and the myth of swift justice." Windsor Review of Legal and Social Justice, 27: 117-144.

Martschuk, N., Sporer, S. L., & Saunderland, M. (2019). Confidence of older eyewitnesses: Is it diagnostic of identification accuracy? Open Psychology, 1, 132–151. https://doi.org/10.1515/psych-2018-0010

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. British journal of psychology, 94(3), 339-354.

Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. Applied Cognitive Psychology, 20(7), 973-985.

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The

confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. Journal of Experimental Psychology: Applied, 19(1), 55.

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and individual differences*, 23(1), 125-133.

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The effects of screen size and message content on attention and arousal. Media psychology, 1(1), 49-67.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 161–168. doi:10.1037/0278-7393.20.1.161

Riddle, H. (2012). "The Stop Delaying Justice! Initiative." The Law Gazette (19th January), accessed at: http://www.lawgazette.co.uk/in-practice/practice-points/the-stop-delaying-justice-initiative

Roediger, H. L., McDermott, K. B., & Goff, L. M. (1997). Recovery of true and false memories:Paradoxical effects of repeated testing. In M. A. Conway (Ed.), Recovered memories and falsememories (pp. 118–149). Oxford: Oxford University Press.

Ryan, R.H. and Geiselman, R.E. (1991). "Effects of biased information on the relationship between eyewitness confidence and accuracy." Bulletin of the Psychonomic Society, 29: 7-9. http://dx.doi.org/10.3758/BF03334752

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. Law and Human Behavior, 34, 337-347.

Shaw, J. S., & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. Law & Human Behavior, 20, 629–653.

Shepherd, J. W., Ellis, H. D., & Davies, G. M. (1982). Identification evidence. Aberdeen: Aberdeen University Press.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychological Bulletin, 118, 315–327. doi:10.1037/0033-2909.118.3.315

Turtle, J.W. and Yuille, J.C. (1994). "Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hyperamnesia." Journal of Applied Psychology, 79: 260-271. http://dx.doi.org/10.1037/0021-9010.79.2.260

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers criteria for judging eyewitness accuracy? Journal of Applied Psychology, 68, 347–362. doi:10.1037/0021-9010.68.3.347

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. Psychological bulletin, 99(3), 320.

Wells, G. L., and Olson, E. A. (2003). Eyewitness testimony. Annu. Rev. Psychol. 54, 277–295.

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station. American Psychologist. 55, 581-598.

Wells, G.L. (1985). "Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy." Journal of Applied Psychology, 70: 619-626. http://dx.doi.org/10.1037/0021-9010.70.4.619

Wheatcroft, J. M., Wagstaff, G. F., & Manarin, B. (2015). The influence of delay and item difficulty in criminal justice systems on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research*, 1, 1-9.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 18(1), 10-65.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation.