

Fall 1938

Time-Spaced Reporting of a Crime Witnessed by College Girls

Katherine Vickery

Lee M. Brooks

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/jclc>

 Part of the [Criminal Law Commons](#), [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

Recommended Citation

Katherine Vickery, Lee M. Brooks, Time-Spaced Reporting of a Crime Witnessed by College Girls, 29 *Am. Inst. Crim. L. & Criminology* 371 (1938-1939)

This Article is brought to you for free and open access by Northwestern University School of Law Scholarly Commons. It has been accepted for inclusion in *Journal of Criminal Law and Criminology* by an authorized editor of Northwestern University School of Law Scholarly Commons.

TIME-SPACED REPORTING OF A "CRIME" WITNESSED BY COLLEGE GIRLS

KATHERINE VICKERY* AND LEE M. BROOKS**

Although the problem of accurate witnessing and testimony has received considerable attention here and abroad for many years, the results about to be described include a few points that are new or variant as compared with the earlier studies.¹ The central purposes here are not only to attack the question of original accuracy in reporting an observed incident or crime but also, and more importantly, to test the witnesses' reliability with the passage of time.²

Out of a total of 150 sophomore college girls, 80 furnished sufficiently complete and usable reports of the same incident—a swift-moving and noisy affair involving three high school boys—to point toward significant conclusions. The reports were made spontaneously by the girls, whose afternoon class in a large semi-public auditorium was interrupted. The experiment was concerned primarily with what was actually seen and with judgment of time elements. The present article omits the available details having to do with audition.³

The study was based upon the hypothesis that not only is immediate testimony by the individual witness, even in relatively

* Alabama College.

** University of North Carolina.

¹ Cf. Moore, E. H., "Elements of Error in Testimony." *Journal of Applied Psychology*, Vol. XIX (1935), pp. 447-462. This is a review of the literature, including 50 titles.

² The many details of the present study have been treated statistically with great care by Dr. Vickery and are available in fair completeness.

³ The setting of the "crime" was as follows: Students at Alabama College for Women (year 1936) occupying center and left section of the auditorium, rows B to R, eighteen and fifty-five feet respectively from the open stage; stage is thirty-four feet wide; day is bright and clear; footlights have been turned off; stage has a grand piano, a bench, two big chairs, four light chairs, and a table with books; an intermediate curtain shuts off back stage. The class, started at 3:30 P. M., has been listening to the instructor for fifteen minutes when, from the inner side door at the front-left and in full view of everybody, a pursued boy dashes into the side aisle a step or two, exclaims confusedly, and then followed by his two pursuers mistakenly gets onto the stage, where, fully visible to the girl students, all three rush about amid overturning furniture, general clatter, and shouting. The incident, consuming only one minute and thirty seconds, was all prearranged and rehearsed with the three boys with due attention to numerous details of appearance, dress, action, and stage properties. The instructor followed and rebuked the intruders, then returned to ask the class to write out freely a description of the occurrence.

simple situations, likely to be highly untrustworthy but that as time passes the unreliability of the witness increases. The implications in connection with criminal court procedure are quite apparent.

The method of gathering and checking the data was as follows:

- a. Immediately after the "crime" and while excitement was rather highly pitched,—a freely written voluntary account on note paper was handed in;
- b. One week later, unannounced, the first objective test for recall was given, using mimeographed forms including true-false, completion, and multiple choice items;
- c. Seven weeks after the "crime," again unannounced, the second objective test was submitted, using mimeographed forms exactly as before.

It is in the three reports so given and spaced that the experiment offers its contribution as compared with other studies in this field.

As for methods of scoring and correlation: the written description, immediately attended to, was scored in somewhat the same fashion as the ten year reading test on the Stanford Revision of the Binet-Simon. Each statement was divided into as many units as there appeared to be separate facts, and the total number of these facts correctly recalled constituted the individual's score. A positive correlation of .078 between the number of correct and the number of erroneous statements made by the witnesses indicates that there is practically no relationship between the amount of information a witness will volunteer and the truth or falsity of the information.⁴

Since height, weight, and age estimates are commonly sought from witnesses, the results of student observations on these characteristics were the first to be tabulated.⁵ Statements made by the witnesses one week after the incident, estimating heights, were consistently lower than the actual heights of the individuals. After seven weeks the statements were still further from the true heights than were those after the first week. The amount of difference in the attempted estimate after one week was .21, .69, and .80

⁴ When the scores of the voluntary response were correlated with the first objective test (one week later) and with the second objective test (seven weeks later), the resulting r 's were .089 and .057 respectively. The scores of the first and second objective tests when correlated revealed r as .444.

The coefficient of variability on the first objective test was 57.7 and that of the second test, 73.8 This indicates that the scores of the first test are 78 per cent as variable as those of the second.

⁵ Numerous detailed tables and correlations support the narrative from this point on, but are omitted from this article for lack of space.

sigma for each of the three boys respectively. Seven weeks after the incident it was .87, .81, and .87 for the three boys respectively.

The weight of the heaviest boy was consistently underestimated, while that of the two smaller boys was overestimated. The differences between actual weights and estimated weights after one week in terms of sigma scores are for the three boys, 1.53, —.20, —.04 sigma, respectively. After seven weeks these differences became 1.48, —.51, —.11 sigma, respectively.

Subjects show greatest confidence in their ability to estimate age and more confidence in making statements about height than weight, if this confidence can be measured by the number who are willing to make a statement. Ages were estimated lower than the actual ages at the end of one week and after seven weeks for all three boys.

With regard to the range of inaccuracy on height, weight, and age, the possibility of individual inaccuracy in testimony is shown in the spread of estimates for height, weight, and age. One week after the "crime," individual estimates of the height of the first boy were from 4 feet 8 inches to 7 feet; of the second boy, from 4 feet to 6 feet 4 inches; of the third boy, from 4 feet to 5 feet 10 inches. After seven weeks these same individual witnesses estimated the heights respectively (boy 1) from 4 feet to 6 feet; (boy 2) from 4 feet 8 inches to 5 feet 10 inches; (boy 3) from 4 feet 8 inches to 5 feet 10 inches.

Weights were as variable: Estimates after one week range from (boy 1) 90 to 170 pounds; (boy 2) 100 to 160 pounds; (boy 3) 85 to 150 pounds. After seven weeks the estimated weights range from (boy 1) 105 to 165 pounds; (boy 2) 95 to 150 pounds; (boy 3) 80 to 145 pounds.

Age estimates range as follows: After one week (boy 1) from 11 to 20 years; (boy 2) from 12 to 19 years; (boy 3) from 10 to 18 years. After seven weeks the estimated ages range from (boy 1) 10 to 19 years; (boy 2) from 11 to 19 years; (boy 3) from 10 to 19 years.

In connection with the exact time and duration of the offense, the "crime" was committed on February 10, 1936, between 3:45 and 3:46½ P. M. After one week 86 per cent of the witnesses were still able to give the correct date; 9 per cent would make no statement; and 4 per cent missed the date by one day. After seven weeks 80 per cent refused to answer, and of those making a statement *none gave a correct date*. Answers varied from Febru-

ary 2nd to March 6th. Two persons responded that Monday was the day; yet all members of the class knew that the incident fell on a Monday since that was the only day of the week on which this entire class met together, but of the dates given none was for a Monday either in February or March. Although the class was in its twentieth joint session, always opening at 3:30 P. M., the estimates of the time of the occurrence varied from 3:00 P. M. to 4:15 P. M. when reported one week later. Seven weeks after the incident the variation was from 3:30 to 4:40 P. M. The Mean of the estimates after one week was 3:58 P. M., and sigma was 10.44. After seven weeks the Mean of the estimates remained 3:58 P. M. with sigma 11.49: After one week 90 per cent of the students made statements, but after seven weeks only 68 per cent replied.

The actual duration of the interruption was 1 minute and 30 seconds. After one week the witnesses were quite willing to estimate the time elapsing during the "crime"; 79 responses ranging from one minute to fifteen minutes; the Mean, 5.6 minutes and sigma 2.95. After seven weeks only 60 made replies and their estimates also varied from one to fifteen minutes; the Mean resting at 5.7 and sigma at 2.77. The actual time-elapse does not fall within one sigma of the estimated duration.

With regard to the rather fine pointed details of scene and action: Not infrequently witnesses in court will be asked questions calling for excessively keen discrimination. The shape or color of cephalic features are among the common materials out of which questions are constructed for the winning of cases. In this experiment it should be emphasized that the front rows of students were about twenty feet away and the back rows some fifty feet from the actors, that with the footlights off, the afternoon light was probably not sufficient to reduce the pupil of the eye and bring out the eye color.

The actual width of the stage was 34 feet. After one week, with students seated in their regular places from which they had viewed the action, only 45 of them ventured an estimate of the visible width of the stage although they were looking directly at it. *Estimates range from 5 feet to 100 feet!* The Mean of their estimates was 33.38 feet with a sigma of 19.89. Seven weeks afterward the range was the same but only 35 students attempted to make a

statement, the Mean of this estimate resting at 34.5 feet and sigma at 19.20.⁶

There were six chairs on the stage at the time of the chase. After one week 45 witnesses varied in their estimates of the number from 2 to 10 chairs. Seven weeks afterward they were more variable, estimating from 1 to 15 chairs,—in this instance only 34 students venturing a response. The Mean of the estimates made within one week was 4.92, with a sigma of 1.73, and after seven weeks the Mean was 4.41 with a sigma of 2.47.

The inquiry about eye-color was presented in completion test form. The two experimenters upon close study agreed upon the eye color of each boy. Boy number one: Brown eyes. After one week 82 per cent of the class could make no statement; 12 per cent thought they were brown; 3 per cent, blue; 1 per cent each stated dark, grey, and black. Seven weeks afterward, 94 per cent gave no answer; 4 per cent called them brown; and 1 per cent each stated blue and black. Boy number two: Light brown eyes. After one week: 87 per cent, no statement; 6 per cent, blue; 5 per cent, brown; and 1 per cent each grey and light. Seven weeks afterward: 93 per cent, no statement; 4 per cent, blue; 2 per cent, brown; and 1 per cent, grey. Boy number three: Hazel eyes. After one week: 93 per cent, no answer; 6 per cent, blue; and 1 per cent, light. Seven weeks afterward: 95 per cent, no answer; 4 per cent, blue; and 1 per cent, grey.⁷

The interpretation of attitudes and motives is not without interest. The witnessing students had attended classes taught by the participating experimenter for three hours a week since early September, 1935. When asked to interpret his attitude during the

⁶In view of this inability to give accurate estimates even for such a stationary object as a stage, it is clear that testimony for moving objects of any size at any distance is likely to be so erroneous as to be practically worthless. The fact that the Mean estimate approximates exactness does not, of course, carry any value whatsoever, since court cases do not depend upon averageable numbers of witnesses usually, but on the contrary are as likely to rely upon a witness of low descriptive accuracy as upon one of fine discrimination.

⁷The answers about color of hair and eyes were so few and so erroneous that the experimenters decided to check the results with a control experiment, using 10 children in a timed observation by the same students in the same hall, but under conditions of orderliness and quiet. To problems of illumination and distance is added the question as to individual ability to describe people accurately under favorable conditions. The 10 children were trained to walk slowly across the stage, one by one, their faces turned toward the students. Not one witness gave an absolutely correct statement as to height, weight, and color of eyes and hair. Estimated height for each child, when averaged, was less than the child's actual height. The same is true of estimates of weight for all but two children.

brief scramble they were not at all consistent. One week afterward when responding to a multiple choice inquiry 40 per cent stated he was angry; 27 per cent, calm; 13 per cent, excited; 9 per cent, amused; 1 per cent, annoyed; 1 per cent, puzzled; and 9 per cent refused to judge.

As to the attitudes of the three boys, the following judgments were tabulated from the multiple choice inquiry:

FIRST BOY (PURSUED)

| | <i>Judgment After 1 Week Per Cent</i> | <i>Judgment After 7 Weeks Per Cent</i> |
|----------------|---|--|
| Calm | 5 | 7 |
| Excited | 75 | 70 |
| Laughing | 12 | 15 |
| No answer..... | 8 | 8 |

SECOND BOY

| | | |
|----------------|----|----|
| Angry | 71 | 76 |
| Afraid | 1 | 0 |
| Calm | 3 | 2 |
| Amused | 10 | 12 |
| No answer..... | 15 | 10 |

THIRD BOY

| | | |
|----------------|----|----|
| Angry | 28 | 34 |
| Afraid | 10 | 4 |
| Calm | 8 | 14 |
| Amused | 12 | 17 |
| No answer..... | 42 | 31 |

Asked by means of multiple choice items why the pursued boy was running from the other two boys, 50 per cent of the witnesses after one week stated that the motive was not evident, 41 per cent thought he had stolen from them the bag he carried, 7 per cent made no answer, 1 per cent thought it was all in fun, and 1 per cent that the first boy had stolen a cap. Seven weeks after the incident, 34 per cent thought the motive was not evident, 46 per cent thought the first boy had stolen the bag he carried, and 20 per cent declined to answer.

In order to determine whether distance from the stage influ-

enced the reliability of observation, the reports given in the control experiment (ten children) were scored by the distance each witness sat from the stage. Group I (15 persons) sat within 20 feet of the stage; Group II (22 persons) within 35 feet; Group III (10 persons) within 45 feet; Group IV (6 persons) within 55 feet of the stage. All responses by these individuals were tabulated as right or wrong. In reporting about color of hair and eyes, other terms which might also be correctly used in place of those employed by the experimenters were tabulated as correct. Reports of height within two inches of the actual height of the subject were counted as correct. Estimates of weight within ten pounds of the correct weight were also considered right. Below are the results of the tabulations in percentage right and wrong.

| | <i>Height</i> | | <i>Weight</i> | | <i>Hair</i> | | <i>Eyes</i> | |
|-----------|---------------|--------------|---------------|--------------|--------------|--------------|--------------|--------------|
| | <i>Right</i> | <i>Wrong</i> | <i>Right</i> | <i>Wrong</i> | <i>Right</i> | <i>Wrong</i> | <i>Right</i> | <i>Wrong</i> |
| Group I | 23 | 77 | 28 | 72 | 82 | 18 | 49 | 51 |
| Group II | 14 | 86 | 27 | 73 | 80 | 20 | 53 | 47 |
| Group III | 32 | 68 | 21 | 79 | 76 | 24 | 50 | 50 |
| Group IV | 20 | 80 | 25 | 75 | 78 | 22 | 43 | 57 |

It would appear from the above tabulation that a difference of 25 feet in distance has no effect on the ability of witnesses to report facts. The percentages of correct responses do not decrease noticeably from Group I, seated within 20 feet of the stage, and Group IV which was about 55 feet from the action.

CONCLUSIONS

1. The results of this experiment indicate that there is very little agreement between the number of correct statements made in the voluntary testimony of a witness and the correct answers given under direct questioning when this questioning occurs within either one week or seven weeks after the "crime."

2. Witnesses who recall many items about the crime are as likely to make correct or false statements as are those who are able to recall only a few facts. In other words, the volubility of the witness bears no relationship to the correctness of his statements.

3. There is practically no relationship between the mental

ability of the witness and the amount of information he volunteers immediately after the crime.⁸

4. The degree of neuroticism of the individual has no influence on the amount of evidence he will give voluntarily.⁹

5. When an objective direct questioning method is used, the degree of intelligence seems to be more of a factor in the number of correct responses than when the voluntary response method is employed. While all correlations are low, the relationship between intelligence and ability to testify correctly under direct questioning is somewhat higher when the testimony is taken one week after the crime than when taken seven weeks afterward. Correlations are too low to warrant any conclusions other than that intelligence, as measured in this instance, has no relationship to the ability to give accurate information. This does not justify a belief that intelligence is never a factor in the ability of a witness to testify. The witnesses to the "crime" herein investigated were all college students, and the incident they observed was very simple. A more complex crime witnessed by a group with more variable abilities should yield a higher correlation.

6. There is a slight inverse relationship between neuroticism and the ability to respond to direct questioning within one week after the incident. This relationship becomes slightly positive when the testimony is taken seven weeks later. Correlations are too low to justify any conclusions other than that neuroticism, as here measured, is not a potential factor in testimony.

7. The correlation between the two objective tests indicates low reliability. The conclusion is warranted that even with an objective method of measurement, the reliability of reports taken within one week and within seven weeks after the crime is too low to be considered significant. It would be interesting to determine whether a number of questions could be found to which the witnesses would respond with any degree of reliability. The reliability of the individual questions on this test had not, of course, been determined. Some of the low scores may be due to the unreliability of the question. Nevertheless it can be concluded that questions made and asked in so objective a manner are probably more reliably answered than are those asked by lawyers when the

⁸ This and later references to intelligence refer to the results of the Thurstone Psychological Examination.

⁹ This and later references to personality are based upon results from the Thurstone Personality Schedule.

witness is under emotional strain and influenced by suggestions which the lawyer purposely includes in his questions.

8. The difference in the Means of the two objective tests is not statistically significant. The Mean of the scores of the first one is somewhat higher than the Mean of the scores of the second. Since the highest possible score on the test was 53, and the Mean score for each objective test was 8.62 and 8.16 respectively, there was very little information correctly recalled either time. Scores ranged from one correctly recalled item to twenty-two correct recalls, and this range of correctly recalled items holds for both series, but while only ten individuals were limited to one correct response one week after the "crime," twenty-one witnesses registered only one correct statement seven weeks afterward.

9. Testimony given seven weeks after the incident is much more variable than that given within one week when this variability is measured objectively.

10. While the number of correct answers to all types of questions is decidedly low, the ability to respond concerning action is superior to the ability to describe the people concerned. The ability to describe the scene of the "crime" is poorest although the scene was a familiar one to all the witnesses.

11. Witnesses tend to underestimate height; and responses are more accurate after one week than after seven weeks. There was a tendency to underestimate the weight of the heaviest boy, but the estimates for the two smaller boys were very nearly correct. The changes in estimates of weight made seven weeks after the "crime" were not significant. Ages of the boys were rather accurately given for both objective inquiries.

12. If confidence in ability to estimate height, weight, and age may be determined by the percentage of witnesses making statements, there is more confidence on the part of witnesses in their ability to estimate age than height. They have least confidence concerning weight. They have more assurance within one week than after seven weeks have elapsed.

13. Individual witnesses vary as much as 2 feet and 4 inches in estimates of the height of the same individual; as much as 80 pounds in their guesses on the weight of one person; and as many as 9 years in their estimates of the age of one individual. The difference in the reporting of the various witnesses is great enough to cause their statements to confuse the judgment of any highly intelligent jury.

14. Ability to measure the lapse of time during an occurrence is very poor. The Mean of the estimated length of time was greater by four minutes than was the time actually consumed. Individual estimates varied from one to fifteen minutes. Although the actual elapsed time during the crime was one minute and thirty seconds, only ten witnesses gave two minutes or less for an answer on the first objective test with seven so responding on the second test.

15. Estimates of the width of the stage, even with the witnesses looking directly at it as they responded, were very variable. The average of the estimates was about correct but individual reports varied as much as 95 feet. The Mean of the estimates regarding the number of chairs on the stage at the time of the crime was less than the actual number of chairs. Again the witnesses were more willing to make statements within one week than seven weeks afterward.

16. After one week 86 per cent of the witnesses remembered or calculated the date with accuracy. After seven weeks 80 per cent refused to answer and of those who responded none gave the correct date.

17. Those making statements about hair and eye color vary to such an extent that there is no consistency in their judgments. Most of the witnesses refused to testify about the color of the offenders' eyes and the color of hair was given by only 50 per cent. Only one time in the six opportunities offered for giving correct response did the majority of those responding indicate the right color, this being the first statement made regarding the color of hair of the second boy.

18. Interpretations of attitudes show little agreement. The amount of disagreement is greater in the instance of the one person, the instructor, whom the entire class had known for five months than in the instance of the three boys whom most members of the class had never met.

19. The control experiment indicates, in common with many other laboratory tests, that witnesses whether excited or not cannot describe people very accurately even when they are looking at them and when all concerned are unhurried. Height and weight tend to be underestimated. Color of hair and eyes is more accurately described than is either height or weight when estimates within two inches of the actual height or within ten pounds of the actual weight are considered correct. Many terms are used in describing color and the variability of witnesses' answers is due

in part to this fact. No single witness gave an accurate description of any one of the ten control children even when a variation of terms was considered permissible and when the two-inch, ten-pound allowance was made.

20. Distance of witnesses from the scene, varying as it did from eighteen to fifty-five feet, did not seem to be a factor in accuracy of report.

Individual testimony is unreliable at any time and the unreliability increases markedly with the lapse of time. If it were possible to assemble, after any crime, as many as eighty eye-witnesses and quiz them objectively, the court might expect an average of eight correct responses on fifty-three objective questions, according to the results of this study. The entire group, then, is unreliable with regard to accuracy of detail. No single witness should be expected to make more than twenty-two out of a possible fifty-three correct responses per witness. Many individuals are limited in ability to three or four right answers out of fifty-three.

The testimony taken one week after the crime is more reliable and less variable than that taken after seven weeks have elapsed, and witnesses are more willing to testify at the end of the shorter period. Seven weeks after the incident, twenty-one individuals were able to make only one correct response out of a possible fifty-three.

Action is more reliably described than is the criminal or the place of action. Poor descriptions of the criminal may be due in part to variability of terms used rather than to failure of the witness to observe.

Time sense and interpretation of attitudes in an exciting situation are very poor. Estimates of duration of an incident are highly variable. To recall a specific date—even when seemingly easy to calculate under conditions quite in contrast to the hammering of the witness stand—is not reasonably to be expected especially after the lapse of weeks. None of the college girls seven weeks afterward gave correctly, whether they could or not, the date of the occurrence.

Finally, if justice in our courts is to depend upon the testimony of witnesses, that testimony must be brought in early and even then accepted with wide allowance for error explainable in terms of faulty sense perception and memory. In our present system, days, weeks, and too often months intervene between an accident or crime and the witnesses' appearance in court. The grim humor

sometimes directed at the medical profession whose mistakes are "buried" can perhaps be more properly applied to the machinery of law where the senses of honest witnesses are likely too often to perjure the facts unwittingly and thus send innocent men to prison or to death. Add to this unreliability an excess of emotion, often deliberately provoked by attorneys, and the court might as well do its own guessing as to the accident or crime. Fallible human witnesses will, of course, function in our courts for a long time to come, but they will recede in importance as science advances and is accepted in the interests of justice, all of which will mean less zeal directed merely to the winning of cases.