

A JURY EXPERIMENT REANALYZED

Shari Seidman Diamond*

Researchers in the behavioral sciences have watched with some pride as the courts have given increased attention to social science studies. Judicial interest in empirical studies is a desirable development but one not quite free of danger. The courts are not yet fully accustomed to dealing critically with such evidence.

The United States Supreme Court ruled recently, in *Colgrove v. Battin*,¹ that six-member juries in civil cases meet the seventh amendment requirement of trial by jury. This decision was not surprising in light of *Williams v. Florida*,² in which the Court ruled that six jurors were sufficient to satisfy the sixth amendment jury trial guarantee in a criminal trial in state court. In both decisions, the Court claimed to be convinced that there would be no difference in trial outcomes if the size of the jury were cut in half. In the *Colgrove* decision, four empirical studies were cited as "convincing empirical evidence"³ in support of this position; two of these studies had been published in the *University of Michigan Journal of Law Reform*.⁴ It has been shown elsewhere,⁵ in summary fashion, why none of the four studies supplied valid information concerning the issue of whether jury size affects trial outcome. This article considers one of these studies in detail, because it provides a good example of the inappropriate use of social science data.⁶ Conclusions are presented which may mislead the unwary reader and important information has been omitted, however unintentionally, from the study report. Because these data were missing, all of the tables in this review had to be developed from materials not presented to readers of the original report. This critique is designed primarily to help future researchers avoid repetition of these errors.

* Assistant Professor of Criminal Justice and Psychology, University of Illinois, Chicago Circle. B.A., 1968, University of Michigan; M.A. (psychology), 1971, Ph.D. (social psychology), 1972, Northwestern University.

The author wishes to thank Hans Zeisel for his careful reading of this article in manuscript form and his helpful comments.

¹ 413 U.S. 149 (1973).

² 399 U.S. 78 (1970).

³ 413 U.S. at 159-60 n.15.

⁴ Note, *An Empirical Study of Six- and Twelve-Member Jury Decision-Making Processes*, 6 U. MICH. J.L. REFORM 712 (1973) [hereinafter cited as *Laboratory Study*]; Note, *Six-Member and Twelve-Member Juries: An Empirical Study of Trial Results*, 6 U. MICH. J.L. REFORM 671 (1973).

⁵ Zeisel & Diamond, "Convincing Empirical Evidence," U. CHI. L.R., forthcoming.

⁶ *Laboratory Study*, *supra* note 4.

I. THE REAPPRAISAL

A. *The Study Design*

In her laboratory experiment on jury decision-making, Kessler randomly assigned college students to six- or twelve-member juries, creating eight experimental juries of each size. All juries viewed the same videotaped trial of a personal injury case. After viewing the trial, each juror privately recorded his own predeliberation verdict. The respective juries then deliberated until a verdict was reached or until the jurors reported that they could not resolve their differences. The author then made a number of comparisons between the six-member and twelve-member juries in order to test the impact of jury size on verdicts, deliberation time, issues discussed, juror participation, and juror satisfaction.

At this point, one might ask what could be wrong with such a well-designed experiment, where random assignment keeps comparable all influences except the experimental difference in jury size. As will be seen, experimental designs pose problems that go beyond these basic requirements.

Each of the variables considered in the study will be treated in turn, beginning with the most important one, the verdict.

B. *Verdicts*

The study correctly found no statistically significant difference between the verdicts of the two different-sized juries in this experiment.⁷ The question is whether this "no-difference" finding allows one to draw any general conclusion regarding jury verdicts in general by six-member as opposed to twelve-member juries. On further inspection, no such conclusion seems warranted. The no-difference finding was clearly caused by the peculiar attributes of the trial used in this experiment; hence, nothing could be learned from the experiment about the effect of jury size in other cases.

The videotaped trial was so slanted in favor of the defendant that nearly 80 percent of all jurors indicated in their predeliberation "verdicts" that they favored a verdict for the defendant. Fourteen of the sixteen juries had a majority favoring the defendant before deliberations began (Table 1).⁸ Ten of these fourteen juries began deliberation with the five-sixths consensus sufficient for a verdict,⁹ thus rendering any deliberation superfluous. Since three juries hung (two six-member and one twelve-

⁷ *Id.* at 724.

⁸ Information for Table 1 was omitted from the report of the original study and was supplied by Joan Kessler, the author of the study report in response to this reviewer's request.

⁹ The Michigan statute requires a five-sixths majority for a verdict, and the laboratory study conformed to this decision rule. MICH. COMP. LAWS ANN. § 600.1352 (Supp. 1972).

TABLE 1
 JURY SIZE, PREDELIBERATION VOTE, AND
 DELIBERATION OUTCOME

<i>Six-Member Juries</i>			<i>Twelve-Member Juries</i>		
Jurors initially favoring:		Verdict for:	Jurors initially favoring:		Verdict for:
<i>Plaintiff</i>	<i>Defendant</i>		<i>Plaintiff</i>	<i>Defendant</i>	
0	6*	def.	1	11*	def.
0	6*	def.	1	11*	def.
1	5*	def.	1	11*	def.
1	5*	def.	1	11*	def.
2	4	def.	2	10*	def.
2	4	hung	2	10*	def.
2	4	hung	5	7	def.
3	3	def.	7	5	hung
total	11 37		20 76		

*Majority necessary for verdict present before deliberation begins

member), only three juries resolved the differences among their members through deliberation, and they too ended up with a verdict for the defendant. It is, therefore, quite impossible to draw any conclusion about six-member versus twelve-member juries from this one extreme case in which juries of any size would have reached identical verdicts.

C. Deliberation Time

It has been suggested that six-member juries might not deliberate as long as twelve-member juries.¹⁰ Although it is not clear whether a shorter deliberation time represents a saving which the judicial system should welcome, jury deliberation time is an important criterion to monitor. Again, the study shows no statistically significant differences between the deliberation times of the two different-sized juries, but this finding, too, is suspect. The deliberation time is, of course, also affected by the unusually slanted fact situation of the case, which made a verdict for the defendant a foregone conclusion. In addition, there is a methodological problem in using the behavior of student jurors as a proxy for that of real jurors. The students participated as part of a course requirement. Whereas the end of deliberation for a real jury generally means return to the jury pool to await reassignment, the end of deliberation in the laboratory study meant completion of a course obligation and freedom to leave.¹¹

¹⁰ Cf. Augelli, *Six-Member Juries in Civil Actions in the Federal Judicial System*, 3 SETON HALL L. REV. 281 (1972).

¹¹ Information supplied to this reviewer by Kessler.

Not one of the deliberations lasted even an hour; fifty-three minutes was the longest deliberation time and that was a hung jury. The other two hung juries deliberated forty-five and forty-one minutes respectively, and all other deliberations were of shorter duration. More importantly, the reported average deliberation times for the two different-sized juries are misleading. The primary determinant of deliberation time is the predeliberation vote (Table 2).¹² If that vote shows that a "verdict" has already been reached, there is no need for deliberation. Only six genuine disagreement-resolving deliberations occurred in the laboratory study, and they included three hung juries whose deliberation times could be arbitrarily cut off or extended.

D. Issues Discussed

The study also reports no statistically significant difference in the number of issues discussed by the two different-sized panels.¹³ The predeliberation vote was the best predictor of the ultimate verdict and the duration of the deliberation. The initial vote distribution also quite accurately predicts the number of issues discussed (Table 3).¹⁴

When such a critical variable as the initial distribution differs across jury size, the only way to make a meaningful comparison between the two different-sized juries is to compare juries with the same proportion of predeliberation plaintiff verdicts. The only juries that meet this criterion are the four juries which began deliberation with a five-sixths majority for the defendant; two of the juries were six-member juries and two were twelve-member juries. As Table 3 indicates, the number of issues

TABLE 2
INITIAL DISTRIBUTION, JURY SIZE, AND
DELIBERATION DURATION

Predeliberation Vote	Six-Member Juries	Twelve-Member Juries
5/6s already agree	7 * (n=4)	8 (n=6)
Initial disagreement	38 (n=4)	36 (n=2)

*Time in minutes

¹² This table was derived in part from Table 2. *Laboratory Study, supra* note 4, at 725. Additional information was supplied by Kessler.

¹³ *Id.* at 727-28 (Table 3).

¹⁴ *Id.* at 728. This table was derived in part from Table 3. *Laboratory Study, supra* note 4, at 728. Additional information was supplied by Kessler.

TABLE 3
INITIAL DISTRIBUTION, JURY SIZE, AND
NUMBER OF ISSUES DISCUSSED

Issues Predeliberation Vote	Six-Member Juries			Twelve-Member Juries		
	irrelevant	relevant	total	irrelevant	relevant	total
Unanimous	3.7	16.7	20.3 (n=2)	no data	no data	no data
11/12	no data	no data	no data	3.7	19.8	23.4 (n=4)
5/6 or 10/12	3.3	17.0	20.3 (n=2)	5.3	22.0	27.3 (n=2)
Other	8.2	39.8	47.9 (n=4)	10.7	34.7	45.3 (n=2)

discussed in the twelve-member juries was slightly higher. This result is merely a suggestion of difference, since the sample is far too small for any real conclusion.

E. Juror Participation

The study correctly reports that jurors on six-member juries were less likely to remain silent than those on twelve-member juries.¹⁵ This conclusion regarding simple presence or absence of participation is amply supported by the data from the study. Fewer jurors on six-member juries were silent throughout the deliberation. The question is, what does this mean? Given the jury's function as a fact-finder, it should represent a multiplicity of viewpoints and consider as many relevant issues as possible.

A juror may find that his opinion is being well represented by another juror and that the points he wishes to raise have been adequately covered by others. When that occurs, the juror may remain silent, and the jury deliberation will not suffer in its fact-finding function. The chance of finding such a fellow juror is, of course, greater in a twelve-member jury than in a six-member jury. This inference emerges from an analysis of the data from the laboratory study. As Table 4 indicates,¹⁶ the average percent of participation by minority members on a jury decreased as the number of jurors holding that minority position increased. Since a twelve-member jury is more likely to have at least two jurors whose views coincide than

TABLE 4
NUMBER OF MINORITY JURORS, JURY SIZE,
AND AVERAGE MINORITY JUROR PARTICIPATION

	<i>Six-Member Juries</i>	<i>Twelve-Member Juries</i>
1 minority juror	26%* (n=2)	22% (n=4)
2 minority jurors	23% (n=3)	18% (n=2)
3 or more minority jurors	13% (n=1)**	8% (n=2)

*Average percent of participation by jurors holding the minority position on the predeliberation questionnaire

**As in the original study, the three jurors on the single evenly split jury who eventually changed their votes are considered minority jurors.

¹⁵ *Id.* at 729.

¹⁶ *Id.* at 731. This table was derived in part from Table 5. *Laboratory Study*, *supra* note 4, at 731. Additional information was supplied by Kessler.

is a six-member jury, the larger jury is more likely to have jurors who find no need to add their opinions.

A second issue regarding juror participation concerns the role of the minority juror. One of the Court's arguments, based on another article which it cited,¹⁷ in favor of the six-member jury was that "the decrease in the size of the jury from 12 to six is conducive to a more open discussion among the jurors, thereby improving the quality of the deliberative process."¹⁸ The laboratory study reported "a tendency for six-member minority jurors to participate more than twelve-member minority jurors."¹⁹ This conclusion is based on an index purporting to measure participation by minority jurors.

Participation by each minority juror was measured by dividing the number of comments made by that juror by the total number of comments made by all members of that jury. Minority jurors on twelve-member juries made 13 percent of all deliberation comments, while minority jurors on six-member juries made 21 percent of the comments.²⁰ But consider the participation index that would be generated by an equal amount of participation by each member on each size jury: $(1/12) \times 100 = 8.33$ percent for the twelve-member jury and $(1/6) \times 100 = 16.67$ percent for the six-member jury. Thus if a minority member on a six-member jury talked twice as often as a minority member of a twelve-member jury, the raw percentage index comparison would show them participating exactly the same amount; if a minority juror contributed the same proportion of comments on each size jury, the index would show him to be twice as active on the six-member jury.

The appropriate way to correct the participation comparison is simply to double the twelve-member average (or to halve the six-member average) in order to account for the difference in jury size. Using this technique, the twelve-member jury minority member average participation is 24 percent greater than the six-member average $(13 \times 2)/21 = 1.24$. Furthermore, this relationship holds for each number of minority jurors (Table 5).²¹

Whether there are one, two, or more minority members, the average participation of minority members exceeds the expected participation by a greater amount in twelve-member juries than in six-member juries. Unfortunately, the report of the laboratory study did not disclose the jury number of each minority juror. Thus, it was possible to perform this computation comparing expected and actual levels of participation only after additional information was supplied. The missing data were not of minor import. The initial vote distribution has a great impact on deliberation

¹⁷ Note, *Reducing the Size of Juries*, 5 U. MICH. J.L. REFORM 87 (1971).

¹⁸ 413 U.S. at 159 n.15.

¹⁹ *Laboratory Study*, *supra* note 4, at 734.

²⁰ *Id.* at 730.

²¹ *Id.* at 731. This table was derived in part from Table 5. *Laboratory Study*, *supra* note 4, at 731. Additional information was supplied by Kessler.

TABLE 5
NUMBER OF MINORITY JURORS, JURY SIZE, AND EXPECTED
AND OBTAINED MINORITY PARTICIPATION

	<i>Six-Member Juries</i>	<i>Twelve-Member Juries</i>
	Average minority juror participation divided by expected juror participation (16.7)	Average minority juror participation divided by expected juror participation (8.3)
1 minority juror	$\frac{26}{16.7} = 1.6$ (n=2)	$\frac{22}{8.3} = 2.7$ (n=4)
2 minority jurors	$\frac{23}{16.7} = 1.4$ (n=3)	$\frac{18}{8.3} = 2.2$ (n=2)
3 or more minority jurors	$\frac{13}{16.7} = .8$ (n=1)	$\frac{8}{8.3} = 1.0$ (n=2)

time, verdict outcome, and participation—precisely those variables which the study highlights.

In addition to the inaccurate analysis of minority juror participation, perhaps the most important issue involving participation was not analyzed at all; that is, what percentage of all comments are contributed by jurors holding the minority position? Analysis of this jury performance measure should disclose how well the minority position as a whole was represented in the jury's discussion. Table 6²² shows the total percentage of comments contributed by minority jurors on the six- and twelve-member juries, when their votes were needed for a verdict and when they were not. Again there is no evidence that the minority position is better represented on the six-member jury than on the twelve-member jury.

F. Juror Satisfaction

A frequently observed finding in small group research is that a group member's satisfaction tends to be positively related to the extent of his participation in the group.²³ This result was not obtained in the labora-

²² *Id.* at 731. This table was derived in part from Table 5. *Laboratory Study*, *supra* note 4, at 731. Additional information was supplied by Kessler.

²³ Strodtbeck, James & Hawkins, *Social Status in Jury Deliberations*, 22 AM. SOC. REV. 716 (1957).

TABLE 6
INITIAL DISTRIBUTION, JURY SIZE, AND
TOTAL MINORITY PARTICIPATION

	<i>Six-Member Juries</i>	<i>Twelve-Member Juries</i>
When minority votes not needed (initial 5/6 agreement)	26%* (n=2)	27% (n=6)
When minority votes needed (initial disagreement)	44% (n=4)	41% (n=2)

*Percent of total comments contributed by minority jurors

tory study, but the reason may lie in an inadequate analysis. The author examined juror satisfaction for all of the jurors in both size juries and recorded the average percent of participation at each of four levels of satisfaction.²⁴ As shown earlier, however, there was a difference in the meaning of percent of participation for the two different jury sizes. Since no correction was made to equate the two measures, the satisfaction and participation analysis should have been done twice: once for the twelve-member juries and once for the six-member juries.

Another prediction made in earlier jury research was that juries which have greater difficulty reaching agreement should show less juror satisfaction with the deliberation than juries in which disagreement is easily resolved.²⁵ When the data from the study are presented according to degree of initial disagreement, this prediction is borne out (Table 7).²⁶

The impact of consensus on satisfaction is quite visible, even in this small sample. No juror on any of the eight juries that began deliberations with a lone minority juror or full consensus rated his satisfaction below a two. All of the other eight juries, which began with less agreement, had at least one dissatisfied juror, and seven of the eight had at least two members who were dissatisfied (Table 8).²⁷ This finding held true for both jury sizes: half of the juries in each category were six-member juries.

From Table 9,²⁸ it appears that while majority jurors do not differ in satisfaction across jury size, minority jurors may be more satisfied with twelve-member juries. This increased satisfaction may imply that minority

²⁴ *Laboratory Study*, *supra* note 4, at 733.

²⁵ *Cf.* Allen, *Situational Factors in Conformity*, in 2 *ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY*, 133-76 (L. Berkowitz ed. 1965).

²⁶ *Laboratory Study*, *supra* note 4. This table is derived from information supplied by Kessler.

²⁷ *Id.* This table is derived from information supplied by Kessler.

²⁸ *Id.* This table is derived from information supplied by Kessler.

TABLE 7
INITIAL DISTRIBUTION, JURY SIZE, AND
AVERAGE JUROR SATISFACTION

Initial Distribution	Six-Member Juries	Twelve-Member Juries
Unanimous	1.2* (n=2)	no data
11 to 1	no data	1.4 (n=4)
5 to 1 or 10 to 2	1.7 (n=2)	1.6 (n=2)
Greater disagreement	2.5 (n=4)	2.2 (n=2)

*average satisfaction (jurors graded themselves)
 1 = very satisfied
 2 = satisfied
 3 = unsatisfied
 4 = very unsatisfied

TABLE 8
INITIAL DISTRIBUTION AND NUMBER OF
DISSATISFIED JURORS

Minority Jurors at Beginning of Deliberation	At least 1 juror dissatisfied	No juror dissatisfied
Unanimous or 1 minority juror	0	8
More than 1 minority juror	8	0

views are given a better opportunity for expression and evaluation in twelve-member juries. Thus, a more accurate statement of the satisfaction findings for the study would be that there is a tendency toward greater satisfaction among twelve-member juries, particularly among minority jurors.²⁹

²⁹ It should be noted that no statistical tests were run on any of the reanalyzed data presented here. In view of the small sample of juries with meaningful deliberations (six of them) and the nonindependence of juror ratings from the same jury, such tests would imply conclusions which are unreliable.

TABLE 9
 JUROR POSITION, JURY SIZE, AND
 JUROR SATISFACTION

	<i>Six-Member Juries</i>	<i>Twelve-Member Juries</i>
Majority jurors	1.7 (n=37)	1.5 (n=78)
Minority jurors	2.8 (n=11)	2.3 (n=18)

II. CONCLUSIONS

The picture which emerges from reanalysis of this laboratory study is quite different from that found in the original report. Consider some of the conclusions presented there:

On the basis of statistical comparisons, the study found: that there were no significant differences between the verdicts, times of deliberation, and numbers of issues discussed in the two different-sized panels. . . .³⁰

These no-difference findings are quite correct as far as this particular experiment was concerned but the distinctive features of the study make it illegitimate as a basis for conclusions about juries in general.

First, these three no-difference findings may be due to the biased trial, which perhaps masked true differences that varied jury sizes might produce in a more balanced situation. It is as if one were to test the cutting quality of two kitchen knives by applying them to a frozen piece of beef. If, as one might expect, both knives perform equally poorly, there would be a finding of no difference. Yet if the test were tried on a normal piece of meat, one knife might be far superior to the other.

The second shortcoming in the design of this study is the sample size. With only eight juries of each size, a difference between the two different-sized juries would have to be overwhelming to be detected. To illustrate, if all of the six-member juries found for the defendant, at least five of the eight twelve-member juries would have had to return verdicts for the plaintiff in order to show a statistically significant difference between the two. If three of the six-member juries favored the plaintiff and five favored the defendant, all eight twelve-member juries would have had to find for the plaintiff in order to show a jury size effect.³¹ Clearly, less dramatic differences are also of concern, and they cannot be detected with this sample size. When no-difference findings are reported, as in this

³⁰ *Laboratory Study*, *supra* note 4, at 734.

³¹ Mainland & Murray, *Tables for Use in Four-Fold Contingency Tables*, 116 SCIENCE 591-94 (1952).

study, such results can be convincing only if differences have been given a strong chance to emerge. If samples are small, the likelihood of concluding that there is no difference when one actually exists is great.

Jury research is costly. Kessler used 144 subjects to obtain only sixteen jury deliberations and verdicts. But in order to draw the conclusion that six- and twelve-member juries do not differ from each other on important dimensions, a larger sample of juries is imperative.

Finally, while the original sample is itself too small, the sample-size problem is exacerbated by the distribution of predeliberation verdicts and the five-sixths decision rule followed in the study. The sample shrinks from sixteen to six when the juries that had reached a verdict before deliberation are removed (Table 1). Since jury size cannot influence predeliberation verdicts, only six juries remained for examination of jury size effects.

In view of the great importance of the jurors' initial position in determining verdict, the appropriate design for the study would have been to assign jurors to juries on the basis of their predeliberation vote, thereby creating two groups of juries with similar sets of vote distributions. Thus, there would have been the same number of juries with four to eight predeliberation defendant votes in the twelve-member jury conditions as there were juries with two to four votes in the six-member jury condition.³² This "stratified" sample would have ensured that differences between jury sizes could not be attributed to differences in the initial vote distributions.

Although comparable percentages of jurors in each condition favored the plaintiff on the initial vote, 23 percent versus 21 percent (Table 1), six (75 percent) of the twelve-member juries had reached a verdict for the defendant at that point, while only four (50 percent) of the six-member juries had. Thus, although the results are not statistically significant, the study reports shorter deliberation times³³ and greater satisfaction³⁴ in the twelve-member juries and suggests these results occurred "because less controversy occurred than in the six-member condition, where more people participated and the diverse ideas of the minority were more clearly drawn."³⁵ In fact, these tendencies could easily have been produced by the greater number of six-member juries not yet at verdict at the time of the predeliberation vote.

While the study design precluded a good test of jury size effects, oversights in analysis also occurred. Consider one of the remaining conclusions:

While not significant, there is a tendency for six-member minority jurors to participate more than twelve-member minority

³² There would be equal numbers of three to three and six to six initial distributions as well. This would also make it possible to detect differences due to jury size which may operate only with particular initial vote distribution proportions.

³³ *Laboratory Study*, *supra* note 4, at 725.

³⁴ *Id.* at 733.

³⁵ *Id.* at 733.

jurors. From a small group communication view-point, the six-member jury may be superior to the larger group, as the smaller size may encourage greater overall juror participation.³⁶

An inappropriate index was responsible for this "finding" and the direction of the statistical result actually reverses, favoring the twelve-member jury, when properly computed. This error also rendered meaningless the test of the effect of participation on satisfaction.

Furthermore, the omission of significant analyses and critical data from the report meant that the interested reader could not easily evaluate the author's conclusions.

Finally, although a caveat is placed at the beginning of the conclusion ("Because of the small number of student juries analyzed in this study, the conclusions can hardly be applied generally.³⁷"), the author summarizes her conclusions and ends the article with: "Although not conclusive, these findings are certainly relevant to an examination of the present trend toward the use of smaller juries."³⁸ This review suggests that the reported inferences not only are not relevant, but are misleading as well. The most accurate, if unsatisfying, conclusion to be drawn from this reanalysis is that methodological problems preclude any valid inferences from this study.

This example should serve as a warning to social scientists offering their wares to the courts. If the proffered research is not of the highest level, the presently improving relationship between behavioral research and the legal system may be endangered. Judges may argue convincingly that since the available research is inadequate, it may be better to ignore it altogether.

³⁶ *Id.* at 734.

³⁷ *Id.* at 734.

³⁸ *Id.* at 734.