Eyewitness Accuracy Rates in Police Showup and Lineup Presentations: A Meta-Analytic Comparison

Nancy Steblay, Jennifer Dysart, Solomon Fulero, and R.C.L. Lindsay

Meta-analysis is used to compare identification accuracy rates in showups and lineups. Eight papers were located, providing 12 tests of the hypothesis and including 3013 participants. Results indicate that showups generate lower choosing rates than lineups. In target present conditions, showups and lineups yield approximately equal hit rates, and in target absent conditions, showups produce a significantly higher level of correct rejections. False identification rates are approximately equal in showups and lineups when lineup foil choices are excluded from analysis. Dangerous false identifications are more numerous for showups when an innocent suspect resembles the perpetrator. Function of lineup foils, assessment strategies for false identifications, and the potential impact of biases in lineup practice are suggested as additional considerations in evaluation of showup versus lineup efficacy.

KEY WORDS: eyewitness; lineup; showup; meta-analysis.

An eyewitness to a crime quickly becomes a potentially critical factor in the apprehension and conviction of the perpetrator. Subsequent to a criminal event in which the perpetrator and witness are strangers, an identification procedure provides a memory test of the witness that can aid police in ascertaining whether a suspect is in fact the perpetrator. The most common police identification test procedures (Lindsay, 1999) are multiperson photo or live displays (lineups) and presentation of a single person to the witness (showup). In recent years, eyewitness researchers have identified flaws in police identification practices and have explored corrective avenues through comparative tests of alternative procedures. This line of research has contributed to the determination of best practices for obtaining and preserving eyewitness evidence (Technical Working Group for Eyewitness Evidence, 1999). The

---

1Preliminary results for this paper were presented at the American Psychology-Law Society Conference, Austin, 2002.
2Department of Psychology, Augsburg College, Minneapolis, Minnesota.
3Department of Psychology, Queen’s University, Kingston, Ontario, Canada.
4Department of Psychology, Sinclair College, Dayton, Ohio.
5To whom correspondence should be addressed at Department of Psychology, Augsburg College, 2211 Riverside Avenue, Minneapolis, Minnesota 55454; e-mail: steblay@augsburg.edu.

523
importance of this applied research—and the techniques that have been developed for reducing the likelihood of false identification—is highlighted by dramatic news reports of DNA exoneration of convicted persons and the contribution of eyewitness identification errors in such cases (Wells et al., 2000).

Research to date has focused primarily on photo and live lineups, with less attention given to showup procedures despite evidence of their frequent use. Flowe, Ebbesen, Burke, and Chivabundit (2001) report that showups were used for 55% of identifications conducted in 488 sampled cases between 1991 and 1995 in a large U.S. metropolitan area. McQuiston and Malpass (2001) document a showup use rate of 30% for identification attempts by police in El Paso County, Texas. Gonzalez, Ellsworth, and Pembroke (1993) enlisted the help of an Illinois detective to record all identifications (lineups and showups) in which he was involved over a designated period of time. Results from this field study indicated that 77% of identification tasks were showups. Thus, showups are a common and sometimes favored police identification procedure.

The small amount of research attention given to showups compared to lineups may be due to an expectation that a showup is simply an abbreviated lineup. A showup in fact may present a cognitive task quite similar to that of a lineup—and procedural recommendations for lineups (e.g., unbiased instructions) should logically be extended to showups. Alternately, a showup task may tap a slightly different cognitive strategy or set of situational influences, thus demanding a separate assessment of strengths and weaknesses. The correctness of either of these two positions is unclear at present, giving rise to the need for an empirical and evaluative comparison of showups with other identification techniques.

Researchers have empirically explored lineup formats and developed theoretical models of how lineup presentation is likely to affect witness decision-making. For example, lineups may be conducted either in simultaneous or sequential manner, and choice of format has been demonstrated to make a significant difference in level and type of ensuing errors (Steblay, Dysart, Fulero, & Lindsay, 2001; Wells et al., 1998). Simultaneous lineup presentation involves presentation of a group of photos or persons all at once to a witness, requiring the witness to decide if one of the displayed lineup members is the perpetrator. This technique allows a witness to compare lineup members and then to select the person who most closely resembles his or her memory for the culprit (i.e., a “relative judgment” strategy; see Wells, 1984). A simultaneous procedure will yield acceptable results when the perpetrator is in fact in the lineup, as the witness’s comparison of the lineup members will often lead to the choice of the perpetrator as the closest match to memory. However, when the culprit is absent from the lineup, many witnesses continue to use the relative judgment strategy, resulting in an increase in the selection of an innocent lineup member, or “false alarm” (Steblay et al., 2001; Wells et al., 1994; Wells et al., 1998).

Lindsay and Wells (1985) proposed an alternative identification procedure, designed to restrict a witness’s ability to use the relative judgment strategy. This technique, known as the sequential lineup, involves presentation of lineup members one at a time, requiring a yes/no identification decision for each member before the next one is shown. With this method, a witness must compare each lineup member to his/her memory of the culprit (i.e., the witness must make absolute judgments).
Lindsay and Wells found support for the superiority of the sequential strategy over the traditional simultaneous technique. Their results showed that the sequential and simultaneous techniques produced nearly equivalent correct identification rates in target-present lineups, yet the sequential technique produced close to a 25% lower false identification rate than the simultaneous method.

A recent meta-analysis comparing the effectiveness of the simultaneous and sequential lineup techniques confirmed what has been dubbed the “sequential superiority effect” (Steblay et al., 2001). The results support the reasoning of Lindsay and Wells (1985) that the sequential lineup forces eyewitnesses to use a more absolute judgment criterion rather than a relative judgment strategy. More specifically, participants in sequential lineup conditions were less likely to choose from the lineup, thus lowering their decision effectiveness for target present lineups but also reducing false identifications in target absent conditions. The reverse was true of the simultaneous lineup witness: An increased tendency to choose generated greater hit rates in the target present condition but also increased false identification errors by 23% in the target absent array, including a 200% increase in false identification of a designated innocent suspect. These outcomes reveal the complexity of lineup presentation issues. For example, the 15% increase in accuracy found for target-present simultaneous lineups appears desirable, but may in fact be due to calculated guesses. Also, under conditions approximating real-life, benefits of the target-present simultaneous presentation were found to diminish whereas the target-absent advantages of sequential lineups remained stable.

Given that the lineup identification procedure is a one-photograph technique requiring only one “yes-no” judgment, it should logically provide the benefit of absolute judgment (i.e., fewer false alarms). This line of thinking suggests that an eyewitness faced with a lineup will be less likely to choose than when viewing a lineup, thereby reducing both correct and false identifications. One might further predict that a lineup would be at least as effective as a sequential lineup and superior to a simultaneous format.

However, one of the benefits of a lineup—either simultaneous or sequential—is that there is some protection for the innocent suspect in the presence of lineup foils. An unreliable eyewitness or absence of the true perpetrator in the lineup can be signaled by a witness’s selection of a foil. The lineup does not offer such protection. In addition, while the eyewitness to a lineup can correctly assume that there will be more than one choice in the task, a lineup is understood by the eyewitness to be a single opportunity to identify the perpetrator. The lineup reveals police suspicions about the single suspect, and the witness is aware that only one person will be shown. Thus the procedure may be considered an “inherently suggestive one” (Lindsay & Wells, 1980; Phillips, McAuliff, Kovera, & Cutler, 1999). Indeed, using the same term, the United States Supreme Court (Stovall v. Denno, 1967; United States v. Wade, 1967) and many state courts (Bradley v. State, 1980; Commonwealth v. Carter, 1979; Holden v. State, 1979) have acknowledged that showups are suggestive. This suggestiveness may affect outcomes by generating more choosing from showups than lineups. Offering some support to this speculation, Behrman and Davey (2001) found that in actual criminal cases, 76% of witnesses in showup circumstances made identifications, whereas only 48% of witnesses in photo lineups did so. If choosing
is increased, the showup procedure may generate an increase in both correct and incorrect choices or simply make witnesses more likely to identify an innocent suspect as the perpetrator without affecting the rate of correct choices. In either case, the benefit gained by an absolute judgment strategy may be balanced or negated by pressure to choose and the fact that identification errors cannot be spread across foils (known errors). Following this logic, one might predict that a showup would be particularly dangerous for innocent suspects and thus less desirable as an identification procedure.

This project is an extension of the past work that compared sequential and simultaneous lineup formats (Steblay et al., 2001). Meta-analysis will be used to compare showup to lineup presentation strategies. Most showup researchers have tested the hypothesis that a one-person showup increases the likelihood of misidentification compared to a full lineup. A recent survey of experts (Kassin, Tubb, Hosch, & Memon, 2001) found that 74% of respondents considered that finding to be reliable, and 85% reported that their opinion was based on published, peer reviewed scientific research.

A preliminary review of past research highlights four intriguing points relevant to this survey finding. First, there is very little available research that explicitly compares showup to lineup performance. Only eight articles, with 12 tests, have been located after extensive investigation. Second, the available research on showup identifications has yielded inconsistent results. A quick tally shows four reports of the negative impact of showups (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997; Wagenaar & Veeffkind, 1992; Yarmey, Yarmey, & Yarmey, 1994, 1996, one that suggests that showups produce more accurate identifications (Beal, Schmitt, & Dekle, 1995), and two reports (Dekle, Beal, Elliott, & Huneycutt, 1996; Gonzalez, Ellsworth, & Pembroke, 1993) that indicate equivocal or no difference in decision outcomes. This variability in study outcome highlights a third issue: Interpretation of outcome is somewhat a function of the dependent measure of interest—positive identifications, choosing rates, or false identifications. Outcomes of prior lineup research suggest that exploration of multiple dependent measures will provide a more complete picture of this complex phenomenon. A final point, as noted above, is that reasonable extrapolation from existing theory and empirical work may lead one to opposing predictions about eyewitness choosing and accuracy levels in showups compared to lineups. For these reasons, a summary report of showup performance is necessary.

A central purpose of meta-analysis is to search the data for any underlying pattern, a consistent display of an effect despite surrounding noise. Subsequent exploration of theoretical and methodological variables that moderate an effect often highlights and clarifies nuances of a complex phenomenon. Essential commonality of hypothesis is critical to the studies that make up a meta-analysis, yet diversity in method addressing that hypothesis typically affords access to more complete knowledge. Despite the small number of empirical studies available on the topic of showup performance, this meta-analysis is anticipated to provide useful supplementary knowledge to our growing understanding of eyewitness performance. This expectation is based on the high quality of studies available—seven of the eight are published—and the attention within these studies to relevant theoretical questions.
and issues central to eyewitness identification practice. The studies present a desirable variety of approaches and samples. For example, Wegenaar and Veeckind (1992) provide two studies, the first a laboratory test utilizing a slide sequence stimulus and including 548 citizen subjects; their second experiment tapped a more realistic scenario involving a staged crime and a full week delay between crime and identification task during which the college subjects did not know that they would be called back to a lineup task. Gonzales, Ellsworth, and Pembroke (1993) in their first study staged a classroom incident that involved their participants in a cross-racial identification task. In a second study, these researchers explored variations of crime (theft in a restaurant) and modality (video) as well as a change to same-race identification in the context of lineup foils of high, medium, or low similarity to the perpetrator. Lindsay et al. (1997) explored subject sample differences (preschool, school-age, and college students) in a noncrime event with a lineup selection procedure that used every member of the target-absent lineup in the target-absent showup condition. This diversity of researchers’ approaches provides the potential for meaningful exploration of the parameters of showup/lineup performance even within the small sample.

Consistent with the Kassin et al., survey of experts, this meta-analysis begins with the primary hypothesis that a showup will lead to increased false identifications compared to a lineup. Additional complexities of the showup-lineup comparison also will be explored, with expectation of less dramatic differences between showups and lineups in target-present scenarios. The research will compare showups and lineups on three primary outcomes: Rates of overall correct identification decisions; correct identifications of perpetrators from target present arrays; and misidentification errors from target absent arrays. The evaluation of misidentification errors is more complicated than at first may appear. In the case of a perpetrator-absent array, a clear comparison can be made in the laboratory between the rates of correct rejection from showups versus lineups. It is more difficult to compare false identification rates. False identification and false positive selection rates are identical for showups (only one choice is available), while in a perpetrator-absent lineup, the innocent suspect may be chosen (false identification) or a foil selection is possible. Care will be taken to distinguish between these two choices (Lindsay et al., 1997).

Additionally, the predicted tendency for false identification to occur more often with showups may depend on how the innocent suspect is selected. Many, but not all, researchers select innocent suspects based on their similarity to the confederate. The result when innocent replacements are determined by other means is not clear and there may be insufficient data at this time to test this effect. However, it is hoped that the available data allows testing of three outcomes of target-absent arrays: Correct rejection rates for showups versus lineups; false identification rates of designated innocent suspects selected on the basis of similarity to the confederate for showups versus lineups; and false identification rates of designated innocent suspects that do not resemble the confederate for showups versus lineups. The specific goals of this meta-analysis are (1) to generate a quantitative and theoretical summary of research findings that compare showup and lineup performance, (2) to ascertain the state of the research literature, and (3) to provide direction for future research efforts.
METHOD

Sample

A computer search of the PsycINFO database provided an initial sample of studies relevant to the hypothesis. Direct contact with lineup researchers provided access to additional tests and more complete data. In order to be included in the sample, the experimental study must have compared showup to lineup performance and provided a statistical test of the relationship between presentation format and identification accuracy. Both sequential and simultaneous lineup formats were included as lineup tests. Multiple dependent measures of accuracy were available in the sample, and the review incorporated performance frequencies of the following: (1) overall correct decisions, collapsed across target-present and target-absent presentation (correct identifications plus correct rejections); for target-present presentations, (2) correct identifications, (3) false rejections, and (4) choice of a foil (a known error); for target-absent formats, (5) correct rejections, (6) identification of any foil, and (7) identification of a designated innocent suspect or target.

Eight papers were located (seven published and one unpublished), providing 12 tests of the hypothesis. The data set included studies completed between 1977 and 2002, representing 3013 participants. Both male and female participants were included in all tests. Sample sizes ranged from 59 to 565, with a mean of 251.08. The set includes data from 1127 community residents (41%), 1320 undergraduates (44%), and 459 (15%) children.

Study Characteristics

Methodological and theoretical variables were coded as part of the data set. Methodological variables included researcher, year of publication, source (published or unpublished), number of hypothesis tests per study, sample size, subject sex, sample makeup (children, undergraduate students, adult sample, mixed), lineup size, lineup mode (live, photo, video), design (between-subject, within-subject), type of crime (robbery, vandalism, non-criminal), event stimulus (video, live, slides), and procedural blinds (double-blind, no double-blind). Variables of more theoretical import included time of delay between event and identification task (immediate, 2 days to 1 week), number of perpetrators, race and gender of perpetrator, inclusion of a verbal description task (present, not present), instruction (biased, unbiased) lineup construction (biased, unbiased), lineup type (sequential, simultaneous), lineup construction strategy (match-to-description, match-to-target), choice of target replacement (best match to target, rotation of foils), and exposure time in seconds.

All 12 studies provided a between-subject design, a lineup of size 6, unbiased lineup instructions, and a single perpetrator. Time of exposure to the perpetrator ranged from 2 to 90 s, with a mean of 57 s. Only one test specifically reported use of a double-blind procedure (Wegenaar et al., Experiment 2, 1992); the remaining articles included no comment regarding double-blind precautions.
Two authors (JD & NS) independently recorded data from each paper, and then compared information to check for oversights. Design variables were coded by a team of student researchers. These codes were derived directly from the papers, with minimal interpretation necessary. Multiple coders were employed simply to assure that available information was recorded correctly. Thus ultimate agreement among coders was 100%.

Statistics

Following the work of Rosenthal (1991), the Pearson correlation coefficient $r$ was used as the measure of effect size. The mean effect size for a group of hypothesis tests is referred to in subsequent discussion simply as $r$. A meta-analytic $Z$ ($Z_{ma}$) was calculated by combining $Z$-scores of individual tests of the hypothesis using the Stouffer method (Rosenthal, 1991). This method produces an overall probability level associated with the observed pattern of results. A fail-safe $N$ ($N_{fs}$) was calculated to estimate the number of additional tests averaging null results that would be needed in order to bring the significance level attained through the meta-analysis to a value larger than .05.

RESULTS

Twelve tests of the hypothesis were available to examine the status of the effect, that is, that lineup presentation fosters better eyewitness performance than does a showup format. Positive $r$ and $Z$ values denote support of this hypothesis. Negative $r$ and $Z$ values indicate results in the opposite direction, that is, that subjects in the showup condition performed with greater accuracy than subjects in the lineup condition. Comparisons are considered as one-tailed tests.

Overall Frequency of Correct Decisions

The first pass through the data set was to ascertain the overall level of correct identification decisions by eyewitness subjects. These figures, from 12 tests, represent the frequency of correct identifications in target-present presentations plus correct rejections in target-absent presentations. Showup presentation produced a mean of 69% correct decisions; lineups generated 51% correct decisions, a significant difference, $Z_{ma} = -9.31$, $p < 0001$, $N_{fs} = 372$, with an effect size $r = -.18$, favoring the showup. This calculation, however, does not distinguish type of error committed. That analysis requires consideration of a critical moderator variable in lineup research: whether the perpetrator is present in or absent from the array (e.g., see Steblay, 1997 and Steblay et al., 2001). The next calculations attempt to assess the impact of showup versus lineup presentation for target-present and target-absent presentations separately (see Table 1).
Table 1. Identification Performance: Showup Versus Lineup

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Showup (%)</th>
<th>Lineup (%)</th>
<th>r</th>
<th>Seq (%)</th>
<th>Sim (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall correct decisions</td>
<td>12</td>
<td>69</td>
<td>51</td>
<td>-.18</td>
<td>56</td>
<td>48</td>
</tr>
<tr>
<td>Target present display</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct ID</td>
<td>12</td>
<td>47</td>
<td>45</td>
<td>-.02</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>Miss</td>
<td>12</td>
<td>53</td>
<td>55</td>
<td>65</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>False rejection</td>
<td>10</td>
<td>58</td>
<td>34</td>
<td>.26</td>
<td>46</td>
<td>26</td>
</tr>
<tr>
<td>Foil ID</td>
<td>10</td>
<td>—</td>
<td>24</td>
<td>19</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Target absent display</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct rejection</td>
<td>11</td>
<td>85</td>
<td>57</td>
<td>-.32</td>
<td>72</td>
<td>49</td>
</tr>
<tr>
<td>Miss</td>
<td>11</td>
<td>15</td>
<td>43</td>
<td>28</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Miss minus Foil IDs</td>
<td>5</td>
<td>—</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False identifications of innocent suspect (minus foil IDs)</td>
<td>2</td>
<td>23</td>
<td>17</td>
<td>.07</td>
<td>09</td>
<td>27</td>
</tr>
</tbody>
</table>

<sup>a</sup>Sequential and Simultaneous Lineups, from Steblay et al., 2001.
<sup>b</sup>\(Z_{ma} > 1.65, p < .05\).

**Frequency of Choosing**

A primary empirical question is whether presentation format affects choosing behavior of witnesses. Collapsed across target-present and target-absent conditions, the data indicate that the witness is twice as likely to choose from a lineup as from a showup (54% vs. 27%). In target-present conditions, 71% of subjects viewing a lineup made a choice from the array (either a correct or a foil ID) and 46% of showup subjects made a choice, in this case, a correct ID. (These data are from a subset of nine tests with the necessary information.) For 11 studies with target absent displays, lineups again produced a higher choosing rate: 43% versus 15%, lineups versus showups, respectively. Accuracy of the choosers is addressed in subsequent sections.

**Decision-Making in Target-Present Conditions**

For the eyewitness presented with a target-present showup or lineup, two outcomes are possible: Correct identification of the perpetrator (a “hit”) or a failure to identify (a “miss”). A miss can take the form of an incorrect rejection of the display, an “I don’t know” (DK) response, or in the case of a lineup, selection of a foil. The data (Table 1) demonstrate that correct identification is slightly more likely in the target-present showup presentation than in the lineup format, \(Z_{ma} = -1.38, p = .08, r = -.02\) (based on \(N = 12\)), with a 2% performance advantage (47% vs. 45%, showup versus lineup, respectively). Inversely, the overall miss or error rate of showups compared to lineups is 53% versus 55%. Effect sizes for correct identifications compared between lineups and showups are displayed on Table 2.

<sup>“Choosers” in this analysis represent those participants who select a member of the array, correctly or incorrectly. Nonchoosers are those who reject the lineup, correctly or incorrectly, or report that they “don’t know.” This definition differs from the Gonzales et al. (1993) “choosers” who were defined as those who were confident enough to “decide” as to the presence or absence of the perpetrator in the array. The current analysis extracted and used the Gonzales data consistent with our definition of choosers.**
A more precise breakdown of error type can be determined in a subset of 10 tests. False rejections when the target is present (including DKs) are significantly fewer in the lineup condition, $Z_{ma} = 7.41, p < .0001, N_{fs} = 132, r = .26$ (58% vs. 34%, showups versus lineups). Foil identifications account for the remaining misses in the lineup condition: Twenty-four percent of subjects in the target-present lineup condition chose a foil, a known error.

Another way to view this outcome is to consider only “choosers.” Showup presentations generate a significantly lower rate of choosing than do lineups, 46% versus 71% in target-present conditions. In a target-present showup, just making a choice assures a hit (100% true positive identification), while lineup choices allow for distribution across foils, thus potentially reducing true positives. Target-present lineup accuracy for choosers is 64%, a significantly lower hit rate compared to showups, $Z_{ma} = -10.18, p < .0001, N_{fs} = 336, r = -.42, N = 9$. Thus, lineups produce higher choosing (71%) with a lower hit rate (64%), and showups produce lower levels of choosing (46%) with a higher hit rate (100%). Overall in target-present presentations, the showup and lineup will produce approximately the same results (46% vs. 45% correct identifications).

Decision Making in Target-Absent Conditions

Two outcomes are possible for an eyewitness confronted with an identification task that does not include the perpetrator: correct rejection of the array (which may be in the form of “I don’t know”) or false identification. In this case, showups produced a significantly higher level of correct rejections compared to lineups (85% vs. 57%), $Z_{ma} = 11.76, p < .0001, N_{fs} = 552, N = 11, r = -.32$. Inversely, the showup produced 15% errors, compared to 43% in the lineup (see Table 1). Effect sizes for the comparison of correct rejections between lineups and showups are displayed in Table 3.
The 15% showup error rate represents a "dangerous" error: Identification of an innocent suspect as the perpetrator. In the case of a lineup, this same error may occur. In addition, however, a lineup may generate a foil selection—a known error. The 43% lineup error rate mentioned above includes both false identifications and foil identifications. Teasing apart these two types of error produces a more precise indicator of dangerous lineup error rate. Five tests allowed separation of foil and suspect choices (Dekle et al., 1996; Gonzales et al., 1993, Experiments 1 and 2; Yarmey et al., 1994, 1996). These five tests appear representative of the larger data set, in that showup vs. lineup error rates of the five average 15% and 41%, respectively (compared to the full sample rates of 15% and 43%).

There are two ways to consider lineup error rates in these five tests. The 41% error rate can be divided into foil (31%) and suspect (10%) identifications. Thus, 10% of lineup decisions result in "dangerous" false IDs, compared to 15% of showup decisions. An alternative procedure is to subtract foil choices from the analyses (reducing the overall number of subjects in the analysis). With foil choices excluded, lineups generate an 84% correct rejection rate and 16% "dangerous" false identification rate, virtually the same as showups (85% correct rejection, 15% false ID), \( r = -0.03 \).

In this subset of five studies, the error rate of choosers from the target-absent showup is 100% (a choice is automatically a false identification), but a smaller percentage of subjects are choosers, 16%. The target-absent lineup-choosing rate is 44%, with a 25.2% false identification rate (based on five tests). The overall error rate is 11% versus 16%, showups versus lineups.

Three of the research teams (Dekle et al., 1996; Yarmey et al., 1994, 1996) have further explored the perpetrator-absent scenario by planting a suspect in the lineup or showup who closely matches the description of the perpetrator. This person becomes an "innocent suspect" in the perpetrator-absent lineup and showup. As mentioned above, overall error rates are higher in lineups (23% vs. 45%, showups vs. lineups, respectively). However, dangerous false identification in these cases is
higher in showups than lineups, 23% versus 10%. With lineup foil choices excluded from analysis, this “dangerous” false identification rate, showups to lineups, is 23% versus 17%, $Z_{ma} = 1.57, p = .06, r = .07$.

Comparison of Showups to Sequential and Simultaneous Formats

A related question for this investigation of showup/lineup performance is the comparison of outcomes for showups versus simultaneous and sequential lineup presentations. Steblay et al., (2001) reported that participants in the sequential lineup condition are less likely to make a lineup choice. This lower choosing rate results in false rejection errors if the target is in the lineup and reduces false identification errors if the target is absent. The reverse pattern occurs with the simultaneous lineup: An increased tendency to choose favors this participant if the target is indeed present while increasing false identification errors, particularly of a designated innocent suspect, in a perpetrator absent lineup. The two right-most columns of Table 1 report figures for sequential and simultaneous lineups from the Steblay et al. meta-analysis.

Showup/lineup comparisons in this study echo some patterns seen in the sequential/simultaneous comparison. Like sequential lineups, showups produce fewer choices, and in doing so lead to more false rejections in the target-present condition and more correct rejections in the target-absent condition. An area of difference, however, is apparent for correct identifications. In the current study, showups and lineups are approximately equal in true positive identifications, while in the earlier work, simultaneous lineups produced significantly better eyewitness performance (15%) than sequential lineups. The majority of lineups represented in the current data set are of simultaneous format. Of the 12 tests included in this meta-analysis, three used the sequential lineup format for some or all of the lineup data, and two of the three were studies repeated from the Steblay et al., (2001) meta-analysis. Eliminating these three from the data set produces a lineup comparison group that consists of only simultaneous format. Effect sizes are similar to the overall group, .03 and -.32 in target-present and absent conditions, respectively.

An additional difference emerges for the showup/lineup comparison when the planted innocent suspect in a target-absent lineup is considered. Steblay et al. report that simultaneous lineups generated three times more false identifications than sequential lineups. In this data set showups produce more false identifications, 12% versus 5%, when compared to simultaneous lineups, however this outcome should be considered tentative, as it is based on only two tests.

Moderator Variables

As the stem and leaf displays indicate, there is essential commonality in target absent performance, rs ranging from -.17 to -.48. Table 4 illustrates the consistency of effect sizes in target absent conditions and suggests minimal impact of moderator factors.

The Target present column of Table 4 for the most part replicates the earlier analyses—indicating just small differences between showup and lineup performance regardless of moderator variable. However, target present performance indicates
Table 4. Effect Size Analysis by Moderator Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Target present: Correct identification</th>
<th>Target absent Correct rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r (N)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool to kindergarten</td>
<td>-.36 (3)*</td>
<td>-.27 (2)*</td>
</tr>
<tr>
<td>Children 8–10 years</td>
<td>.00 (1)</td>
<td>-.35 (1)*</td>
</tr>
<tr>
<td>Children 11–15 years</td>
<td>.03 (1)</td>
<td>-.31 (1)*</td>
</tr>
<tr>
<td>Undergraduates</td>
<td>.07 (8)</td>
<td>-.34 (7)*</td>
</tr>
<tr>
<td>Citizens 18–65 years</td>
<td>-.16 (2)*</td>
<td>-.22 (2)*</td>
</tr>
<tr>
<td>Lineup construction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbiased</td>
<td>-.02 (11)</td>
<td>-.31 (10)*</td>
</tr>
<tr>
<td>Biased toward foil</td>
<td>.07 (1)</td>
<td>-.41 (1)*</td>
</tr>
<tr>
<td>Lineup/showup type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photo</td>
<td>-.04 (11)</td>
<td>-.31 (10)*</td>
</tr>
<tr>
<td>Live</td>
<td>.23 (1)</td>
<td>-.48 (1)*</td>
</tr>
<tr>
<td>Delay between event and identification task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate (no delay)</td>
<td>-.03 (8)</td>
<td>-.33 (7)*</td>
</tr>
<tr>
<td>2 days to 1 week</td>
<td>.10 (3)</td>
<td>-.34 (3)*</td>
</tr>
<tr>
<td>Verbal description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>.00 (9)</td>
<td>-.31 (9)*</td>
</tr>
<tr>
<td>No</td>
<td>-.05 (3)</td>
<td>-.40 (2)*</td>
</tr>
<tr>
<td>Event stimulus mode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live</td>
<td>-.03 (6)*</td>
<td>-.30 (5)*</td>
</tr>
<tr>
<td>Slides or transparencies</td>
<td>-.04 (5)</td>
<td>-.36 (5)*</td>
</tr>
<tr>
<td>Video</td>
<td>.18 (1)</td>
<td>-.26 (1)*</td>
</tr>
<tr>
<td>Event</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robbery/theft</td>
<td>.03 (7)</td>
<td>-.36 (7)*</td>
</tr>
<tr>
<td>Smashed equipment</td>
<td>.24 (1)</td>
<td>-.17 (1)</td>
</tr>
<tr>
<td>Noncriminal event</td>
<td>-.16 (4)*</td>
<td>-.28 (3)*</td>
</tr>
<tr>
<td>Perpetrator gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>.08 (6)</td>
<td>-.34 (5)*</td>
</tr>
<tr>
<td>Female</td>
<td>-.12 (6)*</td>
<td>-.31 (6)*</td>
</tr>
<tr>
<td>Publication status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published</td>
<td>-.03 (10)*</td>
<td>-.30 (9)*</td>
</tr>
<tr>
<td>Not published</td>
<td>.03 (2)</td>
<td>-.43 (2)*</td>
</tr>
</tbody>
</table>

*Z_{ma} > 1.65, p < .05.

The one test (Gonzales et al., 1993) represented in this category is also the only test involving a cross-racial identification. Thus the individual impact of these two factors cannot be separated.

lack of consistency on two levels. First, Table 4 reveals some variability across subsets within a variable, e.g., age of sample produces effect sizes ranging from -.36 to +.07. Second, not evident on Table 4 is the actual variability in effect sizes that underlies average effect sizes hovering around zero. It is important to note that these average outcomes conceal tests with both negative and positive signs (sometimes lineup, sometimes showup superiority). Target present outcomes merit attention, to explore conditions under which effects are most pronounced or constrained. Unfortunately, analysis of moderator variables in this data set is limited by the small number of tests available and the uniformity of some design components across studies.

In target present conditions, showup performance is elevated for child participants, non-criminal stimulus events, and when the perpetrator was female. As these factors are confounded within studies, it is not possible to separate out their impact. There is also an increased showup superiority associated with adult (citizen)
populations. Somewhat superior levels of correct identification in lineups are associated with undergraduates, longer time delays between event and identification task, male perpetrators, and one study in which subjects saw a staged crime in their classroom ("smashed equipment"). Lineup performance is also better in the one study that used a live lineup and a cross-racial identification, factors that could not be separated for analysis.

DISCUSSION

When overall identification decisions are tabulated, showups produce an accuracy advantage over lineups (69% vs. 51%). This initial result is qualified by subsequent analyses. As anticipated, a consideration of specific subject choices provides a more complete picture. Correct identification (hit) rate within the context of a target-present condition is nearly identical for the two types of procedures: Approximately 46% of witnesses shown either a lineup or a showup correctly identified the perpetrator when he or she was present. False suspect identification rates in a target-absent display are also approximately equal between showups and lineups, at about 16%. Analysis of error type provides a reason for the discrepancy between the initial overall showup accuracy advantage and the hit and false identification outcomes just described. Witnesses who choose from a target-absent lineup produce more errors, but also divert their erroneous choices across foils. If foil identifications are categorized as errors, the error rate of lineups increases dramatically. Alternately, when foil identifications instead are folded into the category of nonidentification of the suspect, showup and lineup outcomes converge. Overall, the results present surprising commonality in outcome between presentation formats, and—specific to target-absent displays—an apparent contradiction of the ambient knowledge that showups are more dangerous for innocent suspects than are lineups. Additional factors will also inform a comparison of showups versus lineups. These involve our understanding of the function of lineup foils, assessment of false identifications, and the potential for biases in lineup practice.

Lineup Foils

Following the logic of the above discussion, the role of lineup foils is a first consideration. Used effectively, a lineup will serve two purposes: To determine whether a suspect is in fact the perpetrator observed by the witness and to assess the reliability of the witness. A foil selection suggests unreliable witness memory and discredits the witness rather than the suspect. The lineup witness who selects a foil may rightly be considered an unreliable source for subsequent identification evidence. Only witnesses who reject the lineup by choosing no one may be considered a credible source in a subsequent identification task. On the other hand, the showup witness has no foil options. A witness who rejects the showup retains police trust as a reliable witness, even in the case of a false rejection of a target-present showup. This is potentially dangerous in the face of a subsequent identification attempt with a new and innocent suspect. Therefore, if foil choices are considered useful indications that witnesses are
willing to identify innocent people, lineups have an advantage. (Current data show a 24% foil identification rate in target-present lineups and 31% foil identification rate in target-absent lineups.)

The use of foils to detect an unreliable witness is of particular interest when one considers very young children. A substantial amount of research literature has examined the eyewitness reporting accuracy of children compared to adults. Wells, Wright, and Bradfield (1999) summarize this literature specific to lineup performance, by pointing out that “The primary condition for concern in eyewitness identification from lineups and photo-spreads is the condition in which the actual perpetrator is not present . . .” (p. 60). Dekle et al., (1996) similarly report the literature as showing that while child witnesses make correct identifications from target-present lineups at the approximate level of adults, children are more likely than adults to choose someone from a target-absent lineup (a false ID), even when warned directly that the perpetrator may not be in the lineup. In this data set, children exhibit better target-absent performance for showups than lineups, as did adults. In target-present arrays young children (preschool and kindergarten) also performed significantly better on showups than lineups, a finding that deserves attention in future work. Part of that analysis must be a differentiation of target identification (choices of the target among multiple responses from an individual child) versus correct identification (a single and correct identification). Young witnesses have a tendency to make multiple choices from a lineup, thus impeaching their own testimony. As noted by Lindsay et al. (1997), the reliability of eyewitness identification is thus “seriously compromised by the tendency for children to guess” (p. 401). This impeachment through multiple choices cannot occur in a showup, thus perhaps affecting the appearance of better showup performance.

The counterargument to this framework for consideration of identification tasks—that foil choices that discredit the witness represent a problem for police and the solution of crimes—and the choice of whether such witnesses should be considered a source of identification evidence is a policy, not empirical, issue.

**False Identification Rates**

Experts who responded to the Kassin et al. (2001) survey expressed specific concern with false suspect identification rates of showups. This is reasonable, given that only false identifications lead to the risk of false accusation and wrongful conviction. As reported above, this meta-analysis has identified approximately equal false identification rates from showups versus lineups (16%), deriving that figure by direct tabulation of errors in target-absent lineups. Previous research teams have attempted to estimate false identification rates through two other means. First, the overall rate of false positive choices can be divided by the nominal size or number of people examined during the identification procedure to generate an expected false identification rate. This approach (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997) is based on the assumption that the innocent suspect is no more likely than any other lineup member to resemble the criminal if the lineup has been constructed based on matching foils to the description of the criminal provided by the witness. The lineups in this meta-analysis had a nominal size of 6, thus the expected false
identification rate would be 43%/6 = 7.2%. This compares to the substantially higher rate of 15% for showups. Based on the sequential lineup meta-analysis, 6-person sequential lineups have an expected false identification rate of 5.33% (Steblay et al., 2001). If this approach is accepted, then showups do represent a greater risk of false identification.

A related measure, the diagnosticity ratio, is employed by some researchers to establish utility of a lineup procedure (Wells & Lindsay, 1980). The advantage to this perspective is that police know whether or not the witness chooses someone but do not know if the identification procedure is criminal-present versus criminal-absent. For showups 42% of witnesses choose the suspect in the present condition and 15% in the absent condition, generating a diagnosticity ratio of 2.80 (Wells & Lindsay, 1980). For lineups, 42% of witnesses choose the suspect from the present lineup and 16% from the absent lineup, generating a diagnosticity ratio of 2.6.

A second approach to estimating false identification rates is to designate a specific, criminal-absent lineup member as the innocent suspect. The rate at which the designated individual is identified is considered the false identification rate. Within this tradition, two approaches have been taken. One assigns the innocent suspect role randomly (or perhaps haphazardly) to the six lineup members, while the other and more common approach assigns the role of innocent suspect to the absent lineup member deemed most similar to the criminal. In the current data set, the innocent suspect not explicitly selected based on similarity to the criminal, produced almost identical rates of false identification (15 and 16%, showups to lineups, respectively). If the innocent suspect was selected based on similarity to the criminal, showups generated more choices of that designated innocent suspect (23% vs. 17% respectively). The showup then may be equivalent in risk for an innocent suspect only to the extent that the innocent suspect does not strongly resemble the true criminal.?

Bias

The vulnerability of an innocent suspect who matches the description of the perpetrator illustrates a third factor for consideration. It is reasonable to wonder if other means of influencing a decision criterion, for example, clothing, instruction, and foil biases (Lindsay, Wallbridge, & Drennan, 1987; Lindsay & Wells, 1980; Malpass & Devine, 1981; Steblay, 1997) may increase false identifications differentially for showups and lineups. Although foil bias will not be a factor with showups,

7 Differences in practice are apparent in the studies at two points during formation of a target-absent array. At a first point, researchers identify appropriate lineup foils. Six of the hypothesis tests in this data set indicate a “match to general description” strategy for foil determination, and five tests use a “match to target” method (one test not reported). A match-to-target was typically used as a means to construct high/medium/low foil similarity for exploration of that variable. Analysis of witness error rates indicate small differences associated with construction strategy: Match-to-description produced 13% showup and, 44% lineup errors; Match-to-target produced 17% showup and 43% lineup errors.

As a second step, researchers decide on a target replacement for the target-absent showup or lineup. In four tests, the authors used a strategy that essentially rotated the lineup foils through the position of target replacement for the target-absent showup. For six tests, the target replacement was the foil most resembling the perpetrator. Again, witness error rates differed slightly based on strategy: Use of a target match produced 15% showup errors, 42% lineup errors; Rotation of foils produced 12% showup errors, 44% lineup errors.
clothing and instruction bias may be influential. All studies in this set used nonbiased instructions, thus the effect of instructional bias on showup performance remains an open question. The high rate of correct rejection for criminal-absent showups may reflect reactance to the suggestiveness of the procedure. Combining biased instructions with the showup procedure may therefore result in two distinct outcomes: biased instructions may increase reactance and thus further decrease false positive choices, or biased instructions could alleviate witness concerns that the procedure is biased and dramatically increase false identifications. Current research being conducted on this issue favors the latter explanation (Dupuis, Dysart, & Lindsay, 2001).

Clothing bias may be of particular concern with showups as the procedure is used shortly after the crime and frequently in the field rather than at police stations. Apprehension of suspects for showups is generally based on the combination of a match to the description provided by the witness and proximity to the crime. As a result, suspects will generally be wearing clothing that resembles the witness' description of clothing worn by the criminal during the crime. The fact that the showup generally occurs shortly after the crime may further convince witnesses that the suspect is unlikely to be innocent. They may ask themselves “How many people can there be in this area that look like that and are wearing clothes like that?” The less time between the crime and the showup, the stronger this intuition may be.

Dysart, Dupuis, and Lindsay (2001) have recently found strong evidence of clothing bias with showups, indicating that the type of clothing worn by the perpetrator may interact with other factors, such similarity of the innocent suspect. Although the results from this study are compelling, the data did not include a lineup comparison. We are left with an incomplete picture of showup vulnerability to bias, but reason to speculate that several known lineup biases may influence showups as well.

**Theoretical and Future Research Considerations**

Significantly lower levels of choosing behaviors for witnesses presented with a showup versus a lineup suggest that, even though decision outcomes may be similar, differential decision processes may be attendant to the two identification formats. Given this, it is appropriate to ascertain what we can about witness reliability and strategy from these available data.

As discussed by previous researchers (Lindsay & Wells, 1985), an absolute decision process is desirable, particularly as a means to reduce false identifications. The lower level of choices in showup conditions may be construed as an indication that subjects are in fact using, at least more so than in the lineup condition, an absolute judgment strategy. The increased rejection rates (false rejections in target-present conditions and correct rejections in target-absent conditions) suggest that showup subjects have attained some benefit of absolute judgment, perhaps due to a showup's similarity to a “one-person” sequential lineup. On the other hand, it is apparent that lineup foil options provide a deflection of error away from an innocent suspect and a valuable vehicle to identify the unreliable witness. These benefits help to equalize lineup and showup performance, at least under the rather favorable conditions of these studies.
Evaluation of this data set must include concern regarding the small number of studies available. In fact, one key outcome of the investigation is to alert the research community to the paucity of data and to the need for more deliberate attention to showups. The analyses exposed performance variability yet to be explored in target-present scenarios. The small number of hypothesis tests in this data set deterred analysis of some potentially fruitful variables, as exemplified by the Gonzales et al., study. These authors included in their method two relevant and intriguing components—cross-racial identification and use of a live lineup—that could not be independently examined because they are confounded within that study and not available in other tests. Therein lies direction for future research.

Finally, the showup’s potential for suggestibility—which worries legal professionals and eyewitness experts—is evidenced in this data set, although in a small number of studies. The data currently available leave us with residual concern regarding potential dangers of showups and with a strong appreciation of the need for research that will specifically address showup accuracy under realistic conditions comparing competent practice with biased procedure.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.


