Forensic Science International: Synergy 12 (2026) 100658

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: www.sciencedirect.com/journal/forensic-science-international-synergy

Strengthening operational performance in canine detection teams with double-blind certification testing

Adele Quigley-McBride [a,*], Paola A. Prada-Tiedemann [b], Fred Helfers [c]

[a] Simon Fraser University, Department of Psychology, Canada
[b] Texas Tech University, Department of Environmental Toxicology, USA
[c] Canine Detection Services, LLC, USA

ARTICLE INFO

Keywords:
Double-blind testing
Proficiency
Canine detection
Training
Accreditation

ABSTRACT

The present study evaluated the performance of professional canine–handler teams on narcotics detection certification trials conducted under single-blind and double-blind conditions. Across six years of annual testing (2012–2017), we analyzed 667 first-attempt trials and 132 second-attempt trials from 133 distinct canine-handler teams. Teams demonstrated high accuracy under single-blind conditions (94% pass rate for vehicle searches and 100% for luggage searches), but performance dropped substantially under double-blind conditions (72% pass rate for vehicle searches and 88% for luggage searches), where neither handlers nor evaluators knew the number or location of the target odors. Many teams that failed an initial double-blind trial passed on a second attempt, suggesting that at least some observed deficits in performance may be easily remedied with additional practice participating in double-blind trials. A follow-up survey of 20 handlers indicated generally positive perceptions of double-blind testing—although double-blind trials are more difficult, handlers believe that these types of trials increase their confidence, improve training strategies, and more closely reflect real-world scenarios. Incorporating routine double-blind exercises into certification and maintenance training may provide agencies with a reliable means of preparing teams for unpredictable real-world scenarios. Thus, double-blind testing represents a straightforward, cost-efficient strategy for enhancing the accuracy, credibility, and overall integrity of canine detection.

1. Introduction

Humans and canines have a long history of working together. Canine-detection disciplines have emerged out of this close relationship to provide aid in a variety of disciplines—wildlife conservation [1], medical testing [2], and, of course, law enforcement [3]. In all of these contexts, canine-handler teams work together to find odors that indicate the presence of substances, animals, or other items of interest. Thus, the dog serves as a living "screening tool" with exceptional, natural olfactory abilities that their handlers can rely on to identify substances that handlers themselves are unable to perceive. This is a form of *distributed cognition*—where people offload some of the more difficult aspects of a task onto technology or another tool to accomplish their goal faster and with higher accuracy [4,5].

Within investigative contexts, canine-handler teams contribute mainly through substance detection, such as explosives, illicit drugs, human remains, or weapons, and by providing leads or aiding in the development of associations between people and locations—connecting victims or offenders to evidence and crime scenes. Although there have been critics of detection dogs (e.g., Ref. [6,7]), there is plenty of scientific evidence to suggest that canine olfactory systems are much more sensitive than that of humans (e.g., Ref. [8–10]). Dogs can detect a wide range of odors that humans cannot, even in very small quantities. There is also a long history of people training dogs to perform all kinds of behaviors. In forensic and investigative work, this means that dogs can be taught to discriminate target odors from non-target odor sources with proper training and then perform that skill in operational scenarios. This means that they can learn to recognize a specific odor that is associated with a substance of interest and provide a trained response—called an "alert"—to signal when such an odor is present.

The problem lies in the ability of the handler and canine to communicate. Dogs are not machines, and the contexts in which they are performing their duties are not controlled or pristine, so the dog's behavior will vary. Sometimes an "alert" is obvious and other times

* Corresponding author. Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6, Canada.
E-mail address: aquigley@sfu.ca (A. Quigley-McBride).

https://doi.org/10.1016/j.fsisyn.2026.100658
Received 26 November 2025; Received in revised form 2 January 2026; Accepted 7 January 2026

not—whether an "alert" has occurred is the decision of the handler. Dogs will show signs that an odor is detected (called a "change in behavior") before performing a "trained final response". Trained final responses are clear behaviors canines are trained to perform when they detect a specific odor, such as sitting facing the location of the odor. These are behavior changes that a handler will pick up on when they train and work with the same dog for a long time, and they might decide that those behaviors are sufficient to call an alert without a trained final response. This means that, ultimately, whether an investigative lead is generated by a canine-handler team relies on the handler's subjective interpretation of the canine's behavior, and this interpretation may not always be correct [11]. So, even if the dog is able to perfectly detect a particular odor, the handler still needs to accurately determine that the dog has detected something relevant.

Because the main outcome of interest relies on a person to make a subjective judgment that does not have clear criteria specified in advance, what a handler determines to be an alert could vary over time and will vary between handlers. The latter is less of an issue, because each handler is responsible for what behaviors they train the dogs to perform. But, if a handler is generous in their interpretation of their dog's behavior some of the time and more conservative other times, this could result in inconsistencies in forensic contexts, which could set law enforcement on an incorrect investigative path or cause inefficiencies during the investigation.

Therefore, there needs to be methods that provide some assurance that canine-handler teams can accurately and consistently identify forensically-relevant odors when they are there (a "hit" or "true positive") and also determine that a place or person has no forensically-relevant odors when there are none to be found (a "true negative"). This usually involves participating in regular proficiency tests where "ground truth" (e.g., what odors are and are not present) is known. Many states in the USA also require certification before canine-handler teams can work in operational contexts. Successful certification testing means that when practitioners—in this case, canine-handler teams— complete a series of tests and met certain criteria, these tests and criteria indicate who is and is not ready to be working in operational settings.

Not all tests are made equal though. A test that is too hard will prevent good canine-handler teams from working on real cases, whereas a test that is too easy could result in many misses (odor was present and the handler said the dog did not alert) and false alarms (the handler said that the dog alerted, but there was no evidence of a relevant odor) once the team is deployed. Difficulty does not necessarily equate to complexity either. For example, odor-recognition tests (ORTs) are critical because they help determine whether a dog has the sensory acuity, motivation, and behavioral consistency required for more advanced scent-training work, but these need not be complex scenarios.

ORTs are a type of controlled assessment designed to evaluate a detection dog's fundamental ability to perceive, discriminate, and reliably indicate odors before being trained on specific operational scenarios. During an ORT, the dog is presented with target odors in a testing environment with multiple distractors and "blanks" (control items—empty versions of containers that usually contain odors). The goal is to measure whether the dog can pick out the target odor, maintain focus throughout the search, and offer a clear, repeatable alert behavior without handler influence or bias [12]. One common format for ORTs is the odor lineup, in which several discrete odor ports, containers, or vessels are arranged (often in a straight "line") allowing the dog to systematically investigate each. Only one of these contains the target odor, while the rest contain either blanks or distractors. By randomizing which position holds the target across trials, trainers can assess the dog's discrimination capabilities, its tendency for false alerts, and how consistently the dog checks each sample.

A more sophisticated and widely used design for ORTs is the odor carousel (also called a scent wheel). In this setup, odor-bearing containers are mounted on a circular rotating platform, each "arm" containing either a target scent, a blank, or a distractor. As the carousel rotates between trials, or is reloaded, the location of the target odor changes. The dogs are then directed to search around the wheel, sniffing each port in turn until they find the target and perform their trained final response. This method makes it easy to conduct many trials in a single session, standardizes the presentation of the odor, and minimizes how much the handler needs to interact with the odor itself [13].

Overall, these ORTs form a rigorous training foundation that ensures a detection dog is genuinely ready to move forward, especially when implemented using structured lineups or rotating carousels. When a canine consistently succeeds during ORTs, this validates their olfactory sensitivity, consistency, and decision-making in a controlled environment, thereby improving the quality and reliability of any subsequent target-odor training. An ORT where no one present knows where the target odors are located is very similar to other scientific tests that are conducted to determine whether a person or machine can consistently sort between items of interest and other distractors. That said, it is important to ensure the ORT is testing the skill of the canine-handler team with regard to *odor detection*, not just their ability to interpret other information available to them.

This type of controlled test—where the information that is relied on by the test taker is as important as their performance—is commonly used in most scientific disciplines [14]. For instance, when generating eyewitness identification evidence, another type of forensic evidence, police investigators will conduct a test of the eyewitness's memory. What they want to know is whether their suspect is the person that the eyewitness remembers from the crime event. However, research shows that there are lots of reasons why an eyewitness might identify someone, especially if the suspect is presented alone. They might assume that the police have other reasons to believe this person is the suspect and want to be helpful, so they may positively identify the suspect even if they do not recognize them.

To prevent this, if they are following research-based best practices, they make a lineup where the suspect is embedded among other people who are definitely innocent but resemble the suspect. Now, the eyewitness cannot just pick anyone and be helpful—they have to know who to pick [15,16]. The test still is not perfect, though. If the person guiding the eyewitness through the lineup procedure knows who the suspect is, they might inadvertently indicate to the eyewitness with their body language or words who the suspect is. Thus, one of the most important parts of a lineup procedure is that it must be conducted **double-blind—**both the eyewitness and the lineup administrator cannot know who the suspect is [14]. Under these circumstances, provided other best practices are followed, the police can assume that the reason for any identification of the suspect is because that suspect matches the eyewitness's memory [16].

The same issue can arise in the testing of canine-handler teams. Some common ways that clues might be inadvertently known to the handlers include a test that is set up in a way that makes it obvious where the odor is hidden (i.e., an empty room but for one suitcase), the handler is told that the target odors are in particular locations (e.g., lockers, bags, boxes), or the evaluator is present and inadvertently signals to the handler when they are on the right track (e.g., picks up the clipboard and pays more attention to the team when they get close to an area or object [17]). Thus, if the goal is to test the skill of the canine-handler team in certification tests, then they should be required to successfully complete several testing scenarios double-blind before being certified. Success under those testing conditions provides a higher level of assurance that the canine-handler team is able to detect forensically-relevant odors without additional, helpful information.

This is not a new idea. However, there are several arguments that are often made against the use of double-blind testing in canine-handler certification trials. First, critics claim that it is too difficult and resource intensive to set up double-blind trials. Second, handlers claim that they can ignore the extra information available to them and focus on the task because they are experienced professionals. Third, evaluators claim that they are experienced and can avoid revealing anything to

the test-takers. Finally, some argue it is unnecessary to include double-blind trials as there is extra information available in real cases all the time. These are similar arguments to those made by police when double-blind administration was first suggested by researchers for police lineups. However, double-blind testing can be accomplished with remarkably few resources—just because it is scientific does not mean that it needs to be complicated. Finally, people cannot prevent themselves from being biased or from being demonstrative with their language and body movements, as this is a normal part of human behavior and decision-making that happens outside of our awareness [18,19].

Some existing studies suggest that single-blind testing in canine detection may be insufficient, but these studies have been controversial in the canine-detection community. Lit and colleagues [17] showed that handler expectations affected the performance of the team (increased false alarms). Yet, the sample size was very small ($N = 18$) and there are features of the trials that do not represent best practices, typical certification trials for canine-handler teams, or operational scenarios. Additionally, the design lacked a baseline control condition to evaluate each team's proficiency outside of the experimental manipulations. Another study by DeChant and colleagues [20] found evidence of different behaviors and outcomes when handlers were told how many target odors they could expect to find in a scenario but found no difference between single- versus double-blind conditions. However, the accuracy rates in this study were fairly low overall—even under single-blind conditions—raising concerns about the quality of the trial setup and target odors. These issues may have been compounded by a reliance on sport detection teams rather than professional canine-handler teams who actively assist law enforcement in real cases.

So, although the general findings in these studies are consistent with findings in other forensic disciplines (e.g., showups in eyewitness identification, e.g., Ref. [21]; incriminating contextual information [22]) and even disciplines outside of the forensic world, the existing work directly relevant to canine detection has been met with skepticism. The fact remains, though, that there is a wealth of research in the broader scientific literature suggesting that our beliefs and expectations frame how we interpret new information and situations. Therefore, it seems likely that a handler's expectations about a trial might affect their dog's behavior as well their interpretation of their dog's behavior in locations where they are expecting to see an alert.

The current study compared professional canine-handler team performance during typical narcotics detection certification trials under double-blind and single-blind conditions. We also conducted an optional survey in 2024 with a group of canine-handler teams that participated in the same type of certification trial to get a better sense of the costs and benefits practitioners perceive to be associated with double-blind testing and training scenarios. A primary goal of this research was to show that double-blind conditions can be arranged with very little effort and with very few resources. In addition, we hypothesized that canine-handler teams would find the double-blind testing component more difficult than the single-blind testing component, and this would be reflected in how many of the teams passed those trials. However, we also anticipated that canine-handler teams participating in multiple double-blind tests over time would improve with practice.

2. Method

2.1. Participants

Between 2012 and 2017 (six years total), data was collected from 133 distinct canine-handler teams during annual narcotics detection certification testing, many of which took part in the testing over multiple years (M=1.77 years, SD = 1.18 years). There was an average of 39 teams tested per year (SD=3.66 teams), and three certification testing sessions per year. Table 1 reports some descriptive statistics relevant to the trials and the canine-handler teams. This sample should be representative of the canine handler community in the USA for the purpose of this study even though participants were not randomly selected from canine handler community. Participants consented to the use of their certification data for research purposes so long as their identity and that of their dog were kept confidential. Use of these secondary data for this publication was reviewed and approved by the Institutional Review Board at Simon Fraser University, Canada (Protocol #: 30002020).

Table 1
Summary of information about the canine-handler teams.

Descriptive Statistic	Value
# of teams participating over the years	133 unique teams
# of individual trials	
First attempts	667 trials
Second attempts	132 trials
Handler experience	
Mean	3.85 years
Standard Deviation	3.57 years
Canine experience	
Mean	3.34 years
Standard Deviation	2.06 years
Canine Age	
Mean	4.98 years
Standard Deviation	2.19 years

2.2. Design

These data were collected using a repeated-measures, quasi-experimental design. That is, test administration type (double-blind or single-blind trial) was manipulated within participants so that all canine-handler teams participated in as many of the single-blind trials and the double-blind trials as they could during the years and certification sessions they were present for. There were other factors that varied between years (e.g., the exact set up of each trial during each session over the years), but these were systematically varied to provide variety of testing in conditions while holding constant the variables of interest. There were some covariates collected too that will be incorporated into some of the models to see if they alter the relationship between test administration type and pass rates (e.g., handler/canine experience, canine age).

2.3. Trials - certification testing procedure

These data were collected as part of regular certification testing offered to canine-handler teams, managed and run by Fred Helfers, founding president of the Pacific Northwest Police Detection Dog Assocaition (www.fredhelfers.com). Canine-handlers would sign up for certification testing after participating in training so that they could demonstrate their readiness to participate in real case work. During certification testing, each canine-handler team participated in multiple trials, some of which were conducted single-blind, and others were conducted double-blind. This resulted in data from a total of 667 first attempts (312 single-blind trials and 355 double-blind trials) and 132 second attempts (3 single-blind trials and 129 double-blind trials). There were three main types of scenarios used for the trials, which are scenarios that are typical in certification trials because they are operationally relevant but also can be tightly controlled for the purpose of evaluating each team.

2.3.1. Vehicle searches

Some were vehicle searches where four cars were parked in a row inside a large warehouse. These searches were always conducted double-blind and there were two set ups—Vehicle Search A and Vehicle Search B. The odors were hidden in the vehicles by rolling a dice or using a random number generator. The following aspects of target odor placement were each randomized: how many target odors would be hidden, which of the four cars the target would be hidden in, as well as which quadrant of the car (placed within sniffing distance). Canine-

handler teams would wait outside the warehouse and an administrator would roll a dice or use a random number generator to determine whether they were assigned to A or B.

Then, the canine-handler team would enter the relevant warehouse space, conduct their search alone, and their answers were written down and whether that was correct was not scored until later. This way, the test administrator could not influence the canine or their handler, ensuring that it was double-blind. Sometimes, the trials were video recorded to provide the handlers with feedback and training opportunities. If the canine-handler team did not successfully locate the target odor, they were permitted to attempt the vehicle search again later in the day, but they needed to complete the version that they were not familiar with (i.e., if they completed A originally, their second attempt would be in B).

2.3.2. Luggage or container search

There were also luggage searches, which were sometimes conducted single-blind (2012 and 2013) and other times double-blind (2014–2017). This involved placing a random assortment of suitcases, bags, or containers in a room—usually between six and 10 items total. There was either one target odor concealed within one of the bags or containers in the room, or no target odors in any of the bags (a "blank" trial).

2.3.3. Residence search

The final scenario was a residence search, which was always single-blind. This involved entering a hotel room or other room/building that mirrors a space that is lived in. Between one and three target odors could be hidden in the furniture, cabinets, and appliances. Target odors were always hidden so that they were not visible to the handler or canine.

2.4. Certification testing outcomes

For this certification testing, the canine-handler teams needed to find all of the target odors in the scenario to pass. Even one false alarm or one miss would mean that the canine-handler team failed that trial. For some trials, if the canine-handler team failed their first attempt (at least one miss and/or at least one false alarm), they were allowed a second attempt. If the second attempt was on the double-blind vehicle search trial, then the canine-handler team ran their second attempt on the scenario they had not yet completed (i.e., first attempt was Vehicle Search Version A, so the second attempt would be Version B – see above for more details).

3. Certification trial results

3.1. Analytic approach

Each trial with each canine-handler team was coded so that there was a column for each certification outcome type for first attempts ("pass" = 1 or "fail" = 0). "Retries" or second attempts were assessed separately and coded as a pass (1) or a fail (0). Inferential statistics were run using a multilevel logistic regression approach with whether each canine-handler team passed or failed each trial as the outcome variable. Test administration type was entered as the main predictor in these models, but some models were run with other covariates (e.g., canine age, canine/handler experience). First and second attempts were assessed in separate models. All data entered into these models was nested within canine-handler team ($ICC_{1st\ attempt} = 0.02$, 95 % CI [0.00, 0.96]; $ICC_{2nd\ attempt} = -0.02$, 95 % CI [−0.03, 0.92]) and the year of the trial ($ICC_{1st\ attempt} < 0.01$, 95 % CI [0.00, 0.87]; $ICC_{2nd\ attempt} = 0.08$, 95 % CI [−0.01, 0.99]) so that these factors were controlled for in these analyses.

3.2. The effect of administration type

There was a clear difference in outcomes observed between the certification testing outcomes under double-blind versus single blind conditions. When aggregated across certification trial years, canine-handler teams during their first attempts passed 94% of trials under single-blind conditions whereas, under double-blind conditions, there was only a 72% passing rate among participating canine-handler teams. This was a significant difference in passing rates, with substantially lower passing rates under double-blind than single-blind conditions ($\beta = 2.64$, $p < .001$). In fact, canine-handler teams were more than seven times more likely to pass a trial if it was single-blind rather than double blind (OR = 7.54). Refer to Fig. 2 for a graph of pass rates for first and second attempts with each administration type.

Of the 28% of double-blind trials where the canine-handler team failed ($n = 100$), most chose to do a second attempt (89%) and only 18% failed this second attempt (refer to Fig. 1). Thus, many were able to successfully complete the certification on their second attempt even if they struggled during the first attempt. There were only three single-blind trials attempted a second time (in 2017), and two out of three resulted in a pass.

3.3. The effect of scenario and administration type

First, we ran models to examine the impact of test type (i.e., Vehicle, Luggage, or Residence scenario) on pass rates. The vehicle search was associated with significantly lower pass rates (64%) than either the luggage search (92%; $\beta = 2.42$, $p < .001$, OR = 7.57) or the residence search (92%; $\beta = 2.54$, $p < .001$, OR = 7.62). However, this was always a double-blind test, so the difficulty may have been due to the way the test was administered rather than due to the difficulty of the search scenario. Refer to Fig. 2 for a graph of the pass rates for each scenario type, for each type of test administration.

Because the luggage search was sometimes a single-blind trial and other times (post-2014) a double-blind trial, we were able to compare the difficulty of the same type of task when conducted single-versus double-blind. For the two years that the luggage search was conducted single-blind (2012 and 2013), the pass rate was 100%—all canine-handler teams passed those trials, suggesting that the luggage search was the easiest part of the certification test when conducted as a single-blind trial. However, when the test was changed to a double-blind trial beginning in 2014 through to 2016, the pass rate was 88% during these years, suggesting that double-blind administration will show weaknesses in canine-handler team's even for very easy tasks that every trained team usually passes. So, the vehicle search is likely a harder task in general, but the difficulty is increased because here it was also a double-blind trial.

3.4. Trends associated with characteristics of the canine-handler team

These models were also run with several covariates included in the model: handler experience (in months), canine experience (in months), and canine age (in months). None of these demonstrated a significant interaction with test administration type suggesting that the experience levels of the canine-handler team did not predict whether they would pass double-blind certification tests or not. There was also no significant effect of any of these variables on pass rates, even when test administration type was removed (handler experience: $\beta = -0.06$, $p = .868$, OR = 1.00; canine experience: $\beta = -0.78$, $p = .317$, OR = 0.99; canine age: $\beta = 0.64$, $p = .408$, OR = 1.01).

The only effect of interest related to the canine-handler teams was associated with the number of these certifications they participated in (ranging from 1 to 6). There was an interaction suggesting with each additional certification the canine-handler team participated in, the more likely they were to pass both the double-blind and single-blind testing scenarios, but this effect was stronger for double-blind sce-

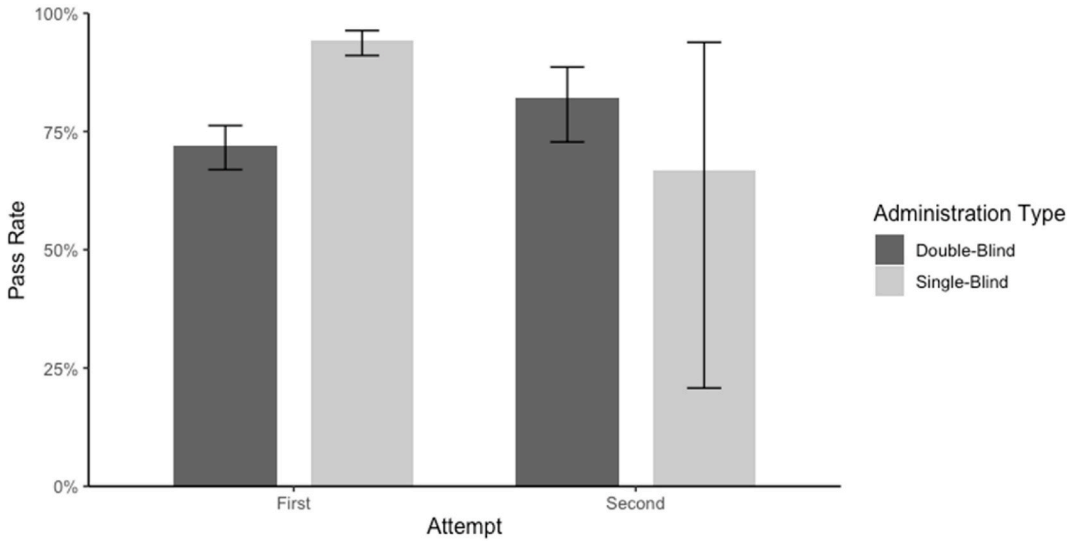

Fig. 1. Canine-handler teams pass rates in double-blind and single-blind trials.
Notes. Error bars represent 95% Confidence Intervals. There were only three second attempts for single-blind trials, so the error bars are very large.

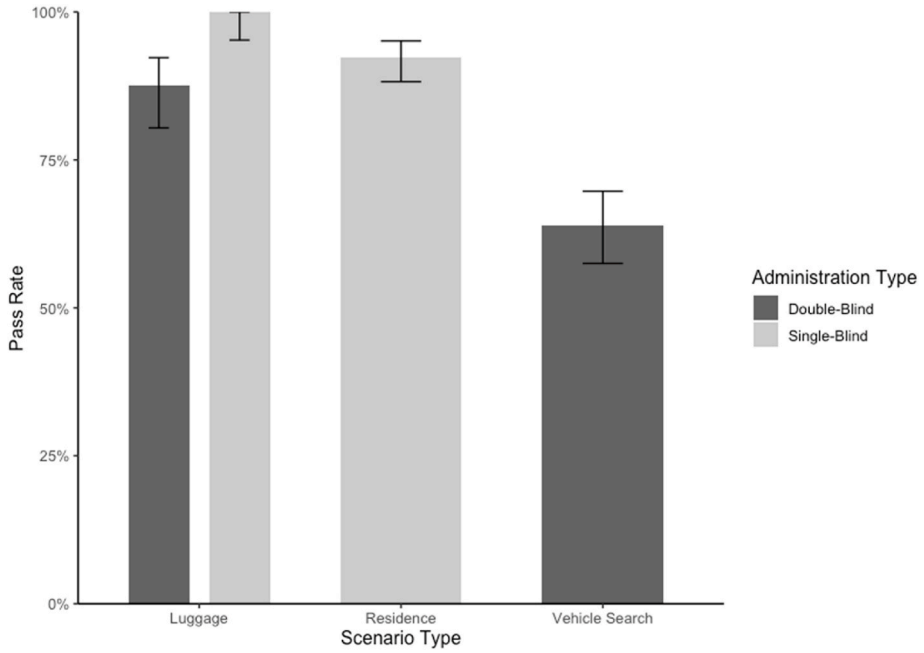

Fig. 2. Canine-handler teams pass rates as a function of scenario type and test administration type.
Notes. Error bars represent 95% Confidence Intervals.

narios (interaction effect: $\beta = 0.88$, $p = .064$, OR = 1.61). Perhaps more experience participating in double-blind testing over the years improved their pass rates, or the teams that participated in the certifications were more likely to update their training to ensure they passed the double-blind components. This was a non-significant effect, though, as there was only a small number of teams that had participated in most of these trials ($n_{all} = 3$; $n_{all\text{-}1} = 4$, $n_{all\text{-}2} = 4$), limiting our statistical power to detect effects related to these variables.

3.5. Trends over time

We also examined passing rates for double-blind trials over the years. In Fig. 3, we present the pass rate within each administration type for each certification year (first attempts only). Pass rates for second attempts for each certification year can be found in Table 2. These data suggest that there was an interaction between when the canine-handler team took part in the double-blind component of the certification testing and their likelihood of passing that component of the certification. Specifically, double-blind trials during the later certifications were associated with significantly higher pass rates than during earlier years, but with only a very small change in pass rates over the years for single-blind tests (interaction effect: $\beta = -2.13$, $p = .001$, OR = 0.59).

4. 2024 survey method

4.1. Participants and design

Survey data were collected via an optional, anonymous paper survey distributed to 20 handlers who participated in the 2024 version of this type of certification trial. Use of these secondary data for this publication

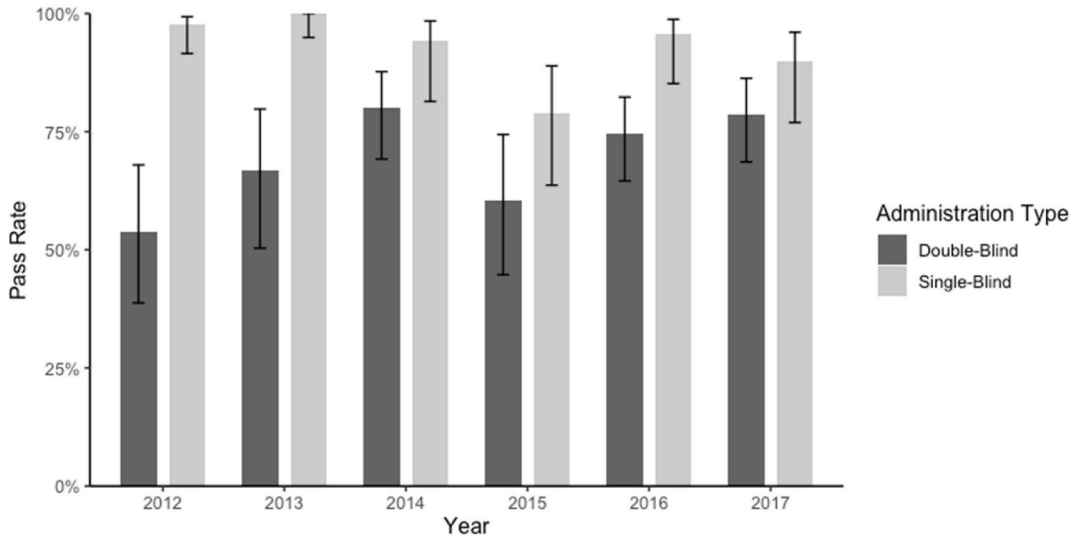

Fig. 3. Canine-handler teams' pass rates for first attempts in double-blind and single-blind trials for each certification year.
Notes. Error bars represent 95% Confidence Intervals.

Table 2
A summary of certification testing outcomes by year.

Attempt	Year	Administration Type	Pass Rate	# of Canine-Handler Teams	# of Trials
First	2012	Double-Blind	54%	41	41
		Single-Blind	98%	41	82
	2013	Double-Blind	67%	36	36
		Single-Blind	100%	36	72
	2014	Double-Blind	80%	35	70
		Single-Blind	94%	35	35
	2015	Double-Blind	61%	38	38
		Single-Blind	79%	38	38
	2016	Double-Blind	74%	45	90
		Single-Blind	96%	45	45
	2017	Double-Blind	79%	40	80
		Single-Blind	90%	40	40
Second	2012	Double-Blind	68%	19	19
		Single-Blind	–	–	–
	2013	Double-Blind	100%	12	12
		Single-Blind	–	–	–
	2014	Double-Blind	83%	9	12
		Single-Blind	–	–	–
	2015	Double-Blind	88%	8	8
		Single-Blind	–	–	–
	2016	Double-Blind	70%	21	23
		Single-Blind	–	–	–
	2017	Double-Blind	100%	14	15
		Single-Blind	67%	3	3

was reviewed and approved by the Institutional Review Board at Simon Fraser University, Canada (Protocol #: 30002020). These survey data were observational only—there were no manipulations between or within subjects.

4.2. Survey procedure and analysis

Canine handlers who completed similar certification trials in 2024 were asked if they would be willing to complete an optional, anonymous paper-based survey about double-blind testing. There were several multiple-choice, quantitative, and open-ended qualitative questions posed that were later examined for useful quotes and content patterns. Handlers who agreed to complete the survey provided information about the age of their dog, the dog's experience, and their own experience as a handler (reported in years). Then, they were asked "*Are double-blind exercises helpful to you?*". If they responded with a "*yes*", they then were asked to explain in their own words "… *how are they helpful?*".

Next, they were asked, "*If you had to do a second attempt for a Double-Blind exercise and pass the second time, why do you think you failed on the first pass, but passed on the second pass? Please be as specific as possible*". They were asked to respond in writing in their own words. Finally, they were asked "*Were you nervous conducting a double-blind exercise? If so, why do you think that was?*" Copies of handler's responses were sent to the first author so that they could conduct some simple quantitative (calculate the number/percentages of different types of responses) and qualitative analyses (identifying quotes featuring similar content/concerns).

5. Survey results and general discussion

The present study examined professional canine-handler team performance on ordinary narcotics detection certification trials conducted under double-blind versus single-blind conditions. With access to six years of annual testing (2012–2017), we analyzed data from 133 distinct teams. Overall, these teams were more than seven times more likely to pass a trial when working under single-blind conditions compared to double-blind conditions. Notably, many teams that initially failed a double-blind trial succeeded on a subsequent attempt within the same certification cycle. By comparing performance on more challenging vehicle searches to easier luggage searches, we also demonstrated that the performance drop under double-blind conditions could not be attributed solely to the difficulty of the vehicle searches—there was a smaller, but still significant, drop in performance observed when luggage searches were conducted double-blind as compared to single-blind. Finally, teams that participated in more recent certification trials demonstrated better performance on double-blind trials than teams in earlier years, suggesting that awareness of and increased experience with double-blind procedures may lead to improved performance under these conditions.

These results are consistent with research in other forensic disciplines demonstrating the importance of double-blind components in both proficiency testing [23–25] and investigative procedures (e.g., eyewitness identification; [16,18]). In real investigations—just as in other domains that require people to make forensically-relevant decisions—canine-handler teams operate under substantial uncertainty: there may be nothing to detect, or there may be multiple target odors located in unexpected places. Because no one on the scene knows the correct number or location of targets, teams must base their judgments solely on their observations of their dogs' trained behaviors rather than

on the cues or guidance available under single-blind conditions.

The present results suggest that canine-handler teams perform extremely well under single-blind conditions, where subtle cues from someone who knows the locations of any target odors may influence the search. However, when working under double-blind conditions, only 72% of teams passed their first attempt at the vehicle searches and 88% of teams passed their first attempt at the luggage searches. Most teams succeeded on a second attempt, suggesting that many can succeed under double-blind conditions with additional practice. Still, some teams were unable to complete the double-blind component even after a second

What Are the Important Points That Canine Handlers Can Take Away from This Study?
1. ***What is a double-blind test? How is it different from what I normally do?*** A double-blind test is when no one present during the test knows where the target odors are located—both the handler and the evaluator are "blind" to this knowledge. Usually, the evaluator is in the room and knows where the target odors are located, but the handler does not, which is called "single-blind".
2. ***The evaluators, trainers, and I are all careful not to cue each other or my dog improperly during training and tests. Do I still need to do double-blind searches?*** People find it really difficult to control their gestures, body language, facial expressions, tone, and other behavioral indicators—it is an innate form of communication that is incredibly useful in interpersonal situations. These are clues that give other people insight into their thoughts and feelings and all happens without us realizing. So, even the most experienced, caring, hard-working evaluators and trainers might cue handlers during a test without realizing they have done so.
3. ***What are the benefits of double-blind tests in certifications and training?*** When a canine-handler team successfully finds a target odor during a double-blind test, this confirms that the dog understands the scent profile of the target odor, the dog produces the correct physical response when they detect that target odor, and the handler is able to correctly interpret the dog's behavior as an alert. It also improves the handler's confidence in their ability to work with their dog and understand their dog's signals. Finally, double-blind tests can identify training deficiencies that can be corrected moving forward.
4. ***Are double-blind searches only useful in certifications or would it also be useful to incorporate double-blind searches into my routine maintenance training?*** Double-blind conditions offer clear benefits for both training and certification. Training regularly under these conditions helps handlers interpret their dog's behavior more effectively in settings that more closely mirror real operational work—where no one present during the search knows whether target odors are present or where they are located. Perhaps the most important benefit of this approach, though, is that handlers who routinely participate in double-blind searches report increased confidence in both their dog's detection abilities and their own capacity to recognize and interpret alert behaviors.
5. ***How do I include double-blind searches in my training? Do I need lots of equipment/resources?*** In fact, many canine-handler teams may already be participating in training activities similar to this without realizing it. You do not need a lot of equipment or resources to run double-blind searches. The main point is that no one present can know where the target odor is located. It can be as simple having your trainer or another individual go into a room and flip a coin to determine whether to place a target odor or leave the room "blank". They should conceal the target odor somewhere random if the trial is not "blank" and leave immediately after. The canine-handler team then runs the scenario and writes down the location of any alerts. The correctness of the alerts can be assessed later by the evaluator.
6. ***When should I incorporate double-blind searches into my training?*** Double-blind searches are fantastic for building confidence and solidifying foundational training. So, it is not recommended that canine-handler teams incorporate them in foundational training but, once both the dog and the handler have demonstrated they are proficient, they should incorporate regular double-blind searches during maintenance training.
7. ***I'm worried about the implications of rewarding my dog during double-blind searches because I won't know if their alerts are right or wrong. How do I work around this issue?*** Dogs are very intelligent, so one instance where they are rewarded incorrectly will not undo all of the hard work spent training them before that. One way to ensure that these types of instances are not an issue for ongoing training and performance is to start to introduce a variable-interval reward system whereby the canine is not tangibly rewarded every time they perform a task as intended. Instead, sometimes they are given the reward and other times they are simply encouraged and rewarded with praise (i.e., "Good dog!"). This will allow for slight variations in how rewards are delivered and when without disrupting their learning.
8. ***Is there anything else that you would recommend for someone who is interested in incorporating double-blind searches into their regular training and tests?*** It is highly recommended that canine-handlers video-record themselves (including both the canine and the handler in the footage) while they complete double-blind searches and review those later. This can help the handlers, and their trainers, determine whether there are behaviors that the handler is engaging in that they were not aware of and that may be resulting in the dog producing behavior changes or trained final responses in locations that they should not.

Fig. 4. Take home messages for practitioners and canine handlers.

attempt. In operational contexts, second attempts may not be feasible—teams must be capable of detecting target odors accurately on their first pass through a search scenario.

These findings carry several important implications for certification practices, training programs, and encouraging operational preparedness. First, double-blind scenarios should be viewed as an essential component of the maintenance training and testing once basic competences have been achieved and teams are beginning to transition toward operationally-relevant work. Although we did not explicitly test the impact of double-blind training exercises on performance, the current results, combined with other existing research and consensus knowledge, support this claim. For instance, double-blind scenarios are more similar to real operational searches, as no one present knows whether, where, or when a target odor may be present—that is ultimately why the canine-handler team has been deployed. Both the canine and handler, therefore, need to be accustomed to working under conditions of uncertainty, where the number and location of targets are unknown. The search strategies and communication with their dog must remain effective despite the lack of external cues.

Moreover, consistent success under double-blind conditions provides strong evidence that a team possesses core competencies across three foundation aspects of canine detection work: 1) the dog has been trained to identify the correct scent profile(s), 2) the dog reliably exhibits the correct behavioral response to the target odor(s), and 3) the handler consistently recognizes and correctly interprets their dog's alert behaviors. Furthermore, a failed double-blind test attempt should not be interpreted as a failure of the team. Instead, it should be viewed as an opportunity to diagnose habits and weaknesses in their current training and approaches that would not be revealed by a single-blind test. Recording or filming teams during double-blind trials can further enhance the training value of double-blind scenarios by enabling a detailed review of search strategy, pacing, the canine-handler relationship, and other subtle behaviors. However, the impact of incorporating double-blind training exercises in canine-handler teams regular training regimens should be directly tested in a future experiment so that the effect of such training is well understood and concrete implementation recommendations can be made for trainers and handlers.

Another important implication of the present findings is that double-blind testing and training need not be complicated, resource intensive, or costly. In fact, many handlers and trainers may be incorporating scenarios that approximate double-blind tests already, even if they do not label them in this way. Implementing true double-blind procedures can be as simple as having a trainer or another designated individual enter a room and flip a coin or roll dice to determine whether to place a target odor or leave the room "blank". If the result of the coin flip or dice roll requires that they place a target odor, the target odor should be concealed somewhere random. Importantly, the person setting up the double-blind scenario should only meet with the team after they have completed their run and written down the number of targets detected and their location(s). Double-blind administration is a scientific concept with many benefits, but it is fortunately not a complicated one to implement in practice. In Fig. 4, we provide practical guidance to canine-handler teams and their trainers for incorporating double-blind scenarios.

Although these data illustrate the value of double-blind certification components, it is also important to understand how handlers themselves view these more challenging scenarios. To gain insight into handlers' perceptions, Fred Helfers (co-author) conducted a survey of canine-handler teams who participated in a certification test similar to those described here. Performance patterns during this 2024 certification trial mirrored those observed in the present study but, in this instance, the 20 handlers who participated were also asked to describe their thoughts and feelings about the double-blind component of the test.

Nine handlers (45%) indicated that they found double-blind scenarios helpful because the scenarios forced them to learn and rely on their dog's behavioral changes, rather than on subtle cues or contextual information available in single-blind tests. Handlers also noted that double-blind conditions reduced distractions and allowed them to focus on the dog's behavior. As one handler described, "*it's just you and the dog. You are not relying on anyone but yourself.*" Another commented that double-blind administration presented a "*new challenge that helped me to learn my dog's behavior,*" and a third stated, "*I can trust my dog as it validates my training.*"

Similarly, many handlers recognized that double-blind practice helped build their confidence in their ability to work effectively when faced with real-world uncertainty. Eight of the 20 handlers (40%) reported that double-blind participation enhanced their confidence in their teams' detection skills. One handler with five years of experience stated that it "*helps me gain confidence in real world scenarios,*" and another with just over a year of experience described it as an "*excellent confidence boost.*" In addition, five handlers (25%) reported that double-blind tests felt more similar to the operational searches they performed professionally. One handler with five-and-a-half years of experience noted, "*they are as close to a real life search as you can get,*" and another with three years of experience described double-blind tests as "*the closest to a real scenario where it's just you and the dog.*"

Handlers were also asked whether they felt nervous about completing the double-blind portion of the certification. Seven (35%) reported feeling nervous, while an equal number said they were not. Some attributed their nerves to normal test anxiety, but others cited specific concerns: that double-blind scenarios were unfamiliar ("*I don't normally practice this exercise*"), or that the test might reveal weaknesses in their skills ("*it will show if you need more training*" or require "*relying fully on your handling … and how you work as a team*"). Handlers who were not nervous often noted that they already incorporate double-blind elements into routine training or felt confident in their dog's reliability, even without additional cues.

Four handlers also expressed concern about rewarding their dogs during double-blind trials, fearing that an incorrect alert might lead them to reinforce an error. As one handler said, "*I like to have confirmation from a trainer before I reward my dog,*" and another explained, "*my only concern was if my dog had a false response, I didn't want to reward her.*" These concerns are understandable and reflect the seriousness with which handlers approach training and reinforcement. However, dogs are highly resilient learners, and occasional, isolated instances of inadvertent reinforcement are unlikely to undermine a strong training foundation. Many experts recommend incorporating an intermittent reinforcement schedule in which handlers increasingly reduce the frequency of rewards for correct responses, reinforcing only select instances, which in turn produces behavior that persists even when reinforcement is absent [26]. Habituating dogs to intermittent reinforcement is particularly important when preparing them for operational scenarios that make it difficult to reinforce the dog regularly and creates natural variation in reinforcement delivery that accommodates the uncertainty inherent in double-blind scenarios without disrupting learning [27–29].

Overall, the present study demonstrates that double-blind testing is not only achievable, but valuable, practical, and strongly aligned with the realities of operational search work (see Fig. 4 for a summary). The marked performance differences between single-blind and double-blind conditions demonstrate that traditional approaches may overestimate operational readiness, whereas double-blind tests more accurately capture the challenges of real searches where no one knows the number or location of target odors. The fact that most teams improved quickly, and that handlers themselves value the experience, shows that double-blind testing is both feasible and welcomed by practitioners. By integrating regular double-blind exercises into certification and ongoing training, agencies can better prepare teams for the unpredictability of real investigations while also identifying areas for growth before they become operational weaknesses. As the field continues to advance, double-blind administration represents a straightforward, cost-effective way to strengthen the accuracy, credibility, and integrity of canine-detection

work.

CRediT authorship contribution statement

Adele Quigley-McBride: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Paola A. Prada-Tiedemann:** Writing – review & editing, Conceptualization. **Fred Helfers:** Writing – review & editing, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Given her role as an Editorial Board Member at Forensic Science International: Synergy, Dr. Quigley-McBride had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

These data were collected as part of regular certification testing offered to canine-handler teams, managed and run by Fred Helfers, founding president of the Pacific Northwest Police Detection Dog Association (www.fredhelfers.com).

References

[1] S. Gadbois, L. Elliott, M. Flannery, A critical review of the use of wildlife conservation canines in species-at-risk research, in: L. Lazarowski (Ed.), Olfactory Research in Dogs, Springer, Cham, 2023, pp. 375–440, https://doi.org/10.1007/978-3-031-39370-9_15.

[2] P. Jendrny, F. Twele, S. Meller, A.D.M.E. Osterhaus, E. Schalke, H.A. Volk, Canine olfactory detection and its relevance to medical detection, BMC Infect. Dis. 21 (2021) 838, https://doi.org/10.1186/s12879-021-06523-8.

[3] S. Sharma, B. Bangwal, D. Tripathi, P.K. Yadav, A. Badiye, Canine (K-9) forensics as an aid to police investigation, in: A. Badiye (Ed.), Advances in Forensic Science: Emerging Technologies and Techniques, Springer, Singapore, 2025, pp. 21–36, https://doi.org/10.1007/978-981-96-8971-2_2.

[4] I.E. Dror, S. Harnad, Offloading cognition onto cognitive technology, in: I.E. Dror, S. Harnad (Eds.), Cognition Distributed: How Cognitive Technology Extends Our Minds, John Benjamins Publishing Company, 2008, pp. 1–23, https://doi.org/10.1075/bct.16.02dro.

[5] I.E. Dror, J.L. Mnookin, The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science, Law Probab. Risk 9 (1) (2010) 47–67, https://doi.org/10.1093/lpr/mgp031.

[6] Innocence Project of Texas, Dog Scent Lineups: a Junk Science Injustice, Special Report, Innocence Project of Texas, 2009, Sept. 21.

[7] A.E. Taslitz, Does the cold nose know? The unscientific myth of the dog scent lineup, Hastings Law J. 42 (1990) 15–134.

[8] P. Berg, T. Mappes, M.V. Kujala, Olfaction in the canine cognitive and emotional processes: from behavioral and neural viewpoints to measurement possibilities, Neuroscience and Behavioral Reviews 157 (2024) e105527, https://doi.org/10.1016/j.neubiorev.2023.105527.

[9] L. Issel-Tarver, J. Rine, Organization and expression of canine olfactory receptor genes, Proc. Natl. Acad. Sci. 93 (20) (1996) 10897–10902, https://doi.org/10.1073/pnas.93.20.10897.

[10] A. Lesniak, M. Walczak, T. Jezierski, M. Sacharczuk, M. Gawkowski, K. Jaszczak, Canine olfactory receptor gene polymorphism and its relation to odor detection performance by sniffer dogs, J. Hered. 99 (5) (2008) 518–527, https://doi.org/10.1093/jhered/esn057.

[11] C.C.A. Clark, N.J. Sibbald, N.J. Rooney, Search dog handlers show positive bias when scoring their own dog's performance, Front. Vet. Sci. 7 (2020) 612, https://doi.org/10.3389/fvets.2020.00612.

[12] W.S. Helton, Overview of scent detection work: issues and opportunities, in: W. S. Helton (Ed.), Canine Ergonomics: the Science of Working Dogs, CRC Press, 2009, pp. 83–96.

[13] L. Lazarowski, S. Krichbaum, L. DeGreeff, A. Simon, M. Singletary, C. Angle, L. Waggoner, Methodological considerations in canine olfactory detection research, Front. Vet. Sci. 7 (2020), https://doi.org/10.3389/fvets.2020.00408.

[14] G.L. Wells, C.E. Luus, Police lineups as experiments: social methodology as a framework for properly conducted lineups, Pers. Soc. Psychol. Bull. 16 (1) (1990) 106–117, https://doi.org/10.1177/0146167290161008.

[15] N. Steblay, J. Dysart, S. Fulero, R.C.L. Lindsay, Eyewitness accuracy rates in police showup and lineup presentations: a meta-analytic comparison, Law Hum. Behav. 27 (5) (2003) 523–540, https://doi.org/10.1023/A:1025438223608.

[16] G.L. Wells, M.B. Kovera, A.B. Douglass, N. Brewer, C.A. Meissner, J.T. Wixted, Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence, Law Hum. Behav. 44 (1) (2020) 3–36, https://doi.org/10.1037/lhb0000359.

[17] L. Lit, J.B. Schweitzer, A.M. Oberbauer, Handler beliefs affect scent detection dog outcomes, Anim. Cognit. 14 (3) (2011) 387–394, https://doi.org/10.1007/s10071-010-0373-2.

[18] M.B. Kovera, A.J. Evelo, The case for double-blind lineup administration, Psychol. Publ. Pol. Law 23 (4) (2017) 421–437, https://doi.org/10.1037/law0000139.

[19] A. Quigley-McBride, I.E. Dror, T. Roy, B.L. Garrett, J. Kukucka, A practical tool for information management in forensic decisions: using Linear sequential unmasking-expanded (LSU-E) in casework, Forensic Sci. Int.: Synergy 4 (2022) 100216, https://doi.org/10.1016/j.fsisyn.2022.100216.

[20] M.T. DeChant, C. Ford, N.J. Hall, Effect of handler knowledge of the detection task on canine search behavior and performance, Front. Vet. Sci. 7 (2020) 250, https://doi.org/10.3389/fvets.2020.00250.

[21] S.D. Gronlund, C.A. Goodsell, S.M. Andersen, Lineup procedures in eyewitness identification, in: L. Nadel, W.P. Sinnott-Armstrong (Eds.), Memory and Law, Oxford University Press, 2012, pp. 59–83, https://doi.org/10.1093/acprof:oso/9780199920754.003.0003.

[22] G.S. Cooper, V. Meterko, Cognitive bias research in forensic science: a systematic review, Forensic Sci. Int. 297 (2019) 35–46, https://doi.org/10.1016/j.forsciint.2019.01.016.

[23] C. Hundl, M. Neuman, A. Rairden, P. Rearden, P. Stout, Implementation of a blind quality control program in a forensic laboratory, J. Forensic Sci. 65 (3) (2020) 815–822, https://doi.org/10.1111/1556-4029.14259.

[24] L.C. LaMotte Jr., G.O. Guerrant, D.S. Lewis, C.T. Hall, Comparison of laboratory performance with blind and mail-distributed proficiency testing samples, Publ. Health Rep. 92 (6) (1977) 554–560.

[25] R. Mejia, M. Cuellar, J. Salyards, Implementing blind proficiency testing in forensic laboratories: motivation, obstacles, and recommendations, Forensic Sci. Int.: Synergy 2 (2020) 293–298, https://doi.org/10.1016/j.fsisyn.2020.09.002.

[26] E.O. Aviles-Rosa, L.S. Fernandez, C. Collins-Pisano, P.A. Prada-Tiedemann, N. J. Hall, The use of an intermittent schedule of reinforcement to evaluate detection dogs' generalization from smokeless-powder, Anim. Cognit. 25 (2022) 1609–1620, https://doi.org/10.1007/s10071-022-01648-y.

[27] N.J. Hall, Persistence and resistance to extinction in the domestic dog: basic research and applications to canine training, Behav. Process. 141 (2017) 67–74, https://doi.org/10.1016/j.beproc.2017.04.001.

[28] J.A. Nevin, R.C. Grace, A. Holland, A.P. McLean, Variable-ratio versus variable-interval schedules: response rate, resistance to change, and preference, J. Exp. Anal. Behav. 76 (1) (2001) 43–74, https://doi.org/10.1901/jeab.2001.76-43.

[29] Zoeta Dogsoul, The science of successful dog training: timing, rewards, and reinforcement explained, Zoeta Dogsoul (2025, June 8). Retrieved from, https://zoeta-dogsoul.com/dog-training/dog-training-timing-rewards/#Understanding_Reinforcement_Types. (Accessed 20 November 2025).