# Individual differences in perceptual abilities in medical imaging: The Vanderbilt Chest Radiograph Test

Mackenzie A. Sunday[1]*, Edwin Donnelly[1] & Isabel Gauthier[1]

[1]Vanderbilt University

*Correspondence to:
Mackenzie Sunday
Department of Psychology
Vanderbilt University
226 Wilson Hall
Nashville, TN 37204
USA
Email: mackenzie.a.sunday@vanderbilt.edu

Isabel Gauthier email: isabel.gauthier@vanderbilt.edu

Edwin Donnelly email: edwin.donnelly@vanderbilt.edu

Abstract:

Radiologists make many important decisions when detecting nodules in chest radiographs. While training can result in high levels of performance on this task, there could be individual differences in relevant perceptual abilities that are present pre-training. A pre-requisite to address this question is a valid and reliable measure of such abilities. The present work introduces a new measure, the Vanderbilt Chest Radiograph Test (VCRT), which aims to quantify individual differences in perceptual abilities for radiograph decisions in novices. We validate the relevance of the test to diagnostic imaging by verifying radiologists' superior performance on the test compared to novices. The final VCRT version produces scores with acceptable internal consistency. Then we investigate how the VCRT can be used in future research by evaluating how the test relates to extant measures of face and object recognition ability. We find that the VCRT shares a small but significant portion of its variance with a measure of novel object recognition, suggesting that some aspect of VCRT performance is driven by a domain-general visual ability.

Significance statement:
This work presents a new measure of lung nodule detection ability for use in research investigating radiological expertise and training. Additionally, the work presents evidence that there may be a general visual ability relevant to detecting nodules in thoracic radiographs.

*Keywords: vision, perception, detection, diagnostic*

In the United States, becoming a thoracic radiologist usually requires 4 years of medical school, 1 year of internship, 4 years of residency and 1 additional year of a thoracic radiology fellowship. This training qualifies radiologists to make expert decisions of vital importance in medical treatment, but studies have documented a non-negligible level of errors in these decisions (Goddard et al., 2001; Manning, Ethell & Donovan, 2004). A better understanding of the various influences on these decisions could help to lessen this error rate. The bulk of radiological expertise research has focused on the relation between search patterns and nodule detection (specifically to address whether radiologists engage in holistic processing, Donovan & Litchfield, 2013; Drew et al., 2013; Kundel et al., 2007). Most of this research has investigated visual search patterns of radiologists to show that radiologists scan radiological images differently from novices (Kundel, Nodine & Carmody, 1978; Mello-Thoms et al., 2005; Bertram et al., 2013). However, other work has shown that experts can rapidly identify nodules at above chance rates in short durations that would only allow a few eye movements (durations as short as 200 millisecond in Kundel & Nodine, 1975; for other work see Oestmann et al., 1988; Mugglestone et al., 1995; Kundel et al., 2007; Carmody, Nodine & Kundel, 1981), suggesting that expertise may partly rely on aspects of perceptual processing that do not require visual search.

Another question addressed in radiological research has been whether radiological expertise generalizes to other tasks and domains (Beck et al., 2013; Nodine & Krupinski, 1998; Sowden, Davies & Roling, 2000). The results of this work have been inconclusive so far, with some work showing that lower level perceptual abilities such as contrast sensitivity are enhanced in radiologists (Sowden et al., 2000), but more complex skills

like visual search (Nodine & Krupinski, 1998) and visual working memory (Beck et al., 2013) are unaffected by acquiring radiological expertise.

Within all of this work, individual differences in performance are sometimes noted (Donovan & Litchfield, 2013) but rarely discussed. Variability in radiologists' performances may occur for several reasons, including differences in decision-making (Donovan & Litchfield, 2013) and perceptual abilities (Bass & Chiles, 1990). In turn, these abilities may be influenced by variability in training and experience or pre-existing individual differences in perceptual abilities. These influences of individual differences have remained unexplored, and even when radiological performance is explicitly measured, most studies do not focus on the psychometric properties of the task, including its reliability (Harley et al., 2009; Bass & Chiles, 1990). The general goal of our study is to develop a test capable of measuring such pre-existing individual differences to then determine how these individual differences might relate to object recognition abilities.

Because the study of individual differences in high-level vision is a recent development, it is unsurprising that pre-existing individual variability in the field of radiology has not been considered. People likely underestimate the extent to which individuals in the normal population vary in perceptual ability, but recent studies have shown large individual differences in perceptual processing of faces, of various familiar object categories, and even of novel objects (Duchaine & Nakayama, 2006; McGugin et al., 2012; Dennett et al., 2012; Richler, Wilmer & Gauthier, in press). Given the recentness of these findings, individual differences in novice radiological detection abilities have been overlooked.

Our goal in creating a measure of perceptual abilities relevant to the domain of chest radiographs is to examine whether variability in novice perceptual abilities determines how much an individual can benefit from experience, and ultimately, how they will differ as experts. Critically, we design our test for use with a novice population. To that end, we use a 4-alternative forced-choice method with a single nodule present on each target to tap into perceptual processing, in contrast to more complicated tasks in which subjects do not know how many nodules may be present, involving more complicated decision-making processes (Donovan & Litchfield, 2013; Bass & Chiles, 1990). A great deal of research in radiology comes from the tradition of visual search (Drew et al., 2013; Bertram et al., 2013; Carmody et al., 1981; Kundel, Nodine & Carmody, 1978). In most classic visual search studies, the target is well-specified and the difficulty comes from localizing it among distractors that possess similar features. The present work was inspired by a different tradition, studies in category learning and object recognition (Palmeri & Gauthier, 2004), in which categories are more probabilistically defined, and in some cases, have to be learned by subjects through trial and error. Therefore, here we are less interested in the ability to localize nodules following instruction on what they look like, and more interested in subjects' abilities to learn the features of suspicious nodules from examples. Ultimately, the processes involved in category learning and in visual search are both likely to be relevant to real world radiological training.

In addition, because most people have little to no familiarity with nodule detection in chest radiographs (compared to recognition of faces, cars, planes, etc.), we measure the extent to which nodule detection is predicted by performance in novel object

recognition. A recent test of novel object recognition memory (Novel Object Memory Test or NOMT; Richler, Wilmer & Gauthier, in press) measured an ability distinct from general intelligence, which generalized across visually different novel objects categories ($r^2$ = .23) and was distinct from face or car recognition abilities ($r^2$ = .10, Richler, Wilmer & Gauthier, in press). Given this, we correlate our chest radiograph test with tests of novel object recognition and of face recognition ability (the Cambridge Face Memory Test or CFMT, Duchaine & Nakayama, 2006) to determine if our measure of nodule detection ability shares more variance with a novel object measure than a face recognition measure (as we predict). To be clear, our goal is not to determine whether expert radiological detection is the same as expert object recognition. Rather, our main aim in testing these relations is to demonstrate how our new test might be used in future research to determine if a domain-general object recognition ability is relevant to radiological expertise. If the new measure we create is very highly correlated with performance on the NOMT (which is possible since chest radiographs are to some extent novel objects to novices), this would suggest a domain-specific test like the one we have created is not necessary to measure pre-existing perceptual abilities relevant to these decisions.

In three studies, we present our new nodule detection test and then begin to explore important properties of the test. We honed our new test to produce acceptable reliability in Study 1, and then assessed the test's validity by measuring how well medical professionals performed on the test (Study 2). In Study 3, we asked if there was any shared variance between our nodule detection test and a face and object recognition

measure, to see if our test might be useful in determining how a domain-general ability may contribute to real-world skills like nodule detection.

## Study 1

To create a measure of lung nodule detection ability, we developed the Vanderbilt Chest Radiograph Test (VCRT). Because low reliability can attenuate the observed correlation between two measures (Nunnally, 1970), it is crucial that we develop a test that produces reliable scores (keeping in mind that reliability is not a test property and must be evaluated with each new dataset). Through several iterations we honed the test to produce reliable scores with a novice population.

### *Methods*

#### *Subjects*

For the first version of the VCRT, 50 subjects were recruited online from Amazon Mechanical Turk and compensated $0.50. For all experiments, only subjects with U.S. IP addresses and at least 95% of their previous Amazon Mechanical Turk tasks accepted were eligible to participate. Subjects were asked to rate their expertise with “chest x-rays” on a scale from 1-9. Two subjects were excluded for failure to follow instructions, leaving 48 subjects for analysis (18 male, mean age = 35.33). For the second VCRT version, 49 subjects were recruited and compensated $0.50. One subject was excluded for incorrectly answering both catch trials, and of the 48 remaining subjects, 16 were male (mean age = 38.65). One hundred and nineteen subjects were recruited to complete the final VCRT followed by an additional test discussed in Study 3 (Novel Object Memory

Test) and were compensated $0.75 for completing both. Ten of these subjects were excluded for failure to follow instructions, leaving 108 subjects (39 male, mean age = 38.51). This study and all following studies were conducted under approval by the Vanderbilt University Institutional Review Board and informed consent was obtained for each subject.

*Stimuli*

Stimuli were chest radiographs of 212 individuals (with any identifying information removed). Of these, 106 chest radiographs contained cancerous nodules (no image contained multiple nodules) and 106 were nodule free. All nodules were confirmed in a follow up CT scan to be non-calcified nodules and the nodules had a mean diameter of 25.3 mm (SD = 12.6 mm, range = 7.0-67.5 mm). Nodules were identified by one of the authors, a thoracic radiologist with over 20 years of experience reading chest radiographs. Images were cropped to a 1.3:1 ratio and converted to grayscale. Other than this, images were not altered, so any inorganic elements (pacemakers, surgical screws, shadows from bed gurneys) were included. In this way, we hoped to keep the chest radiograph images as similar as possible to images seen by radiologists in the field, thereby maximizing the test's construct validity. Nodules appeared 49 times in the left lung and 31 times in the right lung. Though this may have produced a slight left-bias, because each individual sees the same stimuli, this left-bias would not confound the measured individual differences.

*Procedure*

The initial test began with instructions and two practice trials, followed by 106 total trials (two of which were catch trials). Each practice trial was identical to the experimental trials except the feedback was accompanied with text saying "here is the

nodule" so that the subjects understood the feedback. Other than these two practice trials, subjects were given no specific instructions about the nodules but were told they could learn from the feedback. On each trial, subjects viewed two chest radiographs, presented horizontally (Figure 1). Subjects were instructed to "guess which of the four lungs has a cancerous nodule" and to indicate their response by clicking on the location where they believed the nodule was. Responses were un-speeded. Subjects were scored as correct if they click on the correct quarter of the screen (as divided horizontally into four vertical sections, corresponding to each of the four lungs). We did not record the exact location of the click but only the selection of the chosen lung. Because of this, and the fact that we did not purposefully manipulate any of our stimuli properties (number of targets, contrast, etc.), we did not design our task to be a standard search task, though subjects were asked to search through images for a nodule. A chance level of 25% was considered sufficient to measure individual differences among untrained novices. Following each response, subjects received feedback for 2000 milliseconds, during which the correct radiograph image (right or left) was outlined in red and the nodule circled in red (Figure 1). Our decision to include feedback on every trial was in part based on pilot data showing that when no feedback was given to subjects, performance was at chance[1]. We did not intend for the test to be a training task, but instead wanted to measure how well novice subjects can learn from exposure and feedback to detect lung nodules. We did not design the task to be used for training or to measure the efficacy of a training protocol, in fact the task may be too easy for experts and may be more useful in assessing whether individual differences before training predict how well individuals benefit from training. On the two

[1] For the no-feedback 3AFC pilot, the average accuracy of 100 subjects was 36.96% (SD = 5.34%).

159 catch trials, one chest radiograph was presented along with a landscape scene, and any
160 response choosing the chest radiograph (either lung) was coded as correct. The procedure
161 used in second VCRT version and the final VCRT were the same as the initial version.
162 Trials were always presented in the same order to reduce contribution of order variance in
163 the measurement of individual differences.

164 A limitation of testing people online is that some of the variability in performance
165 could be attributed to different testing conditions (not only screen differences, but also
166 ambient light and other factors such as noise in the room, presence of other people, etc).
167 This is a tradeoff against the lack of subject variability that arises when only
168 undergraduate students are tested in the laboratory (in this case, there may be less
169 variability due to testing conditions, but there may also be less person variability). To
170 reduce variability due to testing conditions, we instructed subjects not to complete the test
171 on a handheld screen, and, in the final VCRT version, had subjects perform a contrast
172 check before the test. This contrast check consisted of three trials preceding the practice
173 trials that required subjects to choose a low contrast diagonal Gabor patch from a set of 3
174 patches (the other two being solid). This test was meant to ensure that subjects were
175 completing the test on a screen with sufficient contrast. If subjects did not correctly
176 answer this contrast check, they were instructed to increase their screen contrast. Prior
177 research with the CFMT reveals that tests can perform similarly online and in the lab
178 (even at the level of individual trial information, Cho et al., 2015) and that person
179 characteristics (such as gender) that predict differential performance on some tests do so
180 in both online and laboratory samples (Ryan & Gauthier, 2016), demonstrating that

testing condition variance does not overshadow true individual differences. Nonetheless, future work should validate the VCRT under more controlled conditions.

---PLACE FIGURE 1 ABOUT HERE---

*Results*

The initial VCRT version had an average accuracy of 45.00% (SD = 8.15%) and produced an internal consistency of α = .736. We examined items having a low correlation between item responses and subjects' total scores (which were thus relatively uninformative). For 44 items, we replaced the distractor images (chest radiographs with no nodules), with a different distractor image. This second version of the test had an average accuracy of 48.70% with less variance than the first version (SD = 6.65%). In addition, internal consistency was also lower than the initial version (α = .609).

We further sought to improve reliability by taking the 78 trials from the second version that produced the highest correlation between item responses and subjects' total scores (i.e. the most informative trials), while attempting to maximize the range of difficulty in test items, to create the final VCRT. For this final VCRT, we also ordered trials from easiest to most difficult based off of the item accuracies produced in the second VCRT version. The final VCRT version has 80 trials total (including two catch trials), and takes approximately twenty minutes to complete. This final version had average accuracy of 53.00% (SD = 10.13%) and we observed acceptable reliability in our sample (α = .799). This final version is available online at http://gauthier.psy.vanderbilt.edu/resources/ and the data are available at *future figshare link*).

*Discussion*

We developed the VCRT to measure the ability of novices to learn to identify lungs that contain suspicious nodules in chest radiographs, based on feedback. Based off of the first two versions of the VCRT, we created the final VCRT, which produces reliable scores of chest nodule detection ability. Though our test has good face validity, it is important to critically evaluate the construct validity to ensure our test is measuring its targeted construct, which we do in Study 2.

## Study 2

To validate our new test of nodule detection in chest radiographs, we asked radiologists and radiological students to complete the test. If the VCRT taps into a construct used by radiologists to make actual determinations about the presence of cancerous nodules, then we would expect medical professionals to perform well on the test.

*Methods*

*Subjects*

We recruited five medical professionals to complete the final VCRT version (hereafter referred to as the VCRT). Subjects who completed the task were given a 1-in-5 chance to win $20.00. Two subjects had completed thoracic radiology fellowships and the remaining three were radiology residents (three male, mean age = 37.4).

*Procedure*

The final version of the VCRT, with 80 trials total, was used.

*Results*

Average VCRT accuracy of the expert group was 81.54% (SD = 6.32%), which was significantly greater than the non-medical professionals' performance reported in Study 1 ($t(3) = 9.0$; $p = .002$, $d = 3.38$) .

*Discussion*

Our medical professionals performed well on the VCRT. We believe these medical professionals performed well mostly because of their extensive training and experience with reading chest radiographs. However, it is certainly possible that their performance could be due to increased motivation, a difference in strategy, or something else. A number of other differences could have contributed to the above average performance of the medical professionals, so further investigation is needed to elucidate possible causes for their superior performance.

Regarding our goal of assessing the validity of our test, we can definitively say that it would have been concerning if these medical professionals performed poorly, or even in the same range as novices on our test. However, this sample of medical professionals was superior in their ability to detect nodules compared to novice observers. This supports the VCRT as a valid measure of nodule detection ability in chest radiographs, although it is impossible to know whether they achieved superior performance using a qualitatively different strategy from novices. For now, these results

serve to better characterize the VCRT and open the door for further research aimed at validating this measure.

## Study 3

With the aim of creating a measure of nodule detection in chest radiographs, we developed the VCRT. Our purpose in creating such a measure is to provide a useful tool for future work studying perceptual individual differences as may be relevant to medical training. However, one possibility is that because chest radiographs are essentially novel objects to novices (as compared with domains like cars and faces) our test will essentially tap into the same ability as existing tests that measure perceptual abilities with novel objects. Therefore, we decided to quantify the overlap between the ability measured by the VCRT and an existing test of recognition for novel objects. Our purpose here is not to draw conclusions about the nature of these domains based on the specific mechanisms involved in each of these tasks but rather to better understand how the ability to learn how to identify suspicious chest nodules based on feedback relates to face and novel object recognition ability. While the VCRT involves a purely perceptual task, the CFMT and NOMT tasks include both a perceptual component (to encode the stimuli) and a memory component. Previous work provided evidence that the NOMT measures a domain-general visual ability that is independent from general intelligence and memory span (Richler, Wilmer & Gauthier, in press). The CFMT and NOMT are existing measures of high-level visual abilities that have been found to correlate with performance on other perceptual tasks in past research – importantly, any correlation between these

each of these tasks and the VCRT cannot be attributed to similarity of task format and would therefore be more likely due to task-general visual ability.

This relation is interesting in light of recent evidence for domain-general visual abilities relevant to object recognition, as expressed by common variance between tests of familiar and novel object recognition (Richler et al., submitted; Van Gulick et al., 2016). Additionally, novel object recognition shows some limited shared variance ($r^2$=.10; Richler, Wilmer & Gauthier, in press) with face recognition, as measured by the Cambridge Face Memory Test (Duchaine & Nakayama, 2006). Because chest radiographs are likely closer to novel than familiar objects within a novice population, we expected VCRT performance to show a stronger correlation with a novel object recognition measure than with a face recognition measure. Finding this would indicate that some of the VCRT performance relies on the same ability relevant to discriminating novel objects across different viewpoints, providing further evidence for a domain-general visual ability. We also expected to replicate the small but significant correlation between the CFMT and NOMT.

## *Methods*

### *Subjects*

One hundred and nineteen subjects were recruited to complete the VCRT followed by the NOMT and compensated $0.75 (as described in Study 1). The 108 subjects (mean age = 38.41, 38 male) who were not excluded from analyses in Study 1 were given the opportunity to complete the CFMT for an additional $1.00. Of the 75 subjects who chose to complete the CFMT, 23 were male (mean age = 39.24).

Additionally, the five medical professionals from Study 2 were given the opportunity to complete the NOMT and four did so and were thus compensated $10.00.

*VCRT*

The final version of the VCRT (also used Study 2) was used.

*Cambridge Face Memory Test (CFMT)*

In the CFMT (Duchaine & Nakayama, 2006), subjects studied six Caucasian grayscale male target faces and then had to correctly identify the target face presented with two foil faces on each trial. The first block showed target faces in the studied viewpoint (18 trials), and in the second block subjects identified the target across variations in lighting and viewpoint (30 trials). For the third block (24 trials), Gaussian noise was added to novel target images. Here, we used the long version of the CFMT (Russell et al., 2009), so there was an additional final block (30 trials), which was designed to be the most difficult, with uncropped faces in profile and additional noise added. Subjects studied the target images between each block and responses were un-speeded.

*Novel Object Memory Test (NOMT)*

The NOMT is a test of object recognition ability that minimizes the influence of experience by using computer generated novel objects with which subjects have no experience. The test has produced reliable scores in a normal population tested online (Richler, Wilmer & Gauthier, in press) and shows convergent validity due to its

correlation with similar tasks with other novel categories ($r^2$ = .23). The test follows a procedure modeled after the CFMT, where six novel objects are learned and then tested with a three-alternative forced choice in subsequent trials. In the NOMT, there are 54 trials following the learning phase (in which feedback is given), in which objects have to be recognized across small variations in viewpoint. Here, we use the novel object category called Ziggerins (Wong, Palmeri & Gauthier, 2009, Figure 2).

---PLACE FIGURE 2 ABOUT HERE---

*Results*

With all 112 subjects (108 from Study 1 plus the four medical professionals) who completed the VCRT and NOMT, average NOMT accuracy was 71.54% (SD = 16.73%). Average total time to complete the VCRT (including instructions and practice trials) for online subjects was 26.97 minutes (SD = 8.9 min) and the average response time on a single trial was 6.93 seconds (SD = 3.06 seconds). The average response time of the four medical professionals was 6.77 seconds (SD = 5.79 seconds), which did not differ significantly from the online subjects ($t(110) = 0.01$, $p = .99$). Self-reported chest radiograph expertise (on a scale from 1-9, M = 3.46, SD = 1.69) from the online subjects did not correlate with VCRT accuracy ($r_{108} = .07$, 95% CI [-.12, .25], $r^2 = .004$, $p = .48$), so all online subjects were included. Both tests produced acceptable reliabilities (VCRT $\alpha = .799$; NOMT $\alpha = .960$). There was a significant correlation between performance on the VCRT and NOMT ($r_{108} = .23$, 95% CI [.04, .40], $r^2 = .05$, $p = .02$, Figure 3). This correlation increased somewhat with the four medical professionals included ($r_{112} = .28$, 95% CI [.10, .44], $r^2 = .08$, $p = .003$, Figure 3).

Average CFMT accuracy was 52.00% (SD = 12.25%) and the CFMT also showed good internal consistency (α = .839). As in prior work, the CFMT and NOMT were significantly correlated ($r_{75}$ = .29, 95% CI [.07, .48], $r^2$ = .08, $p$ = .01). However, the CFMT and VCRT did not correlate significantly ($r_{75}$ = .12, 95% CI [-.11, .34], $r^2$ = .01, $p$ = .3), and moreover, the VCRT was not significantly more correlated with the NOMT than with the CFMT (Steiger $Z$ = .80. $p$ = .42).

---PLACE FIGURE 3 ABOUT HERE---

*Discussion*

With the creation of a reliable measure of lung nodule detection ability in novices, we investigated how this ability relates to other high-level visual abilities measured in recent work. We find that the VCRT shares a small but significant amount of variance with a measure of novel object recognition ability, although we did not have sufficient power to demonstrate that there was more variance than the test shared with face recognition ability. Future efforts should include additional domains and other task formats to better characterize the relation between the ability measured in the VCRT and object recognition abilities. Importantly, given its reliability coupled with the present results, the VCRT appears to measure variation between individuals that is distinct from what is measured in these existing tasks.

Interestingly, and despite the modest correlation between the VCRT and the NOMT, the four medical professionals also performed well on the NOMT. The two radiological residents scored above average (80.56% and 77.78%) and the two subjects who had completed thoracic radiology fellowships scored over one standard deviation

above average (both 94.44%). Given the small sample size of medical professionals we have, this is merely an intriguing observation. It could be attributed to superior motivation in our experts, but it is also possible that only individuals with very good domain-general visual skills choose and succeed in medical imaging. More work with larger samples and additional tasks is needed to better understand novel object recognition abilities in expert radiologists, but our work suggests the utility of using tests of object recognition ability in expert radiologists, in addition to the visual search and working memory tasks that have been used in prior research (Donovan & Litchfield, 2013; Nodine & Krupinski, 1998; Beck et al., 2013). Generally, this work provides a starting point for further research investigating how the VCRT relates to other measures.

*General Discussion*

In three studies, we present a new measure of lung nodule detection ability (the VCRT), validate this measure and then assess how the measure relates to object recognition abilities. Our test provided reliable measurements of novices' detection of cancerous lung nodules within chest radiographs. We also found that radiologists performed above average on our test (average z-score = .92), providing some evidence that the test taps into an ability that is high in expert radiologists.

Our long-term goal is to determine whether this test could predict outcomes of diagnostic radiological training. With this goal in mind, we find that our test shares a small amount of variance with a novel object recognition measure, tentatively suggesting that a small but significant amount of variation in VCRT performance may be accounted for by a domain-general recognition ability. Though we might attribute this shared

variance to the fact that chest radiographs can be considered novel to the novice subjects (though how novel chest radiographs are to subjects is an unexplored question), we also find that the small sample of experts show above average NOMT performance. Thus, we are hesitant to conclude that the variance shared between the VCRT and NOMT is entirely driven by the novelty of each domain (since chest radiographs are not novel to radiologists). Instead, we cautiously conclude that some aspect of the ability measured by the NOMT is also relevant to the ability to detect nodules in chest radiographs. Critically, these results highlight the importance of using multiple visual tests when comparing experts and novices. For instance, one study found that experts outperformed novices on a transfer task meant to tap into similar processes as radiograph readings (Sowden et al., 2000), but did not include a control task (like the NOMT) to measure more distant visual processing. Given the result in Study 3, it is difficult to determine whether the experts in that study outperformed novices in the transfer task because of their radiological expertise (as was concluded in the study), because of a domain-general advantage, or a combination of the two. Thus, in addition to providing a new test that can be used to measure chest radiograph nodule detection in novices, this work also suggests that studies comparing novices and experts in domain-specific tasks will benefit from the inclusion of visual tests that tap into a varied set of visual abilities (ideally some visual abilities in which differences are predicted and some in which no differences are predicted).

We already know that experts can demonstrate superior perceptual performance (Russell et al., 2009; Curby, Glazek & Gauthier, 2009) and considerable work in perceptual learning demonstrates that such abilities can be acquired through practice (Gauthier et al., 1998; Jiang et al., 2007; Op de Beeck et al., 2006; Tanaka, Curran &

Sheinberg, 2005; Rossion et al., 2002; Wong et al., 2009; see Sagi, 2011). A new research program rooted in individual differences could help us understand whether some individuals can learn faster than others, and whether pre-training abilities like that measured by the VCRT places a limit on one's ultimate level of performance.

Figure 1. Example trials from the VCRT. Subjects responded by clicking on the nodule and were then given feedback and shown the nodule for 2000 ms. The upper trial is an example of an easy trial and bottom trial is an example of a more difficult trial.

Figure 2. Examples of six Ziggerin stimuli used on the Novel Object Memory Test.

Figure 3. Scatterplot of NOMT and VCRT accuracies (N = 112, medical professionals' data points marked with X's). Shaded region indicates 95%.confidence intervals.

Declarations:

*Ethics approval and consent to participate:*

All of the work reported here was done with the approval of the Vanderbilt University Institutional Review Board (IRB Protocol 050082)

*Consent for publication:*

Not applicable.

*List of abbreviations:*

VCRT (Vanderbilt Chest Radiograph Test); CFMT (Cambridge Face Memory Test); NOMT (Novel Object Memory Test)

*Availability of data and material:*

All data is available upon request to the corresponding author.

*Competing interests:*

None of the authors declare any competing interests.

*Funding:*

This work was supported by the National Science Foundation (SBE-0542013 and SMA-1640681). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (1445197).

*Authors' contributions:*

All authors contributed to the conception and design of the work and data collection. MS and IG contributed to data analysis and manuscript drafting. All authors contributed to the revision of the work.

*Acknowledgements:*

Not applicable

References:

Bass, J. C., & Chiles, C. (1990). Visual Skill Correlation with Detection of Solitary Pulmonary Nodules. *Investigative Radiology*, *25*(9), 994-997.

Beck, M. R., Martin, B. A., Smitherman, E., & Gaschen, L. (2013). Eyes-on training and radiological expertise: An examination of expertise development and its effects on visual working memory. *Human Factors*, *55*(4), 747-763.

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. *PloS ONE*, *8*(6), e66169.

de Beeck, H. P. O., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience*, *26*(50), 13025-13036.

Carmody, D. P., Nodine, C. F., & Kundel, H. L. (1981). Finding lung nodules with and without comparative visual scanning. *Attention, Perception, & Psychophysics*, *29*(6), 594-598.

Cho, S. J., Wilmer, J., Herzmann, G., McGugin, R. W., Fiset, D., Van Gulick, A. E., ... & Gauthier, I. (2015). Item response theory analyses of the Cambridge Face Memory Test (CFMT). *Psychological assessment*, *27*(2), 552.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in

format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587-605.

Donovan, T., & Litchfield, D. (2013). Looking for cancer: Expertise related differences in searching and decision making. *Applied Cognitive Psychology*, *27*(1), 43-49.

Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images?. *Radiographics*, *33*(1), 263-274.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576-585.

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J.W. (1998). Training 'greeble' experts: A framework for studying expert recognition processes. *Vision Research, 38,* 2401-2428.

Goddard, P., Leslie, A., Jones, A., Wakeley, C., & Kabala, J. (2001). Error in radiology. *The British Journal of Radiology*, *74*(886), 949-951.

Harley, E. M., Pope, W. B., Villablanca, J. P., Mumford, J., Suh, R., Mazziotta, J. C., ... & Engel, S. A. (2009). Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. *Cerebral Cortex*, *19*(11), 2746-2754.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape-and category-selective human neural plasticity. *Neuron*, *53*(6), 891-903.

Kundel, H. L., & Nodine, C. F. (1975). Interpreting Chest Radiographs without Visual Search 1. *Radiology*, *116*(3), 527-532.

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, *13*(3), 175-181.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: gaze-tracking study 1. *Radiology*, *242*(2), 396-402.

Manning, D.J., Ethell, S.C., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from posteroanterior chest radiograph. *British Journal of Radiology*, 77, 231–235.

McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research*, *69*, 10-22.

Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., ... & Maitz, G. (2005). Effects of Lesion Conspicuity on Visual Search in Mammogram Reading 1. *Academic Radiology*, *12*(7), 830-840.

Mugglestone, M. D., Gale, A. G., Cowley, H. C., & Wilson, A. R. M. (1995). Diagnostic performance on briefly presented mammographic images. *Medical Imaging 1995* (106-115). International Society for Optics and Photonics.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic Radiology*, *5*(9), 603-612.

Nunnally, J.C. (1970). *Introduction to psychological measurement*. McGraw-Hill, New York.

Oestmann, J. W., Greene, R., Kushner, D. C., Bourgouin, P. M., Linetsky, L., & Llewellyn, H. J. (1988). Lung lesions: correlation between viewing time and detection. *Radiology*, *166*(2), 451-453.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*(4), 291-303.

Richler, J.J., Tomarken, A.J., Vickery, T.J., Ryan, K.F., Floyd, R.J., Sheinberg, D., Wong, A.C.N., Gauthier, I. (in submission). Individual Differences in Object Recognition.

Richler, J.J., Wilmer, J.B., & Gauthier, I. (in press). General object recognition is specific: Evidence from novel and familiar objects.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, *13*(3), 250-257.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252-257.

Ryan, K. F., & Gauthier, I. (2016). Gender differences in recognition of toy faces suggest a contribution of experience. *Vision research*, *129*, 69-76.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, *51*(13), 1552-1566.

Sowden, P. T., Davies, I. R., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity?. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 379.

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*(2), 145-151.

Van Gulick, A. E., McGugin, R. W., & Gauthier, I. (2016). Measuring nonvisual knowledge about object categories: the semantic Vanderbilt expertise test. *Behavior research methods*, *48*(3), 1178-1196.

Wong, A. C. N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for facelike expertise with objects becoming a ziggerin expert—but which type?. *Psychological Science*, *20*(9), 1108-1117.