The Eye of the Beholder: Can Patterns in Eye Movement Reveal Aptitudes for Spatial Reasoning?

Victoria A. Roach,[1] Graham M. Fraser,[2] James H. Kryklywy,[3] Derek G.V. Mitchell,[1,3] Timothy D. Wilson[1]*

[1]Department of Anatomy and Cell Biology, Corps for Research of Instructional and Perceptual Technologies, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

[2]Department of Medical Biophysics, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

[3]Departments of Psychiatry and Psychology, The Brain and Mind Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

Mental rotation ability (MRA) is linked to academic success in the spatially complex Science, Technology, Engineering, Medicine, and Mathematics (STEMM) disciplines, and anatomical sciences. Mental rotation literature suggests that MRA may manifest in the movement of the eyes. Quantification of eye movement data may serve to distinguish MRA across individuals, and serve as a consideration when designing visualizations for instruction. It is hypothesized that high-MRA individuals will demonstrate fewer eye fixations, conduct shorter average fixation durations (AFD), and demonstrate shorter response times, than low-MRA individuals. Additionally, individuals with different levels of MRA will attend to different features of the block-figures presented in the electronic mental rotations test (EMRT). All participants ($n = 23$) completed the EMRT while metrics of eye movement were collected. The test required participants view pairs of three-dimensional (3D) shapes, and identify if the pair is rotated but identical, or two different structures. Temporal analysis revealed no significant correlations between response time, average fixation durations, or number of fixations and mental rotation ability. Further analysis of within-participant variability yielded a significant correlation for response time variability, but no correlation between AFD variability and variability in the number of fixations. Additional analysis of salience revealed that during problem solving, individuals of differing MRA attended to different features of the block images; suggesting that eye movements directed at salient features may contribute to differences in mental rotations ability, and may ultimately serve to predict success in anatomy. Anat Sci Educ 00: 000–000. © 2015 American Association of Anatomists.

Key words: spatial ability; spatial reasoning; mental rotation ability; eye tracking; gaze patterns; gross anatomy education; dynamic visualization; 3D structures

INTRODUCTION

Spatial ability, the capacity to understand and remember spatial relationships between objects, is thought to be a key factor that dictates how individuals perceive and interact with their surroundings (Thurstone, 1938; McGee, 1979; Carroll, 1993). Furthermore, the role of spatial ability influences not only how learners succeed in science, technology, engineering, medicine, or math (STEMM) disciplines (Shea et al., 2001) but also specifically the anatomical sciences (Rochford, 1985; Lufler et al., 2012; Nguyen et al., 2012). Gross anatomy is a visually complex topic, wherein students must learn to recognize anatomical features in different orientations, planes of section, and through different visualization modalities, through the application of visual cues, their spatial relationship to other structures (Zumwalt et al., 2015). Indeed, despite the variety of methods available to teach anatomy, the role that an individual's spatial ability cannot be

*Correspondence to: Dr. Timothy D. Wilson; Department of Anatomy and Cell Biology, MSB 490, CRIPT Lab, Western University, London, Ontario N6A5C1, Canada. E-mail: tim.wilson@uwo.ca

Received 12 August 2015; Revised 2 October 2015; Accepted 27 October 2015.

Published online 00 Month 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ase.1583

understated; particularly when utilizing resources that display anatomical features from varying viewpoints (Garg et al., 2001). With this in mind, one must consider the possible spatial-ability-based pedagogical techniques that could be designed to bolster this trait, and yield enhancements in the training of gross anatomy (Lufler et al., 2012).

Commonly used as an umbrella term, spatial ability is not monolithic, but rather composed of several discrete, but interrelated subabilities (Carroll, 1993). One of these subabilities is mental rotations ability (MRA): the capacity to rotate two- or three-dimensional figures rapidly and accurately (Linn and Petersen, 1985). For decades, MRA has occupied a niche in cognitive psychology, and has been linked to a number of other domains, including skill acquisition, knowledge transfer, and academic performance in spatially complex disciplines, such as surgical training and anatomical science (Wanzel et al., 2002a; Grober et al., 2003; Brandt and Davies, 2006; Van Herzeele et al., 2010; Lufler et al., 2012; Nguyen et al., 2012). Typically, MRA is measured by performance on standardized tests of mental rotations, such as tests employing the line-images of Shepard and Metzler (Shepard and Metzler, 1971) and the Vandenberg and Kuse Mental Rotations Test (Vandenberg and Kuse, 1978; Peters et al., 1995). These tests can serve to identify individuals as high-, intermediate-, or low-MRA based on individual score (Geiser et al., 2006). It is accepted that in timed conditions, individuals with higher MRA complete these tests in less time and with greater accuracy than those with lower spatial ability (Nguyen et al., 2014).

Researchers have attempted to investigate the cognitive processes that underlie MRA, and its relationship with skill acquisition and anatomical knowledge acquisition, but conclusive answers have yet to be determined (Leek et al., 2004). One hypothesis suggests that mental rotation may manifest through the movements of the eye, as fixations, that is, maintaining gaze on a single location (Carpenter, 1988), are intimately involved in our ability to visually encode spatially distributed information (Just and Carpenter, 1976; Shepard and Cooper, 1986). Foundational experimentation has demonstrated that individuals' gaze patterns are under cognitive control, and tailored to the task at hand (Buswell, 1935; Yarbus, 1967). Subsequently, investigations have shown that the order and duration of fixations are tightly linked to the specific target task (Triesch et al., 2003; Land, 2004; Hayhoe and Ballard, 2005; Kowler, 2011; Lai et al., 2013).

In a pioneering study, Just and Carpenter (1985) explored how eye movements may relate to spatial reasoning. Under their paradigm, significant differences in the eye movement metrics of individuals of high and low MRA were identified while participants answered questions composed of Shepard and Metzler line-images of blocks (Just and Carpenter, 1985). On average, low-MRA individuals exhibited longer trial response times, and conducted more fixations per trial. These results have thus encouraged further inquiry into the fundamental differences that exist between high- and low-spatial individuals, and how these intrinsic human factors can predefine success in mental rotations in terms of comprehension and apprehension of spatially salient structures. Regions of spatial salience are the areas that possess perceptual qualities that make them stand out relative to the their surroundings (Itti et al., 1998). In the case of the line-drawn blocks of Shepard and Metzler test, spatially salient structures are hypothesized as regions of the figures that convey depth and positional information divulging the orientation of the structure in space.

This study aims to explore eye movements and MRA, during the completion of an adapted, electronic mental rotations test (EMRT) where no time limits are imposed. The goal is to elucidate both temporal and salience patterns associated with MRA. It is hypothesized that MRA score will be negatively correlated with average fixation duration, average response time, and number of fixations occurring during the performance of the EMRT. Furthermore, individuals with different levels of MRA will attend to different features of the block-figures presented in the EMRT as they solve spatial questions. Finally, it is predicted that individuals of high MRA will demonstrate more variation in question-response time across the performance of the EMRT indicative of cognitive flexibility in solving spatially challenging visual problems. It is thought that through this line of investigation, differences between low- and high-MRA individuals will be revealed, and serve as a foundation for future eye-movement directed spatial ability training protocols. Such protocols would serve to enhance spatial reasoning in low-MRA individuals on MRA tasks, and potentially lead to enhanced performance in both the anatomical science and the STEMM disciplines.

MATERIALS AND METHODS

Participants

Students at The University of Western Ontario with normal, or corrected to normal vision by way of contact lenses, were invited to participate in this exploratory study (7 males and 16 females), under approval from the institution's Research Ethics Board. Individuals with EMRT scores exceeding one standard deviation above the mean were considered to be high MRA, and those with EMRT scores less than one standard deviation below the mean were considered to be low MRA. All other individuals who demonstrated scores within one standard deviation of the mean in either direction were considered to have intermediate MRA. This approach was adopted, rather than a median split, to exacerbate the distinction between high- and low-MRA individuals (Kozhevnikov et al., 2007). That is, individuals of high- and low-MRA are separated by a degree of two standard deviations of MRA score.

Experimental Design

Participants completed the EMRT while monocular gaze was monitored. Measurements of gaze were obtained from movements of the right eye, collected at a rate of 1,000 Hz using EyeLink 1000 eye-tracking equipment (SR Research Ltd., Mississauga, Ontario, Canada). On a question-by-question basis, eye movement metrics consisted of average fixation duration (AFD), number of fixations, and the region of highest salience (Table 1). Additionally, the eye-tracking equipment also collected the average question-response time per participant to supplement analysis. Target images were viewed from a distance of 40 cm, so that each figure subtended approximately 10° of visual angle, and the center-to-center distance between the two figures subtended approximately 15°. Ambient light conditions were kept constant in the testing room at all times.

Target Images

The target images presented to the participants constituted an EMRT. This visual test requires participants to view two three-dimensional (3D) block figures (a "block pair"), and indicate if the pair was the same, or different (Fig. 1) by responding using two keys on the keyboard as quickly, and

Table 1.

Definition of Eye Movement Measurements

Measurement	Definition
Average fixation duration	The mean length, in milliseconds, of a fixation performed by an individual. This value is calculated for each question presented, for each individual participant.
Question response time	The time required in milliseconds (from the onset of image presentation, to button-press) for the participant to respond to the question.
Number of fixations	The number of fixations completed by a participant during the course of a single question.
Region of highest salience	The region on the presented image that was attended to with the greatest frequency and duration. This is represented by Gaussian distributions that are scaled by duration, where multiple fixations are summed (frequency).

accurately as possible. A button-press of "1" indicated a "same" pair, while a "2" indicated a "different" pair.

The design and execution of the EMRT is based on the original line drawings of Shepard and Metzler, used to test MRA (Shepard and Metzler, 1971). Unlike the original question battery used by Shepard and Metzler, which was composed of a large number of block pairs, and presented as a paper and pencil test, the adapted EMRT consists of 16 unique block pairs that were each presented three times throughout the course of the test with the presentation order randomized for each participant. This adaptation yielded a total of 48 image presentations per participant. Both the original Shepard and Metzler question battery, and the EMRT held the same proportion of "same" and "different" questions, where 50% of questions were of the "same" condition and the other 50% were of the "different" condition.

The EMRT was selected for this study, over other tests of mental rotations for its clarity and ease of use in the context of eye tracking. The observational task requires a comparison of only two images making analysis according to region of salience more feasible, unlike the case of the Vandenberg and Kuse MRT, which requires the comparison of a target image and four possible answers. Within the original 16 unique images, the angular disparity between each block pair was varied across these two unique 3D objects. Angular disparity was increased in 20° clockwise increments, from 20° to 80°. Participants' time performing the test battery was recorded, but no time limit was applied to ensure that each participant was exposed to the full battery of EMRT questions.

The use of eye tracking enabled the quantification of individuals' gaze locations during the presentation of each of the 16 block pairs. Fixation maps for the right and left block for each image were created for each participant using a Gaussian distribution to represent visual acuity (Lee et al., 2011). The magnitude of each resulting Gaussian was scaled by the fixation duration resulting in a fixation map for each trial that represented both the spatial distribution of fixations and the relative durations. Each trial fixation map was then normalized to the magnitude of the Region of Highest Salience (the peak representing the combination of both spatial

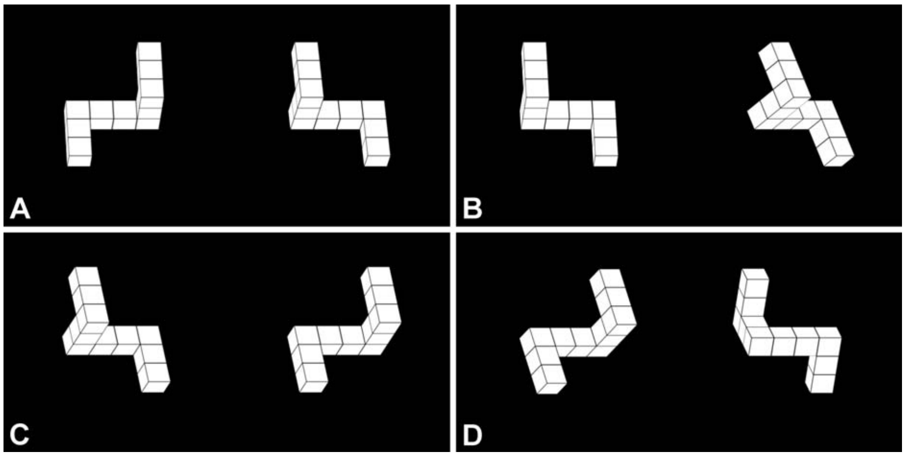

Figure 1.

Four sample questions based on the Shepard and Metzler block pairs used in the electronic mental rotations test (Shepard and Metzler, 1971). Each pair represents one question. Participants used a keyboard to indicate if shapes were the same or different. Answer key: images A and C are different from each other; images B and D are identical to each other.

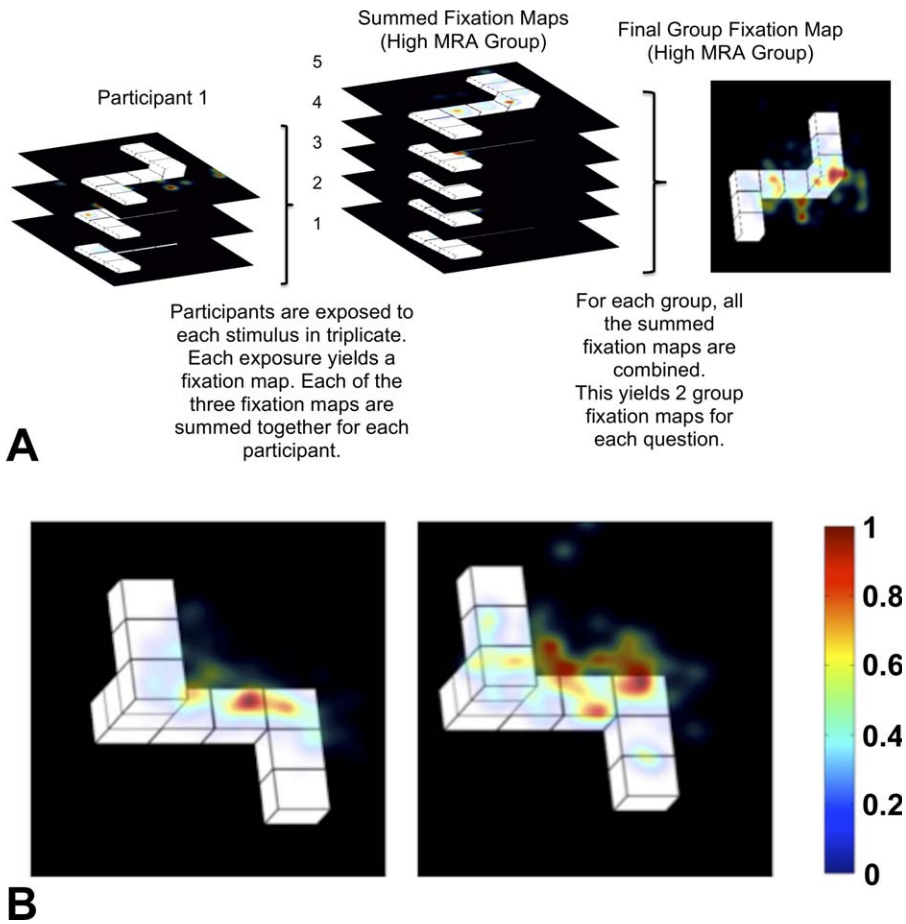

Figure 2.

A schematic representing the creation of salience maps. (A) A representation of the protocol employed to interpret spatial data employing the summation of individual fixation maps to the development of a group level heat map. (B) The dichotomy between high- and low-mental rotation ability (MRA) group heat maps indicative of the position of highest salience. High MRA, on the left, fixate predominantly on one location and on little else, while low MRA on the right attend to several points of interest. The color bar indicates the salience of the region, where red (1) represents areas of highest salience, and blue (0) represents areas of lowest salience in arbitrary units.

attention and duration) for that trial. Normalized maps for each image were combined to produce overall visual acuity maps for high- and low-spatial groups (Fig. 2B).

The region of highest saliency occurring on each heat map was then compared between groups according to location. Six location-based categories for the region of highest saliency were created *ad hoc* based on the features of the blocks (Fig. 3). This identification system served to enable the classification of which regions of the blocks drew the most attention, or spatial salience, from the participants during the problem-solving process. The areas colored red represent the most attended region of the image, and are indicative of the highest salience across the group.

Data Analysis

As eye tracking yields eye movement metrics in the form of both gaze time and location, the data analysis is separated accordingly.

Temporal Analysis. The collection of eye movement metrics facilitated a correlational analysis of MRA score with average fixation duration, question response time, and number of fixations per question. Additional comparisons in terms of response times for correct and incorrect answers for all participants were also conducted by way of the paired Student *t*-test.

Analysis of Salience. The location-based classification of highest saliency facilitated a between group comparison across each category by frequency using the nonparametric Fisher Exact test. This test was employed as the Fisher Exact Test is robust to smaller sample sizes, and is specific to categorical data (Fisher, 1922; McDonald, 2014), such as the locations employed for our salience metrics. An additional comparison of question-by-question agreement was then conducted using Cohen's kappa to determine how often the two groups attended to the same location on a given question (McHugh, 2012).

For all analysis, a significance value of less than $P = 0.05$ was considered to be statistically significant.

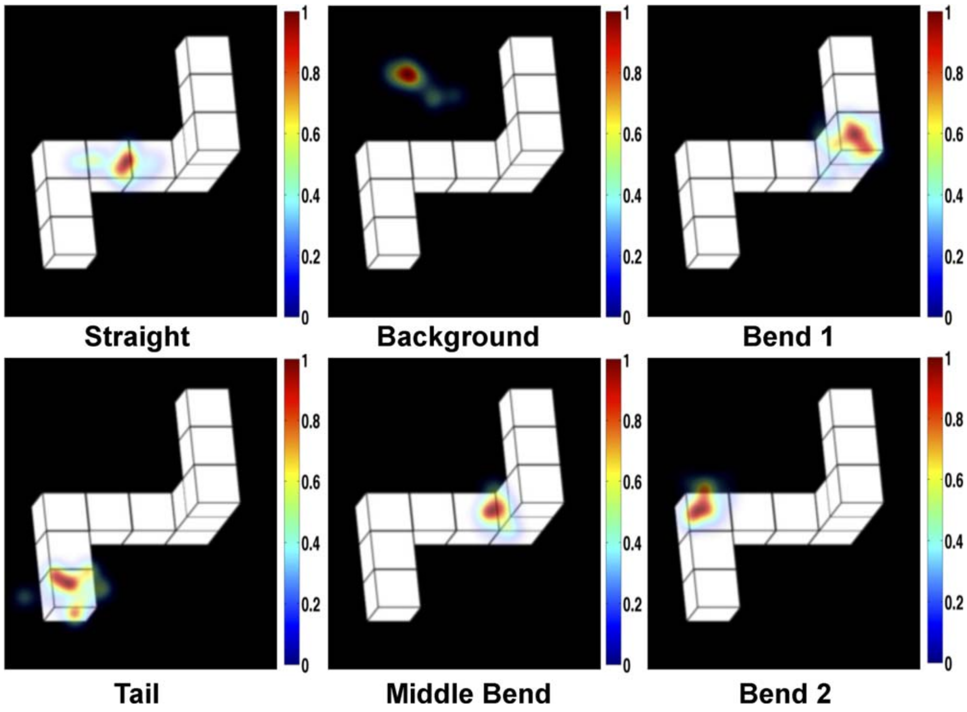

Figure 3.

A representation of the six categorizations for the location of highest spatial salience; scales to right of diagrams are arbitrary units.

RESULTS

Mental Rotation Ability as Defined by the Electronic Mental Rotations Test (EMRT)

Twenty-three individuals participated in the study. The mean age of participants was 25 ± 5 years (male 26 ± 6 and female 25 ± 5 years). Participants with EMRT scores exceeding one standard deviation above the mean (scores in excess of 44/48) were classified as high MRA (M:F is 1:4), and those with EMRT scores less than one standard deviation below the mean (scores of 34/48 or less) were classified as low MRA (M:F is 2:3). All other individuals who demonstrated scores within one standard deviation of the mean in either direction were classified as intermediate MRA (M:F is 4:9) (Kozhevnikov et al., 2007).

The Electronic Mental Rotations Test and the Inclusion of the Low Mental Rotation Ability Group

In studies of performance, often only the individuals of the highest performance ability are studied, and used as exemplars for the behavior (Just and Carpenter, 1985).

However, as the goal of this study, and of many other studies of MRA (Just and Carpenter, 1985; Gages, 1994; Geiser et al., 2006; Kozhevnikov et al., 2007; Nguyen et al., 2012; Wanzel et al., 2002a,b), was to discern how high- and low-MRA individuals differ behaviorally and how it may affect performance, it was prudent to include this low-ability group of individuals.

This methodology has received scrutiny, as the EMRT is considered a "two-alternative forced-choice" (2-AFC) test, in which the participant must make a selection of "same" or "different" when presented with a question (Fechner, 1966/1966). On 2-AFC tests, a score of <50% is indicative of a failure to complete the test, as the score is no better than that incurred by guessing, or by chance (Ulrich and Miller, 2004). As some of the participants populating the low-MRA group demonstrated scores approaching 50%, further investigation was conducted to ensure that the group performance was indeed different from that expected by chance. That is, evidence was required to ensure that the low-MRA group was relying on their limited ability to reason spatially, rather than "guessing" on each question. A Binomial test (Howell, 2007) was performed to reach this end, and it was found that the individuals of the low-MRA group were performing higher than that expected by chance; low group (0.65) was higher than that expected by chance (0.5), $P = 0.03$ (1-sided). This finding confirmed that the group was not guessing as they completed the test, and reaffirmed our inclusion of the data derived from the low-MRA group. This finding was critical to this study as these individuals show significant shortcomings during the completion of these spatial tasks; shortcomings which could be further exemplified through additional experimentation. If additional differences can be observed between high and how individuals, these differences may be capitalized upon to develop a guided approach to spatial problem-solving for the low-MRA individuals.

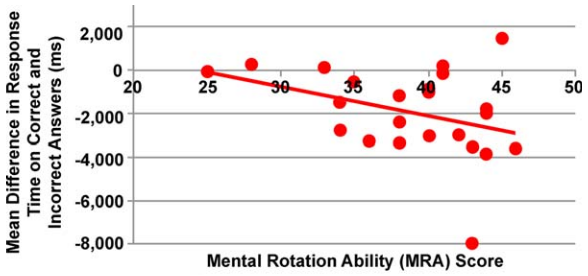

Figure 4.

The relationship between mental rotation ability (MRA) score and the mean difference of response time for questions answered correctly and incorrectly. High-MRA individuals dedicate more time to incorrect answers, and less time to correct answers than low-MRA individuals do, thus creating a larger mean difference.

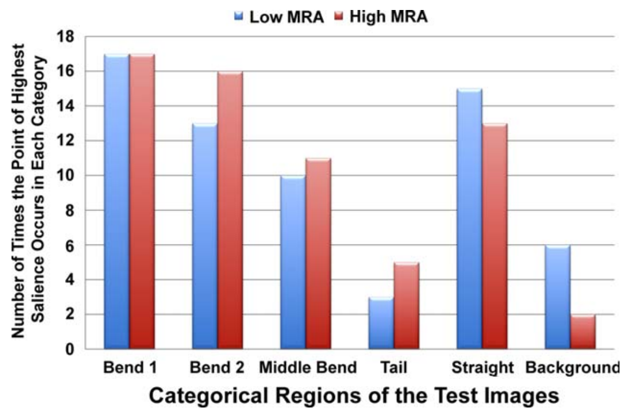

Figure 5.

The frequency distribution of highest saliency by region, for both high- and low-mental rotation ability (MRA) groups. No significant difference between the two groups was observed, suggesting that the groups attend to the regions in the same proportion.

Temporal Measurements

To better understand the relationship between temporal eye movements during the completion of the EMRT, this study first conducted an investigation to discern if differences existed for these variables (average fixation duration, question-response time, and number of fixations) based on whether the question was answered correctly or incorrectly. This was achieved through a paired Student's *t*-test, in which each participant's correct and incorrect mean values were contrasted. No significant differences were observed for the incorrect and correct answers for the measures of AFD and number of fixations; but there were differences observed for response time, which aligns with the findings of MRA and response time (Fig. 4).

Additional analysis with regard to MRA score and the duration of time dedicated to question solving were conducted. A Pearson correlation was employed to elucidate this relationship, ($r = -0.35$, $n = 23$, $P = 0.044$). Individuals of high MRA showed a greater mean difference between response times dedicated to correct and incorrect answers (Fig. 4). This finding suggests that during the process of solving a given question, individuals of high MRA will dedicate more time to solving a question they perceive to be challenging than a low-MRA individual would.

Additionally, an investigation into the hypothesized relationship between the temporal variables and MRA score was conducted by way of a correlational analysis. A Pearson correlation showed no significant relationship between AFD ($r = 0.16$, $n = 23$, $P = 0.457$), number of fixations ($r = 0.17$, $n = 23$, $P = 0.445$), and response time ($r = 0.26$, $n = 23$, $P = 0.228$) with MRA.

An additional Pearson correlation was employed using within-participants standard deviations to elucidate any intraparticipant differences that exist between the temporal variables and MRA. This approach is commonly employed in physical task performance analyses to elucidate patterns of variability between groups, to demonstrate consistent performance on a given task, such as reaching or grasping (Khan and Franks, 2000; Khan et al., 2003; Heath, 2005). In this case, the within-participant analysis of individual variance was completed in effort to observe how consistent individuals of high and low MRA were (in terms of response time) as they completed all 48 questions. This was achieved through analysis of the intraparticipant standard deviations for each of the variables, for each participant. The within-participant standard deviations, when correlated with MRA, demonstrate a significant positive correlation between individual question-response time variability and MRA ($r = 0.49$, $n = 23$, $P = 0.018$). No other correlations were observed between the within-participant variation of the other two variables of interest and MRA scores (average fixation duration: $r = 0.09$, $n = 23$, $P = 0.699$) and number of fixations ($r = 0.32$, $n = 23$, $P = 0.128$).

Salience Measurements

In order to address the second aim of this study, to determine where individuals of high and low MRA attend on the images during spatial reasoning, only the five highest and five lowest scoring individuals' eye movement metrics were analyzed ($n = 10$). Each block pair question was subdivided into right- and left-side blocks and fixation maps were generated for comparison. The regions of greatest saliency were calculated based on the combined group fixation maps, and contrasted per question (Fig. 2A). The analysis indicated that the parts of the image with the highest salience occur in the same frequency for both groups, overall (Fig. 5).

However, when a question-by-question analysis was completed to establish the agreement between the two groups, it was observed that in 65% of questions, the region of highest salience was different ($K = 0.21$). Indeed, the two groups attend to the same region on a given question only 35% of the time, representing a poor agreement between the two groups. That is, the timing of when and where high- and low-MRA subjects attended differed significantly on a question-by-question basis.

DISCUSSION

This study correlated measurements of eye movements to MRAs in an effort to distinguish if gaze patterns are associated with successful completion of a mental rotation test. In previous studies, individuals with higher MRAs completed the original Shepard and Metzler test questions (Shepard and Metzler, 1971) faster, and with fewer errors than low-MRA individuals (Just and Carpenter, 1985); thus, it was hypothesized that MRA would be negatively correlated with average fixation duration, average response time, and number of fixations. Additionally, it was predicted that MRA would be positively correlated with a higher AFD in spatially salient regions of the block image pairs of the EMRT.

Through the observation and quantification of the eye movements of high- and low-MRA individuals, it was thought that a greater understanding of the processes that lend to success on spatial tasks could be revealed. The findings of this analysis are twofold as the first half pertains to temporal measurements (average fixation duration, question response time, and number of fixations), to distinguish how low- and high-MRA individuals differ temporally during the EMRT. The second half of analysis focuses on spatial information pertaining to both the duration of time and location of individuals dedicate to salient regions of the presented images on the EMRT.

Temporal Measurements

From a temporal perspective with this untimed test, there appears to be a considerable lack of distinction between the high- and low-MRA individuals with regard to overall time to complete the EMRT. Closer examination of eye movements reveals the relationship between average fixation duration and EMRT score ($r = 0.16$) and trial response time and EMRT score ($r = 0.26$). These findings suggest that, when unencumbered by a time limit, individuals of high MRA tend to spend more time in fixation, and spend more time in answering overall, than individuals of low MRA. The attention devoted by the high-MRA individuals to salient regions is not supported by weak relationship between the number of fixations performed, and EMRT score ($r = 0.17$). Given that fixation is related to cognitive processing(Just and Carpenter, 1976; Shepard and Cooper, 1986), results of this study suggest that individuals of high MRA tend to spend more time assessing spatially salient features of the blocks on average, than their low-MRA peers. It is hypothesized that the high-MRA individuals implement these features to assist in correctly identifying if the block pairs are the same or different. However, the lack of significant correlational relationships between the time-related measures and MRA score encouraged a subsequent within participants comparison, to elucidate patterns in variability that are specific to high- and low-MRA individuals.

The within-participant analysis of variability for each of the time-related measures, coupled with the mean difference analysis of response time for correct and incorrect answers, was more descriptive in establishing a dichotomy between high- and low-MRA individuals. Despite the observation of very little relationship between AFD variability and variability in the number of fixations, there was a significant relationship between response time variability and MRA score ($r = 0.49$, $n = 23$, $P = 0.018$), and between mean difference in response time for the ratio of correct/incorrect answers and MRA score ($r = -0.36$, $n = 23$, $P = 0.044$). This finding suggests that individuals of high MRA demonstrate much more variability in response time throughout the course of the EMRT while low-MRA individuals are more rigid in their response times, answering each question after approximately 5 sec regardless of the inherent visual properties of the question. This may be evidence of the phenomenon known as "learned helplessness" that is typically exhibited by low-ability individuals. Learned helplessness is a phenomenon, in which an individual establishes that the outcome associated with a response to a task is unpredictable, and becomes debilitated and unable to complete the task (Maier and Seligman, 1976). In this study, low-MRA individuals may have been confronted with questions they perceived as very challenging, perhaps overwhelming their visual working memory, and rushed to an answer, rather than taking the required time required to solve it(Mayer and Sims, 1994)

Although indirect, these observations may indicate an increased working memory capacity for individuals possessing higher MRA (Huk, 2006). Much of the literature on performance and training suggests that with increased proficiency comes reduced variability, and improved consistency (Ericsson, 2004). The current data suggest that the consistency is based on correct answers and not the process undertaken to arrive at the correct answer, and that additional factors are at play when high-MRA individuals completed this test. That is, greater variability observed in high-MRA individuals may relate to increased flexibility in underlying cognitive processing linked to increased working memory. For example, the observation of high variability of response times in the high-MRA group may be representative of enhanced conflict monitoring. The conflict hypothesis posits that monitoring of response conflict may serve as a signal that activates control mechanisms required to overcome conflict and perform effectively (Botvinick et al., 2001). The conflict hypothesis suggests that behavioral adjustments and the engagement of cognitive control follow exposure to a response conflict. Thus, differences in response time are attributable to the high level of conflict associated with "incongruent" (or "different") questions, that yield a greater recruitment of cognitive control and attention for the following question (Botvinick et al., 2001).The act of conflict monitoring, and subconsciously devoting more time to more challenging questions may be responsible for the dichotomy of performance between our two groups.

Salience Measurements

Secondary to the analysis of the time-related measurements associated with EMRT completion, this study also set forth to discern the relationship between MRA and regional apprehension patterns during test completion. The analysis of the regions of visual salience provided perspective into apprehension approaches typical to both high- and low-MRA individuals. The results demonstrated that across the entire EMRT, both groups attend to features of the blocks at approximately the same frequency. However, as this measure only refers to the overall distribution of the regions of highest salience, little information can be garnered as to how the two groups behave on a question-by-question basis. Individuals with higher spatial ability may demonstrate different visual search patterns compared to lower spatially able individuals (Wheeler and Treisman, 2002). This may also be the direct result of limitations of working memory in the low MRA group (Huk, 2006); which could influence a more dispersed, less focal, searching of the images for comparison, due to their reduced ability to hold an exemplar image in one's working memory during the process of spatial reasoning (Baddeley, 2003) (Fig. 2B).

The difference in approach between the two groups is further illustrated through the application of the Cohen's kappa coefficient to evaluate the agreement between them. Through this analysis, it was possible to observe the likelihood of the two groups attending to the same location on a given question (McHugh, 2012) was quite low. High- and low-MRA individuals attended to the same location on a given question only 33% of the time. This dichotomy of visual apprehension between high and low groups is mirrored in the work of other who contrasted the visual search patterns of novice and expert laparoscopists (Wilson et al. 2011) and experience with images in anatomy students (Zumwalt et al., 2015). In these paradigms, expert laparoscopists directed their gaze to very specific regions of a visual familiar surgical scene, while novices directed their gaze nonspecifically over a broad range of visual areas without apparent direction or focus (Wilson et al., 2010). In the student population, as familiarity grew students attended to "cognitively salient regions" with more fixations and longer observation times overall.

If individuals of high and low MRA approach or "view" identical images in different ways, and reach different conclusions, then eye movement data-driven approaches and gaze-directed instructional methods may present an opportunity for education (Wilson et al., 2011; Vine et al., 2012). In domains of high spatial complexity such as anatomy, informing low-MRA populations where and when to "look" during task completion could serve to improve their spatial reasoning and potentially improve task performance overall (Neider et al., 2010; Wilson et al., 2011). The data derived from this study also lends indirect support to Vorstenbosch's suggestions that using images on anatomical examination changes the item difficulty and may jeopardize the validity of the assessment itself (Vorstenbosch et al., 2013). Implications of this study suggest persons with widely differing spatial ability approach spatially challenging questions quite differently. Furthermore, high- and low-MRA participants shared common approaches only 35% of the time further indicating differing strategies linked to spatial ability that significantly affects performance on an anatomical task (Nguyen et al., 2014). Whether a strategy is related purely to sensory input, that is gaze alone, or other factors potentially related to memory, is yet to be determined, but this study suggests gaze to be a significant contributing factor.

Limitations

Unlike other research evaluating MRA, this study did not limit the duration of time that individuals were permitted to complete the EMRT. This augmentation was implemented in an effort to ensure that all participants gained exposure to all the image pair stimuli. This decision may have served to limit this study as the speed of problem solving may be a factor predicting success on tests of mental rotations (Delgado and Prieto, 1996; Nguyen et al., 2014; Resnick, 1993). Without the pressure of a temporal "cut-off," participants of all levels of MRA may have spent a greater duration of time deciphering the image pairs, and "double-checking" their choice prior to answering. In fact, this study observed that without a cut-off time, participants spent up to 20 times longer per question than that typically observed in timed tests of mental rotations.

The paradigm employed image replication to decrease variability in the temporal eye metrics. Each image pair was presented in triplicate, a decline in duration required to solve the question on each subsequent presentation may have occurred. Thus, individuals may have been reliant on recollections of previous answers, rather than on active spatial reasoning to solve the problems. Theoretically, this could thus yield shorter response times for each subsequent viewing, and hearken more to the participants working memory capacity than their abilities.

Finally, this study may have been limited by sample size. As the analysis of salience was conducted on data derived from a subset of the overall sample, there is a possibility that greater differences may have been observed if a larger sample size was examined.

Future studies should seek to examine if similar patterns exist in time-sensitive environments that are more reflective of traditional MRTs and typical assessment in anatomy and the STEMM disciplines. The application of a time-limitation may serve to exacerbate the dichotomy between the high and low MRA individuals, compelling participants to rely on their innate cognitive abilities relating to speeded rotation, rather than on potential strategy. Such a modification would likely yield lower average MRA scores, and a reduction of the positive kurtosis noted in the scores of this study (Caissie et al., 2009).

Future Directions

The current approach explores a previously unaddressed participant-centered, eye movement-based, approach to analyzing spatial test completion. The implications of future research along this trajectory may inform eye movement guided strategies for the instruction of spatially relevant information (Wilson et al., 2011), and possibly extend to spatially complex disciplines including, but not limited to anatomical sciences, surgical skill training, and other science, technology, engineering, medical, and mathematical (STEMM) disciplines.

CONCLUSIONS

The findings of this study suggest further analysis under the constraint of a time limitation and perhaps with a greater number of visual elements, to better understand the role that eye movements play during spatial reasoning. Additionally, as the current work delves into the underlying mechanisms that govern spatial reasoning, future work aims to better illustrate the complex cognitive processes, such as conflict monitoring, that underpin the innate aptitudes for success in mental rotations. If additional differences can be observed between high and how individuals, these differences may be capitalized upon to develop a guided approach to spatial problem solving for the low-MRA individuals.

ACKNOWLEDGMENTS

The authors would like to thank and acknowledge the assistance of the Dr Michael Peters, Ph.D., for the provision of the MRT test images; Drs Ngan Nguyen, Roy Eagleson, and Sandrine de Ribeaupierre for their consultation and assistance during study design; as well as the graduate students of the Mitchell's Laboratory for their support, patience, and endless trouble-shooting assistance.

NOTES ON CONTRIBUTORS

VICTORIA A. ROACH, M.Sc. (Clin Anat), is a doctoral (Ph.D.) candidate in the CRIPT Laboratory, Department of Anatomy and Cell Biology at the University of Western Ontario, London, Ontario, Canada. She is investigating the patterns of visual apprehension that are associated with complex spatial reasoning.

GRAHAM FRASER, Ph.D., is an adjunct assistant professor in the Department of Medical Biophysics at Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada. He teaches biophysics, problem solving, and gross anatomy and his research interest resides in microvascular transport and regulation.

JAMES KRYKLYWY, M.Sc., is a doctoral candidate in the Department of Psychiatry, at the Brain and Mind Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada. He investigates how emotions affect perception and attention using MRI, eye-tracking, and audio simulated environments.

DEREK G.V. MITCHELL, Ph.D., is an assistant professor in the Departments of Psychiatry, and Anatomy and Cell Biology and a member of the Brain and Mind Institute at Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada. His research is principally aimed at determining how dissociable neural systems integrate emotion with cognition and behavior. His research also aims to determine how the human brain bias attention toward the most behaviorally relevant information.

TIMOTHY D. WILSON, Ph.D., is an associate professor in the Department of Anatomy and Cell Biology at Western University, London, Ontario, Canada. He directs and founded the Corps for Research of Instructional and Perceptual Technologies (CRIPT) laboratory investigating digital anatomy development, deployment, and efficacy in pedagogy.

LITERATURE CITED

Baddeley A. 2003. Working memory: Looking back and looking forward. Nat Rev Neurosci 4:829–839.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. Psychol Rev 108:624–652.

Brandt MG, Davies ET. 2006. Visual-spatial ability, learning modality and surgical knot tying. Can J Surg 49:412–416.

Buswell GT. 1935. How People Look at Pictures: A Study of the Psychology of Perception in Art. 1st Ed. Chicago, IL: University of Chicago Press. 198 p.

Caissie AF, Vigneau F, Bors DA. 2009. What does the mental rotation test measure? An analysis of item difficulty and item characteristics. Open Psychol J 2:94–102.

Carpenter RHS. 1988. Movements of the Eyes. 2nd Ed. London, UK: Pion Ltd. 593 p.

Carroll JB. 1993. Human Cognitive Abilities: A Survey of Factor-Analytic Studies. 1st Ed. Cambridge, UK: Cambridge University Press. 819 p.

Delgado AR, Prieto G. 1996. Sex differences in visuospatial ability: Do performance factors play such an important role? Mem Cognit 24:504–510.

Ericsson KA. 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 79:S70–S81.

Fechner GT. 1860/1966. Elements of psychophysic. Volume1. In: Howes DH, Boring EG (Editors). Elements of Psychophysic. 1st Ed. New York, NY: Holt, Rinehart, and Winston Inc. 286 p.

Fisher RA. 1922. On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. J Roy Stat Soc 85:87–94.

Gages TT. 1994. The Interrelationship Among Spatial Ability, Strategy Used, and Learning Style for Visualization Problems. 1st Ed. Columbus, OH: Ohio State University 302 p.

Garg AX, Norman G, Sperotable L. 2001. How medical students learn spatial anatomy. Lancet 357:363–364.

Geiser C, Lehmann W, Eid M. 2006. Separating "rotators" from "nonrotators" in the mental rotations test: A multigroup latent class analysis. Multivariate Behav Res 41:261–293.

Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA. 2003. Validation of novel and objective measures of microsurgical skill: Hand-motion analysis and stereoscopic visual acuity. Microsurgery 23: 317–322.

Hayhoe M, Ballard D. 2005. Eye movements in natural behavior. Trends Cognit Sci 9:188–194.

Heath M. 2005. Role of limb and target vision in the online control of memory-guided reaches. Mot Contr 9:281–311.

Howell DC. 2007. Statistical Methods for Psychology. 6th Ed. Belmont, CA: Wadsworth Publishing. 739 p.

Huk T. 2006. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn 22:392–404.

Itti L, Koch C, Niebur E. 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259.

Just MA, Carpenter PA. 1976. Eye fixations and cognitive processes. Cogn Psychol 8:441–480.

Just MA, Carpenter PA. 1985. Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychol Rev 92:137–172.

Khan MA, Franks IM. 2000. The effect of practice on component submovements is dependent on the availability of visual feedback. J Mot Behav 32: 227–240.

Khan MA, Lawrence G, Fourkas A, Franks IM, Elliott D, Pembroke S. 2003. Online versus offline processing of visual feedback in the control of movement amplitude. Acta Psychol (Amst) 113:83–97.

Kowler E. 2011. Eye movements: The past 25years. Vision Res 51:1457–1483.

Kozhevnikov M, Motes MA, Hegarty M. 2007. Spatial visualization in physics problem solving. Cognit Sci 31:549–579.

Lai ML, Tsai MJ, Yang FY, Hsu CY, Liu TC, Lee SWY, Lee MH, Chiou GL, Liang JC, Tsai CC. 2013. A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ Res Rev 10:90–115.

Land LM. 2004. Eye movements in daily life. In: Chalupa LM, Werner JS (Editors). The Visual Neurosciences. 2 Volume Set. 1st Ed. Cambridge, MA: MIT Press. p 1357–1368.

Lee WF, Huang TH, Yeh SL, Chen HH. 2011. Learning-based prediction of visual attention for video signals. IEEE Trans Image Process 20:3028–3038.

Leek EC, Johnston SJ, Atherton CJ, Thacker NA, Jackson A. 2004. Functional Specialisation in Human Premotor Cortex: Visuo-Spatial Transformation in Pre-SMA During 2D Image Transformation. Tina Memo No. 2004-015. 1st Ed. Manchester, UK: University of Manchester. 37 p. URL: http://tina.wiau.man.ac.uk/tina-knoppix/tina-memo/2004-015.pdf [accessed 13 September 2015].

Linn MC, Petersen AC. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Dev 56:1479–1498.

Lufler RS, Zumwalt AC, Romney CA, Hoagland TM. 2012. Effect of visual–spatial ability on medical students' performance in a gross anatomy course. Anat Sci Educ 5:3–9.

Maier SF, Seligman ME. 1976. Learned helplessness: Theory and evidence. J Exp Psychol Gen 105:3–46.

Mayer RE, Sims VK. 1994. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. J Educ Psychol 86: 389–401.

McDonald JH. 2014. Fisher's exact test of independence. In: McDonald JH. Handbook of Biological Statistics. 3rd Ed. Baltimore, MD: Sparky House Publishing. p 77 –85.

McGee MG. 1979. Human Spatial Abilities: Sources of Sex Differences. 1st Ed. New York, NY: Praeger Publishers Inc. 144 p.

McHugh ML. 2012. Interrater reliability: The kappa statistic. Biochem Med 22:276–282.

Neider MB, Chen X, Dickinson CA, Brennan SE, Zelinsky GJ. 2010. Coordinating spatial referencing using shared gaze. Psychonomic Bull Rev 17:718–724.

Nguyen N, Mulla A, Nelson AJ, Wilson TD. 2014. Visuospatial anatomy comprehension: The role of spatial visualization ability and problem-solving strategies. Anat Sci Educ 7:280–288.

Nguyen N, Nelson AJ, Wilson TD. 2012. Computer visualizations: Factors that influence spatial anatomy comprehension. Anat Sci Educ 5:98–108.

Peters M, Chisholm P, Laeng B. 1995. Spatial ability, student gender, and academic performance. J Eng Educ 84:69–73.

Resnick SM. 1993. Sex differences in mental rotations: An effect of time limits? Brain Cognit 21:71–79.

Rochford K. 1985. Spatial learning disabilities and underachievement among university anatomy students. Med Educ 19:13–26.

Shea DL, Lubinski D, Benbow CP. 2001. Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. J Educ Psychol 93:604–614.

Shepard RN, Cooper LA. 1986. Mental Images and Their Transformations. 1st Ed. Cambridge, MA: The MIT Press. 376 p.

Shepard RN, Metzler J. 1971. Mental rotation of three-dimensional objects. Science 171:701–703.

Thurstone LL. 1938. Primary Mental Abilities: Psychometric Monographs, No. 1. 1st Ed. Chicago, IL: University of Chicago Press. 121 p.

Triesch J, Ballard DH, Hayhoe MM, Sullivan BT. 2003. What you see is what you need. J Vis 3:86–94.

Ulrich R, Miller J. 2004. Threshold estimation in two-alternative forced-choice (2AFC) tasks: The Spearman–Kärber method. Percept Psychophys 66:517–533.

Van Herzeele I, O'Donoghue KG, Aggarwal R, Vermassen F, Darzi A, Cheshire NJ. 2010. Visuospatial and psychomotor aptitude predicts endovascular performance of inexperienced individuals on a virtual reality simulator. J Vasc Surg 51:1035–1042.

Vandenberg SG, Kuse AR. 1978. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 47:599–604.

Vine SJ, Masters RS, McGrath JS, Bright E, Wilson MR. 2012. Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. Surgery 152:32–40.

Vorstenbosch MA, Klaassen TP, Kooloos JG, Bolhuis SM, Laan RF. 2013. Do images influence assessment in anatomy? Exploring the effect of images on item difficulty and item discrimination. Anat Sci Educ 6:29–41.

Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED, Cusimano MD. 2002a. Effect of visual-spatial ability on learning of spatially-complex surgical skills. Lancet 359:230–231.

Wanzel KR, Matsumoto ED, Hamstra SJ, Anastakis DJ. 2002b. Teaching technical skills: Training on a simple, inexpensive, and portable model. Plast Reconstr Surg 109:258–263.

Wheeler ME, Treisman AM. 2002. Binding in short-term visual memory. J Exp Psychol Gen 131:48–64.

Wilson M, McGrath J, Vine S, Brewer J, Defriend D, Masters R. 2010. Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. Surg Endosc 24:2458–2464.

Wilson MR, Vine SJ, Bright E, Masters RS, Defriend D, McGrath JS. 2011. Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study. Surg Endosc 25:3731–3739.

Yarbus AL. 1967. Eye movements during perception of complex objects. In: Yarbus AL. 1967. Eye Movements and Vision. 1st Ed. New York, NY: Plenum Press. p 171–211.

Zumwalt AC, Iyer A, Ghebremichael A, Frustace BS, Flannery S. 2015. Gaze patterns of gross anatomy students change with classroom learning. Anat Sci Educ 8:230–241.