Time Limits in Testing: An Analysis of Eye Movements and Visual Attention in Spatial Problem Solving

Victoria A. Roach,[1] **Graham M. Fraser**,[2] **James H. Kryklywy**,[3] **Derek G.V. Mitchell**,[4,5] **Timothy D. Wilson**[5*]

[1]*Department of Foundational Medical Studies, William Beaumont School of Medicine, Oakland University, Rochester, Michigan*
[2]*Cardiovascular Research Group, Division of BioMedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada*
[3]*Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada*
[4]*Department of Psychiatry, Brain and Mind Institute, Schulich School of Medicine and Dentistry, The University of Western Ontario, London, Ontario, Canada*
[5]*Department of Anatomy and Cell Biology, Corps for Research of Instructional and Perceptual Technologies, Schulich School of Medicine and Dentistry, The University of Western Ontario, London, Ontario, Canada*

Individuals with an aptitude for interpreting spatial information (high mental rotation ability: HMRA) typically master anatomy with more ease, and more quickly, than those with low mental rotation ability (LMRA). This article explores how visual attention differs with time limits on spatial reasoning tests. Participants were assorted to two groups based on their mental rotation ability scores and their eye movements were collected during these tests. Analysis of salience during testing revealed similarities between MRA groups in untimed conditions but significant differences between the groups in the timed one. Question-by-question analyses demonstrate that HMRA individuals were more consistent across the two timing conditions ($\kappa = 0.25$), than the LMRA ($\kappa = 0.013$). It is clear that the groups respond to time limits differently and their apprehension of images during spatial problem solving differs significantly. Without time restrictions, salience analysis suggests LMRA individuals attended to similar aspects of the images as HMRA and their test scores rose concomitantly. Under timed conditions however, LMRA diverge from HMRA attention patterns, adopting inflexible approaches to visual search and attaining lower test scores. With this in mind, anatomical educators may wish to revisit some evaluations and teaching approaches in their own practice. Although examinations need to evaluate understanding of anatomical relationships, the addition of time limits may induce an unforeseen interaction of spatial reasoning and anatomical knowledge. Anat Sci Educ 00: 000–000. © 2017 American Association of Anatomists.

Key words: spatial ability; mental rotations ability; spatial reasoning; eye tracking; salience; STEMM; spatial working memory; gross anatomy education

INTRODUCTION

To interact with the three-dimensional world, humans must have an understanding of the positional relationships occurring between objects around them in space. Such an understanding is loosely referred to as one's "spatial ability." One component of spatial ability refers specifically to an individual's ability to interpret and comprehend objects as they are rotated. This individual capability is referred to as mental rotation ability (MRA). Specifically, MRA is the ability to maintain a mental representation of a two-dimensional or three-dimensional object while it is turned in space

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

Received 3 August 2016; Revised 13 March 2017; Accepted 13 March 2017.

Published online 00 Month 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ase.1695

Table 1.

A Summary of Key Terms and Accompanying Definitions

Term	Definition
Salience	A quality, or a set of characteristics held by a region of an image that is perceived by the viewer as being conspicuous relative to its surroundings. These regions "stand out" or command visual attention.
Fixation Density Map	A visual representation (map) illustrating the location of each fixation conducted by a given individual, on a given question of the EMRT, with no consideration of fixation duration.
Salience Map	A visual representation (map) illustrating the location of each fixation conducted by a single individual, on a single question of the EMRT, overlaid with a Gaussian distribution that has been scaled for fixation duration.
Salience Distribution	A visual representation (map) illustrating the location of each fixation, conducted by a given individual over all three exposures to a single question on the EMRT; overlaid with a Gaussian distribution that has been scaled for fixation duration.
Group Salience Distribution	A visual representation (map) illustrating the location of each fixation produced by a group of individuals, over all three exposures to a single question, overlaid with a Gaussian distribution that has been scaled for fixation duration.
Region of Highest Salience	The location on a "group salience distribution" that represents the area that has been fixated most frequently, and for the longest duration ("salience center").
Frequency Distribution	A representation of how often the "region of highest salience" occurs in each of the pre-defined areas: Bend 1, Bend 2, middle bend, tail, straight, or background.

EMRT, Electronic Mental Rotations Test.

(Spearman, 1927). Psychometricians have developed different measures to quantify MRA, including timed, standardized tests such Vandenberg and Kuse's mental rotations test (VKMRT) (Vandenberg and Kuse, 1978; Peters et al.,1995a,b), the card rotation test (Ekstrom et al., 1976), and more recently, the timed electronic mental rotations test (Timed EMRT) (Roach et al., 2016).

In recent years, MRA has emerged in medical and anatomical education research publications for its role as an "indicator of success" in domains requiring an understanding of spatially complex information and relationships, such as anatomy and surgery (Wanzel et al., 2002; Lufler et al., 2012; Nguyen et al., 2012; Lai et al., 2013). Literature suggests that MRA is directly linked to success in surgical skill acquisition in novices (Wanzel et al., 2002, 2003; Brandt and Davies, 2006) and in knowledge acquisition and comprehension in anatomy (Lufler et al., 2012; Nguyen et al., 2012). As MRA is related to success in the comprehension of spatially distributed information, more research is required to explain the cognitive processes that underlie this ability.

In an effort to better understand MRA, research has sought to explore other related neurologic and sensory factors. One correlational hypothesis suggests that an aptitude for mental rotation may be revealed by the movements of the eyes (Just and Carpenter, 1976). The theory suggests that as an individual looks at elements of spatial stimuli, the stages of problem solving are manifest in each fixation of the eye (the act of maintaining visual gaze on a single location) (Just and Carpenter, 1976; Carpenter, 1988). As such, patterns occurring in an individual's fixations may reveal the underlying cognitive processes responsible for driving gaze and processing spatial information (Just and Carpenter, 1976; Grant and Spivey, 2003; Thomas and Lleras, 2007; She and Chen, 2009).

In a foundational study, Just and Carpenter demonstrated significant differences in mean response latency and in the frequency of back-and-forth eye movements (comparisons between two presented images) between individuals of high mental rotation ability (HMRA) and low mental rotation ability (LMRA) as they performed an untimed test of MRA (Just and Carpenter, 1985). While the work of Just and Carpenter demonstrated differences in eye movements and in approaches to mental rotation between ability-based groups, the eye movement experiments by Roach et al. (2016, 2017) further confirmed and dissected the dichotomy between HMRA and LMRA. Through the use of specific performance measures (mean fixation duration, response latency and mean fixations per question) and an attentional salience measure, differences between the HMRA and LMRA groups were revealed (Roach et al., 2016, 2017). The "salience" measure (Table 1) quantified where groups directed their attention as they performed a spatial test, the EMRT.

While previous studies have demonstrated important salience differences between HMRA and LMRA in isolated timing conditions (timed and untimed exclusively), little is known regarding how the addition of a time restriction impacts an individual's performance on an MRA test. In Lohman's analysis and subdivision of spatial ability, MRA is renamed "speeded rotations" (Lohman, 1979; Shepard and Cooper, 1986; Lohman, 1988), with the implication that individuals who are able to mentally rotate structures effectively, do so markedly faster than those who are less proficient (Hegarty and Waller, 2005). Many accepted and validated tests of MRA have incorporated this assumption into their protocol. Stringent time limits on test completion are enforced on both the timed electronic mental

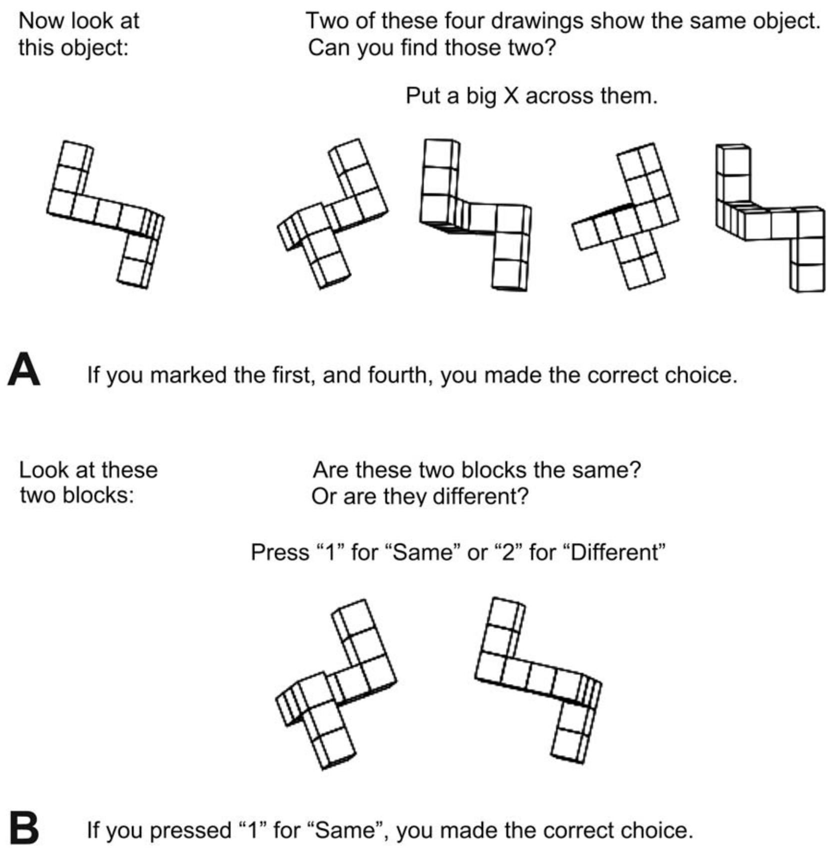

Figure 1.

Visual representations of different mental rotation tests. **A**) A typical Vandenberg and Kuse Mental Rotation Test (VKMRT)-style question (Vandenberg and Kuse, 1978) recreated using the digital block figure image bank redrawn by Peters et al. (1995a,b). Participants must identify if the block images are the same or mirrored images of each other. The VKMRT consists of 24 questions, with a maximum total time allowance of 6 minutes. **B**) A typical timed electronic mental rotation test (EMRT)-style question (Roach et al., 2016) recreated using the digital block figure image bank originally created by Peters et al. (1995a,b). Participants must judge if the two block figures are the same (rotations) or different (reflections) within a span of 6 seconds. The EMRT consists of 48 questions, and has a maximum time allowance of 5 minutes (Roach et al., 2016).

rotations test (Timed EMRT) (Roach et al., 2016) and the gold-standard of MRA evaluation, Vandenberg and Kuse's mental rotations test (VKMRT) (Vandenberg and Kuse, 1978; Peters, et al., 1995a,b) (Fig. 1). While fundamentally different in appearance, and in their execution, both of these MRA tests are considered to be "speeded" tests, as each are so temporally constrained that most test-takers will not have sufficient time to consider and answer every question (Bridgeman et al., 2004). By imposing such strict time limits, speeded tests demand that test-takers work at the highest rate possible, and serve to evaluate performance in terms of test-taker work rate (Lu and Sireci, 2007). Literature suggests that speeded tests favor high ability individuals, as low ability individuals often have difficulty assessing the difficulty of presented questions. As low ability individuals struggle to assess difficulty, they often fail to allocate their time appropriately to challenging questions (Schnipke, 1995). With this in mind, it is possible that the application of strict time limits could differentially impact attention, the processing of visual information, and subsequently performance in high and low ability individuals.

By tracking eye movements, the current study aims to reveal the relationships that exist between mental rotation ability, test time limits, and visual salience. It is predicted that both MRA groups will score lower on the spatial tests when a time limit is applied. Further, it is predicted that the two MRA groups will attend to different elements of the presented images while problem solving, and that these elements will differ according to each timing condition.

MATERIALS AND METHODS

Participants

This experiment received approval from the University of Western Ontario's Institutional Review Board (IRB). The IRB ensures that all research conducted on human subjects is conducted in accordance with all federal, institutional, and ethical guidelines.

Participants were volunteer graduate students in the allied health sciences and anatomy and cell biology with normal, or corrected to normal vision by way of contact lenses. Prior to testing, participants ($n = 20$; 14 female and 6 male) were classified according to their MRA, through the completion of a timed, electronic test of mental rotations ability, the Vandenberg and Kuse's mental rotations test (VKMRT)

(Vandenberg and Kuse, 1978; Peters and Battista, 2008) (Fig. 1A). The VKMRT was chosen because it displays high internal consistency (Kuder–Richardson 20 = 0.88) and test–retest reliability (0.83) (Vandenberg and Kuse, 1978). Individuals with VKMRT scores in excess of one standard deviation of the sample mean were considered to be HMRA ($n = 7$; 3 female and 4 male) while those with VKMRT scores less than one standard deviation below the sample mean were considered to be LMRA ($n = 7$; 6 female and 1 male). Individuals scoring within one standard deviation of the sample mean were classified as having intermediate MRA ($n = 6$; 5 female and 1 male) and were not included in this study. This approach was adopted to emphasize the distinction between HMRA and LMRA individuals (Kozhevnikov et al., 2007).

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

As this study sought to identify and distinguish attentional salience differences between HMRA and LMRA groups in timed and untimed test conditions, it was necessary to include lower performing individuals in the analysis. A Binomial Test (Howell, 2007) was employed to demonstrate that the included LMRA individuals performed poorly as a result of failed attempts at questions, and not because of guessing. The binomial test indicated that the proportion of correct answers obtained by the LMRA group on the timed condition (0.63) was higher than that expected by chance (0.5), $P < 0.05$ (1-sided). This calculation was also performed for the untimed condition (0.82) and similar results were obtained ($P < 0.001$; 1-sided). These findings affirmed the inclusion of the LMRA group and confirmed that their low scores were not the result of guessing, but due to the impaired ability to solve spatial problems.

Experimental Design

Individuals allocated to the HMRA or LMRA groups, based on their VKMRT score entered the next phase of experimentation. All participants completed two additional tests of MRA while their eye movements were collected. Participants completed both a timed and untimed version of the electronic mental rotations test (EMRT: a secondary, novel task of mental rotations ability, distinct from the initial VKMRT) (Fig. 1B), while monocular gaze was monitored from the right eye. The order of tests (timed vs. untimed) was randomized across participants. Eye movement metrics (fixation location (x,y), and the duration of each fixation) were collected using EyeLink 1,000 eye-tracking equipment (SR Research, Mississauga, Canada), with sampling occurring at 1,000 Hz. All test images were presented at a consistent viewing distance of 40 cm. The ambient light in the testing room was held constant throughout all testing events. All participants completed the timed and untimed tests in random order, without feedback on their performance, on the same day, over a twenty-minute period. A five-minute break was allotted between the first and second test iterations.

Target Images Used in the Electronic Mental Rotation Test

The target images presented to the participants in both the timed and untimed conditions constituted the electronic mental rotations test (EMRT) described initially in Roach et al. (2016, 2017). The EMRT required participants to view two three-dimensional (3D) block figures (a "block pair") and determine if the pair was the same (a rotation) or different (a nonsuperimposable reflection) as quickly and accurately as possible. Participants logged their responses using a keyboard. A button-press of "1" indicated a "same" pair, while a "2" indicated a "different" pair.

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. As the task requires a comparison of only two images, it facilitated a streamlined analysis of participants' fixation locations, and subsequent analysis of salience.

Quantifying Salience

Salience distributions were generated according to the methodology described by Roach et al. (2016, 2017). Roach's quantitative analysis of salience was conducted using a methodology similar to the technique described by Lee et al. (2011). Lee suggests that the magnitude and locus of an individual's attention is measured by fixation density, collected via eye tracking technology. Thus, fixation density maps were used to represent the salient regions of the images. In the current study, fixation density maps were generated throughout the data collection phase for each individual, within each group, for each image presented in each test, where each image held discrete fixation points $\left(\left(x_n^f,\ y_n^f\right),\ n=1,\ \ldots,\ N\right)$ and where N represented the total number of fixation points conducted by a participant on a single question, and where $\left(x_n^f,\ y_n^f\right)$ is the location of the nth fixation point (Lee et al., 2011). The fixation points for each image were interpolated using a Gaussian function, and scaled by fixation duration to yield a salience map $s(x,y)$,

$$s(w,l)= \sum_{n=1}^{N} t_n^f \cdot \exp\left(-\frac{\left(w-w_n^f\right)^2+\left(l-l_n^f\right)^2}{2\sigma_s^2}\right)$$

where x and y denote the horizontal, and vertical positions of the fixated pixel, σ_s represents the standard deviation of the Gaussian, and t_n^f is the fixation duration in milliseconds. The value of σ_s is calculated based on the accuracy of the eye tracking system using L, the viewing distance from the display, and d, the display pixel density:

$$\sigma_s=L \times \tan\frac{0.5\pi}{180} \times d$$

The salience map was, thus, obtained by summing the time weighted Gaussian function of all fixations for each question. The salience map allowed for discrete fixation pixels identified by the eye tracker to be represented as regions of fixation based on the visual accuracy of the eye tracking system (Lee et al, 2011). As a result, a pixel that resides in a heavily fixated region is more "salient" than one in a scarcely fixated area. The salience maps for right and left blocks from each question were then normalized by dividing by the maximum pixel value in the map. Salience distributions for each block were then created for each group by summing the normalized salience maps and normalizing again to the maximum pixel value (Roach et al., 2016, 2017). This salience, or group salience distribution, was then represented visually

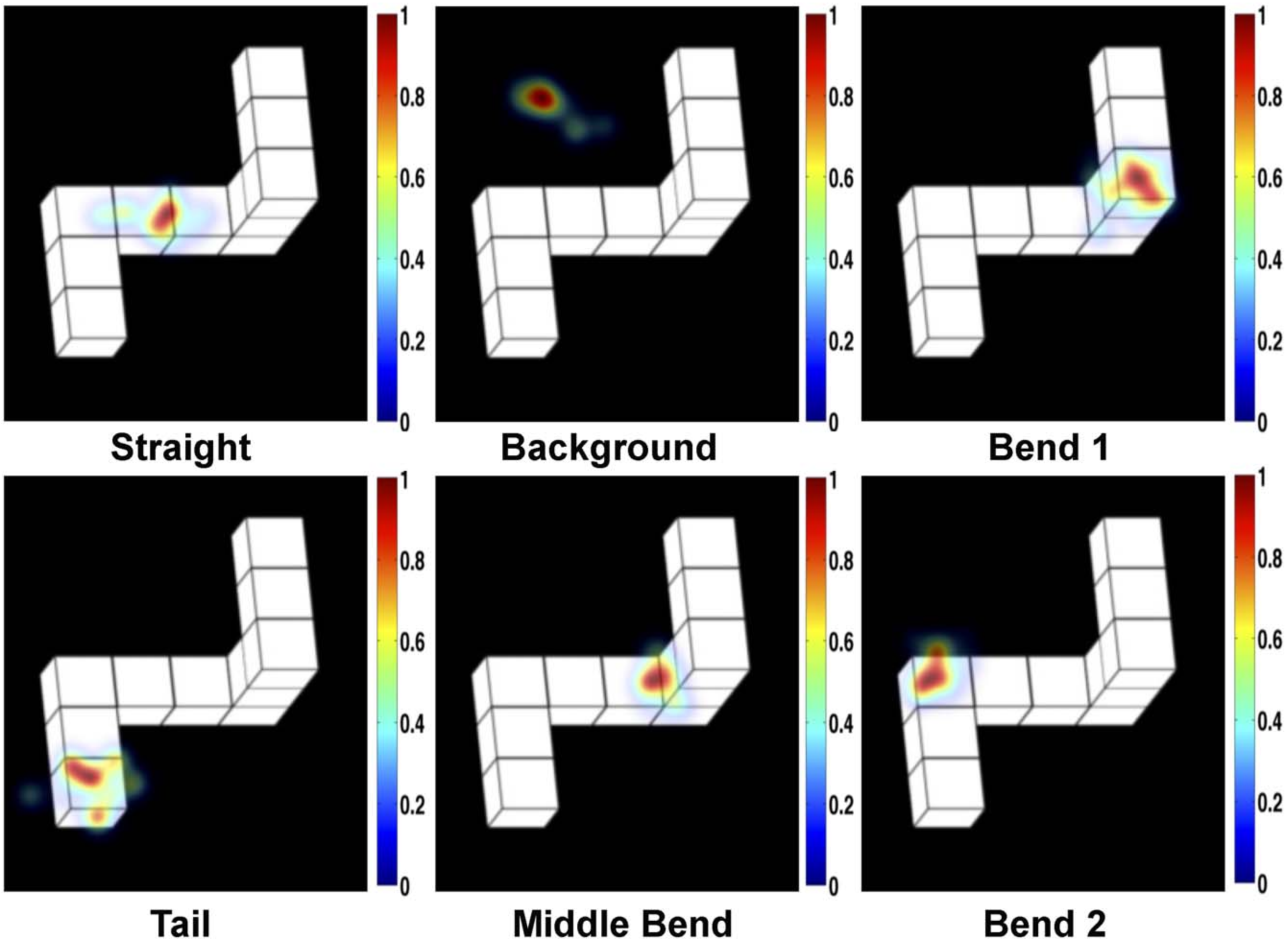

Figure 2.

Observers classified each region of highest salience according to its location on the presented block figure. This collection of images depicts exemplars for each categorical classification. In these exemplars, typical salience is illustrated according to color intensity, whereby areas approaching red indicate increasing salience. The regions of highest salience represent the locations that garner the most visual attention, for the longest duration. Reprinted from Roach et al. (2016) with permission of John Wiley & Sons on behalf of the American Association of Anatomists © 2015.

through the use of a color intensity scale. These data were then sorted by question (1 through 16), by timing condition (timed vs. untimed) and by MRA group (low vs. high).

These group salience maps thus represented where high and low individuals attended with the most frequency, and for the greatest duration, on each question, during the two testing conditions. The region of highest salience for the group means, denoted by the highest intensity of color, was then classified according to its location on the image. Six discrete classifications for the location of the region of highest salience existed: Bend 1, Bend 2, Middle bend, Tail, Straight, and Background. Exemplars of each categorization can be viewed in Figure 2. Three observers carried out the location-based classification of each region of highest salience (ICC: 0.84). A frequency distribution was then created for each timing condition, and MRA group, based on the number of times (or frequency) the region of highest salience occurred in each category.

The Fisher Exact test was used to distinguish differences in the distribution of salience across the six categories. Cohen's Kappa was used to identify differences in question-by-question agreement, between the groups, and within the groups, across timing conditions (Cohen, 1960). Cohen's Kappa, and question-by-question agreement is interpreted using the following classification structure (Table 2). This approach enabled a quantitative analysis of where these groups were looking on timed, and untimed testing conditions.

Data Analysis

All data analyses were completed using SPSS statistical software package (IBM Corp., Armonk, NY), and P-values less than 0.05 were considered statistically significant.

Performance analysis. Timed EMRT scores were compared with untimed EMRT scores to illustrate the role of timing on spatial reasoning performance. Analyses were conducted via a 2 (Timing) × 2 (MRA Group) Mixed ANOVA.

Analysis of salience distribution. The location-based classification of the regions of highest salience enabled comparisons between groups, between timing conditions, and across each region using the Fisher Exact test (Fisher, 1922; McDonald, 2014). This test is robust to small sample sizes and specific to categorical data (McDonald, 2014) such as salience. Comparisons of question-by-question agreement were conducted using Cohen's kappa (Cohen, 1960) to determine how often HMRA and LMRA groups attended to the same location on a given question, in each timing condition (McHugh, 2012).

RESULTS

Performance Effects of Time Limits

The VKMRT served as a diagnostic test to allocate individuals into two MRA groups (Vandenberg and Kuse, 1978; Peters et al., 1995a,b). In line with current literature, significant differences in score on the subsequent EMRT were observed between the two MRA groups' overall $F(1.12) = 14.41$ ($P = 0.003$) (Partial η^2: 0.546) (Roach et al., 2016). The average EMRT score (Mean ± SD) was 41.29 ± 4.65 in the HMRA and 34.71 ± 6.93 in the LMRA group, out of a possible maximum score of 48 (Fig. 3A).

As predicted, both groups demonstrated significantly higher EMRT scores in the untimed condition 42.71 ± 3.95 and 39.29 ± 4.61, than in the timed condition 39.86 ± 5.15 and 30.14 ± 5.84 for HMRA and LMRA respectively, $F(1.12) = 9.09$ ($P = 0.011$) (Partial η^2: 0.431; Fig. 3B). No interaction was observed between group and timing condition $F(1,12) = 2.49$ ($P = 0.14$) (Partial η^2: 0.172).

Distribution of Salience

An analysis of the salience distribution across the MRA groups under the two timing conditions demonstrated significant differences with respect to where the groups directed visual attention on the block diagrams during problem solving (Fisher Exact Test = 122.18, $P < 0.001$) (Fig. 4).

In untimed conditions, the HMRA and LMRA showed similar salience distributions on a question-by-question basis (Fisher Exact Test: 1.31, $P = 0.95$), with their regions of highest salience overlapping in 75% of questions ($\kappa = 0.67$; Fig. 4).

However, in the timed condition, the two group salience distributions diverged (Fisher Exact Test: 33.24, $P < 0.001$), as the question-by-question salience agreement decreased to 42%, representing a slight agreement between the groups ($\kappa = 0.20$; Fig. 4).

The HMRA group demonstrated a fair agreement on 38% of questions across the two timing conditions ($\kappa = 0.25$; Fig. 4). In contrast, the analysis of agreement for the LMRA group revealed that across the two timing conditions, the LMRA group showed agreement on only 13% of questions. According to literature, this value suggests that nearly no

Table 2.

The Interpretation of Kappa Values for Agreement

Value of Kappa	Level of Agreement
≤ 0	Less than chance agreement
0.01–0.20	Slight Agreement
0.21–0.40	Fair Agreement
0.41–0.60	Moderate Agreement
0.61–0.80	Substantial Agreement
0.81–0.99	Almost Perfect Agreement

Interpretation based on Viera and Garret (2005).

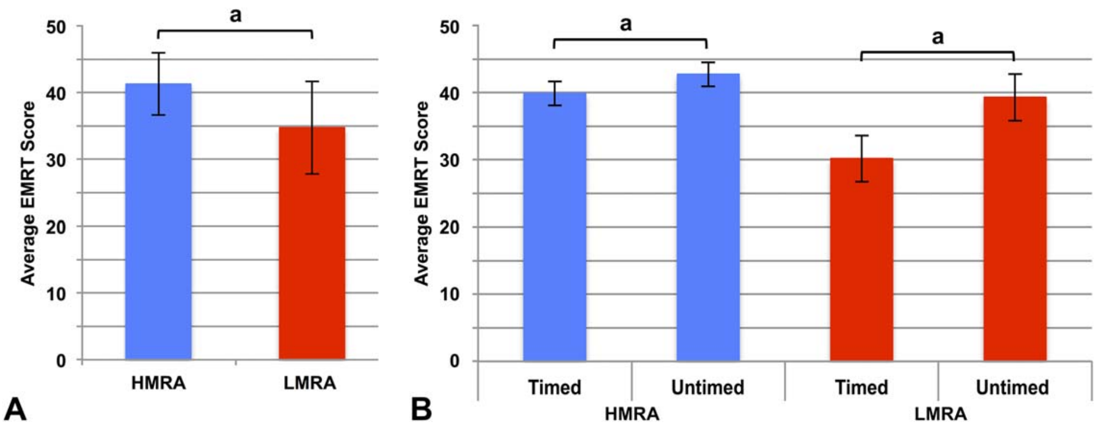

Figure 3.

Contrasts of performance on the electronic mental rotation test (EMRT). **A**) A between group comparison of performance (average EMRT score ± SD), regardless of timing condition. The high mental rotation ability (HMRA) group outperforms the low mental rotation ability (LMRA) group consistently and significantly $F(1.12) = 14.41$ ($P = 0.003$). **B**) The performance (average EMRT Score ± 95% Confidence Interval) of both MRA groups for both timing conditions. Both groups show lower EMRT scores in the timed condition, compared to the untimed condition. The presence of the lower case letter "a" indicates a significant difference ($P < 0.05$) between the groups.

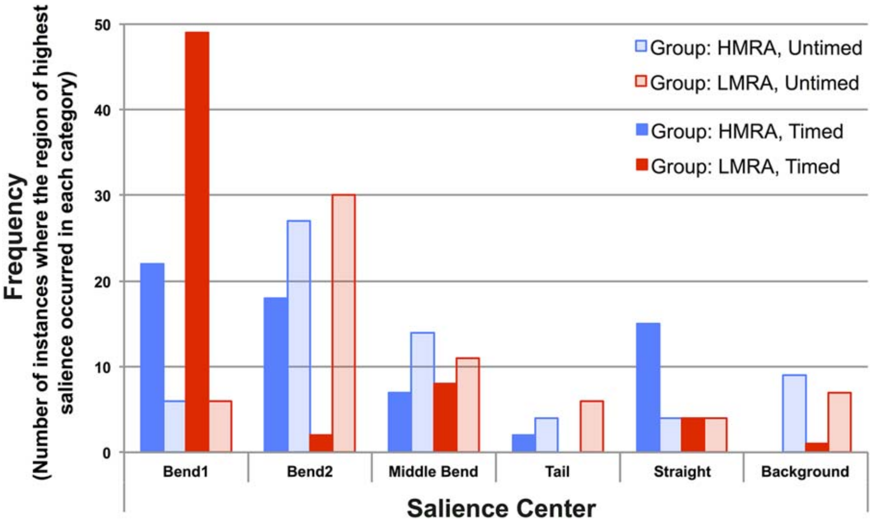

Figure 4.

The between-group comparison of salience distribution frequency, for both timing conditions representing how often attention was devoted to pre-defined areas: Bend 1, Bend 2, middle bend, tail, straight, or background. There is a significant difference between the four groups: (1) High mental rotation ability (HMRA): Timed; (2) High mental rotation ability: Untimed; (3) Low mental rotation ability (LMRA): Timed; and (4) Low mental rotation ability: Untimed (Fisher Exact Test = 122.18, $P < 0.001$).

agreement exists in the LMRA group across the testing conditions ($\kappa = 0.013$) (Viera and Garret, 2005) (Fig. 4).

DISCUSSION

The goal of the current study was to characterize how differences in time limits manifest in the salience distributions of high and low mental rotation ability (MRA) individuals. An analysis of where each MRA group devoted visual attention during testing revealed similarities between the groups in the untimed condition, but significant differences between the two groups in the timed condition. More specifically, on a question-by-question basis, the HMRA group was more consistent across the two timing conditions ($\kappa = 0.25$), than the LMRA ($\kappa = 0.013$).

The analysis of salience distribution (Roach et al., 2016) between HMRA and LMRA groups across the two timing conditions suggest that individuals with different MRA attend to different visual elements of the images as they solve complex spatial problems (Fig. 4). Further, where individuals devoted their attention was contingent on the time available to solve the problem. These results are supported by those observed by Wilson et al. (2011), who monitored gaze and task performance in novice and expert laparoscopists. In Wilson's experiment, experts (akin to the HMRA group in the current experiment) demonstrated a more economical gaze pattern than their novice counterparts (like the LMRA group). The experts made fewer movements and arrived at their decisions faster. Wilson referred to salience indirectly, by referring to the participants' time spent fixating on their "tools" and on the "target" of the surgical procedure. Wilson noted that experts attended specifically to their "target," while novices fixated heavily on their tools, and their surroundings (Wilson et al., 2011). However, in the context of mental rotation tests, HMRA and LMRA individuals replace experts and novices, and the "tools" and "targets" of surgical simulation are exchanged for the regions of the black and white block figures of the EMRT. In the current study, when the untimed condition was considered exclusively, the influence of MRA was not as apparent in the analysis of salience. Individuals of HMRA and LMRA inspected the block pairs in roughly the same fashion, yielding similar salience distributions and devoting attention to the same salient regions ($\kappa = 0.67$). Specifically, both groups attended preferentially to one specific feature of the block image (Bend 2) during the untimed condition. However, when a time limit was applied, the salience distributions diverged between the two groups ($\kappa = 0.20$) as their focus shifted to different areas of the images. When time was limited, the HMRA group expressed more flexibility in their visual assessment of the image. Instead of relying preferentially on the inspection of Bend 2, they directed their attention to different regions of each block figure on each question across the test. However, unlike the HMRA in timed testing, the LMRA individuals continued to devote their attention to a single feature in the timed condition (Bend 1).

In essence, it appears that the LMRA group attempted to focus their visual search around a single feature, regardless of the orientation of the image. This observation suggests that the LMRA individuals are not cognizant of the fact that the relevant pieces of the block pair will change according to the blocks' orientation in space. Instead, the LMRA adopt a "feature-matching," or analytic approach to problem solving. The analytic approach, first characterized by Geiser et al. (2006) is described as a nonspatial approach to mental rotation—as no real mental rotation takes place. Geiser suggests this approach is the least effective method to complete tasks of mental rotation, and is often adopted by the lowest performing individuals (Geiser et al., 2006). This finding aligns with literature that suggests that experts and novices attend to different structures as they problem-solve (Kundel and Nodine, 1983; Krupinski, 1996; Chapman and Underwood, 1998; Haider and Frensch, 1999; Charness et al., 2001; Nodine et al., 2002; Graesser et al., 2005; Van Gog et al., 2005; Jarodzka et al., 2010).

Further, it is possible that the application of a time limit may exacerbate differences in salience distributions, because the HMRA may be equipped with enhanced resources for visual processing, and thus may have surplus time available to inspect more questions (Sternberg, 1984; Schweizer and Ren, 2013). Literature suggests that higher processing rates in the HMRA may be attributed to differences in individual working memory capacity (Wilhelm and Schulze, 2002). This is supported by research that suggests a very strong correlation between an individual's ability to reason spatially, and their working memory capacity (Kyllonen and Christal, 1990; Kyllonen, 1996; Engle et al., 1999; Oberauer et al., 2000). Individuals who are proficient in reasoning will typically have high working memory capacities. With this considered, when visuospatial tasks are speeded, the burden on working memory is elevated and the rate of mental processing becomes critical to performance. Thus, it is possible that the application of a time constraint directly influences attention during problem solving, and reduces the ability to perform mental rotation tests accurately. The within-group analysis of salience supports this hypothesis, as the LMRA group demonstrated notable differences in salience across the timing conditions. The HMRA group showed a fair agreement between the two conditions ($\kappa = 0.25$), while the LMRA group diverged dramatically; with agreement on only 13% of questions between the two tests ($\kappa = 0.013$). With this observation in mind, it is clear the two groups respond to changes in time limits differently, and that difference is particularly evident when considering how these individuals visually inspect images while problem solving.

As the majority of existing spatial tests are administered with time limits, it was unsurprising that the High MRA individuals consistently outscored the Low MRA individuals on both timed, and untimed spatial tests. In contrast, the lack of interaction between group and timing was unanticipated. The lack of interaction suggests that despite MRA group divisions, all individuals completed the EMRT with greater success in the absence of time limits. This finding can be explained for both groups, as theories exist that support extra time for individuals at both ends of the ability spectrum (Bridgeman et al., 2004; Lu and Sireci, 2007).

Weaver (1993) suggests that individuals of low ability are handicapped by strict time limits, and that allowances of extra time enable individuals to process questions more completely, and then appropriately demonstrate their knowledge. As a result, the application of strict time limits may mask LMRA individuals' ability to perform on tests requiring spatial reasoning. With this in mind, the observed decline in score may not be directly attributed to a deficiency in ability or the knowledge being tested, but instead on their ability to process information quickly.

In contrast, the observation that the HMRA individuals also suffered decreased scores as a result of the time constraint is supported by literature on scholastic assessment test (SAT). Work by Bridgeman et al. (2004) suggests that high-scoring individuals tend to benefit more than low-scoring students when additional time is allotted on tests (Bridgeman et al., 2004). Bridgeman rationalizes that by allowing high-ability individuals more time on a test, they are able to address more questions, and subsequently reap the benefits of higher scores. In the context of performance on a spatial problem-solving task, it appears that all individuals are advantaged by relaxed time constraints, and all are disadvantaged by strict time limitations.

Limitations

This experiment may have been limited by several factors; namely the use of repetitive image pairs in the EMRT, sample size, and a low angular disparity between image pairs. As each image pair was presented in triplicate to increase the signal to noise ratio of eye-tracking parameters, there is a remote possibility that participants may have based each their response on the conclusions drawn from their first exposure to a question, relying on memory and recognition, rather than actively solving the problem presented. However, retrospective analysis of data did not show any patterns associated with improved accuracy, or consistency in answer responses as individuals progressed through the test. Second, the specificity of the experimental design may have limited the conclusions. Some conclusions derived from the data collected refer to linkages with other cognitive process beyond mental rotation (including spatial working memory). This study did not sample any cognitive processes beyond mental rotation, thus these linkages are inferential, and further studies employing tasks such as the *N*-Back Task (Kirchner, 1958) for working memory could be explored to corroborate these inferences.

Finally, this experiment may have been limited by the design of the main test metric, the EMRT. Block pairs in the EMRT were staggered at intervals ranging from ten to ninety degrees of angular disparity in a single axis to provide challenging visual stimuli for interpretation. However, it is possible that this range of disparity was too narrow, and stimuli could have been made more challenging with greater disparity and multiple axes of disparity (Niall, 1997). This being said, by incorporating greater angular disparity and/or multiple axes of rotation, it is likely that greater dichotomies would be evident between the HMRA and LMRA groups.

CONCLUSIONS

This experiment describes a relationship between mental rotation ability, time limits, and visual salience distribution. These three factors contribute to an individual's ability to solve spatial problems. It is apparent that a learner's mental rotation ability dictates their approach to solving spatially complex questions (Rehder et al., 2009). In this eye-tracking paradigm using the EMRT, it was observed that the HMRA individuals devoted their attention to different regions of the block images, when they were presented in different spatial orientations. This effect was not evident in the LMRA group, who appeared to use a single feature as a landmark for all questions, regardless of the landmark's task-relevance, and regardless of the block image's orientation in space. With this in mind, it is possible that "where" individuals attend during problem solving may be essential to their ability to reason spatially.

The research presented in the current experiment has direct implications for the spatially complex discipline of anatomy. Mastery in anatomy relies on a sound understanding of the interactions between three-dimensional structures in the visually complex environment of the human body

(Lufler et al., 2012; Nguyen et al., 2014; Zumwalt et al., 2015). As many anatomical evaluations occur in time-limited, "bell-ringer" environments, anatomy educators may wish to consider how their student's spatial ability may impact their ability to assess, and interpret anatomical structures and relationships under time-constraints. If instructors have concerns about equity across the spatial ability spectrum, they may seek to apply instructional techniques that minimize the negative aspects of extraneous cognitive load imposed on a student's cognitive resources (Mayer and Moreno, 2003), potentially by relaxing question-by-question time limits, or by employing visual guidance strategies (Lai et al., 2013) and visual signaling (Wilson, 2015) to better orient low-ability individuals on time-restricted evaluations.

NOTES ON CONTRIBUTORS

VICTORIA A. ROACH, Ph.D., is an assistant professor in the Department of Foundational Medical Studies at Oakland University William Beaumont School of Medicine, in Rochester, Michigan. She teaches medical gross anatomy, and is investigating the role that spatial reasoning plays in the acquisition of anatomical and surgical skill.

GRAHAM M. FRASER, Ph.D., is an assistant professor in the Cardiovascular Research Group, Division of BioMedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada. His teaching includes biophysics, cardiovascular physiology, problem solving, and gross anatomy and his research focuses on microvascular geometry, blood flow, and regulation.

JAMES H. KRYKLYWY, Ph.D., is a postdoctoral fellow in the Department of Psychology, at the University of British Columbia, Vancouver, British Columbia, Canada. He investigates the impact of emotion on spatial processing using MRI, eye-tracking, and behavioral testing, as well as the efficacy of procedural learning in both real world and 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, Schulich School of Medicine and Dentistry at the University of Western Ontario, 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 the Schulich School of Medicine and Dentistry at the University of Western Ontario, London, Ontario, Canada. He founded and directs the Corps for Research of Instructional and Perceptual Technologies (CRIPT) laboratory investigating digital anatomy development, its deployment, and its efficacy in pedagogy.

LITERATURE CITED

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

Bridgeman B, McBride A, Monaghan W. 2004. Testing and Time Limits. 1st Ed. Princeton, NJ: Educational Testing Service, R&D Connections. 6 p.

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

Chapman PR, Underwood G. 1998. Visual search of driving situations: Danger and experience. Perception 27:951–964.

Charness N, Reingold EM, Pomplun M, Stampe DM. 2001. The perceptual aspect of skilled performance in chess: Evidence from eye movements. Mem Cognit 29:1146–1152.

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46.

Ekstrom RB, French JW, Harman HH, Dermen D. 1976. Manual for Kit Factor-Referenced Cognitive Tests. 1st Ed. Princeton, NJ: Educational Testing Service. 224 p.

Engle RW, Tuholski SW, Laughlin JE, Conway ARA. 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. J Exp Psychol Gen 128:309–331.

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

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.

Graesser AC, Lu S, Olde BA, Cooper-Pye E, Whitten S. 2005. Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. Mem Cognit 33:1235–1247.

Grant ER, Spivey MJ. 2003. Eye movements and problem solving: Guiding attention guides thought. Psychol Sci 14:462–466.

Haider H, Frensch PA. 1999. Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. J Exp Psychol Learn 25: 172–190.

Hegarty M, Waller DA. 2005. Individual differences in spatial abilities. In: Shah P, Miyake A (Editors). The Cambridge Handbook of Visuospatial Thinking. 1st Ed. Cambridge, UK: Cambridge University Press. p 121–169.

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

Jarodzka H, Balslev T, Holmqvist K, Nyström M, Scheiter K, Gerjets P, Eika B. 2010. Learning perceptual aspects of diagnosis in medicine via eye movement modeling examples on patient video cases. In: Ohlsson S, Catrambone R (Editors). Cognition in Flux: Proceedings of the 32nd Annual Conference of the Cognitive Science Society (COGSCI 2010). Portland, OR, August 11–14, 2010, p 1703–1708. Austin, TX: Cognitive Science Society.

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.

Kirchner WK. 1958. Age differences in short-term retention of rapidly changing information. J Exp Psychol 55:352–358.

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

Krupinski EA. 1996. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 3:137–144.

Kundel HL, Nodine CF. 1983. A visual concept shapes image perception. Radiology 146:363–368.

Kyllonen PC. 1996. Is working memory capacity Spearman's g? In: Dennis I, Tapsfield P (Editors). Human Abilities: Their Nature and Measurement. 1st Ed. Mahwah, NJ. Lawrence Erlbaum Associates Inc. p 49–75.

Kyllonen PC, Christal RE. 1990. Reasoning ability is (little more than) working-memory capacity?! Intelligence 14:389–433.

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 10:90–1015.

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.

Lohman DF. 1979. Spatial Ability: A Review and Reanalysis of the Correlational Literature. 1st Ed. Redwood City, CA: Stanford University, Aptitude Research Project, School of Education. 204 p.

Lohman DF. 1988. Spatial abilities as traits, processes, and knowledge. In: Sternberg RJ (Editor). Advances in the Psychology of Human Intelligence. Vol. 4. 1st Ed. Hillsdale, NJ: Lawrence Erlbaum Associates. p 181–248.

Lu Y, Sireci SG. 2007. Validity issues in test speededness. Educ Meas 26:29–37.

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.

Mayer RE, Moreno R. 2003. Nine ways to reduce cognitive load in multimedia learning. Educ Psychol 38:43–52.

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

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

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.

Niall KK. 1997. 'Mental rotation', pictured rotation, and tandem rotation in depth. Acta Psychol 95:31–83.

Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP. 2002. Time course of perception and decision making during mammographic interpretation. AJR Am J Roentgenol 179:917–923.

Oberauer K, Süß HM, Schulze R, Wilhelm O, Wittmann WW. 2000. Working memory capacity—Facets of a cognitive ability construct. Pers Indiv Differ 29: 1017–1045.

Peters M, Battista C. 2008. Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. Brain Cogn 66:260–264.

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

Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C. 1995b. A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. Brain Cogn 28:39–58.

Rehder B, Colner RM, Hoffman AB. 2009. Feature inference learning and eye tracking. J Mem Lang 60:393–419.

Roach VA, Fraser GM, Kryklyw JH, Mitchell DG, Wilson TD. 2016. The eye of the beholder: Can patterns in eye movement reveal aptitudes for spatial reasoning? Anat Sci Educ 9:357–366.

Roach VA, Fraser GM, Kryklywy JH, Mitchell DG, Wilson TD. 2017. Different perspectives: Spatial ability influences where individuals look on a timed spatial test. Anat Sci Educ (in press; doi: 10.1002/ase.1654).

Schnipke DL. 1995. Assessing speededness in computer-based tests using item response times. In: Proceedings of the Annual Meeting of the National Council on Measurement in Education; San Francisco, CA, April 19–21, 1995. Princeton, NJ: Educational Testing Service, Princeton, NJ and Graduate Record Examinations Board. p 2–31.

Schweizer K, Ren X. 2013. The position effect in tests with a time limit: The consideration of interruption and working speed. Psychol Test Assess Model 55:62–78.

She HC, Chen YZ. 2009. The impact of multimedia effect on science learning: Evidence from eye movements. Comput Educ 53:1297–1307.

Shepard RN, Cooper LA. 1986. Mental Images and Their Transformations. 2nd Ed. Cambridge, MA: MIT Press. 376 p.

Spearman C. 1927. The Abilities of Man: Their Nature and Measurement.1st Ed. London, UK: MacMillan and Co., Ltd. 484 p.

Sternberg RJ. 1984. Toward a triarchic theory of human intelligence. Behav Brain Sci 7:269–287.

Thomas LE, Lleras A. 2007. Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. Psychonomic Bull Rev 14:663–668.

van Gog T, Paas F, van Merriënboer JJ, Witte P. 2005. Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. J Exp Psychol Appl 11:237–244.

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

Viera AJ, Garret JM. 2005. Understanding interobserver agreement: The kappa statistic. Fam Med 37:360–363.

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

Wanzel KR, Hamstra SJ, Caminiti MF, Anastakis DJ, Grober ED, Reznick RK. 2003. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. Surgery 134:750–757.

Weaver SM. 1993. The validity of the use of extended and untimed testing for postsecondary students with learning disabilities (extended testing). Doctorate of Philosophy Dissertation. Toronto, Canada: University of Toronto. 168 p.

Wilhelm O, Schulze R. 2002. The relation of speeded and unspeeded reasoning with mental speed. Intelligence 30:537–554.

Wilson MR, McGrath JS, Vine SJ, Brewer J, Defriend D, Masters RS. 2011. Perceptual impairment and psychomotor control in virtual laparoscopic surgery. Surg Endosc 25:2268–2274.

Wilson TD. 2015. Role of image and cognitive load in anatomical multimedia. In: Chan LK, Pawlina W (Editors). Teaching Anatomy: A Practical Guide. 1st Ed. New York, NY: Springer International Publishing. p 237–246.

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