Correlating spatial ability with anatomy assessment performance: A meta-analysis

Victoria A. Roach¹, Misa Mi¹, Jason Mussell², Sonya E. Van Nuland², Rebecca S. Lufler³, Kathryn DeVeaup, Stacey M. Dunham⁵, Polly Husmann⁶, Hannah L. Herriott⁶, Danielle N. Edwards⁷, Alison F. Doubleday⁸, Brittany M. Wilson⁹, Adam B. Wilson⁹,*

¹Department of Foundational Medical Studies, Oakland University William Beaumont School of Medicine, Oakland University, Rochester, Michigan
²Department of Cell Biology and Anatomy, School of Medicine, Louisiana State University Health Sciences Center, New Orleans, Louisiana
³Department of Medical Education, Tufts University School of Medicine, Boston, Massachusetts
⁴Department of Anatomy and Cell Biology, School of Medicine and Health Sciences, George Washington University, Washington, District of Columbia
⁵Department of Anatomy, Cell Biology, and Physiology, Indiana University School of Medicine, Bloomington, Indiana
⁶Department of Anatomy, Cell Biology, and Physiology, Indiana University School of Medicine, Indianapolis, Indiana
⁷Department of Cell, Developmental and Integrative Biology, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama
⁸Department of Oral Medicine and Diagnostic Sciences, College of Dentistry, University of Illinois at Chicago, Chicago, Illinois
⁹Department of Cell and Molecular Medicine, Rush College of Medicine, Rush University, Chicago, Illinois

Abstract

Interest in spatial ability has grown over the past few decades following the emergence of correlational evidence associating spatial aptitude with educational performance in the fields of science, technology, engineering, and mathematics. The research field at large and the anatomy education literature on this topic are mixed. In an attempt to generate consensus, a meta-analysis was performed to objectively summarize the effects of spatial ability on anatomy assessment performance across multiple studies and populations. Relevant studies published within the past 50 years (1969–2019) were retrieved from eight databases. Study eligibility screening was followed by a full-text review and data extraction. Use of the Mental Rotations Test (MRT) was required for study inclusion. Out of 2,450 screened records, 15 studies were meta-analyzed. Seventy-three percent of studies (11 of 15) were from the U.S. and Canada, and the majority

*Correspondence to Dr. Adam B. Wilson, Department of Cell and Molecular Medicine, Rush University, Armour Academic Center, Suite 505A, 600 S. Paulina St., Chicago, IL 60612; telephone: 312-942-5903; adam_wilson@rush.edu.
(9 of 15) studied professional students. Across 15 studies and 1,245 participants, spatial ability was weakly associated with anatomy performance ($r_{pooled} = 0.240$; CI at 95% = 0.09, 0.38; $p = 0.002$). Performance on spatial and relationship-based assessments (i.e., practical assessments and drawing tasks) was correlated with spatial ability, while performance on assessments utilizing non-spatial multiple-choice items was not correlated with spatial ability. A significant sex difference was also observed, wherein males outperformed females on spatial ability tasks. Given the role of spatial and non-spatial reasoning in learning anatomy, educators are encouraged to consider curriculum delivery modifications and a comprehensive assessment strategy so as not to disadvantage individuals with low spatial ability.

**Keywords**

gross anatomy education; meta-analysis; spatial abilities; anatomy performance; mental rotations test

**INTRODUCTION**

Spatial ability and the acquisition of anatomical knowledge are thought to be closely interconnected (Garg et al., 2002; Provo et al., 2002; Cohen and Hegarty, 2007; Levinson et al., 2007; Hegarty et al., 2009; Hoyek et al., 2009; Stull et al., 2009; Keedy et al., 2011; Brewer et al., 2012; Nguyen et al., 2012; Tan et al., 2012; Plumley et al., 2013; Nguyen et al., 2014; Sweeney et al., 2014; Jang et al., 2017; Langlois et al., 2017; Langlois et al., 2020). However, anatomy educators rarely assess or attempt to improve students’ spatial abilities in order to strengthen anatomy performance outcomes. Likewise, tips and recommendations from the spatial ability literature are not often incorporated into the design of anatomy curricula or assessments. These practices, or lack thereof, are likely a consequence of impracticalities, time constraints, and/or the opaqueness of the scientific literature. To bring the literature into clearer focus and to help educators make informed decisions about the role/impact of spatial ability on anatomy assessment outcomes, this study summarizes the links between spatial ability and its implications for anatomy education.

**Historical Context**

The precise definition of spatial ability in the literature is contentious; likely a byproduct of its many facets which have been recognized since its inception by Francis Galton (1880). Within the Cattel-Horn-Carroll (CHC) Theory of Cognitive Abilities, “the most comprehensive and empirically supported psychometric theory regarding the structure of cognitive abilities to date” (Flanagan and Dixon, 2014), spatial ability is categorized as ‘visual processing’.

Visual processing is defined as “the ability to perceive, analyze, synthesize, and think with visual patterns, including the ability to store and recall visual representations” (McGrew, 1997). According to the CHC theory, visual processing consists of eleven related abilities including: visualization, speeded rotation, closure speed, the flexibility of closure, visual memory, spatial scanning, serial perceptual integration, length estimation,
perceptual illusions, perceptual alternations, and imagery (McGrew, 1997); all of which may be considered facets that contribute to an individual’s ‘spatial ability’.

The Mental Rotations Test (MRT) and other measures of visual processing have been routinely utilized to provide empirical evidence of a significant positive relationship between spatial ability and educational performance (Shea et al., 2001; McGrew and Evans, 2004; Wai et al., 2009; Hoffler, 2010; Lubinski, 2010). This positive relationship is particularly evident in the science, technology, engineering, and mathematical (STEM) domains; yet no causal explanation for these relationships exists at present (Stieff and Uttal, 2015; Ramey and Uttal, 2017; Buckley et al., 2018). Regardless, visual processing abilities are malleable and efforts to develop visual processing skills in students have shown positive educational effects (Uttal et al., 2013; Buckley et al., 2018).

**Spatial Ability in Anatomy and Surgery**

The pioneering work by Rochford (1985) paved the way for spatial ability investigations in the early stages of medical training, particularly in the study of gross anatomy. The theoretical association between spatial ability and gross anatomy comprehension is supported by the notion that higher achievement in anatomy is reliant on a firm understanding of the three-dimensional relationships between anatomical structures; especially when viewed from different perspectives (Fernandez et al., 2011; Lufler et al., 2012; Nguyen et al., 2014; Zumwalt et al., 2015). This premise also translates to the acquisition of surgical skills. For example, spatial ability is correlated with overall surgical skill performance, as assessed through rating scales and the economy of hand motion, on visually complex surgical training tasks in individuals naive to surgery (Wanzel et al., 2002, 2003; Roach et al., 2012). However, this effect was not observed in individuals with established surgical expertise (i.e., surgical residents and attending physicians), for whom practice and surgical experience obviated the impact of innate abilities. This suggests that spatial ability may be most instrumental in the earliest phases of skill acquisition (Wanzel et al., 2003).

Several empirical studies have sought to confirm the connections between spatial ability and the acquisition of anatomical knowledge through correlational analyses, yet a firm consensus on the matter has yet to be reached. Conclusions in the literature are largely mixed, ranging from no correlation existing between spatial ability and anatomy performance (Gutierrez et al., 2018; Knudsen et al., 2018; Lone et al., 2019; Gonzales et al., 2020) to positive correlations of various magnitudes occurring between the two variables (Garg et al., 2002; Provo et al., 2002; Cohen and Hegarty, 2007; Levinson et al., 2007; Hegarty et al., 2009; Hoyek et al., 2009; Stull et al., 2009; Keedy et al., 2011; Brewer et al., 2012; Nguyen et al., 2012; Tan et al., 2012; Plumley et al., 2013; Nguyen et al., 2014; Sweeney et al., 2014; Jang et al., 2017).

Additionally, the reciprocal relationship has also been queried in a set of recent studies (Vorstenbosch et al., 2013; Bogomolova et al., 2020) and meta-analyzed by Langlois and colleagues (2020). The objective of this recent review was to assess how spatial abilities change in response to anatomy education training (Langlois et al., 2020). This small meta-analysis of seven studies revealed that studying anatomy could serve as an interventional
To date, only one systematic review has explored the effects of spatial ability on students’ anatomy assessment outcomes. This systematic review queried the topic from predominately a thematic perspective, and included a small meta-analysis of only three studies (Langlois et al., 2017). As stated by Langlois et al. in their follow-up study, “Future studies will need to validate if [an] enhancement in spatial abilities is related to spatial learning of anatomy” (2020). As such, anatomy education researchers are reminded that, at present, no conclusive evidence exists to support the theoretical association between spatial abilities and outcomes on anatomy assessments.

METHOD

Literature Search

This work was conducted in accordance with PRISMA guidelines (Moher et al., 2009) and recommendations by Cook and West (2012). This study was not eligible for institutional review board approval as it did not directly involve human subjects.

Eight databases were used to search the literature over a 50-year time range spanning from January 1969 to December 2019. Databases included PubMed (U.S. National Library of Medicine, National Institutes of Health, Bethesda, M.D.), CINAHL (EBSCO Industries Inc., Birmingham, A.L.), PsycINFO (American Psychological Association, Washington, D.C.), ERIC (Institute of Education Sciences, Washington, D.C.), EMBASE (Elsevier, Amsterdam, Netherlands), Scopus (Elsevier B.V., Amsterdam, Netherlands), Web of Science (Clarivate Analytics, Philadelphia, P.A.), and the Cochrane Library (Cochrane, London, U.K.). The search query relied on relevant medical subject headings and key search terms to identify pertinent records. Examples of key terms include ‘spatial ability/reasoning’, ‘visualization ability’, ‘anatomy’, ‘models’, etc. (see Supplemental Material 1 for an example PubMed search query). References of included records were also hand-searched to identify additional studies omitted by the electronic search.

Eligibility Criteria

Studies were included for preliminary review if they tested the influence of spatial ability/reasoning (or its sub-facets including visualization ability, mental rotations, spatial
translation, spatial orientation, and/or spatial relations) on anatomy performance. All included studies conducted correlations between learners’ spatial reasoning abilities, as assessed through the Mental Rotations Test, and their performance on spatial anatomy tasks and/or anatomical knowledge examinations. Though many tests of visual perception exist (Flanagan and Dixon, 2014; Castro-Alonso and Atit, 2019), the Mental Rotations Test (MRT) was selected as the gold standard due to its well-established reliability and validity properties (Vandenberg and Kuse, 1978; Peters et al., 1995). To avoid previously published limitations (Langlois et al., 2017), studies that did not utilize the MRT were intentionally excluded to minimize variability in the measurement of spatial ability.

The evaluation of non-expert learner populations was also a requisite for study inclusion to mitigate potential confounding factors common among experienced learners (e.g., surgical residents). Non-expert learners were defined as cohorts of students (e.g., first-year medical students) with collective inexperience in intensive laboratory-based anatomy instruction. No restrictions were placed on student populations. Student populations were allowed to vary from high school and undergraduate college students to allied health (nursing, occupational therapy, physiotherapy, physician’s assistants, etc.), professional (veterinary, medical), and graduate students. Studies originating from outside of the United States were included for analysis so long as they were written in the English language. Studies not meeting all of the above criteria were excluded.

**Study Selection and Data Extraction**

Five teams of paired researchers reviewed an equal share of records retrieved from the literature search (approximately 490 records per team). During the screening process, each researcher made an independent judgment on whether to include or exclude a record for full review based on the record’s title and abstract alone. Records were included for full review if selected by at least one researcher. Records that were disagreed upon by a pair of researchers were automatically assigned to undergo a full-review by a different research team; thus, there was no need for team members to reach consensus during the record screening process. Both percent agreement and Cohen’s $\kappa$ statistics were used to assess the inter-rater reliability of the dichotomous inclusion/exclusion judgments (Cohen, 1960). A Cohen’s $\kappa$ statistic of 0.61 or higher is considered to demonstrate substantial coding agreement between raters (Landis and Koch, 1977).

Each of the five research teams was assigned an equal share of studies to review in-full and to extract data from. Each research team member entered data from their assigned studies independently into a data extraction form. VAR and ABW verified the accuracy of the final data submissions by reviewing each of the included articles and, as necessary, reached consensus to resolve any data extraction discrepancies. Attempts were also made to clarify and acquire additional information from corresponding authors for studies with incomplete/ unusable numerical datasets. If correspondence with an author was unsuccessful after two attempts, the article in question was excluded.
Statistical Analyses

Data were collected from studies using a ‘data extraction form’ generated in Qualtrics XM (Qualtrics, Provo, UT) and were exported to Microsoft Excel® (2013, Microsoft Corporation, Redmond, WA) for organization and accuracy review. Within MedCalc, version 15.0 (MedCalc Software, Ostend, Belgium), a meta-analysis was performed using a Fisher Z transformation of the correlation coefficients (Hedges and Olkin, 1985). The random-effects model used in this study incorporated the heterogeneity statistic to compute the summary correlation coefficient (DerSimonian and Laird, 1986). Studies were weighted as a function of their sample sizes, and effect size estimates are reported as pooled correlation coefficients. Note that some included studies reported outcomes by a subset of participants. Fortunately, each subgroup of participants represented an independent sample so as not to violate the independence assumption of meta-analyses (Petitti, 2000).

Separate subgroup analyses were performed to determine whether the learner population or assessment type uniquely influenced the pooled correlation of the analyzed studies (Wang and Ware, 2013). A separate random-effects meta-analysis was performed on a subset of included studies to explore the effects of gender on both MRT outcomes and anatomy assessment scores. For this analysis, the standardized mean difference (SMD) was reported for the effect size. The magnitude of the summary effect size was interpreted using Cohen’s recommendations for small (0.20 – 0.49), medium (0.50 – 0.79), and large (≥0.80) effects (Cohen, 1988).

Confidence intervals (CI) were reported at the 95% confidence level. A Q-statistic was used to determine the presence of heterogeneity (Huedo-Medina et al., 2006), and an I² statistic estimated the amount of variance between studies. Heterogeneity was considered inconsequential if the variance between studies (I²) was less than 25% and was considered substantial if greater than 75% (Higgins et al., 2003). To determine the presence of publication bias, funnel plots were visually inspected (Duval and Tweedie, 2000).

RESULTS

Search and Screening Outcomes and Study Characteristics

A total of 2,450 records were screened, 59 underwent full-text review, and 15 studies (k) were included in the final analysis. Table 1 reports the inter-rater agreement among raters for the title and abstract screening process. The average percent agreement for including/excluding records across the 5 teams was 98.9%. The mean Cohen’s kappa was acceptable at 0.586. Figure 1 summarizes the full progression of the study inclusion/exclusion process and presents a listing of anatomy-related studies that were excluded for not meeting one or more of the eligibility criteria. All extracted data came from published full-text articles.

Of the included studies, 47% (7 of 15) were from the U.S., the majority (9 of 15) studied professional students, and all studies were published in 1999 or later (Table 2). Anatomical knowledge was assessed through either practical assessments (n = 7), written assessments (n = 6), or through drawing tasks (n = 2; Table 2). All studies used the MRT to assess spatial reasoning.
Meta-analytic Findings: Correlating Spatial Ability with Anatomy Performance

Across 15 studies (21 investigations) totaling 1,245 participants, a weak positive association was identified between learners’ spatial abilities and their anatomy assessment outcomes as demonstrated by a pooled correlation of 0.240 (CI at 95% = 0.09, 0.38; p = 0.002; Figure 2). The heterogeneity across these studies was higher than the optimal level ($I^2 = 81.27\%$; $p < 0.001$). By convention, this represents sizable variation between studies, likely a consequence of variable study designs and outcomes. The relative symmetry of the funnel plot (Supplemental Material 2) suggests publication bias was unlikely.

Subgroup Analyses

Per the outcomes of the subgroup analyses, the overall summary effect was moderated by studies that included professional students (Figure 3). The moderating effects observed for both ‘practical examinations’ and ‘drawing tasks’ suggest these forms of assessment rely more heavily on spatial reasoning than written multiple-choice examinations, for which no moderating effect was detected. A formal subgroup analysis of correlation coefficients by participant gender was not feasible due to the lack of available data.

Gender Effects on the Mental Rotations Test

To provide some additional context, studies that reported male versus female MRT performance data were meta-analyzed to determine the magnitude of gender effects on spatial reasoning performance alone. This analysis included six of the 15 original studies and three additional studies excluded from the meta-analysis above due to an absence of correlational data. A total of 1,292 participants ($n_\text{♂} = 623; n_\text{♀} = 669$) comprised these 9 studies (10 investigations). Higher MRT scores significantly favored males as demonstrated by a medium effect size (SMD = 0.768; CI at 95% = 0.623, 0.913; $p < 0.001$; Figure 4). Collectively, these studies demonstrated relative homogeneity ($I^2 = 22.69\%$; $p = 0.234$) as they did not significantly vary in their outcomes. The funnel plot showed no evidence of publication bias.

Gender Effects on Anatomy Assessment Outcomes

In an even smaller subset of studies ($k = 4; n_\text{♂} = 335, n_\text{♀} = 309$) (Guillot et al., 2007; Lufler et al., 2012; Nguyen et al., 2014; Lone et al., 2019), a separate meta-analysis reported no difference in anatomy performance scores (SMD = 0.272; CI at 95% = −0.47, 0.59; $p = 0.095$) between males and females. These studies were mildly heterogeneous ($I^2 = 63.46\%$; $p = 0.042$). Given the small number of studies, further analysis may be needed.

DISCUSSION

The present study used a meta-analysis to evaluate the association between students’ spatial abilities and their anatomy assessment outcomes. Compared to a smaller meta-analysis ($k = 3$ studies) by Langlois et al. (2017), the present work represents the most comprehensive review of the literature between January 1969 to December 2019 ($k = 15$ studies). Both the Langlois et al. study and the current meta-analysis detected a significant pooled correlation demonstrating a positive association between spatial abilities and anatomy performance. However, the pooled correlation coefficient of the Langlois et al. (2017) study was much
stronger ($r_{pooled} = 0.53$) than that of the current study ($r_{pooled} = 0.240$). Langlois and colleagues also reported higher heterogeneity ($I^2 = 92\%$) across studies compared to the present work ($I^2 = 81\%$) (Langlois et al., 2017). Collectively, these findings substantiate the positive association between students’ spatial abilities and their anatomy performance; though the strength of the positive association may be weaker than initially suspected.

**Implications for Anatomy Assessments**

A prior descriptive analysis concluded that spatial abilities are not related to anatomical knowledge gains when anatomy is assessed using essays or non-spatial multiple-choice examinations (Langlois et al., 2017). Conversely, significant relationships between spatial abilities and anatomical knowledge have been identified when measuring knowledge through “practical examinations, three-dimensional synthesis from two-dimensional views, drawing tasks, and cross-sections” (Langlois et al., 2017). The present study objectively confirms these thematic findings. The subgroup analysis of assessment types reported a significant positive pooled correlation between spatial abilities and anatomy performance ($r_{pooled} \geq 0.330; p \leq 0.004$; Figure 3) when practical examinations and drawing tasks were used to assess anatomical knowledge. This suggests that these examination modalities rely more heavily on spatial reasoning. No correlation ($p = 0.458$) was detected when outcomes were aggregated from studies that utilized written non-spatial multiple-choice examinations to assess anatomy performance. Collectively, these findings confirm and reinforce the importance of assessing anatomical knowledge from both spatial and non-spatial perspectives. Assessing anatomy through practical examinations alone, which arguably requires greater spatial aptitude, may exacerbate the identified gender effects in favor of males.

**Gender Effects on Spatial Abilities.**—The spatial visualization literature is rich with studies and meta-analyses reporting that males outperform females across a variety of tests measuring spatial ability (Linn and Petersen, 1985; Voyer et al., 1995; Maeda and Yoon, 2013; Lauer et al., 2019). The present study was no exception, as males demonstrated significantly higher MRT scores than females (Figure 4). According to work by Halpern and Collaer (2005), Reilly and colleagues (2017), and a recent meta-analysis on gender differences in spatial reasoning (Lauer et al., 2019), a significant gender difference exists between the spatial abilities of males and females. Consequently, the underrepresentation of women in STEM fields may be partly explained by the inherent spatial ability differences between genders (Casey et al., 1995; Reilly et al., 2017). The origins of gender differences in spatial abilities are thought to be multifaceted ranging from biological and social factors to cultural influences and life experiences (Reilly et al., 2017; Lauer et al., 2019). While gender effects vary in magnitude, they are often moderated by a number of factors including age, whereby gender differences incrementally increase from childhood to adolescence to adulthood (Voyer et al., 1995; Reilly et al., 2017; Lauer et al., 2019). Male advantage on spatial reasoning tasks is also moderated by the use of dynamic (as opposed to static) visualizations (Castro-Alonso et al., 2019), the implementation of time limits on spatial tasks (Maeda and Yoon, 2013), and masculine identification (Reilly and Neumann, 2013).
Gender Effects on Anatomy Assessments.—What remains uncertain is whether similar gender effects translate to anatomy assessments. The current study \((k = 4; \text{n}_{\text{total}} = 644)\) found no differences between males and females among studies that reported anatomy performance outcomes by gender. A more comprehensive meta-analysis on this topic is warranted as the scope of the present findings was limited to the spatial ability literature, and anatomy performance scores were not segregated by assessment type. According to one theory, males may have an inherent advantage over females on practical examinations, in particular. Arendasy and colleagues suggest that spatial ability tasks may be easier for males due to their skills in visually identifying and interpreting task-relevant visual cues \((\text{Arendasy et al.}, 2011)\). This theory is also supported by gaze-tracking studies which illustrate that individuals with low spatial skills tend to adopt uneconomical visual search strategies when asked to answer complex spatial questions under time constraints \((\text{Roach et al.}, 2017a, b, 2019)\). Contextually, this phenomenon may inadvertently disadvantage students (i.e., predominantly females) with low spatial abilities when testing using practical (‘bell-ringer’) examinations, characterized by time-bound problem solving that is wholly reliant on visual searching and the identification of anatomical structures using diagnostic relationships. To help minimize the effects of possible assessment disparities, it is important to assess anatomical knowledge through both spatial and non-spatial mechanisms. Further, it may be appropriate to assign a lower weight for practical assessments when computing course grades, or, if logistically feasible, consider removing/reducing time restrictions for their administration.

Implications for Anatomy Instruction

Fortunately, instructional design research has identified two learning theories that can be leveraged to minimize the effect that low spatial ability may have on anatomical knowledge acquisition. Both the Cognitive Load Theory \((\text{Sweller}, 1988)\) and the Cognitive Theory of Multimedia Learning \((\text{Mayer and Moreno}, 2003)\) provide instructors with the tools necessary for designing equitable learning experiences for all. From these two theories come three key instructional strategies which include: 1) the signaling principle \((\text{van Gog}, 2014)\), 2) the transient information effect \((\text{Ayres and Paas}, 2007)\), and 3) the redundancy effect \((\text{Kalyuga and Sweller}, 2014; \text{Fraser et al.}, 2015)\).

Instructional approaches that utilize visual ‘signaling’ (e.g., strategic labeling, cueing, etc.) benefit students’ in their learning when images and multimedia serve as core curricular content \((\text{Mayer}, 2009; \text{Wilson}, 2015; \text{Roach et al.}, 2019)\). Visual signaling assists students with lower spatial abilities by modeling more effective search strategies \((\text{Arendasy et al.}, 2011)\). Specifically, the signaling principle suggests that by utilizing visual cues to direct learners’ attention to the most essential learning elements, learners can better recognize the primary structures of interest and their relevant relationships. These visual cues may be additive (e.g., the use of extra visual stimuli such as arrows, pointers, hand gestures, leader lines, frames, alphanumeric characters, labels, etc.) or non-additive (e.g., drawing the learners’ attention to a salient structure through the manipulation of the image itself, via the use of background blurring, lighting, transparencies, overlays, coloring, and contrasts; Figure 5A). Of these techniques, the use of non-additive signals is considered the most effective in guiding individuals with low spatial ability towards salient information, as they
draw attention without obstructing the spatial information present in the original image, thus leading to improved spatial task performance, including anatomy comprehension (Munzer et al., 2009; de Koning et al., 2010; Imhof et al., 2013; Skulmowski and Rey, 2018).

The transient information effect may also be useful for anatomy educators when covering spatially complex information. The transient information effect refers to the inability of individuals with a low spatial ability, and correspondingly a low spatial working memory (Castro-Alonso and Atit, 2019), to keep pace and visually interpret and process information as it is presented. The transient information effect may be overcome by utilizing static or dynamic visual aids (e.g., videos or animations) which include controls permitting students to navigate through the content at their own pace. This effect may also be overcome through the use of content segmentation, whereby information is presented in shorter/smaller chunks, rather than as long, dynamic visualizations. Research has shown that by providing curated pauses during content delivery, students are less likely to become cognitively overloaded, and thus, learning is improved (Ayres and Paas, 2007; Aldahmash and Abraham, 2009; Berney et al., 2015; Castro-Alonso et al., 2018; Kühl et al., 2018; Loftus et al., 2018).

Finally, the redundancy effect is apparent when non-essential spatial information is presented to students. The presentation of “extra” material, whether that be multiple views of anatomical structures or interesting anatomical anomalies in conjunction with essential foundational material, requires students to utilize their working memory resources to process both essential and non-essential information concurrently, and thus presents an increase in cognitive load and a decrease in one’s available working memory resources. Indeed, the literature reveals that redundant information, and particularly redundant visuospatial information, disproportionately hinders low spatial ability individuals in their ability to process visual information (Levinson et al., 2007). For anatomy educators seeking to accommodate individuals with low spatial ability, extra information that is not fundamental for the learning topic should be discarded from the visualizations. This is well achieved by presenting only key views of anatomical structures (Garg et al., 1999; 2002) and eliminating any “interesting but not essential” visual material (Castro-Alonso et al., 2019b), particularly when the audience is learning the content for the first time (Figure 5B).

Of these three instructional approaches, all have the capacity to ameliorate anatomy instruction. However, it is the signaling principle which is perhaps the most valuable in the context of anatomy practical examinations; particularly if the assessment is conducted using still photographic images. For example, anatomy educators may wish to reconsider how arrows, or other obscuring visual cues, are utilized when assessing students’ understanding of anatomical structures and relationships. When compared to non-additive cues such as highlighting, the presence of additive cues (e.g., arrows) may obscure relevant information conveyed by background structures and may impose a greater challenge on students’ working memory during assessment events.

One practical question that lingers is, “Should anatomy educators assess learners’ spatial abilities prior to delivering anatomy content?” The authors collectively agree that testing students’ spatial abilities is not necessary. All student populations are expected to have some proportion of learners with low spatial abilities. Making deliberate delivery adjustments (per
the recommendations above) benefits both low and high spatial ability students (Keehner et al., 2008; Castro-Alonso et al., 2019a), negating the need for spatial ability testing. Spatial ability is also known to be more influential during the initial learning process. Hence, the above recommendations are predominately intended for educators teaching anatomy to novice learners. Over time, with more anatomy experience, the magnitude of spatial disparities wanes and gaze patterns mature (Lufler et al., 2012; Zumwalt et al., 2015). Once students have gained a foundational knowledge of anatomy and its reliance on spatial awareness, they are better equipped to tackle anatomy content with greater spatial complexity (e.g., head and neck anatomy). Refer to Table 3 for a summary of recommendations for instructors.

Study Limitations

In the STEM fields, it is documented that learning through dynamic visualizations is more effective than static visualizations (Castro-Alonso et al., 2019c). The current study was unable to explore these effects between dynamic and static visualizations as the majority of studies used both visualization types and did not report findings separately for each. The present analyses also presumed that items selected by study authors for inclusion in the MRT were of equal difficulty across studies, though there was no mechanism to confirm or control for this. Of note, Peters et al., (1995) report, “The magnitude of effect sizes for sex...does not increase with increasing difficulty of the [mental rotation] task”. This suggests that differences in MRT item difficulty are likely to have had little to no impact on the findings of the current meta-analysis. Studies included in this meta-analysis were not fully inclusive of all anatomical regions which may limit the generalizability of these findings. It is also plausible that the presented outcomes were moderated by other non-investigated factors such as instructional design, cultural influences, working memory capacity, etc.

CONCLUSIONS

The available evidence confirms that learners’ spatial abilities are weakly correlated to anatomy assessment outcomes, with practical examinations and drawing tasks eliciting the strongest correlations. Though males outperform females on the Mental Rotations Test, upon preliminary analysis, no gender differences were detected on assessments measuring anatomical knowledge. A broader meta-analytic evaluation of gender effects on various forms of anatomy assessments is warranted. Given the above evidence, anatomy educators are fundamentally compelled to critically reflect on their assessment and content delivery practices. To further enrich anatomy learning for individuals with low spatial abilities, it is recommended that educators deliver their anatomy content using techniques that employ non-additive visual signaling, minimize redundant visuospatial information, and allow learners to set an individualized pace for content consumption.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
NOTES ON CONTRIBUTORS

VICTORIA A. ROACH, Ph.D., is an assistant professor in the Department of Foundational Medical Studies, and the Department of Surgery at Oakland University William Beaumont School of Medicine in Rochester, MI. She teaches anatomy, and co-directs the Embark Scholarly Concentration Program. Her research focuses on the optimization of learning and surgical skill acquisition.

MISA MI, Ph.D., M.L.I.S., is a professor and health information specialist in the Department of Foundational Medical Studies at Oakland University William Beaumont School of Medicine in Rochester, MI. She designs and provides training programs on evidence-based medicine, program evaluation, systematic review, and health informatics for learners across the spectrum of medical education. Her research interest is in learning design, evidence-based medicine, health literacy, and informatics.

JASON MUSSELL, Ph.D., is an associate professor in the Department of Cell Biology and Anatomy at the Louisiana State University School of Medicine, Health Sciences Center (LSUHSC) in New Orleans, LA. He teaches gross and developmental anatomy, diseases of the musculoskeletal and dermatologic systems, humanities in medicine, and is active in the LSUHSC longitudinal mentoring programs.

SONYA E. VAN NULAND, Ph.D., is an assistant professor in the Department of Cell Biology and Anatomy, Louisiana State University Health Sciences Center, New Orleans, LA. She teaches medical gross anatomy and histology. Her research interests are medical education and e-learning tool effectiveness.

REBECCA S. LUFLER, Ph.D., is an associate professor of anatomy in the Department of Medical Education at Tufts University School of Medicine, Boston, MA. She teaches anatomy to first-year medical, dental, and physician assistant students, and directs the fourth-year medical and physician assistant courses. Her research interest is in medical education and curriculum.

KATHRYN DEVEAU, Ph.D., is an assistant professor of anatomy in the Department of Anatomy and Cell Biology, School of Medicine and Health Sciences at George Washington University in Washington, DC. She teaches anatomy and neuroanatomy to first- and second-year medical students as well as graduate students. Her research interests are in medical education.

STACEY M. DUNHAM, Ph.D., is an academic specialist and coordinator of undergraduate anatomy programs in the Department of Anatomy, Cell Biology, and Physiology at Indiana University School of Medicine, Bloomington, IN. She teaches anatomy to undergraduate students. Anatomy and medical education are her areas of research interest.

POLLY HUSMANN, Ph.D., is an assistant professor in the Department of Anatomy, Cell Biology, and Physiology at Indiana University School of Medicine, Bloomington, IN. She teaches anatomy to medical, graduate, and undergraduate students. Her research focuses on factors outside of the classroom that affect student outcomes in anatomy.
HANNAH L. HERRIOTT, is a graduate (Ph.D.) student in anatomy education in the Department of Anatomy, Cell Biology, and Physiology at Indiana University School of Medicine, Indianapolis, IN. She teaches anatomy, embryology, histology, neuroscience, and physiology to students in various graduate and healthcare professional programs. Her research interests include curricular development, interprofessional education, and professional development.

DANIELLE N. EDWARDS, Ph.D., is an assistant professor of anatomy in the Department of Cell, Developmental, and Integrative Biology at the University of Alabama at Birmingham in Birmingham, AL. She teaches anatomy to medical and health professional students. Her research interests are in medical education.

ALISON F. DOUBLEDAY, Ph.D., is an associate professor in the Department of Oral Medicine and Diagnostic Sciences, and Director of Faculty Development in the College of Dentistry at the University of Illinois at Chicago, Chicago, IL. She teaches anatomy and embryology to first-year dental students. Her research interest is in collaborative learning and the use of technology to support collaborative learning environments.

BRITTANY M. WILSON, Ph.D., is a postdoctoral research fellow in the Department of Cell and Molecular Medicine at Rush University Medical Center in Chicago, IL. She has taught gross anatomy to undergraduate, graduate, nursing, physician assistant, and medical students. Her primary research interest is in skeletal biology.

ADAM B. WILSON, Ph.D., is an associate professor in the Department of Cell and Molecular Medicine at Rush University, Chicago, IL. He is the departmental director of anatomy education, the discipline director for medical gross anatomy, and course director for the musculoskeletal course. His research focuses on topics in educational measurement and evaluation.

LITERATURE CITED


Figure 1: Flow chart of the inclusion/exclusion process.
A total of 15 studies were included for analysis.
Figure 2: Forest plot (random-effects model) of correlations by study.

The summary effect (diamond) shows a significant positive association ($r_{pooled} = 0.240$) between spatial ability (i.e., MRT scores) and anatomy performance outcomes. Boxes denote the correlations (effect sizes) for each study. Boxes are sized according to the weight assigned to each study (a function of a study’s sample size). Bars extending from each box represent 95% confidence intervals.

The following footnotes specify caveats/nuances related to select studies:

- Garg et al., 1999: This study analyzed two independent groups: (A) Males multiple views and (B) females multiple views.
- Luursema et al., 2006: This study analyzed two independent groups: (A) binocular group and (B) stereoptical group.
• Guillot et al. 2007 (B) and Gutierrez et al. 2018: A correlation coefficient of ‘0’ was entered for those studies that reported non-significant findings, yet did not report an exact correlation coefficient value.

• Cui et al., 2017: This study analyzed four independent groups: (A) 3D MRT-low group, (B) 3D MRT-high group, (C) 2D MRT-low group, and (D) 2D MRT-high group. Sample sizes per group were derived from Figure 6 of the original study.

• Lufler et al., 2012: Correlational data was obtained from the corresponding author directly as it was not reported in the original manuscript.
The overall weak positive association between learners’ spatial abilities and their anatomy performance outcomes was moderated by studies that included professional students and assessed learners using practical examinations and/or drawing tasks (p < 0.005). Diamonds denote summary effect sizes. Bars extending from each diamond represent 95% confidence intervals.

**Figure 3: Forest plot (random-effects model) of subgroup analyses.**
Figure 4: Forest plot (random-effects model) of MRT Gender Differences.
The summary effect (diamond) shows a medium-sized difference in MRT scores between genders with males significantly outperforming their female counterparts (SMD=0.768; CI at 95%=0.623, 0.913; p<0.001). *indicates studies excluded from the primary meta-analysis. Boxes denote the correlations (effect sizes) for each study. Boxes are sized according to the weight assigned to each study (a function of a study’s sample size). Bars extending from each box represent 95% confidence intervals.
Figure 5: Preferred strategies (*) for enhancing anatomy learning for low spatial ability students. Images were captured from AnatomyTV (2020, Primal Pictures, United Kingdom). A. Visual signaling approaches. B. Example of a key view versus spatially challenging views.
<table>
<thead>
<tr>
<th>Team (n of raters)</th>
<th>Number of records reviewed</th>
<th>Number of inclusion/exclusion discrepancies</th>
<th>Percent Agreement</th>
<th>Cohen’s / *Fleiss’ Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (2)</td>
<td>491</td>
<td>3</td>
<td>99.39%</td>
<td>0.569</td>
</tr>
<tr>
<td>B (2)</td>
<td>490</td>
<td>5</td>
<td>98.98%</td>
<td>0.833</td>
</tr>
<tr>
<td>C (2)</td>
<td>490</td>
<td>12</td>
<td>97.55%</td>
<td>0.323</td>
</tr>
<tr>
<td>D (2)</td>
<td>491</td>
<td>6</td>
<td>98.78%</td>
<td>0.619</td>
</tr>
<tr>
<td>E (3)</td>
<td>488</td>
<td>2</td>
<td>99.569</td>
<td><strong>NaN</strong></td>
</tr>
<tr>
<td><strong>Total (Averages)</strong></td>
<td><strong>2450</strong></td>
<td><strong>28</strong></td>
<td>(98.9%)</td>
<td>(0.586)</td>
</tr>
</tbody>
</table>

NaN (Not a Number): In this data, two of the three raters agreed 100% of the time. As such, these raters individually have a 100% estimated chance of selecting the ‘exclusion’ category. Thus, the probability of agreement by chance (Pe) is P_{e}=1\times 1=1. The denominator of Fleiss’ Kappa includes 1− Pe, which equals zero in this instance. In R, since the numerator is also zero, the result is 0/0=NaN.
Table 2:

Summary of Study Features

<table>
<thead>
<tr>
<th>Study Features</th>
<th>Number (% of Studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country of Origin</strong></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>4 (26)</td>
</tr>
<tr>
<td>France</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Germany</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Ireland</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1 (7)</td>
</tr>
<tr>
<td>United States</td>
<td>7 (46)</td>
</tr>
<tr>
<td><strong>Learner Population</strong></td>
<td></td>
</tr>
<tr>
<td>Professional Students</td>
<td>9 (60)</td>
</tr>
<tr>
<td>Dental Students</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Medical Students</td>
<td>6 (40)</td>
</tr>
<tr>
<td>Veterinary Students</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Undergraduate College Students</td>
<td>6 (40)</td>
</tr>
<tr>
<td>Other (e.g., Faculty)</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Publication Period</strong></td>
<td></td>
</tr>
<tr>
<td>2011–2019</td>
<td>10 (67)</td>
</tr>
<tr>
<td>2000–2010</td>
<td>4 (26)</td>
</tr>
<tr>
<td>1999 or earlier</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Assessment of Anatomical Knowledge</strong></td>
<td></td>
</tr>
<tr>
<td>Written Assessment</td>
<td>6 (40)</td>
</tr>
<tr>
<td>Specimen-based Practical Assessment</td>
<td>3 (20)</td>
</tr>
<tr>
<td>Model/Digital-based Practical Assessment</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Drawing Task</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Study Features</td>
<td>Number (%) of Studies</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Total</td>
<td>15 (100)</td>
</tr>
</tbody>
</table>

*Some studies included more than one learner population.*

*One study utilized both forms of practical assessment.*
Table 3: Summary of recommendations with examples/rationales.

<table>
<thead>
<tr>
<th>Recommendation for Instructors</th>
<th>Example / Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilize the signaling principle.</td>
<td>When delivering visual material, draw your students’ attention to key spatial relationships using non-additive cues, like coloring and blurred backgrounds.</td>
</tr>
<tr>
<td>Consider the transient information effect.</td>
<td>If dynamic visuals, such as videos and 3D animations, are used in instruction, ensure that students can pause, rewind, and replay content at their own pace.</td>
</tr>
<tr>
<td>Keep the redundancy effect in mind.</td>
<td>For a novice population of students, instructors should aim to display only key views of anatomical structures, and exclude non-essential rotational, or oblique views.</td>
</tr>
<tr>
<td>Avoid assessing anatomy through practical examinations alone.</td>
<td>Play it safe by assessing learners through both practical and written examinations. Anatomy practical examinations require greater spatial aptitude and relying solely on this form of assessment may exacerbate the identified gender effects in favor of males. While preliminary evidence suggests there are no gender effects for anatomy performance, more evidence is needed to confirm this.</td>
</tr>
<tr>
<td>There is no need to assess spatial abilities.</td>
<td>It should be assumed that all learner populations will have some proportion of learners with low spatial abilities. The same content design strategies that benefit students with low spatial abilities also benefit average and high performers, negating the need for spatial ability testing.</td>
</tr>
</tbody>
</table>