Athens Journal of Sciences- Volume 9, Issue 4, December 2022 – Pages 257-272

Using Pixel Counts to Measure Accuracy and Confidence in Representations of Mental Maps

By Andrew T. Johnson *

Mental imagery and cognitive maps are difficult to study because these are so subjective and not easily observable. This exploratory study uses pixel counts to measure accuracy and confidence related to mental imagery. In this study 35 fifth grade students received an outline map of the continental United States on a standard-sized sheet of paper and asked to write state abbreviations as large as possible being 100% confident that the abbreviations would be within the boundary of the respective states. The response sheets were scanned. Adobe Photoshop was used to calculate pixel counts of the area of the abbreviations within and outside of the respective state boundaries. The ratio of In and Out pixel counts provided a measurement of Accuracy, while a ratio of In and Total State pixel counts provided a measurement of Confidence. More abbreviations were attempted for US states that had one or more sides present on the US outline map. The girls showed greater accuracy and higher confidence across all conditions. Similarly, there was a linear relationship between the number of reference sides and the confidence outcomes. The results provide proof of concept that pixel count measurements provide value to measuring mental imagery and spatial cognition.

Keywords: *mental imagery, cognitive maps, spatial cognition, accuracy, confidence index*

Introduction

Historical reference to mental imagery can be found in the Greek tradition with the account of Simonides using the Method of Loci to recall attendees who were killed by a collapse of a banquet hall by visually imagining where they were seated. More recently, cognitive psychologists have taken on the task of quantifying mental imagery. Some mental imagery processes are similar to visual processing (Kosslyn et al. 1978, Shelton and Gabrieli, 2004, Shepard and Metzler 1971). However, there are also systematic distortions within mental maps (Moar and Bower 1983, Thorndyke 1981, Taylor 2005, Tversky 1997, 2000).

Geography is the study of nature and relative arrangement of places. In the United States, fifth-grade students are expected to know the states that comprise the nation (Missouri Department of Elementary and Secondary Education 2016). Furthermore, Moore and Boehm (2011) argue that geography knowledge is vital for a person's politico-economic understanding and should be taught in grade school. Geography for school-aged children includes not only knowledge of the states, but the visual shape/ representation connected to the states. This includes identifying the names of states, information about the state, e.g., capital, prominent

*Professor, Park University, USA.

https://doi.org/10.30958/ajs.9-4-3 doi=10.30958/ajs.9-4-3

cities, area location, etc., as well as storing visual characteristics of the state to locate it on a map.

Literature Review

Maps are not only at the center of geography, but they influence the development of spatial cognition (Gauvain 2000). Maps are universal tools of thought and transcend cultures, as such their role is history is well established (Gauvain 2000). Mental images are not always accurate, but rather portray a world as a person perceives it to be. Moeser (1988) explored cognitive mapping of nursing students in a five-story hospital by having them draw out the floor they typically trained on in the hospital. Moeser found the even after two years of experience nursing students failed to form a global cognitive representation of the hospital. Moeser concluded that internalized cognitive representations are not automatically created and that motivational factors play a key role. If there is a value to building a cognitive map, the individual will do that.

The work of Marchette et al. (2011) used undergraduate participants at Johns Hopkins University to identify buildings on their Homewood campus and test out judgments of relative direction (JRDs). They found support that their participants organized and referenced their memory of campus from a particular point of reference, e.g., North-as-up. Additionally, having participants imagine the campus from another point of reference hindered performance. Marchette et al. (2011) findings provided evidence for the importance of point of reference, but avoided examining any gender effects in their study.

Several studies have examined sex differences for males and females. Chumney and Zinser (2011) examined the performance of undergraduates on identifying several geographical aspects of the United States. Participants were asked to label US states and major cities on outline maps, locate major cities, and reproduce state boundaries on an US outline map. The males and females varied in their performance. The male undergraduates displayed a performance advantage in some of the tasks, e.g., map-labelling and fragmented state identifications, and in context-dependent tasks of the United States. They point out in their article that the geographical limitations (coming from one region of the US) of their participants make their results difficult to generalize to other adults. Furthermore, they speculate that gender differences may be driven by experiences in grade school (Chumney and Zinser 2011).

Vecchi and Girelli (1998) explored gender difference across four experiments in visuo-spatial abilities by using 2D and 3D matrices of varying sizes and asking adult participants to complete specific tasks. For example, in Experiment 1 participants were played sets of direction statements and asked to indicate the final position on a 2D matrix or 3D cube. Males outperformed females in 2D and 3D tasks as well as active tasks across most of the interference manipulations Passive storage tasks did not show evidence for a male advantage (Vecchi and Girelli 1998).

Beatty and Troster (1987) collected data for over 1,800 undergraduates across five studies and several institutions. They asked participants to list locations within the US they had visited and to identify physical features, e.g., lakes, rivers, mountains, and locate cities on outline maps for regions of the United States. Their findings indicated that the male undergraduates outperformed their female counterparts on the tasks. However, the differences were not particularly robust. The males located places on a map of the United States more accurately than females. Overall, gender differences accounted for a range variance between 7% to 31% (Beatty and Tröster 1987). Are these sex differences present at the grade school level? If so, this may provide evidence that boys and girls may benefit from different instructional strategies to remove this difference (Beatty and Tröster 1987).

Current Study

The literature on spatial cognition and mental imagery has featured mostly research on adults and has revealed varying results. There is sufficient evidence that supports that individuals process and respond to mental representations similarly to actual visual stimuli (Kosslyn et al. 1978, Shepard and Metzler 1971). However, there have been conflicting results related to whether males and females handle mental manipulations differently. A previous study of mine (Johnson and Griffith 2016) investigated these topics through the use of a fragmented test of US states applied to grade-school students. Fifty fifth-grade students were given an outline map of the US and a fragmented test of US states. Each state set (44 of 48 continental US states) featured four-sized versions. One of the versions was scaled to fit in an outline map of the US given to each student. The fifth graders were instructed to label the state and use their mental imagery to choose which of the four selections was the appropriate size. For this phase of the study, the students answered a little over 31 of the 44-state size-estimations and identifications. The boys attempted significantly more states than the girls. The boys had higher accuracy for naming the states presented. However, for the accuracy of estimating the correct size, there was no difference between the boys and girls.

Additionally, there was a strong positive correlation between mean number of correct size estimation and mean number of correct state identifications. Those students who had higher numbers of correct state identifications also had higher accuracy of correct size estimations. There was also a strong positive correlation between percent accuracies (percent correct for size estimations and identifications). In other words, students who had a high percent of correct estimations also had high accuracy for state identifications.

The study reported in this article presents the second phase of the Johnson and Griffith (2016). In Phase 2 the students were asked to fill in the US outline map with state abbreviations. We were interested in examining the performance of boys and girls and how they would be able to perform and how “on target” their abbreviations would be. While Phase 1 involved identification of US states and imagining which state selection fit the outline map, Phase 2 involved more specific location of states with an added measurement of confidence. Participants can report confidence of mental images and mental maps, but how accurate are these ratings? Is there a way to objectively quantify mental map confidence? Self-report data may be flawed (King et al. 1980, Moore and Healy 2008) so more objective measurements are necessary. Several researchers have quantified many aspects of visual images/mental maps (Chumakov et al. 2021, Aram et al. 2019, Gardony et al. 2015, ELIAN (www.seldage.com)) and automated calculations such as point counts, line length, area, speed, duration, orientation, pen pressure, hesitations, etc. However, in many instances, these applications rely on human/expert programming.

In this study we are examining a performance measure of confidence. Instead of directly asking students to rate their confidence for each abbreviation attempt, we proposed framing the task to students to draw their abbreviations being 100 percent confident that they would place the abbreviations within the state boundaries. Then we apply pixel counts to calculate areas and derive a confidence measurement. We are interested in exploring how well they will do on this and whether any gender differences will be present.

Method

Participants

A total of 35 fifth-grade students (22 girls and 13 boys) from a Midwestern USA suburban grade school participated in this study.

Procedure

Each of the participants were given an answer sheet, outline reference map of the United States on a standard 8.5” x 11” sheet of paper formatted in a landscape position, and the test booklet in a class context. The study was conducted in two phases. Phase 1 consisted of completion a fragmented test of 44 continental states of the United States with two state sets per half page. The state set consisted of four choices varying from 40% to 160% of the actual state size that corresponded with the outline US reference sheet. Participants were told for each state set to write down the state abbreviation on the answer sheet then select one of the four choices that would accurately fit the reference outline using their mental imagery.

After Phase 1, Phase 2 consisted of having participants using the outline reference map of the US (see Appendix). They were asked to write state abbreviations in capital letters, e.g., MO, on an outline map of the US “as large as possible being 100% sure that the abbreviations would fall within the boundaries of the states (see Appendix for an example). A reference sheet with the abbreviations of the US states was provided. Phase 1 results were presented previously (Johnson and Griffith 2016). The results presented in this paper are for Phase 2.

Materials

We prepared the stimuli by first separately scanning the individual reference map sheets (300 dpi – black and white). Next, within Adobe Photoshop, a layer consisting of the outline US map with state boundaries was added to each individual map sheet. Following this, the Magnetic Lasso tool within Photoshop was used to wrap around the edges of each state abbreviation to make the most regularized shape possible. Using the Histogram outputs, pixel count totals were recorded for each abbreviation. An additional pixel count measurement was taken for abbreviations that were partially inside the appropriate state boundary. Then the pixel count measurements of the state abbreviation total and state abbreviation total within the state boundary were entered into a Google Form. The Google Form consisted of: The Participant Identification number, Sex, State Abbreviation, Pixel Count total, Pixel Count In (state boundary). A separate look-up table with US state pixel count totals was used to calculate Confidence (operationalized as Abbreviation pixel count Total/Abbreviation pixel count within the state boundaries). Accuracy was calculated as State pixel count Total/Abbreviation pixel count within the state's boundary. US states were sorted into groups related to the number of sides that are referenced on the outline map. An example of a three-sides state is Florida. Washington is an example of a two-sides state, whereas its neighbor Oregon is a one-side state example. Examples of zero-reference side states are Missouri, Kansas, Iowa, Nebraska, etc. - ones that are found away from the outside US Continental boundary line. A total of four states were classified as three-sided and five others as two-sided. Twenty-one states were classified as one-sided states leaving the remainder of 18 continental states as zero reference-sided.

Results

Our main research question examined the accuracy of depicting state abbreviations within the respective states and the confidence of the depiction reflected by the size of the state abbreviations. We were also interested in how measurements of accuracy and confidence would differ across the number of reference-sided states. Finally we were interested in investigating whether any sex differences were present.

On average females attempted 19.1 state abbreviations (Range = 3 to 40). On the other hand, males attempted 16 abbreviations on average (Range = 3 to 41). On average the females and males attempted a similar number of state abbreviations across state reference sides groupings, however the average attempt percentage declined across the number of state reference sides. For zero reference sides the mean attempts for females and males are 36.12% and 37.65%, respectively. This slightly increased for one and two reference sides (1 Side 38.74% Female; 39.19%, Male; & 2 Sides 46.59% Female, and 46.15% Male, refer to Figure 1). Finally, the females and males attempted more abbreviations for three reference-sides states – 61.36% and 57.69%, respectively (refer to Table 1). These averages indicate that states with perceptual cues on the outline map were easier to respond to. Moreover, the five of the six top attempted states contained two or three reference sides: Florida, Texas, Washington, Maine, and Michigan (For a listing of the states and the total attempts, see Table 2).

Figure 1. *Mean Number of Abbreviations Attempts across Gender and Number of Reference Sides*

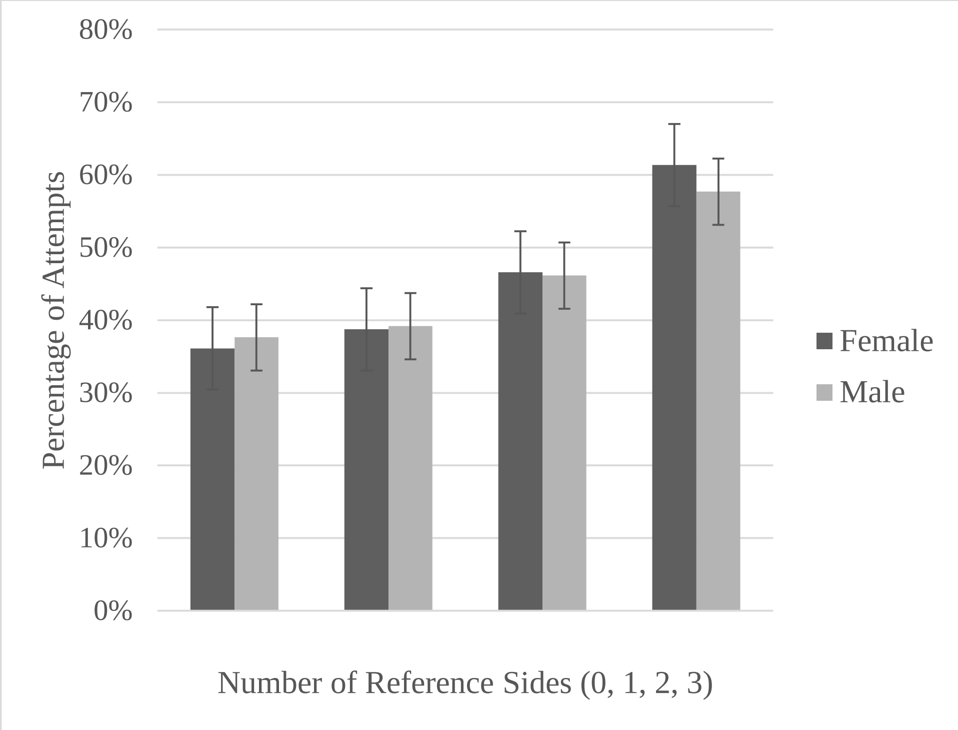

Table 1. *Average Attempt Percent for the Top Three States Attempted Across the Number of State Reference Sides*

Number of Reference Sides	Highest State/ Percentage	Second Highest State/ Percentage	Third Highest State/ Percentage
0	OK / 68.57%	CO & NV / 54.29%	MO / 51.43%
1	CA / 91.43%	AZ / 68.57%	MT / 60%
2	TX / 94.29%	WA / 88.57%	MN & WI / 51.43%
3	FL / 97.14%	ME / 82.86%	MI 71.43%

We used Abode Photoshop to calculate the pixel counts for the state abbreviations within (In) the respective state boundary and outside (Out) of the respective state boundary. Taken together the In and Out pixel counts (In/Out ratios) provide a measurement of Accuracy. The lower the Out-pixel counts are, the more accurate the estimate is. Overall, the mean state In/Out ratios ranged from 583627.900 (TX) to 0.051 (RI). (For a listing of the all of the continental US states and the Mean and Standard Deviations for In/Out total attempts, see Table 2). The outcomes for Texas are outliers because Texas is significantly larger than the other US continental states. The bottom border of Texas easily provides perceptual cues for depicting an abbreviation. Nevertheless, the data for Texas were included in this study. Aside from Texas and the outcomes for Arizona and Washington, the ratios for the remaining states are one digit or less.

Table 2. *Total Number of Abbreviation Attempts, Means and Standard Deviations for the Ratio of Pixel Counts Within the State Boundary to Pixel Counts Outside of the State Boundary, and Ratio of Pixel Counts Within the State Boundary to Total Pixel Counts of the State Across States*

	Ref Sides	Attempts	In/ Out Pixel Count Ratios		In/ State Total Pixel Count Ratios	
State	**N**	**N**	**M**	**SD**	**M**	**SD**
AL	1	13	2.544	4.940	0.080	0.091
AR	0	13	4.208	10.166	0.119	0.110
AZ	1	24	15.740	40.779	0.169	0.180
CA	1	32	1.816	5.503	0.157	0.148
CO	0	19	1.199	3.154	0.076	0.127
CT	1	5	0.134	0.300	0.145	0.199
DE	1	4	1.951	3.902	0.297	0.595
FL	3	34	2.189	4.256	0.181	0.144
GA	1	10	1.651	2.554	0.099	0.076
IA	0	14	1.029	1.841	0.151	0.137
ID	1	20	1.423	1.673	0.114	0.148
IL	1	17	2.172	4.049	0.132	0.187
IN	0	8	0.083	0.222	0.040	0.070
KS	0	17	1.566	4.163	0.076	0.095
KY	0	16	2.280	2.280	0.081	0.092
LA	1	20	7.002	9.163	0.187	0.126
MA	2	4	0.030	0.060	0.025	0.055
MD	0	5	0.429	0.959	0.102	0.229
ME	3	29	2.428	8.589	0.238	0.119
MI	3	25	8.794	25.347	0.236	0.155
MN	2	18	4.238	8.454	0.193	0.144
MO	0	18	2.357	3.460	0.155	0.122
MS	1	18	7.434	17.212	0.152	0.150
MT	1	21	4.719	12.875	0.126	0.088
NC	1	13	2.503	3.534	0.119	0.150
ND	1	14	4.133	10.199	0.113	0.108
NE	0	16	3.035	10.291	0.083	0.171
NH	0	5	0.052	0.116	0.026	0.058
NJ	3	9	1.119	2.765	0.138	0.237
NM	1	20	0.916	2.490	0.079	0.091
NV	0	19	1.879	5.661	0.099	0.080
NY	1	20	3.334	8.385	0.172	0.125
OH	1	16	0.111	0.354	0.210	0.151
OK	0	24	1.511	1.511	0.088	0.117
OR	1	17	3.329	8.291	0.189	0.146
PA	0	13	1.568	3.638	0.113	0.167
RI	1	9	0.051	0.153	0.115	0.346
SC	1	11	1.027	1.027	0.177	0.245
SD	0	12	1.730	2.418	0.160	0.140
TN	0	13	1.419	1.419	0.114	0.173
TX	2	31	583627.9	3248432	0.156	0.140
UT	0	10	3.518	9.957	0.106	0.089
VA	1	15	2.540	8.575	0.120	0.159
VT	1	10	0.452	0.803	0.138	0.211
WA	2	31	10.100	28.636	0.236	0.146
WI	2	18	2.543	9.985	0.110	0.129
WV	0	9	0.286	0.553	0.084	0.120
WY	0	13	2.554	6.601	0.063	0.081

In, Out, and Total Pixel Count Outcomes

We used a two-way between-subjects ANOVA to test any main effects and interaction of Gender and Number of Reference Sides on counts of pixels In, Out of state boundaries, and Total of pixel counts for abbreviations.

For the three-pixel count measurements (In, Out, Total) there were significant differences for Gender. For the In measurement, Females had significantly more pixel counts in their abbreviations, $F(1,666)=8.457$, $p=0.004$, partial $\eta 2=0.013$. Similarly, Females had more pixel counts in the their abbreviations that fell outside on the respective state boundaries, $F(1,666)=8.813$, $p=0.003$, partial $\eta 2=0.013$. Not surprisingly, Females also had more Total pixel counts in their abbreviations, $F(1,666)=13.495$, $p\sim 0$, partial $\eta 2=0.02$. Refer to Table 3 for Means and Standards for the In, Out, and Total measurements across Gender.

Table 3. *Means and Standard Deviations of Pixel Counts within and without State Boundaries as a Function of Gender and Number of Sides Referenced on the Outline Map*

			M	SD	N
In	Male	0	5854.82	6424.079	91
		1	7060.99	11218.264	109
		2	17446.91	23799.815	23
		3	7097.03	4813.746	29
		Total	7577.50	11459.68	252
	Female	0	7766.10	9810.431	142
		1	11743.37	13768.702	183
		2	24596.90	31449.206	39
		3	8750.84	6985.559	51
		Total	11222.64	15317.192	415
	Total	0	7019.64	8681.943	233
		1	9995.49	13054.537	292
		2	21944.48	28853.817	62
		3	8151.34	6303.793	80
		Total	9845.46	14087.716	667
Out	Male	0	4882.12	5109.069	91
		1	3713.30	5291.369	109
		2	4722.30	6634.953	23
		3	1732.55	4969.844	29
		Total	3999.52	5383.325	252
	Female	0	11132.50	12220.617	142
		1	5877.66	8195.041	183
		2	5186.77	9646.249	39
		3	2540.63	4836.646	51
		Total	7200.68	10044.238	415
	Total	0	8691.36	10499.036	233
		1	5069.73	7313.954	292
		2	5014.47	8596.151	62
		3	2247.70	4869.579	80
		Total	5991.25	8720.539	667
Total	Male	0	10799.42	5627.932	91
		1	10953.73	12125.961	109
		2	25569.57	25751.065	23
		3	8829.59	5519.090	29
		Total	11987.55	12456.987	252

	Female	0	18962.75	14084.222	142
		1	17620.86	14037.48	183
		2	29783.67	33674.817	39
		3	11284.41	8613.554	51
		Total	18444.33	16913.074	415
	Total	0	15774.49	12197.359	233
		1	15132.10	13720.322	292
		2	28220.37	30818.68	62
		3	10394.54	7691.821	80
		Total	16004.88	15688.001	667

The ANOVA results also yielded significant results for the In, Out, and Total pixel counts for the Number of Reference Sides, In $F(3,666)=17.544$, $p\sim 0$, partial $\eta 2=0.074$; Out $F(3,666)=11.578$, $p\sim 0$, partial $\eta 2=0.05$, and Total $F(3,666)=17.495$, $p\sim 0$, partial $\eta 2=0.072$. Utilizing Tukey post-hoc comparison significant differences were found between 2 Reference Sides and 0, 1, 3 Side conditions ($p\sim 0$) for In measurements. No other significant comparisons were indicated. For Out of State pixel counts and 0 Side condition, there were significant differences between the 0 and 1 Side conditions ($p\sim 0$), 0 and 2 ($p=0.01$), and 0 and 3 conditions ($p\sim 0$). For Out of State pixel counts and 1 Side condition, there was an additional significant difference between the 1 and 3 Side conditions ($p=0.035$). The Total pixel count measurement the Tukey post-hoc comparison revealed significant differences between the O and 2 Side conditions ($p\sim 0$), the O and 3 Side conditions ($p=0.027$), the 1 and 2 Side conditions ($p\sim 0$), and the 2 and 3 Side conditions ($p\sim 0$). Finally, the ANOVA results yielded a significant result for the interaction of Gender and Number of Sides on Out of state pixel counts $F(3,666)=3.913$, $p=0.009$, partial $\eta 2=0.018$. Females drew larger abbreviations for the 0 Reference Side states than the Males.

In/ Out State and In/ State Total Pixel Ratios Outcomes

We used a two-way between-subjects ANOVA to test any main effects and interaction of Gender and Number of Reference Sides for In/Out State pixel count ratios and In/State Total pixel count ratios. For the In/Out ratios and In/State Total ratios there were significant differences for Gender. For both measures, there was a main effect of Gender for In/Out Ratio $F(1,666)=5.881$, $p=0.016$, partial $\eta 2=0.009$, In/State Total Ratio $F(1,666)=7.575$, $p=0.006$, partial $\eta 2=0.011$.

There were main effects for Number of Sides for both measures, In/Out Ratio $F(1,666)=3.857$, $p=0.009$, partial $\eta 2=0.017$, and In/State Total Ratio $F(1,666)=8.096$, $p\sim 0$, partial $\eta 2=0.036$. The Ratios for 2 Reference Side states (principally Texas) elevated both ratios (refer to Table 4).

Tukey post-hoc comparisons revealed significant differences were found for In/Out Ratio between 2 Reference Sides and 0, 1 Side conditions ($p\sim 0$) and between 2 and 3 Reference Side conditions ($p=0.003$). No other significant comparisons were observed. For In/State Total Ratios, there were significant differences between the 0 and 1 Side conditions ($p=0.038$), 0 and 3 ($p\sim 0$), and 2 and 3 conditions ($p=0.039$). Finally, the ANOVA results yielded a significant result for the interaction of Gender and Number of Sides only for In/Out Ratios

F(3,666)=4.051, *p*=0.007, partial $\eta 2$=0.018. Males had significantly lower In/Out Ratios for 2 Reference Side states than the Females.

Table 4. *Means and Standard Deviations for In/Out Pixel Ratios and In/State Total Pixel Ratios across Gender and Number of State Reference Sides*

			M	SD	N
In / Out Ratio	Male	0	2.888	13.135	91
		1	3.716	10.977	109
		2	1.582	6.249	23
		3	3.948	17.530	29
		Total	3.249	12.323	252
	Female	0	1.976	7.196	142
		1	3.872	15.771	183
		2	150.468	643.543	39
		3	4.250	13.224	51
		Total	17.046	200.033	415
	Total	0	2.332	9.930	233
		1	3.814	14.153	292
		2	95.236	513.094	62
		3	4.141	14.819	80
		Total	11.834	158.036	667
In / State Total Ratio	Male	0	0.091	0.096	91
		1	0.116	0.180	109
		2	0.120	0.118	23
		3	0.188	0.130	29
		Total	0.116	0.146	252
	Female	0	0.125	0.154	142
		1	0.167	0.166	183
		2	0.160	0.157	39
		3	0.231	0.168	51
		Total	0.160	0.164	415
	Total	0	0.112	0.136	233
		1	0.148	0.173	292
		2	0.145	0.144	62
		3	0.216	0.156	80
		Total	0.143	0.159	667

Discussion

Measuring mental imagery is challenging. Chronometric studies (Kosslyn et al. 1978) and mental-rotation studies (Shepard and Metzler 1971) have significantly contributed to the understanding of mental visual imagery. More recent work (Friedman and Montello 2006, Friedman et al. 2002, Kerkman et al. 2003) has advanced the literature with implementing thoughtful measurement approaches. This study adds to the existing literature through the novel application of pixel counts to measure area. Pixel count calculations may also be used to create an index of confidence that does not rely on self-report.

The results from this study indicate that using mental imagery to make estimations is influenced by perceptual cues. The fifth-grade students attempted more abbreviations for states with sides on the outline map. There were no differences between the boys and girls for their abbreviations attempts across reference sides. Abbreviation accuracy was higher for states with sides referenced on the outline map. For the most part the boys and girls performed similarly across the reference-side states except that girls significantly outperformed boys for 2-reference sided states, more specifically for Texas. The girls were significantly more accurate than the boys. Moreover, the girls drew the abbreviations for Texas significantly larger than the boys.

Confidence as measured by the ratio of total pixels of the abbreviations to the total of pixels of the respective state. Similar to the accuracy measure, there was a linear relationship between the number of reference sides and the confidence outcomes. Having only one side referenced on the outline map improved the ratio of in-state pixel counts to the state pixel total. The highest ratios were in the states with three sides referenced on the outline map. Having a perceptual cue boosted confidence. The girls had higher confidence ratios across all four reference side conditions. This includes the zero-reference side condition where no perceptual cues were available. These outcomes call attention to mismeasures of female mental visuo-spatial cognition like previous research (Kerkman et al. 2000).

This study provides a proof of concept that pixel count measurements provide value to the area of mental imagery. The advantage of using a map of the continental United States is that the states can be sorted into different groups depending upon the number of reference sides on the outline map. That allows for useful comparisons. Overall, the boys and girls had a wide range of the abbreviations attempts (from 3 to 40/41) and the girls attempted more state abbreviations on average than the boys. The results for the average state abbreviations attempted show a distinct linear decline in attempts both in the boys and girls. Just over a third of the zero-referenced states were attempted while just under two-thirds of the 3-referenced sides states were attempted.

Furthermore, there were no significant differences between the boys and girls for the abbreviation attempts across referenced-side state groups. This indicates that the boys and girls found the task difficult especially for the internal states where there were no perceptual cues to the location of the state. This is particularly interesting because the participants were tested shortly after they completed their segment on states of the United States – learning about locations, capitals, and other information, e.g., state bird and motto. These results underscore that semantic memory and visuo-spatial memory operate independently (Shelton and Yamamoto 2009).

There were limitations in this study. The US states vary in size, spatial position, and the number of sides that are present on an outline map. The irregular outline afforded sets of comparison state side-referenced groups. An advantage of this is that comparisons may be made between states with no perceptual cues to states with one, two, and three perceptual cues (sides referenced). Then there is Texas. Texas is so large and the perceptual cues are so salient that the abbreviation pixel counts overshadowed the other states for the accuracy measure. Texas was the second most-attempted abbreviation after Florida. We debated removing the Texas data from analysis, but decided that those results were valuable to include. Nevertheless the post-hoc comparisons significances for accuracy and number of reference sides were due to the Texas data. The girls took advantage of perceptual benefits of Texas and drew significantly larger abbreviations than the boys and scored higher accuracy.

We did not include independent self-report measurements for confidence. While self-reported confidence measurements could be used to compare with our performance measure of confidence, self-report confidence may be inaccurate (Moore and Healy 2008). Adding in this estimate would add time and potentially distract from the activity. This study utilized a low-technology approach of paper and pencil with the fifth-grade students. The activity seemed like a game to many of them. Going forward, as this approach moves from exploratory to explanatory, it may be helpful to formalize data collection by online or application-based testing where more technological features may be implemented. If the task was programmed to allow participants to use a mouse or stylus to draw in the area, an easy way to indicate self-report confidence would be to allow participants to use color for confidence, e.g., blue for cold (low confidence) to red (high confidence). This could be correlated with the performance measurement of confidence.

A game or online application (a gamification approach) could offer players higher points for using more mental imagery, e.g., zero-reference side states. This could promote more attempts and could be used as a manipulation to observe response bias effects. For example, if the within-the state area reaches a certain value or if there is no area of the abbreviation outside of the state boundary, players earn bonus points. Perhaps, participants can provide different sizes of abbreviations, for example, a 95% Confident estimation and a 75% Confident estimation.

Moving forward, there have been significant development in software applications to analyze drawings and mental maps, e.g., Wallon's (n.d.) ELIAN; Chumakov et al.'s (2021) Creative Map Studio; Aram et al.'s (2019) Aram Mental Map Analyzer; and Gardony et al.'s (2015) Gardony Map Drawing Analyzer, yet all of these program rely on extensive programing focused on specific types of drawing outputs. Nevertheless, these examples demonstrate that automated scoring and analysis of drawings is possible. Perhaps it would be beneficial to apply this approach to more regularized maps/contexts other than using the continental United States with its idiosyncratic state shapes. Having regularized "states" or areas can simplify measurements. Additionally, having the testing materials based solely on using the United States is severely delimiting for generalizing the results.

Conclusions

Studies like the one presented provide a novel strategy to examine the accuracy and confidence of mental imagery and spatial cognition. Direct measures of confidence have been elusive and we offer performance-based solution. Furthermore, our study demonstrates that perceptual cues improves accuracy and increases confidence. Applying these strategies to online applications will connect to a broader population and take advantage of principles of gamification to motivate responses. Furthermore, the development of computational analysis programs show promise to automate the complexities and variations of hand-drawn maps.

References

Aram F, Solgi E, Higueras García E, Mohammadzadeh SD, Mosavi A, Shamshirband S. (2019) Design and validation of a computational program for analyzing mental maps: Aram Mental Map Analyzer. *Sustainability* 11(14) 3790–3810.

Beatty WW, Tröster AI (1987) Gender differences in geographical knowledge. *Sex Roles,* 16(Jun): 565–590.

Chumakov RV, Ryabinin KV, Belousov KI, Duan J (2021) Creative map studio: a platform for visual analytics of mental maps. *Scientific Visualization* 13(2): 79–93.

Chumney F, Zinser O (2011) Gender, memory, and geographic tasks. *Geographical Education* 24(11): 41–50.

Friedman A, Montello DR (2006) Global-scale location and distance estimates: common representations and strategies in absolute and relative judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(2): 333–346.

Friedman A, Brown N, McGaffey A (2002) A basis for bias in geographical judgments. *Psychonomic Bulletin & Review* 9(1): 151–159.

Gardony AL, Taylor HA, Brunyé TT (2015) Gardony Map Drawing Analyzer: Software for quantitative analysis of sketch maps. *Behavior Research Methods* 48(1): 151–177.

Gauvain M (2000) The instrumental role of maps in the development and organization of spatial knowledge. *Developmental Science* 3(3): 269–270.

Johnson A, Griffith R (2016, May) Fifth graders' recall and mental imagery for names and sizes of US states. Poster Presentation at the *2016 Annual Meeting of the Association for Psychological Science*. Chicago, IL.

Kerkman DD, Friedman A, Brown NR, Stea D, Carmichael A (2003) The development of geographic categories and biases. *Journal of Experimental Child Psychology* 84(4): 265–285.

Kerkman D, Wise J, Harwood E (2000) Impossible "mental rotation" problems: a mismeasure of women's spatial abilities? *Learning and Individual Differences* 12(3): 253–269.

King J, Zechmeister E, Shaughnessy J (1980) Judgments of knowing: the influence of retrieval practice. *American Journal of Psychology* 93(2): 329–343.

Kosslyn SM, Ball TM, Reiser BJ (1978) Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4(1): 47–60.

Marchette SA, Yerramsetti A, Burns TJ, Shelton AL (2011) Spatial memory in the real world: long-term representations of everyday environments. *Memory & Cognition* 39(8): 1401–1408.

Missouri Department of Elementary and Secondary Education (2016) *6-12 social studies grade level expectations*. Available at: https://dese.mo.gov/media/pdf/curr-mls-standards-ss-6-12-sboe-2016.

Moar L, Bower GH (1983) Inconsistency in spatial knowledge. *Memory & Cognition* 11(2): 107–113.

Moeser SD (1988) Cognitive mapping in a complex building. *Environment and Behavior* 20(1): 21–49.

Moore J, Boehm R (2011) Factors strengthening school geography's curriculum position in the nineteenth century. *The Social Studies* 102(1): 49–57.

Moore DA, Healy PJ (2008) The trouble with overconfidence. *Psychological Bulletin* 115(2): 502–517.

Shelton AL, Gabrieli JD (2004) Neural correlates of individual differences in spatial learning strategies. *Neuropsychology* 18(3): 442–449.

Shelton AL, Yamamoto N (2009) Visual memory, spatial representation, and navigation. In JB Brockmole (ed.), *The Visual World in Memory*, 140–177. Routledge.

Shepard R, Metzler J (1971) Mental rotation of three-dimensional objects. *Science* 171(3972): 701–703.

Taylor HA (2005) Mapping the understanding of understanding of maps. In P Shah, A Miyake (eds.), *The Cambridge Handbook of Visuospatial Thinking*, 295–333. Cambridge University Press.

Thorndyke PW (1981) Distance estimation from cognitive maps. *Cognitive Psychology* 13(4): 526–550.

Tversky B (1997) Spatial constructions. In NL Stein, PA Ornstein, B Tversky, C Brainerd (eds.), *Memory for Everyday and Emotional Events*, 181–208. Erlbaum.

Tversky B (2000) Remembering spaces. In E Tulving, FI Craik (eds.), *The Oxford Handbook of Memory*, 363–378. Oxford University Press.

Vecchi T, Girelli L (1998) Gender differences in visuo-spatial processing: the importance of distinguishing between passive storage and active manipulation. *Acta Psychologica* 99(1): 1–16.

Wallon P (n.d.) Seldage: https://seldage.com/.

Appendix

Blank Outline Map of the United States

Outline Map Sample with US State Abbreviations