Spring 2024

Examining Working Memory Precision Estimates as an Individual Difference

Stephen A. Campbell
University of Texas at Arlington

EXAMINING WORKING MEMORY PRECISION ESTIMATES AS AN INDIVIDUAL DIFFERENCE

by

STEPHEN CAMPBELL

THESIS

Submitted in partial fulfillment for the degree of Master of Science in Experimental Psychology
at the University of Texas at Arlington
May 2024

Supervising Committee:

Matthew Robison, Supervising Professor
Hunter Ball
Tracy Greer

Abstract

Working memory (WM) capacity has remained a central topic of individual differences research due to its ability predict performance in various cognitive domains and higher-cognitive abilities. While studies performing individual differences research with WM capacity are common, there are comparatively few studies investigating individual differences with WM precision. The present study examined WM precision as an individual difference, by examining the psychometrics of the modeled precision parameter derived from the Standard Mixture Model, investigating the relationship between precision for different feature types (e.g., color and spatial location), and looking at the relationship between precision and a well-known correlate of WM capacity: fluid intelligence. In two research studies, we found a significant positive correlation between spatial WM precision and fluid intelligence, but not color WM precision. Additionally, we found no significant correlation between color and spatial WM precision and that a latent factor model which loads both precision parameters into one factor did not fit well. However, an examination of the intratask reliability of the WM precision estimate via split-half estimation of the parameters revealed low reliability for the color version of the task, but not the spatial version. WM capacity estimates from the continuous-report task were found to have significant positive correlations with each other, with fluid intelligence, with measures of capacity in other tasks, load well onto a single factor, and were reliable.

Keywords: working memory, capacity, precision, individual differences, fluid intelligence

TABLE OF CONTENTS

Examining Working Memory Precision Estimates as an Individual Difference

Working memory (WM) refers to a short-term storage system which enables the active maintenance of information at the service of ongoing tasks. This system has typically been characterized by its limited capacity, which some behavioral and neural studies suggest being around 3~4 items (Cowan, 2001; Luck & Vogel, 1997). This storage limitation of WM has remained one of the key topics of interest in the cognitive field due to correlations between individual differences in WM capacity and performance with a variety of higher order cognitive abilities and tasks such as reading comprehension, language acquisition, problem solving, and fluid intelligence/reasoning (Daneman & Green, 1986; Kyllonen & Christal, 1990; Fukuda et al., 2010; Unsworth et al., 2014). While capacity has remained of central interest, recent debate regarding the nature of capacity limitations has led to the creation and use of WM recall tasks and models which provide not only a measure of WM capacity, but also a measure for the quality/precision of these WM representations.

The continuous-report task, also known as the delayed-estimation task, involves participants viewing an item array that appears briefly and, after a variable interval delay, the location of a single target item will be probed and the participant will be required to recall and report to the best of their ability the precise feature of the target item to its location on a continuous response field (Prinzmetal et al., 1998; Wilken & Ma, 2004). For example, a common form of the task will present a set of color squares on the screen for a brief interval, followed by a blank delay interval, and then the location of a single square from the previous array will be probed. Participants must recall and report the precise color of the probed item and respond by making a response on a continuous color wheel (Zhang & Luck, 2008). The task is thought to provide a continuous estimate of the quality, or precision, of the memory item. The raw score variable is calculated by recall performance, which is the angular deviation between the participant's response and the true location of the target item's color on the continuous wheel. Put simply, if memory for the target item feature is low then the participant's response will be farther away from the target item (e.g., high angular deviation) and vice versa if memory is high. Using circular statistical analysis, we can analyze the distribution of errors for recall performance. The distribution of responses then can be used to estimate a person's average distance from perfect memory for an item. Precision is calculated as the reciprocal of the circular standard deviation of error responses (Bays & Hussain, 2008). However, this measure of precision assumes that all items are remembered in a given array, when it may be the case that some of these trials may also include guesses when no memory of the item is present. Thus, this model-free estimate of precision presents an issue of distinguishing between the source of memory error, such as when participants are making responses based on a valid but noisy internal representation of the item in memory vs. making a response based on a guess due to recall failure or failure to encode. Additionally, a third possibility is the incorrect report of a non-probed target (i.e., "swap error", Zhang & Luck, 2008; Bays et al., 2009). This has led to the development of mixture models which take into account these behaviors and quantifies them to provide independent measures of WM precision and capacity. Thus, these models attempt to separately parameterize quantity and quality of WM representations.

Much of the work utilizing these mixture models has been to investigate the nature of WM capacity. The debate of this issue has led to the establishment of two camps: one heralding a discrete resource, or fixed-limit, slots hypothesis of WM and another positing a flexible, continuous resource allocation hypothesis (for review, see Ma et al., 2014). Discrete resource models suggest that resources are allocated to only a select few items that can be held at a given time up to a certain limit with items that are failed to be stored when the number of items to be remembered exceeds one's capacity. The items held in memory have a fixed resolution, or precision, independent of the number of items being maintained. This has largely been supported by research which finds that performance declines as the number of items to be remembered increases in WM tasks (Luck and Vogel, 1997; Fukuda et al., 2010). Alternatively, the flexible, continuous resource models suggest that we do not have a fixed limit on the number of items we can distribute resources to. Instead, we distribute resources to all items to be remembered and that when we must remember a larger number of items, each item is given fewer resources which is reflected in worse memory (Bays et al., 2009). While each camp offers different predictions of the nature of WM capacity, the debate of this issue has led to two prominent models for examining data for the continuous report task.

One of the prominent mixture models is the two-component model, or standard mixture model, developed by Zhang and Luck (2008). This model assumes that responses in the task are a probabilistic mixture of the two response behaviors discussed previously: responses based on a noisy memory representation of the target feature and random guessing. If the representation of the target feature is held in memory, then the response will be close to the target. Errors are modeled as a Von Mises Distribution, which is a normal distribution for circular data, centered around the true target value. When the representation is not stored or there is a recall failure, then no feature information is available, and the participant must make a random guess. The error distribution is thus represented as a uniform distribution around the target item. Essentially, the model uses the participant's response value, the true value of the target, and the probability of giving a uniform, guessing response to provide the following parameters: P_u, the probability of a guessing response; P_t, the probability of a target response; and a concentration parameter, κ

(Grange & Moore, 2022). P_t provides an estimate of the *capacity* of WM (i.e., average number of valid representations), whereas κ provides a measure of *precision*, with higher values reflecting more precise memory representations. Results from studies using the two-component model have typically been used as evidence supporting the discrete resources camp (Zhang and Luck, 2008; Zhang and Luck, 2011; Adam et al., 2017). For example, Zhang and Luck (2008) found that P_t was twice as high at set sizes of three compared to set sizes of six, with no significant change in precision. Thus, they argued that we may hold a fixed and small number of item representations with high precision and that, when our limit in WM is reached, other items and their features are not stored at all. Rather, responses to unremembered items are just guesses, consistent with the idea of a discrete resource, fixed-capacity model. Another prominent mixture model is the three-component model, or swap model (Bays et al., 2009), which assumes that not only are responses a mixture of the previously discussed behaviors, but also responses to a non-target feature value. The model considers this a 'swap error', which results when a participant makes a response for a non-probed target. Results from studies using the three-component model tend to lend support for the continuous resource camp and have typically found that we see further decreases in precision as set size increases, consistent with the idea that we may hold all items in memory and these representations are less precise due to needing to allocate more resources to each item.

Both the two-component and three-component models are meant to provide a parameter estimate for precision *when the target is in memory*. While each model offers different predictions for how participants make responses to the continuous report task, they do agree that the task itself is useful in WM research as it provides an estimate of precision that tend to fit the data well (Ma et al., 2014). Likewise, the models agree that precision in WM is variable, with some items represented as clear and precise and others containing little to no information, even if the exact theory behind whether an item-limit which caps precision is divisive (Zhang and Luck, 2008; Bays et al., 2009; Fougnie et al., 2012). Beyond studies investigating competing models and item-limits, some other research on WM precision has focused on the development and growth of WM precision in childhood (Heyes et al., 2012), WM precision in patients suffering from neurological disorders which impair WM (Gold et al., 2010; Xie et al., 2018), WM and long term memory precision (Brady et al., 2013, Bidermen et al., 2018), and more. Research in the area has also expanded on the distinction between capacity and precision in WM, such as evidence for distinct neural correlates of WM precision (Zhao et al., 2020) and studies finding distinct factors which independently affect or relate to precision and not capacity such as negative affect (Xie et al., 2017), perceptual expertise (Lorenc et al., 2014), and lure discrimination (Xie et al., 2024).

However, despite the numerous experimental studies examining WM precision, there is a distinct lack of individual differences research. While the relationship between WM capacity and higher-order cognitive abilities has been thoroughly examined, the same relationships have been scarcely been demonstrated with WM precision. That is, while individual differences in the quantitative aspect of WM have been well-researched, there are comparatively fewer studies for the qualitative aspect of WM. Given the interest in WM capacity and its relationship with a variety of higher-order cognitive abilities and other outcomes, one might expect that a task that provides both a measure of WM capacity and precision to be quite useful for individual-differences research. Indeed, the capacity estimates provided by the mixture models have been found to be related to some of the other higher-order cognitive abilities like fluid intelligence as well as the capacity estimates from other WM tasks (for example, see Unsworth et al., 2014). While mixture models seem to provide an estimate of WM precision when the target is in memory, and research suggests that precision is a distinct aspect of WM from capacity and can be measured independently, the lack of individual differences studies using the modeled WM precision estimates from the continuous report task suggests an open question of the use of these parameters in individual differences research. The current study seeks to contribute to this growing area of study by examining two independent datasets which had participants complete a battery of WM tasks and a fluid intelligence task.

Current Study

The current study investigates the psychometrics of the WM precision parameters, the relationship among precision parameters attained from continuous report tasks with different features, and the relationship between these parameters and a known correlate of WM, fluid intelligence.

Measurement of reliability is essential for individual differences research to perform correlational analyses. If the reliability of our precision estimates were found to be too low, this would not allow us to draw conclusions about the relationships (or lack thereof) in the data. Thus, our first goal was to examine the reliability of the modeled parameters from the continuous report. We predict that the modeled precision parameter, κ, will be a reliable measure such that we expect results from a reliability analysis where we will split trial data in half (odd/even trials) and examine the correlations between the precision parameters using Spearman-Brown split-half correction to the correlation to be above an acceptable level of 0.70. A response to a meta-analysis by Bidermen et al. (2019) examined the reliability of model parameters by using a bootstrap method, where for each subject and experiment, they generated pairs of the capacity and precision estimates based on resampling the trial-by-trial data with replacement over 100 times and used the mean correlation as the reliability estimate (see suppl., Xie et al., 2020). They found that the reliability of the modeled precision parameter, represented by the circular standard deviation of the Von Mises distribution around the target item, after 100 resampling iterations to be roughly ~0.70. They argued that their reliability values were reasonable, as they are not particularly different from the psychometric properties of major personality scales. Thus, though our methods to assess reliability differ, we expect to find at least similar levels of reliability.

Once we confirm our measures are reliable, we can then tackle another question of assessing the feature-generality of WM precision. That is, do separate continuous report tasks requiring the storage and accurate recall of different stimulus features (i.e., color and spatial location) have significant differences in precision performance? And what are the relations among these modeled precision parameters? Continuous report tasks have been used with a variety of stimuli ranging from color, spatial location, shape, oriented bars, motion, and even faces (Zhang & Luck, 2008; Bays & Hussain, 2009; Lorenc et al., 2014). While there are several versions of the continuous report task, there remains a question of how performance with one version of a task (e.g.., color) relates to performance with another version of the task using a different feature (e.g., spatial location). Research within WM capacity seems to support a domain-general view of WM capacity, such that performance in WM tasks that measure separate domains such as verbal and visuo-spatial are highly correlated and when loaded together under a single construct in confirmatory factor analyses, the resulting construct is a strong predictor of higher order cognitive ability (Kane et al., 2004). Likewise, as mentioned earlier, the capacity measures of continuous report tasks have been found to have significant positive correlations with one another and are able to fit well in a latent variable model which loads them under a single capacity construct (Unsworth et al., 2014). This suggests that performance on continuous report tasks with different stimulus features are predictive of each other. While precision and capacity are independent measures, we nevertheless predict that the modeled precision parameters will have significant positive relationship with one another, suggesting precision is likewise domain general. Due to this, we expect that when loading precision estimates from each task to a single precision construct in a latent variable model, the model will have good fit, allowing us to examine the relationships between the latent factors (see Appendix A for a visualization of this model).

The final aim of our study seeks to address the relationship between WM precision and fluid intelligence, a higher-order cognitive ability that is a well-known correlate of WM capacity. While the use of the modeled precision parameters from the continuous report tasks are sparsely used in individual differences research, one study did examine the relationship between working memory quantity (capacity) and quality (precision) with fluid intelligence using a change detection paradigm. Change detection tasks are a well-known WM paradigm which have participants view a stimulus array and, after a brief blank screen delay, they are presented with another stimulus array and asked whether a cued item among the new array was the same or different from the original. The primary dependent variable is a measure of working memory capacity, *K*. In their study, Fukuda et al (2010) measured the precision of WM using a capacity measure from a version of this task where the stimuli were four shapes (two ovals and two rectangles) with lines. After being presented a sample array of the four shapes, participants would have to report after a delay whether the cued shape was the same or different as before. However, the conditions were either a large change (i.e., when an oval changed to a rectangle) or a small change (i.e., changes from one oval to another or from one rectangle to the other). The authors reasoned that to make a successful change judgment in the small change trials, participants would have to have held the cued item with sufficient resolution to discriminate it from the other similar items. Thus, higher capacity on small change trials would be reflective of better WM precision. Using factor analysis to load the separate capacity measures from the large and small change trials to number and resolution factors, respectively, they were able to demonstrate that indeed capacity and precision were distinct aspects of WM. However, the resolution factor and the individual precision estimates did not have a significant relationship with fluid intelligence. A replication of the results of this study has yet to be conducted with the continuous report task. An advantage of using modeled parameters from the continuous report task is that we can distinguish between WM capacity and precision, which you cannot do when you use a single capacity estimate as in Fukuda et al. and can thus investigate the potentially unique contributions and relationships that WM precision has with other variables, independent from WM capacity. Unsworth et al. (2014) utilized several versions of the continuous report tasks in their study, but only used the modeled capacity, P_t * set size, value in their analyses. This project investigates the relationship between the modeled WM precision estimates from the continuous report task and fluid intelligence task performance. We predicted that we will replicate the findings from Fukuda et al. and that whereas WM capacity will correlate with fluid intelligence, WM precision will not.

Study 1

Study 1 was an individual differences investigation of WM capacity, WM precision, and fluid intelligence utilizing multiple measures for each variable of interest to examine the relationships between these variables. Pupillary data was also collected but will not be analyzed in the proposed project.

Method

Participants and Procedure

All participants (N = 223; 148 women, 71 men, 4 non-binary, M_{age} = 19.58, SD_{age} = 2.11) were undergraduate students at the University of Texas at Arlington, who participated in the study in exchange for course credit. All but six participants reported normal or corrected-to-normal vision. After signing an informed consent document, participants completed a demographics questionnaire and then a battery of cognitive tasks. In the two-hour session, participants completed three change detection tasks (color, orientation, and letter) measuring WM capacity, three continuous report (color, orientation, and space) tasks measuring WM precision and capacity, and three tasks (Raven Advanced Progressive Matrices, number series, and letter sets) measuring fluid intelligence. Participants finished the study by completing a strategy questionnaire. Baseline pupil size measurements were also measured at the beginning of the study session though these data will not be analyzed here, and have been reported elsewhere (Robison & Campbell, 2023). All participants completed the tasks in the following order: color change, color continuous, Raven, letter change, space continuous, letter sets, orientation change, orientation continuous, and number series (Robison et al., 2024). The experimental protocol was approved by the Institutional Review Board of the University of Texas at Arlington.

Tasks

Color change detection (Luck & Vogel, 1997)

Target items were colored squares that subtended 3° of visual angle each. Target items appeared in six preselected locations spaced equally around the center of the screen. Colors were sampled randomly from a continuous HSV space, with each color being at least 30° apart in the HSV space. Each task involved participants staring at a fixation cross on the center of the computer screen against a gray background for 1,000 ms. Afterward, four color squares appeared for 250 ms during an encoding phase followed by a 1,000 ms blank screen delay. After this delay, the four squares would reappear with the one item circled by a black ring. Participants' task was to indicate whether the circled item was the same or different color as in the encoding phase by pressing the 'F' or 'J' keys on a keyboard to indicate 'same' or 'different', respectively. After a response, the next trial would begin after a 500-ms blank screen delay. Participants completed 6 practice trials with accuracy feedback, then 100 experimental trials without accuracy feedback. For example, see Figure 1A. The variable measured was working memory capacity estimate *k* (4 * [hit rate – false alarm rate]; maximum score = 4).

Orientation change detection (Luck & Vogel. 1997)

Target items were black oriented bars that subtended 3° of visual angle each. Target items appeared in six preselected locations spaced equally around the center of the screen. Orientations were sampled randomly from a continuous orientation space (0 – 180°), with the requirement that each orientation was at least 30° apart. Each task involved participants staring at a fixation cross on the center of the computer screen against a gray background for 1,000 ms. Afterward, four bars appeared for 250 ms during an encoding phase followed by a 1,000 ms blank screen delay. After this delay, four bars would reappear with the target item containing a white dot at its center. Participants' task was to indicate whether the target item was the same or different orientation as in the encoding phase by pressing the 'F' or 'J' keys on a keyboard to indicate 'same' or 'different', respectively. After a response, the next trial would begin after a 500-ms blank screen delay. Participants completed 6 practice trials with accuracy feedback, then 100 experimental trials without accuracy feedback. For example, see Figure 1B. The variable measured was working memory capacity estimate k (4 * [hit rate – false alarm rate]; maximum score = 4).

Letter change detection (Robison & Brewer, 2020)

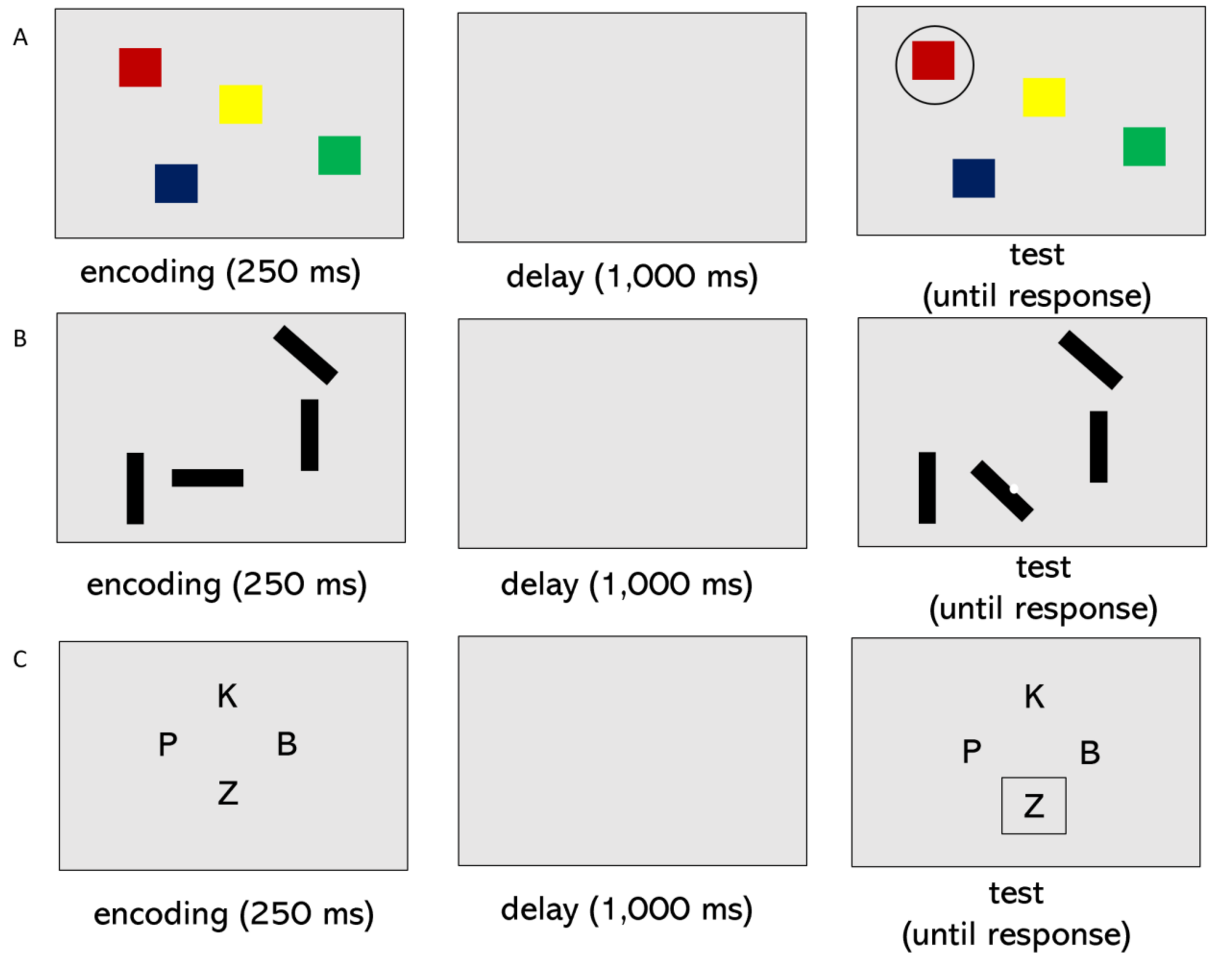
Target items were letter stimuli that subtended 3° of visual angle each. Target items appeared in six preselected locations spaced equally around the center of the screen. Letters were sampled randomly from English consonants. Each task involved participants staring at a fixation cross on the center of the computer screen against a gray background for 1,000 ms. Afterward, four letters appeared for 250 ms during an encoding phase followed by a 1,000 ms blank screen delay. After this delay, four letters would reappear with the one item surrounded by a black box. Participants' task was to indicate whether the tested item was the same or different letter as in the encoding phase by pressing the 'F' or 'J' keys on a keyboard to indicate 'same' or 'different', respectively. After a response, the next trial would begin after a 500-ms blank screen delay. Participants completed 6 practice trials with accuracy feedback, then 100 experimental trials without accuracy feedback. For example, see Figure 1C. The variable measured was working memory capacity estimate k (4 * [hit rate – false alarm rate]; maximum score = 4).

Color continuous report task (Zhang & Luck, 2008)

Four colored squares were presented simultaneously for 250 ms at encoding. Colors were randomly selected from a continuous HSV space of 180 colors that were evenly distributed on a circle in the CIE Lab color space. After a 1000-ms blank delay, a single target was probed by a black outline of a square which appeared in the same location on the encoding phase. Participants were instructed to report with as much accuracy as possible what the color of the probed target was on a color ring compromising 180 colors. The color ring and target probe remained on screen until response was made. Participants completed 6 practice trials with one

Figure 1

Example Trial of Change Detection Task



Note. A) Change detection task for color features. B) Change detection task for oriented bars. C) Change detection tasks for letters.

square and were provided accuracy feedback, then 100 experimental trials with four squares without feedback. For example, see Figure 2A. Variables measured include a capacity parameter indicated by the probability of target item in memory, P_t, and precision parameter, κ, both of which are given by the Standard Mixture Model (Zhang & Luck, 2008).

Space continuous report task (Unsworth et al., 2014)

Four letter stimuli were presented simultaneously for 250 ms at encoding. Letters were randomly selected from all consonants and locations were set to six preselected locations on the visual array with up to 30° of jitter on each trial. After a 1000 ms blank delay, a single letter which appeared during the encoding phase was presented at the center of the screen around a circular response ring. Participants were instructed to report with as much accuracy as possible the location of the letter in the previous encoding phase on the ring by moving a response probe around the circle and clicking to select the location they believe the target previously appeared. The response circle stayed on the screen until participants made a response. Participants completed 6 practice trials with one letter and were provided accuracy feedback, then 100 experimental trials with four letters without feedback. For example, see Figure 2B. Variables measured include a capacity parameter indicated by the probability of target item in memory, P_t, and precision parameter, κ, both of which are given by the Standard Mixture Model (Zhang & Luck, 2008).

Raven Advanced Progressive Matrices (Raven, 1965)

On each trial, a 3 x 3 patterned matrix appeared with the bottom-right piece of the pattern missing. The participants' task was to select from eight possible options the piece that best completed the implicit pattern in the matrix. Participants had 10 minutes to complete as many of the 18 problems as possible.

Number series (Thurstone, 1938)

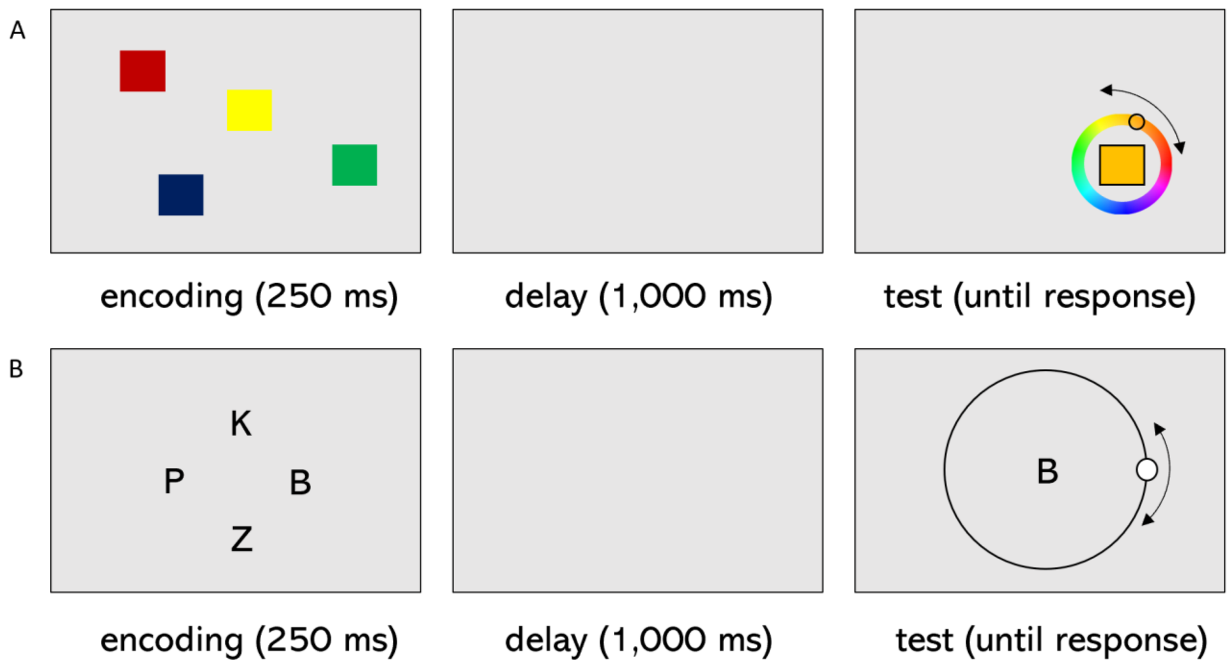
On each trial, a sequence of numbers appeared, and the participants' task was to select from a set of five possible options the number that best continued the sequence. Participants had 4.5 min to complete as many trials of the 15 problems as possible.

Letter sets (Ekstrom et al., 1976)

On each trial, a set of four different four-letter sets appeared. Among the sets, three of the four sets followed an implicit rule, and one of the four sets violated this rule. The participants' task was to select the set of letters that violated the rule. Participants had 5 minutes to complete as many trials of the 20 as possible.

Figure 2

Example Trials of the Continuous Report Task



Note. A) Continuous report task for color features. B) Continuous report task for spatial location feature.

Strategy Questionnaire

Participants completed a strategy questionnaire regarding the use of certain strategies for the change detection, continuous report, and raven tasks. For the change detection and continuous report tasks, participants were asked "In the task shown above, what strategy did you use to complete the task? and were asked to select between four options with: 1) I assigned a verbal label to each color/location/bar to help me remember the during the delay, 2) I tried to maintain a visual image of the items ‘in mind’ during the delay, 3) I did both, or 4) I did neither. Answers to these questions were coded as the following: 1) Verbal label, 2) Visual image, 3) Both, and 4) Neither. Participants were also asked the same question for the Raven task and selected between four options: 1) I used a process of elimination to find the piece among the 8 options that seemed like it best fit the pattern, 2) I tried to visualize the solution in my head, then find a piece that matched that solution in the options, 3) I did both of those, 4) I did neither of those. Answers to these questions were coded as the following: 1) Response Elimination, 2) Constructive Matching, 3) Both, and 4) Neither.

Data Analysis

The data were analyzed in R (R Core Team, 2022) using the *tidyverse* (Wickham et al. 2019), *psych* (Revelle, 2015), *data.table* (Dowle & Srinavasan, 2020), *lavaan* (Rosseel, 2012), and *mixtur* (Grange & Moore, 2022) packages.

To examine the data from the continuous report task, we fit the Zhang and Luck (2008) standard mixture model using the *mixtur* package in R. Recall error (in radians) is calculated on every trial as the angular deviation between the participants’ response and the true target feature location on the continuous response wheel. As discussed before, the mixture model assumes two sources of error contribute to performance: variability, or noise, in reporting the correct feature when the item is in memory or successfully retrieved and the proportion of trials where retrieval fails resulting in a guess. The mixture model fits the response errors with a mixture of two distributions. One distribution is a Von Mises distribution centered around the true target value, with a concentration kappa (κ) parameter. The κ parameter captures the variability in successful target retrieval, with higher values reflecting better precision. Another distribution is a circular uniform distribution with a probability of P_u, representing the probability of a guess response. Thus, the probability of a target response is represented as P_t, or $1 - P_u$.

In many studies using mixture modeling, the MATLAB MemToolbox (Suchow et al., 2013) has been utilized to estimate model parameters. Considering this, we examined the relationship between the model estimated parameters from *mixtur* and MemToolbox to make sure the parameters we used would be equivalent whether we used either program. We found the color and space modeled parameters to be near identical ($r \geq .97$, see Appendix B). However, while we also observed the same correlation between our orientation modeled parameters, the values of κ and P_t were numerically different, unlike the color and space modeled parameters which were numerically similar. In our data, the modeled guessing parameter, P_u, was roughly double in mixtur than in MemToolbox, leading to different P_t estimates from both programs. Orientation data from continuous response task is different from the color and space versions because the feature space is 180°, rather than 360°, and therefore response error must be +/- 90°, rather than +/- 180°. In each program, this can be specified when running the model. Thus, this issue may be due to a difference in how each program utilizes the response error data when it is specified to have come from an orientation version of the task. Therefore, despite the similarity in rank-ordering of the orientation data, we have decided to exclude these data from the present study.

A correlation analysis was conducted to examine the relationship between variables. For the latent variable analyses, we specified a model where the three change detection tasks load onto a WM capacity factor *K*, the two capacity measures from the continuous tasks to load onto a WM factor (P_t), the two precision measures from the continuous tasks to load onto a factor κ, and the three fluid intelligence tasks to load onto a factor (gF) (see Appendix A1). For all measures, reliability was estimated using a split-half (odd/even trials) and the Spearman-Brown split-half correction applied to the correlation. We screened the data for outliers and excluded any data points that fell outside +/- 3 standard deviations of the mean. The number of those excluded on each task is listed on the descriptive tables.

Results

Table 1 lists descriptive statistics for each measure and Table 2 lists zero-order correlations between dependent variables. We first examined the reliability of our tasks, with particular interest in the mixture model parameters. Measures of working memory capacity from the change detection tasks (K) and capacity estimates from the continuous report task (P_t) showed acceptable to good reliability (split-half corrected correlations > .70). Measures of fluid intelligence showed acceptable reliability (split-half corrected correlations > .70). However, for the modeled precision estimates κ, we found only the spatial estimate of the continuous report task showed acceptable reliability (split-half corrected correlation > .70) while the color estimate had poor reliability (split-half corrected correlation < .10).

Next, we examined the relationship between parameters from the working memory tasks to examine their domain generality, with particular interest in the interrelatedness of the continuous report tasks. As expected, we found that WM capacity, K, from the change detection tasks had significant positive correlations with the capacity estimate from the color continuous report task, P_t. Similarly, we found significant positive correlations between change detection K and spatial continuous report P_t. Examining the correlations between P_t and K in all tasks reveals significant high positive correlations with each other. These correlations suggest that those who performed better on a change detection task tended to have better probability of recall for target memory items in the continuous report task. Consistent with prior research, all capacity measures had significant, positive correlations with all measures of fluid intelligence, such that those who performed better on the change detection tasks or had a higher probability of recall for target memory items in the continuous report tasks tended to have better fluid intelligence task performance.

Table 1

Descriptive statistics for Study 1

Measure	N	M	SD	Skew	Kurtosis	# Excluded
Color K	214	2.50	0.66	-0.83	0.96	9
Letter K	215	2.00	0.80	-0.37	-0.43	8
Orientation K	213	1.88	0.80	-0.39	-0.53	10
Color P_t	213	.67	.22	-1.07	0.38	10
Space P_t	214	.79	.13	-1.21	1.28	9
Color κ	213	15.4	12.69	2.36	6.28	10
Space κ	214	29.4	17.93	1.27	1.76	9
Raven	222	8.58	3.41	-0.24	-0.51	0
Number series	222	7.32	2.88	0.01	-0.49	0
Letter sets	220	8.72	3.12	0.13	-0.53	0

Examining correlations between continuous report task estimates revealed a significant positive correlation between color P_t and spatial P_t, suggesting that those who had a higher probability to recall target color items tended to have higher probability to recall target spatial location of items. However, we found no significant correlation between color κ and spatial κ, suggesting that those who had better precision for target color items when the item was in memory, did not tend to have better precision for target item's spatial location when the item was in memory. To further assess continuous response task interrelatedness, we performed a confirmatory factor analysis where we specified a latent variable model allowing the precision

Table 2

Zero-Order Correlations for Study 1

	1	2	3	4	5	6	7	8	9	10
1. Color k	*0.78*									
2. Letter k	**0.47****	*0.82*								
3. Orientation k	**0.46****	**0.53****	*0.83*							
4. Color P*t*	**0.55****	**0.38****	**0.43****	*0.84*						
5. Space P*t*	**0.40****	**0.63****	**0.59****	**0.45****	*0.82*					
6. Color κ (precision)	-0.02	0.00	-0.03	**-0.46****	-0.13	*0.05*				
7. Space κ (precision)	**0.20****	**0.22****	**0.31****	**0.21****	0.12	-0.06	*0.78*			
8. Raven	**0.30****	**0.33****	**0.36****	**0.39****	**0.26****	**-0.16**	**0.25****	*0.75*		
9. Letter Sets	**0.21****	**0.26****	**0.36****	**0.33****	**0.37****	-0.12	**0.14**	**0.26****	*0.79*	
10. Number Series	**0.24****	**0.25****	**0.17**	**0.27****	**0.21****	-0.12	**0.16**	**0.42****	**0.46****	*0.76*

Note. Bold indicates significant at $p < .05$. * - indicates significance at $p < .01$. ** - indicates significance at $p < .001$. Split-half reliabilities are listed along the diagonal estimates to load onto a factor, the three change detection capacity estimates to load onto a factor, the two continuous report capacity estimates to load onto a factor, and the three fluid intelligence tasks loaded onto a single factor had poor fit, χ^2 (29) = 127.09, $p < .001$, CFI = 0.80, RMSEA = .14, 90% CI = [0.11, 0.16], SRMR = .10). An examination of the model parameters reveals that our precision estimates did not load onto a factor, likely due to the lack of a correlation between variables, leading to the poor model fit.

We next examined the correlations between precision estimates and performance on the fluid intelligence tasks to determine whether the pattern of correlations were similar to those found in Fukuda et al (2010). We found that spatial precision had a significant positive correlation with the Raven fluid intelligence task. Likewise, there was also a small significant, positive correlation with letter sets and number series tasks. Thus, these data suggest that those who had better location precision for target items tended to have better performance in fluid intelligence tasks. Conversely, there was a significant, negative correlation between color κ and Raven task performance. However, we found no significant correlation between precision estimates from the color continuous report task and the letter sets and number series tasks.

Finally, we examined responses from the strategy questionnaire. Descriptive statistics can be found in Appendix C. Two one-way ANOVA tests were conducted to examine the relationships between average precision (κ) and probability of target retrieval in memory (P_t) with reported strategy type. We found there were no significant differences between average precision and reported strategy type, $F(2, 194) = .08$, $p = .93$. We also found no significant differences between average color P_t and reported strategy type, $F(2, 194) = .27$, $p = .76$. Two one-way ANOVA were used to also examine these relationships on the spatial version of the task. We found there were no significant differences between average precision and reported strategy type, $F(2, 195) = .24$, $p = .79$. We also found no significant differences between average space P_t and reported strategy type, $F(2, 195) = 2.54$, $p = .08$. These results suggest that reported task strategy did not have an effect on precision or probability of target retrieval in memory on either the color or space continuous report tasks. Responses indicating ‘Neither’ were excluded due to the small number of responses for the color ($N = 1$) and space ($N = 3$) tasks.

Discussion

The aim of Study 1 was to examine the psychometrics of modeled WM precision estimates as individual differences,, the generality or interrelatedness of these estimates, and the relationship between these precision estimates and a well-known correlate of WM capacity, fluid intelligence. We found that while the modeled spatial precision estimate was reliable, the color estimate was not. However, these findings may be limited by the fact that a large number of trials are necessary to obtain reliable estimation of P_t and κ parameters. This number is suggested to be greater than 120 (Zhang & Luck, 2008; Xie & Zhang, 2017) trials per condition, but simulations from Grange & Moore (2022) using their *mixtur* r package for mixture modeling suggests that as few as 50 trials provides good parameter recovery ($r > .80$), with 200 trials being required for good-to-excellent parameter estimation. When we split the number of trials to obtain split-half reliability estimates, we are effectively applying the mixture model on 50 trials and using these estimated parameters to calculate reliability. As mentioned, though our model may have accurately estimated the model parameters based on simulation from Grange and Moore, it may be the case that these parameters themselves were not reliable estimates of P_t and κ and thus we could not be certain that our split-half reliability estimates are accurate indications of the true reliability of the task.

However, since our correlations are based on the estimated parameters from 100 trials, we can be more certain about the conclusion we draw from the correlations between variables. Here, we found that while the capacity parameters from the change detection and continuous report working memory tasks were significantly positively correlated consistent with prior research (Zhang and Luck, 2008; Unsworth et al., 2014), the precision estimates were not. Additionally, likely due to this lack of a relationship between the variables, a latent variable model which loads both precision estimates to a single factor had poor fit. This suggests that a participant's ability to recall more precise representations of a color feature when the target is in memory is not related to their ability to recall more precise representations of an item's location when the target is in memory. This result was contrary to our hypothesis that these variables would be significantly and positively correlated, suggesting that precision ability, or our ability to recall an item feature more accurately when it is in memory, was domain general and participants' performance in versions of the task with different feature types (i.e., color, space, orientation, etc.) would be related.

This was also reflected in the correlations between the two different continuous report task and the three fluid intelligence tasks. Contrary to our hypothesis, we did not replicate the findings from Fukuda et al. (2010), where the resolution of WM was measured using a modified change detection task was not correlated with fluid intelligence. We instead found that spatial precision was positively correlated with all fluid intelligence tasks, but color precision was only significantly correlated with raven and this correlation was negative.

Study 2

Based on our findings from Study 1, we designed Study 2 as a follow-up to investigate the reliability of the modeled WM precision parameters and to see whether we could replicate the findings from our correlation analyses. To this end, we increased the number of trials within the continuous report task to 200 and removed the change detection tasks from the test battery. This allows us to increase the number of trials used when estimating the precision and capacity estimates for both the correlation analyses and split-half reliabilities. We settled on 200 and the exclusion of the change detection tasks to allow adequate time to complete every task within the two-hour study period. An additional task, the Wisconsin Card Sorting task, was included but these data were not analyzed here.

Method

All participants (N = 257; 197 women, 55 men, 5 non-binary, M_{age} = 18.86, SD_{age} = 2.45) were undergraduate students at the University of Texas at Arlington, who participated in the study in exchange for course credit. After signing an informed consent, participants completed a demographics questionnaire and then a battery of cognitive tasks. In the two-hour session, participants completed three continuous report (color, orientation, and space) tasks measuring working memory precision and capacity, and three tasks (Raven Advanced Progressive Matrices, number series, and letter sets) measuring fluid intelligence. Participants also completed the Wisconsin Card Sorting Task. Participants finished the study by completing a strategy questionnaire. All participants completed the tasks in the following order: color, Raven, space, letter sets, orientation, number series, and the Wisconsin Card Sorting Task. The experimental protocol was approved by the Institutional Review Board of the University of Texas at Arlington.

Tasks

The color and space continuous response tasks were like those used in Study 1, except for completing 200 trials per task instead of 100. Raven, letter sets, and number series tasks were identical to those used in Study 1. The strategy questionnaire was also identical to that used in Study 1.

Data analysis

Data analysis procedures were identical to Study 1. In a latent factor model, we specified a model where we loaded capacity estimates from the continuous report tasks to a factor P_t, precision estimates to a factor κ, and fluid intelligence tasks to a factor gF (see Appendix A2). Data from the orientation version of the tasks were excluded for the same issue mentioned in study 1. We screened the data for outliers and excluded any data points that fell outside +/- 3 standard deviations of the mean.

Results

Table 3 lists descriptive statistics for each measure and Table 4 lists zero-order correlations between dependent variables. We first examined the reliability of our tasks. Measures of working memory capacity from the continuous report task (P_t) showed excellent reliability (split-half corrected correlations >.90). Measures of fluid intelligence showed acceptable reliability (split-half corrected correlations >.70). Consistent with Study 1, for the modeled precision estimates κ we found only the spatial estimate of the continuous report task showed good reliability (split-half corrected correlations >.80) while the color estimate had poor reliability (split-half corrected correlations < .50).

We next assessed the domain generality and interrelatedness of the continuous response tasks by examining the relationship between P_t and κ for both versions. Consistent with Study 1, we found a significant positive correlation between color and spatial capacity estimates, suggesting that those who had a higher probability to recall target color items tended to have higher probability to recall target spatial location of items. Likewise, both P_t measures had significant, positive correlations with all measures of fluid intelligence, such that those who had a higher probability of recall for target memory items in the continuous report tasks tended to have better fluid intelligence task performance. However, consistent with Study 1, there was still no significant correlation between color κ and spatial κ, suggesting that those who had better precision for target color items when the item was in memory, did not tend to have better precision for target item's spatial location when the item was in memory. We attempted a confirmatory factor analysis where we specified a latent variable model allowing the precision estimates to load onto a factor, the two continuous report capacity estimates to load onto a factor, and the three fluid intelligence tasks loaded onto a single factor, and found the model had poor fit, χ^2 (29) = 52.43, p < .001, CFI = 0.85, RMSEA = .13, 90% CI = [0.95, 0.17], SRMR = .08).

We then examined the correlations between precision estimates and performance on the fluid intelligence tasks. We found that spatial precision had a significant positive correlation with all three fluid intelligence tasks. Consistent with Study 1, these data suggest that those who had better spatial precision for target items in memory tended to have better performance in fluid intelligence tasks. However, we found no significant correlations between precision estimates from the color continuous report task and the letter sets and number series tasks.

Finally, we examined responses from the strategy questionnaire. Descriptive statistics can be found in Appendix D. Two one-way ANOVA tests were conducted to examine the relationships between average precision and P_t in the color continuous response task with reported strategy type. We found there were no significant differences between average precision and reported strategy type, $F(2, 218) = .29$, $p = .75$. We also found no significant differences between average color P_t and reported strategy type, $F(2, 218) = 2.73$, $p = .07$. These results suggest that reported task strategy did not have an effect on precision or probability of recalling target items on the color continuous report task. Two one-way ANOVA tests were also conducted to examine the same relationships on the spatial version of the task. We found there were no significant differences between average precision and reported strategy type, $F(2, 228) = 1.1$, $p = .34$. Likewise, we found no significant differences between average space P_t and reported strategy type, $F(2, 228) = .67$, $p = .51$. Responses indicating 'Neither' were excluded due to the small number of responses for the color ($N = 1$) and space ($N = 1$) tasks.

Discussion

Study 2 served as a follow-up to Study 1 to assess if and how increasing the number of trials for the continuous report task would affect the reliability for the modeled parameter estimates and whether we could replicate the pattern of results found in Study 1. Regarding reliability, while the magnitude of reliability for color κ increased it was still not at an acceptable level. However, we found that spatial κ was still reliable. Consistent with Study 1, we found that the capacity estimates from the continuous report tasks had a significant positive correlation with each other and with each fluid intelligence task, spatial precision and color precision did not have a significant correlation with each other, and spatial precision had a significant positive correlation with all fluid intelligence tasks. However, we did not find a significant correlation between color precision and any fluid intelligence tasks.

General Discussion

The present study sought to add onto the burgeoning literature involving an individual differences approach using the modeled precision parameters of the continuous report task. Study 1 examined the modeled precision parameters derived from the standard mixture model for continuous report tasks as an individual difference by examining its psychometric properties, investigating the relationship between precision performance for different feature types, and

Table 3

Descriptive statistics for Study 2

Measure	N	M	SD	Skew	Kurtosis	*N* Excluded
Color $P_{t,}$	237	.66	.19	-.91	0.57	13
Space $P_{t,}$	242	.76	.14	-1.22	1.14	8
Color κ	237	14.43	10.51	2.94	12.19	13
Space κ	242	26.73	13.13	0.83	0.57	8
Raven	245	7.50	3.27	0.15	-0.33	5
Number series	247	7.13	2.79	0.25	-0.19	3
Letter sets	246	8.12	2.88	0.31	0.03	4

Table 4

Zero-Order Correlation for Study 2

	1	2	3	4	5	6	7
1. Color P_t	*0.94*						
2. Space P_t	**0.54****	*0.93*					
3. Color κ (precision)	**-0.32****	0.04	*0.49*				
4. Space κ (precision)	**0.38****	**0.27****	-0.09	*0.81*			
5. Raven	**0.37****	**0.29****	-0.03	**0.21****	*0.74*		
6. Letter Sets	**0.18**	**0.31****	-0.01	**0.19****	**0.28****	*0.70*	
7. Number Series	**0.23****	**0.29****	0.00	**0.23****	**0.46****	**0.31****	*0.78*

Note. Bold indicates significant at $p < .05$. * - indicates significance at $p < .01$. ** - indicates significance at $p < .001$. Split-half reliabilities are listed along the diagonal.

looking at the relationship between precision and a well-known correlate of WM capacity, fluid intelligence. Results indicated that only the precision parameter from the spatial version of the continuous report task had acceptable reliability, there was no significant relationship between precision in the two continuous report tasks, and that spatial WM precision parameters had a positive relationship with performance in all fluid intelligence tasks while color precision only had a significant relationship with Raven. Study 2 sought to replicate these results with an increased trial number for the continuous report tasks and found a similar pattern of results except only the spatial precision parameter having a significant relationship with the fluid intelligence tasks. Overall, these results suggest some considerations in the use of modeled precision parameters for individual differences research, feature-related differences in the continuous report task, and the relationship between WM precision and fluid intelligence.

In both studies, the modeled capacity estimate derived from the continuous report task, P_t, was a reliable parameter, consistent with prior research (Unsworth et al., 2014; see supl. Xie et., 2020). However, contrary to Xie et al. which found a reliability of ~.70 for their WM precision estimate, we found that only the modeled precision estimate from the spatial version of the task met or exceeded this threshold. It should be noted that the method to calculate reliability differed between our studies. While they utilized a bootstrap method described earlier, we instead split the trials between even and odd trials, fit the data under the standard mixture model, and then used Spearman-Brown correction to the correlation to calculate reliability. Notably different in our methodology is that we had to fit the model to a smaller number of trials which could affect the model's ability to recover the best fitting parameters (Zhang and Luck, 2008; Grange & Moore, 2022). Thus, the correlations between the modeled estimated parameters for the 50-trial split halves could be dubious due to poor model parameter recovery in Study 1. While

Grange and Moore suggest that one can get good parameter recovery with as few as 50 trials, Zhang and Luck (2008) suggest around 120 or more trials are necessary. This being the case, Study 2 increased the total trial amount to 200, allowing us to fit the model to each split half for 100 trials to increase the likelihood of the model to recover the best-fitting parameters for us to use in calculating reliability. In both studies, we found that only the precision estimate from the spatial version of the task was reliable, while the color version was not.

The current finding proposes a question regarding the intra-task consistency of one's ability to precisely recall a target feature when the target is in memory within trials of the task. That is, if recalling the precise feature representation of an item in memory is a differentiable cognitive ability, then we would expect consistency within task performance reflected in parameters that measure WM precision. However, if the likelihood of obtaining the best fitting parameters is a subject of trial size, it may be the case that even more trials are necessary to make conclusions using a split-half reliability estimate for the modeled estimated parameters. Future research could run different versions of each task with a large number of trials (~1000) and assess split half reliability using a number of trials that we would expect to get excellent parameter recovery for both model estimated parameters. Likewise, while the amount of information seems to be consistent within a task for parameters reflecting the quantity of WM information, regardless of feature type, we only found reliability for parameters reflecting the quality of information for a spatial location feature and not for a color feature. Thus, future research should examine whether there are differences in the number of trials required to obtain the best fitting parameters from the model between feature types.

In addition to this point regarding the different feature types, we found no significant relationship between the precision estimates for color and spatial location in both studies. These results suggest that our ability to precisely recall target features (i.e., color and spatial location) may not be domain-general, such that we would expect performance for one feature type to be correlated with performance across other feature types. As expected of this finding, we found that we could not load these modeled precision estimates to a single factor in a latent factor model, though we are able to do so for capacity estimates. Previous studies have found that a location feature in the continuous report task typically had better performance compared to a color or orientation feature, such that participants had better precision and probability of the target in memory (Cooper et al., 2017; Korkki et al., 2020). Indeed, we found that performance in the spatial version of the task was better (higher precision and probability of target in memory) than the color version in both studies. However, given the lack of correlation between precision parameters between tasks, it is not the case that participants who had higher spatial precision estimates tended to have higher color precision estimates. Additionally, we found in both studies that there were no significant differences in average task performance for both modeled parameters regardless of reported task strategy. Thus, it may be the case there are feature-specific differences, such as how feature information is stored in memory, which may drive these individual differences in performance. However, our strategy questionnaire and options may not have been sensitive enough to adequately capture how participants may truly approach the tasks. Future research should further explore task strategy acquisition and use in the continuous report task to expand our understanding of what and how certain strategies relate to task performance.

This idea that WM precision is not domain-general is reflected in our findings regarding the relationship between WM precision and fluid intelligence. The present study sought to replicate the findings from Fukuda et al. (2010) which found that the quantity of WM relates to fluid intelligence, but not quality. However, our study utilized the modeled precision estimates from the standard mixture model using the continuous report task rather than capacity estimates from a change detection paradigm. Consistent with prior research, we found that the capacity estimates from both tasks correlated with fluid intelligence (Unsworth et al., 2014). However, unlike in Fukuda et al., we found that spatial precision did significantly correlate with all measures of fluid intelligence in both studies. However, color precision only correlated with the Raven task in Study 1, and this did not replicate in Study 2. These results imply that, at least with one feature type, there may be a relationship between our ability to recall a target item feature more precisely and the complex reasoning abilities as assessed in fluid intelligence.

However, our change detection tasks were different from the change detection tasks utilized by Fukuda et al (2010). While we used basic stimulus features (i.e., color, shape, orientation), they used shapes which varied between large changes (reflected in a change in the type of shape) or small changes (reflected in a change within the same type of shape). They reasoned that distinguishing whether a change occurred in a small-change trial would require the target item to have been stored with enough resolution. They also ran a regular color change detection task and found that there was no correlation between the capacity estimates for color and the capacity estimates for the small-change trials (reflecting resolution), but there was a positive correlation between color capacity and capacity estimates for the large-change trials (reflecting quantity). Likewise, in Study 1 we found that precision estimates for color did not correlate with any measure of change detection. However, precision estimates for spatial location were positively correlated with change detection measures. However, the distinction between this change detection paradigm and using modeled parameters in the continuous report task is that the change detection capacity measure for small-change trials can only indirectly infer precision while the continuous report task provides an estimate of precision distinct from capacity. Thus, while it may be the case our ability to make accuracy judgments for an item based on a small or large change is dependent on having a high quality representation of the to-be-remembered item, simply being able to hold and recall higher quality representation to make change judgments may not be the same as our ability to recall the precise target feature value of this representation. Future research should examine the relationship between precision in other feature types (e.g., motion, orientation, etc.) to further explore whether our findings reflect a lack of domain-generality as well as the relationship between WM precision for these other feature types and fluid intelligence.

Conclusion

While individual differences in WM capacity and its relationship with a variety of other higher-order cognitive abilities is a well-researched area, the same cannot be said for WM precision and the use of the continuous report task and modeled precision parameters in individual differences research remains comparatively untapped. If WM capacity and WM precision are indeed distinct aspects of WM that can be independently measured (Awe et al., 2007; Zhang and Luck, 2008; Bays et al., 2009) and contribute separately to other cognitive factors (Fukuda et al., 2010), then individual differences research to fully understand the distinct relationship between WM precision and other cognitive facets is important for furthering our understanding of working memory wholly.

References

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10). https://doi.org/10.1167/9.10.7

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working memory has better fidelity than long-term memory: The fidelity constraint is not a general property of memory after all. *Psychological Science*, *30*(2), 223–237. https://doi.org/10.1177/0956797618813538

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*(6), 981-990. https://doi.org/10.1177/0956797612465439

Cooper, R. A., Richter, F. R., Bays, P. M., Plaisted-Grant, K. C., Baron-Cohen, S., & Simons, J. S. (2017). Reduced hippocampal functional connectivity during episodic memory retrieval in autism. *Cerebral cortex*, *27*(2), 888–902. https://doi.org/10.1093/cercor/bhw417

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dowle, M., Srinivasan, A., & Short, T. (2019). data. table: Extension of 'data. frame'. *R package version*, *1*(8).

Ekstrom, R. B., & Harman, H. H. (1976). Manual for kit of factor-referenced cognitive tests, 1976. *Educational testing service.*

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. https://doi.org/10.1038/ncomms2237

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin and Review*, *17*(5), 673–679. https://doi.org/10.3758/17.5.673

Gold, J. M., Hahn, B., Zhang, W. W., Robinson, B. M., Kappenman, E. S., Beck, V. M., & Luck, S. J. (2010). Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. *Archives of General Psychiatry*, *67*(6), 570–577. https://doi.org/10.1001/archgenpsychiatry.2010.65

Grange, J. A., & Moore, S. B. (2022). mixtur: An R package for designing, analysing, and modelling continuous report visual short-term memory studies. *Behavior Research Methods*, *54*(5), 2071–2100. https://doi.org/10.3758/s13428-021-01688-1

Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science*, *15*(4), 528–539. https://doi.org/10.1111/j.1467-7687.2012.01148.x

Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Korkki, S. M., Richter, F. R., Jeyarathnarajah, P., & Simons, J. S. (2020). Healthy ageing reduces the precision of episodic memory retrieval. *Psychology and aging*, *35*(1), 124–142. https://doi.org/10.1037/pag0000432

Kyllonen, P. C., Christal, R. E., Woltz, D., Ackerman, P., & Jensen, A. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4). 389-433. https://doi.org/10.1016/S0160-2896(05)80012-1

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, and Psychophysics, 76*(7), 1975–1984. https://doi.org/10.3758/s13414-014-0653-z

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nat Neurosci*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: 1. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1) , 261-282. https://doi.org/10.1037/0096-1523.24.1.261

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., https://www.R-project.org/.

Raven, J. C. (1965). *Advanced Progressive Matrices, Sets I and II.* H. K. Lewis London.

Robison, M. K., & Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception, and Psychophysics*, *82*(7), 3273–3290. https://doi.org/10.3758/s13414-020-02077-0

Robison, M. K., & Campbell, S. (2023). Baseline pupil diameter does not correlate with fluid intelligence. *Psychonomic Bulletin and Review*, *30*, 1988-2001. https://doi.org/10.3758/s13423-023-02273-7

Robison, M. K., Celaya, X., Ball, B. H., & Brewer, G. A. (2024). Task sequencing does not systematically affect the factor structure of cognitive abilities. *Psychonomic Bulletin & Review*, 31, 670 – 685. https://doi.org/10.3758/s13423-023-02369-0

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Revelle, W. and Revelle, M.W. (2015) Package ‘Psych’. The Comprehensive R Archive Network.

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin and Review*, *15*(1), 215–222. https://doi.org/10.3758/PBR.15.1.215

Suchow, J.W., Brady, T.F., Fougnie, D., & Alvarez, G.A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 9. https://doi.org/10.1167/13.10.9

Thurstone, L. L. (1938). Primary mental abilities. *Chicago, IL: University of Chicago Press.*

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Vogel, E. K., & Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science, 17*(2), 171-176. https://doi.org/10.1111/j.1467-8721.2008.00569.x

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Xie, W., Berry, A., Lustig, C., Deldin, P., & Zhang, W. (2019). Poor sleep quality and compromised visual working memory capacity. Journal of the International *Neuropsychological Society, 25*(6), 583–594. https://doi.org/10.1017/S1355617719000183

Xie, W., Cappiello, M., Park, H. B., Deldin, P., Chan, R. C. K., & Zhang, W. (2018). Schizotypy is associated with reduced mnemonic precision in visual working memory. *Schizophrenia Research, 193*, 91–97. https://doi.org/10.1016/j.schres.2017.07.046

Xie, W., Park, H. B., Zaghloul, K. A., & Zhang, W. (2020). Correlated individual differences in the estimated precision of working memory and long-term memory: commentary on the study by Biderman, Luria, Teodorescu, Hajaj, and Goshen-Gottstein (2019). *Psychological Science, 31*(3), 345–348. SAGE Publications Inc. https://doi.org/10.1177/0956797620903718

Xie, W., Ma., T., Thakurdesai, S., Kim, I., & Zhang, W., (Preprint) Discrimination of mnemonic similarity is associated with short-term and long-term memory precision.

Xie, W., & Zhang, W. (2017). Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition*, *166*, 73–83. https://doi.org/10.1016/j.cognition.2017.05.025

Xie, W., & Zhang, W. (2018) Mood-dependent retrieval in visual long-term memory: dissociable effects on retrieval probability and mnemonic precision. *Cognition and Emotion*, *32*(4), 674-690. https://doi.org/10.1080/02699931.2017.1340261

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. https://doi.org/10.1038/nature06860

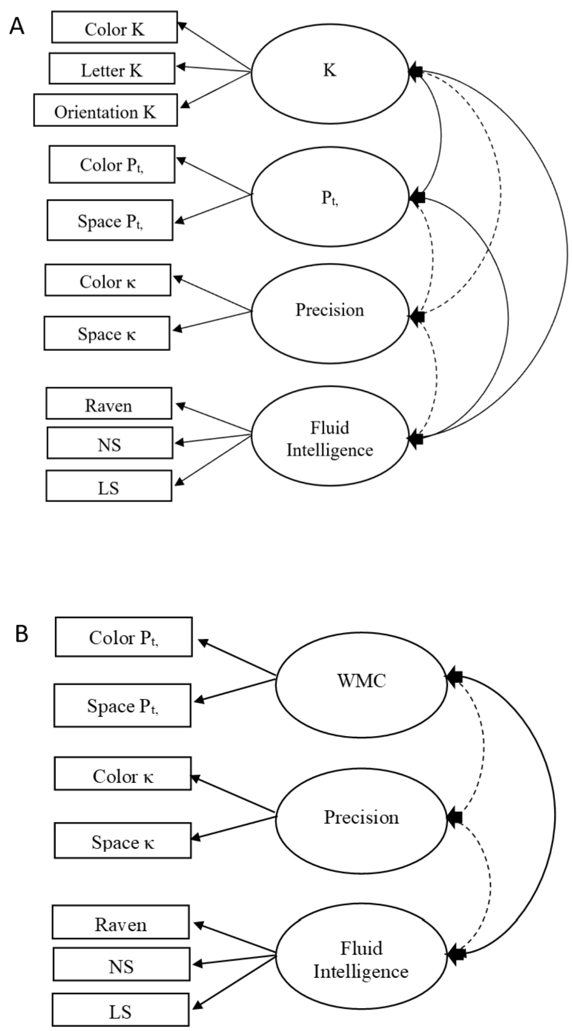
Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434–1441. https://doi.org/10.1177/0956797611417006

Zhao, Y., Kuai, S., Zanto, T. P., & Ku, Y. (2020). Neural correlates underlying the precision of visual working memory. *Neuroscience*, *425*, 301–311. https://doi.org/10.1016/j.neuroscience.2019.11.037

Appendix A

Figure 1

Proposed Latent Factor Models in Confirmatory Factor Analysis

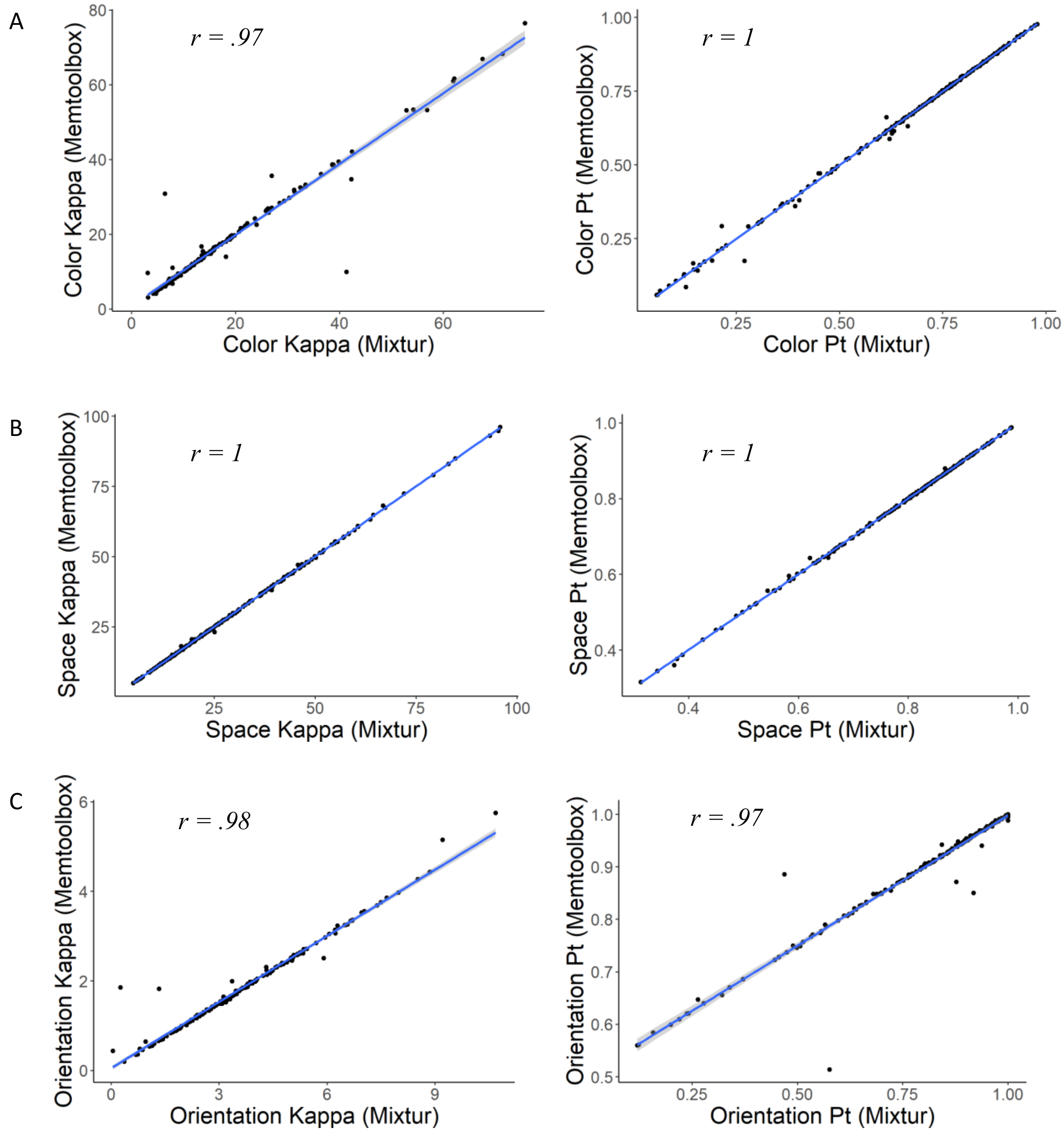


Note. Bold lines indicate significant relationship. Dashed lines indicate non-significant relationship. A) Latent factor model loading change detection scores to a *K* factor representing working memory capacity for change detection, P_t scores to a P_t factor representing working memory capacity for continuous report, κ values from the continuous report task to a precision factor, and fluid intelligence task scores to a gF factor representing a fluid intelligence factor. B) Similar to A, excluding capacity values from the change detection tasks.

Appendix B

Figure 1

Scatterplots of Correlations between Modeled Paramters in Mixtur and MemToolbox for Study 1



Note. A: Color Continuous Report task. B: Spatial Location Continuous Report task. C: Orientation Continuous report task.

Appendix C

Table 1

Descriptive statistics for Modeled Parameters by Reported Strategy for Experiment 1

Strategy		**N**	**M**	**SD**	**Skew**	**Kurtosis**
Verbal Label	Color P_t	27	.66	.18	-.3	-1.17
	Space P_t	30	.79	.12	-1	.65
	Color κ	27	16.15	9.01	0.95	-.11
	Space κ	30	27.15	14.21	.48	-.75
Visual Image	Color P_t	97	.67	.24	-1.13	.3
	Space P_t	83	.77	.16	-1.01	.38
	Color κ	97	15.1	13.58	2.42	5.6
	Space κ	83	29.76	18.31	1.18	1.43
Both	Color P_t	73	.69	.21	-1.15	.74
	Space P_t	85	.81	.11	-1.1	.77
	Color κ	73	15.21	11.86	2.39	7.75
	Space κ	85	29.48	19.27	1.45	1.94

Appendix D

Table 1

Descriptive statistics for Modeled Parameters by Reported Strategy for Experiment 2

Measure		**N**	**M**	**SD**	**Skew**	**Kurtosis**
Verbal Label	Color P_t	35	.7	.2	-1.09	.63
	Space P_t	44	.76	.16	-1.1	.94
	Color κ	35	13.3	8.14	2.11	4.24
	Space κ	44	27.2	12.97	.58	-.47
Visual Image	Color P_t	88	.63	.2	-.68	-.25
	Space P_t	94	.75	.15	-1.03	.65
	Color κ	88	14.26	7.97	1.56	2.01
	Space κ	94	25.55	12.91	.62	-.04
Both	Color P_t	98	.69	.15	-.67	.33
	Space P_t	93	.77	.14	-1.54	1.92
	Color κ	98	13.41	8.83	1.74	2.44
	Space κ	93	28.39	13.59	1.05	1.11