TALENT
2020, Vol 10, Issue 2, 162-175

Kayacan, G, Ateşgöz, N. N., & Sak, U. (2020). A comparative analysis of psychometric properties of memory tasks and their relationships with higher-order thinking skills: Recognition versus recall. *Talent, 10*(2), 162-175.
DOI: https://doi.org/10.46893/talent.847543
Received: September 26, 2020; Accepted: November 28, 2020

Research

A Comparative Analysis of Psychometric Properties of Memory Tasks and Their Relationships with Higher-Order Thinking Skills: Recognition versus Recall*

Gamze Kayacan[1], N. Nazlı Ateşgöz[2] & Uğur Sak[3]

Abstract

This study aims to compare the psychometric properties of recognition and recall task measurements and to examine their relationships with other higher-order thinking skills. Memory measurements made with recognition and recall tasks were based on the Visual Span Memory (VSM) subtest of the ASIS intelligence scale. The participants of the study consisted of 228 students attending first and second grade in a primary school in the city center of Eskisehir. The data were collected by administering the recognition and recall task forms to the students individually. The findings reveal a statistically significant difference between the average scores of recall and recognition tasks. The mean scores of the recognition task measurements are significantly higher than the mean scores of the recall task measurements ($t_{(227)} = 5.79$, $p<0.01$; Cohen d = 0.435, Cohen dz = 0.38). In addition, there is a significant difference between the reliability coefficients of recognition and recall task score in favor of the recall task score ($\chi^2(1) = 6.181$, $p <.02$). It was also found that the mean item-total correlations of the recall task measurements (r=.41) were higher than the recognition task measurements (r=.27), and the item-total correlations of the six items differed significantly in favor of the recall task measurements. The correlation of the recall task score with the other 5 subtests in the ASIS intelligence scale was higher than that of the recognition task score. The findings show that the psychometric properties of the measurement performed with the recall task are stronger.

Key Words: working memory assessment, memory tasks, recall, recognition

Introduction

What would happen if we deactivated our memory for a day? This question can only be answered hypothetically. We would likely turn into unconscious beings trapped in a brief moment, deprived of all kinds of cognitive functions such as thinking, speaking, learning, and gaining experiences. Memory is the basic structure that processes the environmental stimuli and stores this information for later use. Therefore, even the adaptation that ensures the continuity of life is possible only with the healthy functioning of the memory. This is also the case with accomplishing the primary aim of education, which is the "intentional and desired behavioral change" (Ertürk, 2013, p.13).

* *The paper is based on the first author's master thesis*

[1]*Corresponding author, MS, Research assistant, Anadolu University, Faculty of Education, Eskisehir, Turkey, gamzekayacan@hotmail.com.tr,* ORCID: 0000-0001-9378-9175
[2] *PhD., Research assistant, Center for Research and Practice for High Ability Education (EPTS), Anadolu University, Faculty of Education, Eskisehir, Turkey, ORCID: 0000-0001-5944-937X*
[3] *Prof., Director, Center for Research and Practice for High Ability Education (EPTS), Anadolu University, Faculty of Education, Eskisehir, Turkey, ORCID: 0000-0001-6312-5239*

Due to its pivotal role in processing information (Nutley & Söderqvist, 2017) and its relatively more stable structure (Alloway & Alloway, 2010), working memory has been one of the basic structures used in the measurement of cognitive functions. To date, various tasks have been created and used to assess memory. Perhaps the oldest and most frequently used of these are recognition and recall tasks which date back to the 1800s when memory was first measured experimentally. While recall refers to the tasks that require the person to recreate a series of items they have been previously exposed to (Cleary, 2019; Schwartz, 2018), recognition refers to memory tasks that require the person to distinguish a certain stimulus from among other stimuli (Kintsch, 1970). Although both memory tasks measure the retrieval process, whether these processes are similar or not is still a matter of contention.

The process of processing information begins with the creation of a memory trace in the memory for each new piece of information received from the environment. Each time the information is repeated, a thickening is observed in the neurons on this memory trace, and the thicker the myelin sheath surrounding the neurons, the faster the electrical flow between neurons (Fields, 2020; Hasan et al., 2019), which means that the information can be retrieved more easily and quickly. However, information with poor memory trace is more difficult to retrieve. The memory trace must be strong enough for the individual to be able to recall the "old" information. While information with a strong memory trace can be easily recalled without requiring any clue, some hints may be required to retrieve information with a weak memory trace. For changing retrieval tasks (such as free recall, cued recall, recognition), it is necessary to create memory traces of different strengths (Kintsch, 1970; Margolis, 1992; Radvansky, 2017). Recall tasks require a stronger memory trace than recognition tasks. As the memory trace gets stronger, the information can be recalled more easily and quickly, this leaves enough energy for higher-order thinking and actions (Heacox & Cash, 2014).

It is not possible to retrieve every piece of information we process through recall. It is sufficient to recognize some information in certain contexts (when some clues are provided). The use of multiple-choice tests in education is an example of this. However, in the higher levels of education, when the field specialization begins, the knowledge acquired in the relevant field is expected to be reconstructed when needed, that is to be remembered. For example, the doctor who cannot create the information at the right time cannot inspire the necessary confidence in the patients. Therefore, it can be said that the amount of information that can be recalled is of great importance in the formation of expertise.

It is frequently stated in the literature that recall tasks are more difficult than recognition tasks (Andrew & Bird, 1938; Margolis, 1992; Radvansky, 2017). The main reason for this is that these two different memory tasks have different processing requirements. The need for strategic processing is quite high in free recall tasks. On the other hand, the cue, which is at the heart of cued recall and recognition, greatly reduces the need for strategic processing. It can be said that recognition tasks with low processing need measure short-term memory rather than working memory, unless an additional task is used (see Schneider & McGrew, 2018). However, there are varying opinions about retrieval tasks. Some researchers (Chubala et al., 2020; Gisselgård et al., 2007) argue that recall tasks are also affected by long-term memory, so they cannot perform as a robust measurement as recognition tasks. However, Unsworth and Engle (2007) found that recall tasks, which require more processing than recognition tasks, predict working memory as well as complex span tasks, which are commonly used in measurements.

There is a limited number of studies that directly compare recognition and recall tasks (Chubala et al., 2020; Gisselgård et al., 2007). Studies comparing these two memory tasks generally focus on visual, auditory, and semantic similarity (Tse et al., 2011; Chubala et al., 2019), dynamic visual noise (Chubala et al., 2018), and the effect of the related stimulus frequency (Chubala et al., 2019) on the performance in recall tasks. However, no study was found that compared the psychometric properties of these two different memory tasks, which are frequently included in cognitive scales, and their relationships with other higher-order thinking skills. Therefore, the current study aimed to compare the psychometric properties of recognition and recall tasks using the same stimuli on the same participant group, and to examine their relationship with other higher-order thinking skills included in the ASIS scale.

Method

Participants

The participants of the study consisted of 228 first and second-year students who are attending one of the pilot schools within the scope of the "Gifted Education Project" and were diagnosed with the ASIS intelligence scale in the 2018-2019 academic year. Of the 228 students who participate in the study, 123 were girls and 105 were boys. Although the age range of the participants varies between 65-94 months, the average age is 76.39 months.

Measures

The data were collected using the Anadolu-Sak Intelligence Scale (ASIS). The most important factor in choosing the ASIS intelligence scale was based on the fact that the scale draws upon the Turkish culture, and therefore it does not create a cultural bias in measurements. Memory measurements performed with recognition and recall tasks were also based on the Visual Span Memory subtest of the ASIS intelligence scale. The Visual Span Memory subtest (VSM) measures working memory with recognition tasks. Thus, the original form of the VSM subtest was used for recognition tasks. In addition, to make comparisons, an alternative form was prepared in which measurements can be made with recall tasks by sticking to the order, number, and size of the items in the VSM subtest. Recall measurements were carried out by using this alternative form.

Anadolu-Sak Intelligence Scale (ASIS). Anadolu-Sak Intelligence Scale is a test battery used in the cognitive assessment of children between the ages of 4-12. It consists of 7 subtests aimed at evaluating reasoning, memory, attention, perception, and cognitive functions and is administered individually. The CHC taxonomy and Luria's simultaneous-successive processing model constitute the theoretical infrastructure of ASIS. In addition, Baddeley's working memory model was taken into account while creating memory subtests of the ASIS intelligence scale. The norm study conducted by Sak et al. (2016) revealed that the internal consistency reliability coefficients of the index scores of the ASIS intelligence scale varied between .95 and .99. These values indicate that the ASIS intelligence scale has an excellent level of internal consistency. The consistency coefficients of the index scores of the ASIS intelligence scale ranged from .89 to .95. That the retest consistency coefficient of the general intelligence index is .95 is of particular significance. This value indicates that the general intelligence measurement performed at short intervals is largely consistent.

Visual Span Memory Subtest - Recognition task form. The ASIS Visual Span Memory (VSM) subtest aims at measuring short-term memory, memory span, and successive processing skills through recognition tasks. The VSM is composed of test items in which various figures are presented as a series. These series of figures, ranging from 2 to 9 digits, are shown to the participants for 5 seconds. Then, the participants are expected to distinguish the stimuli they saw among many other distractors. Figure 1 shows an example item for the recognition task. Before starting the subtest, a tutorial on sample items is performed. The evaluation is started with the first item in the subtest after ensuring that the participant has learned the task. If the participants make three mistakes in a row or made four mistakes in the last five questions, the subtest is terminated. Participants receive one point for each correct answer they give. The highest score that can be obtained for this 20-item subtest is 20 points.

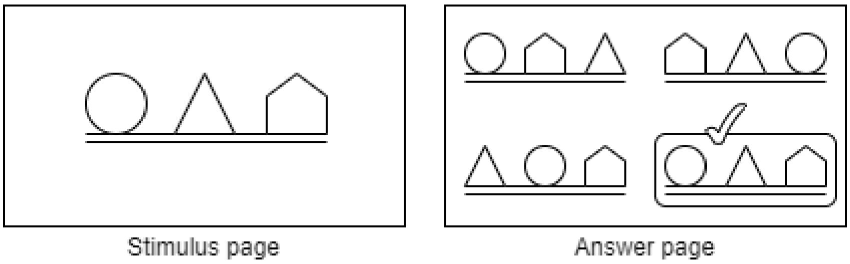

Figure 1. An example of the Recognition Task Item

In the norm study conducted by Sak et al. (2016), the internal consistency reliability coefficient of the subtest was calculated as .88, which is defined by DeVellis (2017) as “very good”. In addition, the retest consistency coefficient of the subtest was calculated as .81 (Sak et al., 2016). Hence, it can be said that the scores obtained from the subtest do not vary over time and are quite consistent. In the pilot study conducted before the research, the internal consistency reliability coefficient was found to be .68. This value corresponds to the “lowest acceptable” value according to the criteria defined by DeVellis (2017). Considering that the alpha coefficient is affected by the number of items and the sample size reached (Abdelmoula et al., 2015; Shevlin et al., 2000), the limited number of items in the subtest and the small sample obtained may have played a role in reliability coefficient being lower than normal. It should also be noted that a higher alpha value would be obtained if the pilot study were to be repeated with a larger sample.

Alternative form - Recall task form. By removing the multiple-choice answer pages in the VSM subtest, an alternative form that allows the participants to create the stimuli they see was prepared. The alternative form consists of exactly the same items as the original VSM form to enable making a comparison between the recall and recognition tasks. It is a recall task form that requires the participant to recreate the stimulus they had previously seen on a line. An example item for the recall task is given in Figure 2. In the recall task, participants were shown figure sequences ranging from 2 to 9 digits for 5 seconds, and then they were given one minute to create the same sequence. Similar to the recognition task, a tutorial on sample items is performed before starting the recall task subtest. After the participant is taught what the task requires, the assessment is started with the first item on the subtest. The subtest is terminated if the participants make three mistakes in a row or made four mistakes in the last five questions. Participants receive one point for each correct answer they give. The highest score that can be obtained for this 20-item subtest is 20 points.

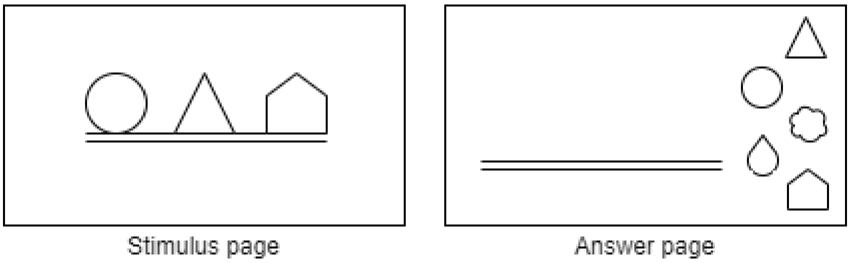

Figure 2. An Example of Recall Task Item

The reliability coefficient of the recall task form was found to be .72 in the pilot study. DeVellis (2017) describes this value as "significant". As in the recognition task form, the number of items in this form is limited and the sample group is small. Therefore, it should be kept in mind that a higher alpha value can be reached when the pilot study is repeated with a larger sample.

Procedure

The ASIS intelligence test, which includes the VSM subtest, was implemented by five research assistants pursuing their postgraduate studies in the field of gifted education and are qualified to administer ASIS. The intelligence tests, which lasted 2 weeks, were carried out in the classrooms and offices approved by the school administration. Applications of the recall task form was carried out by the researcher herself, who was also involved in the first stage of the data collection. Groth-Marnat (2003) states that when the second test is given at least eight weeks later, the memory effect on the responses to the items will be quite low (cited in Gatewood et al., 2011). For this reason, the recall task form was administered nine weeks after the ASIS intelligence scale implementation and lasted three weeks.

Results

Comparison of the Average Scores Obtained from the Measurements

While the average of the students' scores from the recognition-memory task was 5.56, the average of their scores from the recall task was 4.44. In both memory tasks, where a maximum of 20 points could be obtained, students got 0 as the lowest and 13 points as the highest. Although the lowest and highest scores obtained by students in both measures were similar, it was observed that the range or spread of the items answered was dissimilar. While the students reached the 14th item in the recall task, they progressed to the last item (the 20th item) in the recognition task.

The dependent samples t-test was conducted to find out any significant difference between the means of recognition and recall task measurements. The results of the dependent samples t test revealed a significant difference between recognition and recall task scores ($t(227) = 5.79$, $p < 0.01$; Cohen $d = 0.435$, Cohen $d_z = 0.38$). Cohen (1988) suggested using the Cohen d_z formula to examine the differences between paired observations (repeated measures) in a sample group. Evaluation criteria for Cohen d_z are stated as "small (.14), medium (.35) and large (.57) effect" (Lakens, 2017). Therefore, there is a significant and medium effect size difference at $p < 0.001$ level in favor of the recognition task. In other words, the average scores of the recognition task measurements are significantly higher than the average scores of the recall task measurements.

Comparison of the Psychometric Properties of the Measurements

An internal consistency analysis was conducted to reveal to what extent the items in the scale were consistent with each other. The Cronbach alpha coefficient for the recognition task measurements was calculated as .69, and the Cronbach alpha coefficient for the recall task measurements as .77. Considering the value ranges specified by DeVellis (2017), it is observed that the reliability of the recognition task measurement is in the "least acceptable" range, while the recall task measurement is in the "significant" range.

In addition to evaluating the Cronbach's alpha reliability coefficients obtained from the measurements according to certain threshold values, it is also necessary to examine whether there is a significant difference between the two coefficients that is not due to chance factor (Diedenhofen & Musch, 2016). Thus, the Cronbach alpha coefficients obtained from the two measurements were compared statistically as shown in Table 1.

Table 1. Comparison Statistics for the Cronbach Alpha Coefficients

	Cronbach alpha coefficient	Confidence Interval		Chi square	p
Recognition	0.686	0.624	0.742	6.181	0.013
Recall	0.773	0.728	0.814		

As can be seen in Table 1, there is a significant difference between the reliability coefficients of recognition and recall task measurements ($\chi^2(1) = 6.181$, $p < .02$). There is no suitable effect size index for Cronbach alpha coefficients obtained from dependent samples (Diedenhofen & Musch, 2016; Liu & Weng, 2009). However, the recall task measurements are observed to be significantly ($p < .02$) more reliable than the recognition task measurements.

Item analysis of the recognition and recall task measurements were conducted because it contributes to the improvement of test reliability by defining "problematic" items in the scale (Ho, 2014). While some students reached the 20th item, which is the last question in the recognition task, the same group of students were able to progress up to the 14th item in the recall task. The corrected item-total correlations of the items that students could answer were .035 (the lowest) and .476 (the highest) for the task of recognition, while it ranged from the lowest of .07 to the highest of .737 for the recall task. The mean item-total correlations of these items were found to be .27 for the recognition task measurements and .41 for the recall task.

In addition, paired comparisons of equivalent items in recognition and recall tasks were also performed. However, since the students were able to answer the first 14 questions within the starting and ending rule in the recall task, paired comparisons could be made for these items. Since the item-total correlations of the items subject to comparison could be affected by other items not included in the comparison, the item-total correlations were recalculated for only the first 14 items of two memory task measurements.

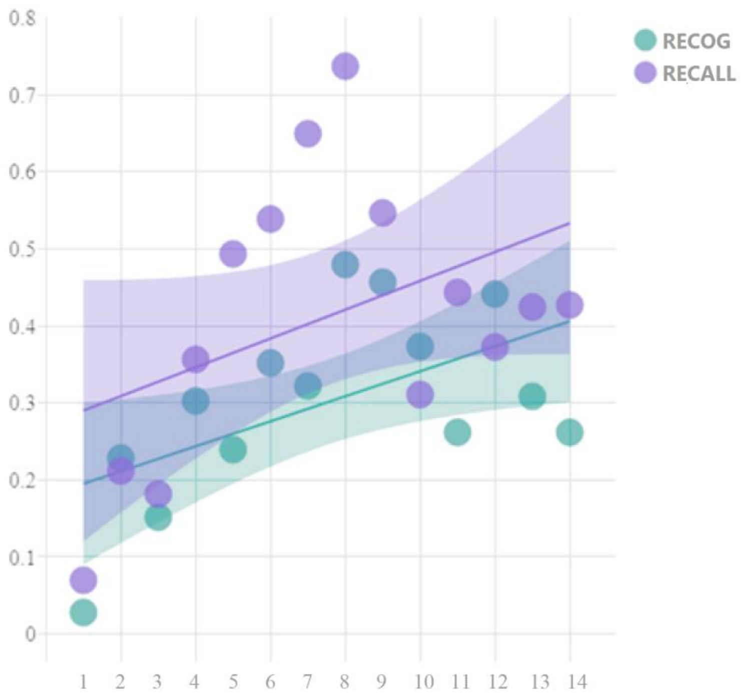

Figure 3. Item-total Correlations and Confidence Intervals for Recognition and Recall Task Items

Although the use of Fisher's z score in comparing the correlation values obtained from independent samples is widely accepted, there is no single widely agreed and used analysis to compare correlations obtained from dependent samples (Ramseyer, 1979). Some studies (Hittner et al., 2003; Silver et al., 2004), examining various analysis methods that are applied to compare correlation scores obtained from dependent samples, show that Dunn and Clark's z score yields better results compared to other analyses. Dunn and Clark's z score provides a reasonable control over Type I error, it displays a statistically good power. Therefore, Dunn and Clark's z score (1969) was calculated in comparing the dependent correlation values obtained from the same sample. In addition, Zou's (2007) confidence interval formula, which can be used to compare dependent correlations, was included. Diedenhofen and Musch's (2015) R based Cocor (Comparing correlations) package program was used in the calculation of Dunn and Clark's z score and Zou's confidence interval. Any potential Type I error due to multiple comparison tests was prevented by reducing the significance level to .005. The analysis results are given in Table 2.

Table 2 shows that there is a statistically significant difference between the item-total correlations of the 5th, 7th, and 8th items (Recognition<Recall). There is a significant difference in the reliability interval of the 5th, 6th, 7th, 8th, 11th, and 14th items.

Table 2. Item-Total Correlation Comparison Statistics by Each Item

		Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14
Dunn and Clark's z score (1969)	z	-0.49	-0.17	-0.19	-0.7	-3.06	-2.49	-4.96	-4.29	-1.69	0.97	-2.2	0.83	1.14	-1.98
	p	0.62	0.866	0.851	0.485	0.002 *	0.013	0.000 *	0.000 *	0.091	0.333	0.028	0.406	0.255	0.048
Zou's (2007) 95% confidence interval		-0.23 / 0.14	-0.19 / 0.158	-0.19 / 0.159	-0.22 / 0.103	-0.40 / -0.09**	-0.33 / -0.04**	-0.48 / -0.21**	-0.38 / -0.14**	-0.25 / 0.019	-0.08 / 0.244	-0.35 / -0.02**	-0.09 / 0.218	-0.06 / 0.226	-0.32 / -0.001**

*p <0.005

** Intervals that do not contain zero are statistically significant (Zou, 2007).

The Relationship between Memory Tasks and Other Higher-Order Thinking Skills

The recognition task currently included in the scale affects the general intelligence and index scores obtained from the ASIS intelligence measures to a certain extent. Therefore, correlation analyses between composite scores (general intelligence, visual intelligence or memory index, etc.) and memory measurements may cause biased measurements. Hence, examining the correlations of recognition and recall task measurements with other higher-order thinking skills that make up these composite scores, rather than the composite scores themselves, will provide clearer data about the general structure. The correlations of recognition and recall tasks with other higher-order thinking skills were examined and the results of this analysis are presented in Table 3.

Table 3. Correlation Values of Recognition and Recall Measurements with the ASIS Subtests

		Recall	Verbal Analogies	Visual Flexibility	Visual-Spatial Analogies	Verbal Short-Term Memory	Visual-Spatial Design Memory	Words and Meanings
Recognition	r	.362**	.147*	.105	.256**	.183**	.489**	.189**
	p	.000	.027	.113	.000	.006	.000	.004
Recall	r	1	.270**	.151*	.363**	.216**	.321**	.284**
	p		.000	.022	.000	.001	.000	.000

* The correlation is significant at the 0.05 level.

** The correlation is significant at the 0.01 level.

As can be seen in Table 3, the correlation of the recall task measurements with the other five subtests in the ASIS intelligence scale was found to be higher than that of the recognition task measurements.

Discussion

Many studies in the literature (Bower, 2000; Margolis, 1992; Radvansky, 2017) confirm that recognition tasks are easier than recall tasks. As expected, the mean recognition task scores of the students participating in the study are significantly higher than the mean scores they got from the recall task. The lowest 0 and the highest 20 points can be obtained from memory tasks. The same group of students got the lowest 0 and the highest 13 points in two different measurements. Since the base score is 0, it is not surprising that the two measurements are similar. However, on a scale with a maximum score of 20, it gives confidence that the same student group consistently gets the highest score of 13 in two different memory tasks.

Although students scored in the 0-13 range in both memory tasks, it was observed that the range or spread of the items the students answered in these was not similar. While the students reached only the 14th item in the recall task, they progressed to the 20th item in the recognition task. This can be explained by the fact that memory capacity limits, which is 7±2 items in short-term memory tasks, is approximately 4±1 items in working memory tasks with high processing load (Cowan, 2001). Because the tasks that increase the central executive burden cause a decrease in the amount of information kept in the storage units specific to the area in memory (Heitz et al., 2005). On the other hand, the fact that the answer item spread is narrower in the recall task measurements provides an advantage to the scale in terms of the ceiling effect. The ceiling effect refers to the lack of a sufficient number of high-order test items to effectively distinguish students in the upper percentile (VanTassel-Baska, 2007). The problem of ceiling effect is often be encountered in the scales administered to gifted individuals. Although the highest scores of the student group for both measurements are consistent, the fact that the range of items answered in the recognition task measurements is larger than the recall task measurements makes the probability of the ceiling effect more likely. The chance factor in multiple-choice items can be shown as the reason for such a situation in the measurement of recognition tasks. The fact that the chance factor was almost non-existent in the recall tasks and the partial rise in the difficulty level may have raised the ceiling a little higher for students.

A close examination of the data revealed that the score of 133 students (58%) obtained from the measurement of the recognition task was higher than the score they obtained from the recall task. This finding is consistent with theories that state that recognition tasks are easier than recall tasks (eg, threshold value theory, generate-recognize model, etc.). While 31 students (14%) got the same score in both measurement types, 64 students (28%) got higher scores from the recall task measurements. This result, which contradicts with the studies and theories indicating that recognition tasks are easier than recall tasks, is explained with Cowan's (1999) embedded processing model and Vygotsky's (1978) zone of proximal development (ZPD) hypothesis. These two explanations are closely related to attention processes, only differing in whether the attention paid is intentional or unintentional.

Some students' higher scores in the recall task measurements may be explained by the changes in the type of measurement. According to Cowan's (1999) embedded processing model, attention is directed in two different ways. The person may bring some stimuli into the focus of attention deliberately and intentionally or some stimuli may enter the focus of attention of the person unintentionally. Generally, new and different stimuli are drawn to the focus of attention more easily (Cowan, 2001). Students are more familiar with multiple-choice, pencil-and-paper measurements, one of the traditional forms of assessment frequently used in the education system. The introduction of different materials into the assessment environment along with the recall task may have facilitated some students' attention to the activity, as proposed by Cowan's working memory model. In addition, the curiosity aroused by the materials and the students' desire to play with the materials may have increased the motivation of some students during the recall task measurements. With the increasing attention level and motivation, it seems possible for the student to get a higher score in the recall task measurements.

Another explanation is closely related to Vygotsky's (1978) zone of proximal development hypothesis. The zone of proximal development offers psychologists and educators a tool to understand the internal development process (Vygotsky, 1978, p. 87). According to this hypothesis, student development is closely associated with the optimal arrangement of learning environments. The effect of such arrangements on the student can be defined from both cognitive and emotional perspectives. From a cognitive perspective, the materials, tools, and equipment used should not be too difficult or too easy. The fact that the material is too difficult or easy causes some emotional reactions in the student. When the activity is very simple, the student may get bored, and when it is too difficult, it can lead to confusion and disappointment. Boredom and confusion may result in distraction, disappointment, and lack of motivation (Murray & Arroyo, 2002). Recall tasks may have increased the difficulty level of the measurement slightly and placed a little more responsibility on the student to create their own answer. In the recognition task tests, the possible answers are given, so it is impossible for the student to be unable to produce any answers to the question. Some students who did not have to come up with an answer may not have directed enough attention to the activity. Conversely, in recall tasks, the student has to produce their own answer. A student who is aware of the responsibility of giving their own answer may have directed their attention to the activity in a controlled manner during the process to answer the questions. In addition to all these, Vogel and Schwabe (2016) stated that while moderate stress enhances memory for the source of stress, it may disrupt the encoding of stressor-unrelated stimuli. In this case, the moderate stress caused by the increased level of difficulty may have caused the students to focus on the task and to turn themselves off from other distracting stimuli in the environment.

A student with a high score in one type of measurement, is expected to get an approximate score in other types of measurement. However, it was observed that a student who got a high score in one type of measurement received a very low score in the other measurement type (Participant 59: Recall = 13, Recognition = 5; Participant 193: Recall = 13, Recognition = 5; Participant 221: Recall = 13, Recognition = 7; Participant 102: Recognition = 13, Recall = 6; Participant 209: Recognition = 13, Recall = 9). According to Dehn (2015), the fact that the recognition task scores are higher than the unusually recall task scores indicates that the piece of information is in long-term storage, but the individual has difficulty in retrieving this bit of information at will. In such cases, it can be assumed that retrieval issues are not storage related. When the same information is recalled with recognition and recall tasks, and in cases where the performance in recognition tasks is not significantly better than the recall tasks, it can be concluded that the information is never actually integrated, is forgotten, and is no longer included in long-term memory (Dehn, 2015).

The correlations of recognition and recall tasks with each other and with other higher-order thinking skills were examined. The correlation coefficient between recognition and recall task measurements was found to be .36. Further, it is observed that the correlation coefficient between the recognition task and another subtest measure measuring visual-spatial working memory with the recognition type questions is .49. Theoretically, the correlation between two forms consisting of the same questions is expected to be the highest. However, as the memory task changes, the same stimuli may be subjected to different types of processing in the mind. However, although composed of different stimuli, the two forms sharing the same memory task have similar cognitive processing. Therefore, it can be said that with the change of the memory task, the cognitive processing differs and the process of the subtests that share similar memory tasks is also similar. Thus, forms that measure visual-spatial working memory with recognition tasks may have shown higher correlation than forms that measure visual working memory with different memory tasks (recognition and recall).

A higher correlation was found between the recall task measurements and the visual analogical reasoning measurements compared to other measures. Visual analogical reasoning form measures fluent intelligence and reasoning. This finding is supported by those of other studies in the literature. There are some studies (Engle et al., 1999; Fry & Hale, 1996; Kane et al., 2005; Kyllonen, 1993; Kyllonen & Christal, 1990) indicating that working memory has a strong relationship with reasoning and fluid intelligence. Moreover, in their meta-analysis study, Süß et al. (2002) found a stronger relationship between working memory capacity and reasoning than other higher-order thinking skills and even some memory factors. Kane et al. (2005) state that working memory is more closely related to fluid intelligence and reasoning than short-term memory. Schneider and McGrew (2018) state that the tests differ according to the "processing" and "storage" requirements, and they described the tests consisting of items requiring low processing and high storage as short-term memory tests, and tests consisting of items requiring high processing low storage as attention control tests. They classified the tests consisting of items requiring simultaneous processing and storage as working memory tests. Due to the increased processing requirement, it can be thought that recall tasks predict working memory better than recognition tasks, and therefore have a higher correlation with fluid intelligence and reasoning tasks. On the other hand, recognition tasks with high storage requirements are a pure measure of short-term memory (Chubala et al., 2020; Gisselgard et al., 2007).

Conclusion

As we have stated before, although the recall and recognition tasks both evaluate the retrieval processes, there is no consensus on how much these processes overlap or differ. Although Unsworth and Engle (2007) found that recall tasks predict working memory as well as complex span tasks, some researchers (Chubala et al., 2020; Gisselgård et al., 2007) argue that recall tasks are influenced by long-term memory and therefore cannot provide as a pure measurement as recognition tasks. On the other hand, recall tasks require a stronger memory trace than recognition tasks. According to Heacox and Cash (2014), the stronger the memory trace, the easier and faster the information is called, and it leaves enough energy for higher-order thinking. Therefore, the ability to successfully recall the information (rather than recognize) may play an important role in facilitating higher-order thinking skills. The findings of our study revealed that the correlation of the recall task score with the other 5 subtests in the ASIS intelligence scale was higher than that of the recognition task score. In addition, the findings also show that the psychometric properties of the recall task measurements are stronger. Considering that the ability to encode information quickly and recall when necessary may play an important role in learning, specialization, and employing higher-order thinking skills, the use of recall tasks in cognitive tests can be recommended.

References

Abdelmoula, M., Chakroun, W., & Akrout, F. (2015). The effect of sample size and the number of items on reliability coefficients: Alpha and Rhô: A meta-analysis. *International Journal of Numerical Methods and Applications*, *13*(1), 1–20. https://doi.org/10.17654/ijnmamar2015_001_020

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Andrew, D. M., & Bird, C. (1938). A comparison of two new-type questions: recall and recognition. *Journal of Educational Psychology*, *29*(3), 175–193. https://doi.org/10.1037/h0062394

Bower, G. H. (2000). A brief history of memory research. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 3–32). Oxford University Press.

Cleary, A. M. (2019). Dependent measures in memory research: From free recall to recognition. In H. Otani & B. L. Schwartz (Eds.), *Handbook of research methods in human memory* (1st ed., pp. 19–35). Routledge.

Chubala, C. M., Guitard, D., Neath, I., Saint-Aubin, J., & Surprenant, A. M. (2020). Visual similarity effects in immediate serial recall and (sometimes) in immediate serial recognition. *Memory & Cognition*, *48*(3), 411–425. https://doi.org/10.3758/s13421-019-00979-5

Chubala, C. M., Neath, I., & Surprenant, A. M. (2019). A comparison of immediate serial recall and immediate serial recognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *73*(1), 5–27. https://doi.org/10.1037/cep0000158

Chubala, C., Surprenant, A. M., Neath, I., & Quinlan, P. T. (2018). Does dynamic visual noise eliminate the concreteness effect in working memory? *Journal of Memory and Language*, *102*, 97–114. https://doi.org/10.1016/j.jml.2018.05.009

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). https://doi.org/10.1017/CBO9781139174909.006

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

Dehn, M. J. (2015). *Essentials of working memory assessment and intervention* (1st ed.). Wiley.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications, Inc.

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Diedenhofen, B., & Musch, J. (2016). cocron: A web interface and R package for the statistical comparison of cronbach's alpha coefficients. *International Journal of Internet Science*, *11*(1), 51–60.

Dunn, O. J., & Clark, V. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, *64*(325), 366–377. https://doi.org/10.1080/01621459.1969.10500981

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Ertürk, S. (2013). *Eğitimde "program" geliştirme* (6. Basım). Edge akademi.

Fields, R. D. (2020, March 1). The brain learns in unexpected ways. Scientific American. https://www.scientificamerican.com/article/the-brain-learns-in-unexpected-ways/

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241. https://doi.org/10.1111/j.1467-9280.1996.tb00366.x

Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2011). *Human resource selection* (7th ed.). Cengage Learning.

Gisselgård, J., Uddén, J., Ingvar, M., & Petersson, K. M. (2007). Disruption of order information by irrelevant items: A serial recognition paradigm. *Acta Psychologica*, *124*(3), 356–369. https://doi.org/10.1016/j.actpsy.2006.04.002

Hasan, M., Kanna, M. S., Jun, W., Ramkrishnan, A. S., Iqbal, Z., Lee, Y., & Li, Y. (2019). Schema-like learning and memory consolidation acting through myelination. *The FASEB Journal*, *33*(11), 11758–11775. https://doi.org/10.1096/fj.201900910r

Heacox, D., & Cash, R. M. (2014). *Differentiation for gifted learners: Going beyond the basics*. Free Spirit Publishing.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61–77). Sage Publications.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

Ho, R. (2014). *Handbook of univariate and multivariate data analysis with IBM SPSS* (2nd ed.). CRC Press.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 331–373). https://doi.org/10.1016/b978-0-12-521350-9.50016-4

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology*, *120*(3), 375–405. https://doi.org/10.1080/00221309.1993.9711154

Kyllonen, P. C., & Christal, R. E. (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

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Liu, H. Y., & Weng, L. J. (2009). An effect size index for comparing two independent alpha coefficients. *British Journal of Mathematical and Statistical Psychology*, *62*(2), 385–400. https://doi.org/10.1348/000711008X315518

Margolis, L. J. (1992). *Do recall and recognition rely on qualitatively different processes?* [Doctoral dissertation, The American University]. American University Digital Research Archive. https://dra.american.edu/islandora/object/thesesdissertations:2673

Murray, T., & Arroyo, I. (2002, June 2-7). Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. Intelligent Tutoring Systems, 6th

International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain. https://doi.org/10.1007/3-540-47987-2_75

Nutley, S. B., & Söderqvist, S. (2017). How is working memory training likely to influence academic performance? Current evidence and methodological considerations. *Frontiers in Psychology*, *8*, 1–12. https://doi.org/10.3389/fpsyg.2017.00069

Radvansky, G. A. (2017). Formal models of memory. In *Human memory* (3rd ed., pp. 325–355). Routledge. https://doi.org/10.4324/9781315542768-10

Ramseyer, G. C. (1979). Testing the difference between dependent correlations using the Fisher Z. *The Journal of Experimental Education*, *47*(4), 307-310. https://doi.org/10.1080/00220973.1979.11011698

Sak, U., Bal-Sezerel, B., Ayas, M. B., Tokmak, F., Özdemir, N. N., Demirel-Gürbüz, Ş., & Öpengin, E. (2016). *Anadolu Sak Zeka Ölçeği: ASİS uygulayıcı kitabı [Anadolu Sak Intelligence Scale: ASIS user manual]*. Anadolu Üniversitesi ÜYEP Merkezi.

Schneider, W. J. ve McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment* (4th ed., pp. 73–163). Guilford Press.

Schwartz, B. L. (2018). Introduction to the study of memory. In B. L. Schwartz (Ed.), *Memory: Foundations and applications* (3rd ed., pp. 1–33). Sage.

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, *28*(2), 229–237. https://doi.org/10.1016/s0191-8869(99)00093-8

Silver, N. C., Hittner, J. B., & May, K. (2004). Testing dependent correlations with nonoverlapping variables: A monte carlo simulation. *The Journal of Experimental Education*, *73*(1), 53-69. https://doi.org/10.3200/jexe.71.1.53-70

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability - and a little bit more. *Intelligence*, *30*(3), 261–288. https://doi.org/10.1016/s0160-2896(01)00100-3

Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *Quarterly Journal of Experimental Psychology*, *64*(12), 2425–2437. https://doi.org/10.1080/17470218.2011.604787

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

VanTassel-Baska, J. (2007). An overview of alternative assessment measures for gifted learners and the issues that surround their use. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 1–15). Prufrock Press.

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: Implications for the classroom. *Npj Science of Learning*, 1–10. https://doi.org/10.1038/npjscilearn.2016.11

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*(4), 399–413. https://doi.org/10.1037/1082-989x.12.4.399