Routledge
Taylor & Francis Group

Journal of Clinical and Experimental Neuropsychology

ISSN: 1380-3395 (Print) 1744-411X (Online) Journal homepage: http://www.tandfonline.com/loi/ncen20

Item analyses of memory differences

Timothy A. Salthouse

To cite this article: Timothy A. Salthouse (2017) Item analyses of memory differences, Journal of Clinical and Experimental Neuropsychology, 39:4, 326-335, DOI: 10.1080/13803395.2016.1226267

To link to this article: https://doi.org/10.1080/13803395.2016.1226267

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY, 2017
VOL. 39, NO. 4, 326–335
http://dx.doi.org/10.1080/13803395.2016.1226267

Routledge
Taylor & Francis Group

Item analyses of memory differences

Timothy A. Salthouse

Department of Psychology, University of Virginia, Charlottesville, VA, USA

ABSTRACT

Objective: Although performance on memory and other cognitive tests is usually assessed with a score aggregated across multiple items, potentially valuable information is also available at the level of individual items. *Method*: The current study illustrates how analyses of variance with item as one of the factors, and memorability analyses in which item accuracy in one group is plotted as a function of item accuracy in another group, can provide a more detailed characterization of the nature of group differences in memory. Data are reported for two memory tasks, word recall and story memory, across age, ability, repetition, delay, and longitudinal contrasts. *Results*: The item-level analyses revealed evidence for largely uniform differences across items in the age, ability, and longitudinal contrasts, but differential patterns across items in the repetition contrast, and unsystematic item relations in the delay contrast. *Conclusion*: Analyses at the level of individual items have the potential to indicate the manner by which group differences in the aggregate test score are achieved.

ARTICLE HISTORY
Received 2 May 2016
Accepted 14 August 2016

KEYWORDS
Group-by-item interactions; Group differences; Memorability analysis; Score decomposition; Systematic relations

Memory performance is typically evaluated with an aggregate score representing the sum of items on a test assumed to assess memory. However, performance on individual items within memory tests can also be analyzed to evaluate how the total test score is achieved. That is, individual items can be considered equivalent to assessments varying in characteristics such as serial position and semantic relatedness, and in a manner analogous to inferences based on comparisons across different types of tasks, comparisons at the level of individual items might be informative about the factors contributing to differences in aggregate performance.

One method of analyzing data at the item level when the primary focus is on differences between two groups is with an analysis of variance (ANOVA) in which item, group, and the interaction of group and item are factors. Of greatest interest in this type of analysis is the interaction because it indicates whether the difference between groups is uniform, or varies across items. Although informative, interactions can be manifested in several ways that are not easily distinguished in an ANOVA. For example, the group difference could be largest on the easiest items, on the most difficult items, or on items that are not systematically related to item difficulty. Fortunately, these alternatives can be distinguished with a complementary analytical procedure known as memorability analysis in which item accuracy in the poorer performing group is plotted as a function of item accuracy in the better performing group (Rubin, 1978).

Figure 1, based on a figure in Stine and Wingfield (1988), schematically portrays possible outcomes from memorability analyses involving two groups. The data points in the figures represent average accuracy for individual items, with one group plotted along the x axis and the other along the y axis. The distinguishing feature of the outcome in Panel A is low R^2 values, indicating little or no systematic relation between the average performance on individual items in the two groups. A pattern such as this could occur if different factors contribute to the item variation in the two groups, possibly with systematic influences of serial position or semantic relatedness in one group, and momentary attentional lapses resulting in the random failure to encode or process items in the other group.

CONTACT Timothy A. Salthouse Salthouse@virginia.edu Department of Psychology, 102 Gilmer Hall, University of Virginia, Charlottesville, VA 22904-4400, USA

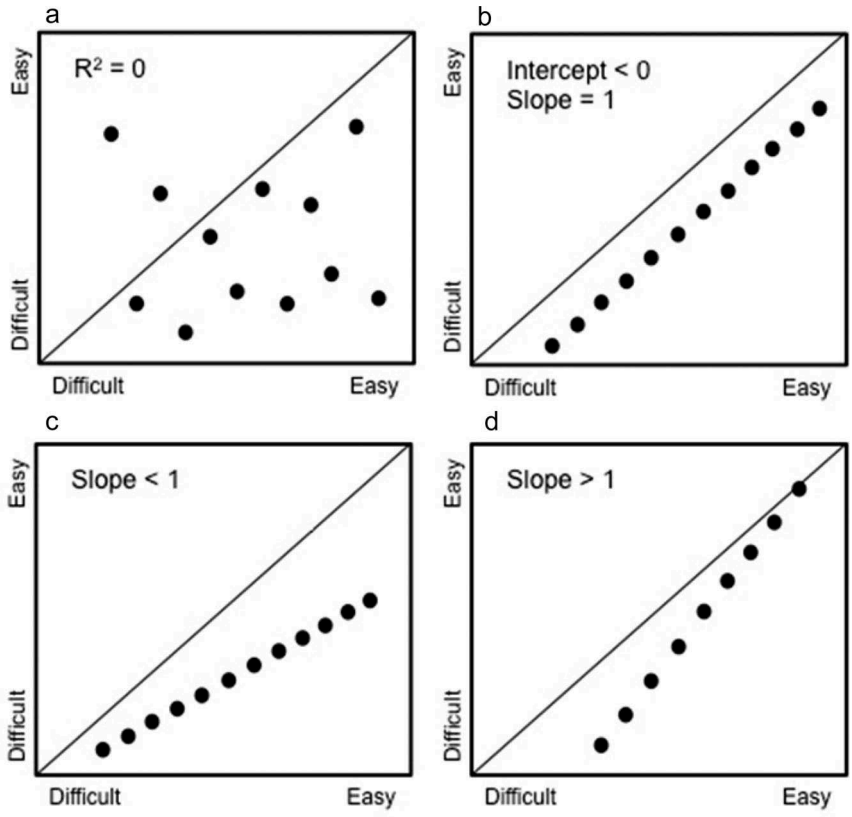

Figure 1. Schematic illustration of possible outcomes from memorability analyses. Each point represents the hypothetical accuracy for an individual item averaged across participants. The *x* axis corresponds to performance in one group (or condition), and the *y* axis correspond to performance in another group (or condition).

If the relation between the performance in the two groups is systematic, as manifested by high R^2 values, examination of the intercept and slope parameters of the regression equation can be informative about the specific nature of the relation. For example, the outcome in Panel B has an intercept less than 0 and a slope close to 1, which implies nearly uniform reductions across items varying in average accuracy. A pattern such as this, particularly when accompanied by little or no group-by-item interaction, would be consistent with a primarily quantitative difference between the groups. The outcome in Panel C has a slope less than 1, indicating that the largest group differences were on the easy items with the highest accuracy levels. This type of pattern, with nonuniform differences across items, would be consistent with a qualitative difference between groups, possibly attributable to a failure in the lower performing group to capitalize on item characteristics (such as semantic relatedness) conducive to high performance. Finally, the outcome in Panel D has a slope greater than 1, indicating that the largest group differences were on difficult items with the lowest accuracy levels. This pattern is also consistent with a qualitative difference, possibly with fewer resources available in the lower performing group to cope with difficult items.

Distinguishing among the four alternatives portrayed in Figure 1 requires moderately large samples of individuals to obtain precise estimates of the average accuracy for each item, and a sufficient number of items to provide sensitive estimates of the parameters of the regression equations. Data with these properties are available in the Virginia Cognitive Aging Project (VCAP), which is a cross-sectional and longitudinal study of adults between 18 and 99 years of age (Salthouse, 2014; Salthouse, Pink, & Tucker-Drob, 2008). A novel feature of this project is that the participants performed alternate versions of two different memory tests on each of three sessions, which results in more relevant data than in situations in which only a single test version is performed.

The two memory tests are commonly used in neuropsychological assessment, and therefore,

unlike experimental investigations, all examinees received the tests and items in the same order. The word recall test (Wechsler, 1997) consists of four auditory presentations of a list of 12 unrelated words with a recall attempt after each presentation, followed by the presentation and recall of a new list (List B), and another attempt to recall the original list without an additional presentation. In the delayed version of the test the participants attempted to recall the initial list after an interval of about 20–30 min occupied by the performance of other cognitive tests.

The logical memory test (Wechsler, 1997) involves the auditory presentation of a story followed by an attempt to recall as many details as possible, with the recall scored in terms of 25 prespecified idea units. The initial story is followed by a second story, which is presented twice with a recall attempt after each presentation. In the delayed version of the test the participants attempted to recall both stories after an interval of about 20–30 min, during which they performed other cognitive tasks.

Five comparisons with the same two memory tests are described in the current report: age, involving a contrast of healthy adults 18–35 and 65–85 years of age; ability, involving a contrast of adults 65 and older who were in the top and bottom quartiles of the distribution of memory performance; repetition, involving a contrast of adults 65 and older on four successive recall trials of the same items, and on the first and second trial of the second logical memory story; delay, involving a contrast of adults 65 and older on the last immediate trial and on the delayed trial; and longitudinal, involving adults 65 and older who experienced moderate memory declines from a first to a second occasion. All of the comparisons except age were based on adults 65 and over to reduce influences of age in the relevant contrasts, and because individuals in this age range exhibited the greatest longitudinal declines.

The dependent variable in all analyses was the average accuracy across participants on the individual items in each test. The ANOVAs included group (or condition) and item (i.e., 12 for recall, and 25 for logical memory) as the primary factors, with session, and trial when relevant, as secondary factors. The group (condition) factor was a between-subjects comparison for age and ability, and a within-subjects comparison for the repetition, delay, and longitudinal contrasts. The item factor was a within-subjects factor in all analyses. Alternative analytical methods could have been used, such as multilevel models, but the group-by-item ANOVAs are sufficient to illustrate the primary type of information available from this approach. Main effects of group (condition) and item were expected in each contrast, with the interactions informative about possible differential performance across items.

The memorability analyses were based on regression analyses examining the relation between performance (averaged across participants) on individual items in the two groups, or conditions. Two sets of regression analyses were conducted in each contrast. In one set, performance on each item was averaged across the three sessions and the n trials, and in the second set, the analyses were based on the $3 \times n$ individual items. The former analyses based on averages have the advantage of greater precision of the estimates of each data point, and the latter analyses have the advantage of a larger number of data points, which should increase sensitivity of the regression parameters. As noted above, parameters of the regression equations were examined to characterize the nature of the group (condition) differences with respect to systematicity (i.e., R^2) and pattern (i.e., intercept relative to 0, and slope relative to 1).

Because prior research by Stine and Wingfield (1988) involving age comparisons revealed systematic functions with a slope close to 1, these results were expected to be replicated in the current study. The longitudinal contrast also involves an age comparison, and thus a pattern similar to that with the age contrast was expected for the longitudinal contrast. However, there were no a priori hypotheses concerning the ability, repetition, and delay contrasts because in each case plausible arguments could be proposed for several of the outcomes portrayed in Figure 1.

Method

Characteristics of the participants in each comparison are reported in Table 1. The participants were recruited by advertisements and referrals from other participants. The primary exclusionary criteria consisted of a diagnosis of dementia, and auditory or visual limitations severe enough to impair performance on the tests.

Scaled scores for the word recall and logical memory tests are adjusted for age, and have means of 10 and standard deviations of 3 in the nationally representative normative sample. With the exception of

Table 1. Descriptive characteristics of the participants in each comparison.

Contrast	Groups		Effect Size
Age 18–35 and 65–85	Age 18–35	Age 65–85	*d*
N	628	669	NA
Age (years)	25.3 (4.8)	73.1 (5.7)	−8.99*
Proportion female	.61	.59	0.04
Years of education	14.9 (2.3)	16.0 (2.8)	−0.43*
Health	2.1 (0.9)	2.4 (0.9)	−0.27*
MMSE	28.8 (1.7)	27.9 (2.2)	0.44*
Logical memory SS	11.4 (3.0)	11.9 (3.1)	−0.18*
Word recall SS	11.7 (3.1)	12.1 (3.3)	−0.11
Ability, age 65–99	Top 25%	Bottom 25%	
N	179	173	NA
Age (years)	74.9 (6.3)	77.8 (7.1)	−0.49*
Proportion female	.73	.47	0.55*
Years of education	16.7 (2.7)	15.4 (2.8)	0.46*
Health	2.4 (0.9)	2.4 (1.0)	−0.06
MMSE	28.7 (1.6)	26.5 (2.8)	0.98*
Logical memory SS	14.9 (1.7)	8.3 (2.7)	2.83*
Word recall SS	15.5 (2.0)	8.1 (2.6)	3.24*
Repetition, age 65–90			
N	712		
Age (years)	74.1 (6.8)		
Proportion female	.59		
Years of education	16.0 (2.9)		
Health	2.4 (0.9)		
MMSE	27.8 (2.2)		
Logical memory SS	11.8 (3.1)		
Word recall SS	12.0 (3.3)		
Delay, age 65–91			
N	530		
Age (years)	72.7 (5.9)		
Proportion female	.66		
Years of education	16.4 (2.8)		
Health	2.2 (0.9)		
MMSE	28.3 (2.0)		
Logical memory SS	12.7 (3.4)		
Word recall SS	12.9 (2.8)		
Longitudinal, age 65–91	Time 1	Time 2	
N	72	72	NA
Age (years)	74.9 (6.9)	77.8 (7.1)	−2.25*
Proportion female	.46		
Years of education	15.5 (2.7)		
T1–T2 interval (years)	2.8 (1.2)		
Health	2.4 (0.9)	2.5 (0.9)	0.13
MMSE	27.6 (2.1)	28.0 (2.0)	−0.13
Logical memory SS	13.6 (2.6)	10.3 (3.4)	1.29*
Word recall SS	14.0 (3.2)	9.3 (3.2)	1.82*

Note. Health is a self rating on a scale ranging from 1 for "excellent" to 5 for "poor." MMSE is the Mini-Mental State Exam (Folstein et al., 1975) used to screen for dementia. Logical memory SS and word recall SS are age-adjusted scaled scores. Effect sizes for the longitudinal comparisons are based on the within-subject formula. NA indicates that the value was not available.

*$p < .01$.

the ability and longitudinal contrasts, the scaled scores averaged about 0.5 standard deviations above the means in the normative sample, indicating that most of the participants were high functioning. However, it is important to note that the variability in the samples, as reflected by the standard deviations close to 3, was similar to that in the nationally representative normative sample.

Compared to young adults, the older adults in the age comparison had poorer self-ratings of health and lower scores on the MMSE test (Folstein, Folstein, & McHugh, 1975) frequently used to screen for dementia, but more years of education, and higher logical memory scaled scores. The two ability groups were formed from individuals in the top and bottom quartiles of the average scaled score. The participants in the high-ability group were younger, with a higher proportion of females, more years of education, and higher Mini-Mental State Examination (MMSE) and scaled memory scores than the individuals in the low-ability group.

The comparisons were based on the initial recall trials of the lists involving different words (i.e., Trial 1 and List B) and on the first trials in the two logical memory stories in the age, ability, and longitudinal contrasts. The within-person repetition contrast was between performance on Trial 1 and performance on Trials 2, 3, and 4 for word recall, and between performance on the first and second trials of the second story for logical memory. The within-person immediate-delay contrast was between performance on the last immediate trial (i.e., Trial 4) and the delayed trial for word recall, and performance on the last immediate trials (i.e., Trial 1 for Story A, Trial 2 for Story B), and the delayed trials for logical memory. In order to focus on individuals exhibiting longitudinal decline, participants in the longitudinal comparison were selected on the basis of a decrease of at least two points from the first to the second occasion in the average scaled scores for logical memory and word recall.

Results

All comparisons are reported in the same format. Results of ANOVAs with group or condition as one factor and item as a second factor are reported in Table 2, and results of regression analyses are reported in Table 3. The left panels in Figures 2 through 6 portray average accuracy (and standard errors) by item input position in the two groups or conditions, and the right two panels of each figure portray the corresponding memorability functions with accuracy in the lower performing group or condition plotted as a function of accuracy in the higher performing group or condition. The top panels in each figure represent results from the word recall task, and the bottom two panels represent results from the logical memory task.

Table 2. Analysis of variance results for the group/condition and item main effects, and Group/Condition × Item interactions.

Factor	Recall F (η^2_p)	Logical memory F (η^2_p)
Age	429.84* (.255)	177.37* (.130)
Item	1211.09* (.491)	443.39* (.271)
Age × Item	34.13* (.026)	11.34* (.009)
Ability	307.98* (.478)	505.81* (.615)
Item	416.67* (.553)	95.50* (.232)
A × Item	2.64* (.008)	2.31* (.007)
Repetition: 2	3040.97* (.816)	5571.58* (.897)
Item	574.47* (.455)	291.25* (.312)
R2 × Item	33.30* (.046)	23.18* (.035)
Repetition: 3	4829.12* (.875)	
Item	537.38* (.438)	
R3 × Item	61.54* (.082)	
Repetition: 4	5853.56* (.895)	
Item	512.24* (.427)	
R4 × Item	86.25* (.111)	
Delay	2048.40* (.862)	1808.12* (.851)
Item	77.35* (.191)	92.44* (.226)
D × Item	36.38* (.100)	124.89* (.283)
Occasion	23.94* (.406)	33.01* (.408)
Item	31.79* (.476)	30.07* (.385)
O × Item	0.87 (.024)	1.24 (.025)

Note. Variance results: F ratio and partial eta^2. A = age; R2 = Repetition 2; R3 = Repetition 3; R4 = Repetition 4; D = delay; O = occasion.
*$p < .01$.

Table 3. Regression parameters for analyses of item accuracy in two groups or conditions.

Contrast	No. data points	R^2	Intercept	Slope
Word recall				
Age	12	.925	−0.241*	1.170
	72	.861	−0.189*	1.075
Ability	12	.973	−0.179*	0.992
	72	.888	−0.150*	0.935
Repetition: 2	12	.934	−0.301*	1.224
	36	.908	−0.262*	1.161*
Repetition: 3	12	.913	−0.630*	1.558*
	36	.866	−0.548*	1.438*
Repetition: 4	12	.910	−0.946*	1.887*
	36	.851	−0.844*	1.748*
Delayed	12	.230	0.008	0.456
	36	.227	−0.038	0.519*
Longitudinal	12	.967	−0.076	1.027
	72	.885	−0.047	0.959
Logical memory				
Age	25	.902	−0.043	0.914
	150	.878	−0.109*	1.022
Ability	25	.915	−0.166*	0.889
	150	.853	−0.142*	0.844*
Repetition	25	.850	−0.423*	1.310*
	75	.874	−0.235*	1.081
Delayed	25	.244	0.287*	0.399*
	150	.774	−0.054	0.917
Longitudinal	25	.887	−0.035	0.938
	150	.890	−0.020	0.894*

Note. Significance of the intercepts was relative to 0, and significance of the slopes was relative to 1.0.
*$p < .01$.

The main effects of group/condition and item were significant and moderately large (i.e., ($\eta^2_p \geq .130$) in each analysis. All of the interactions except those in the longitudinal contrast were significant, but there was a considerable range of effect size estimates (i.e., $\eta^2_p = .007$ to .283).

Age

The effect size for the group-by-item interaction with the recall measure was modest ($\eta^2_p = .026$), and the top panels of Figure 2 indicate that the serial position functions were nearly parallel with the exception of small age differences in the last input items. The memorability functions were systematic, and the regression parameters in Table 3 indicate that the functions were associated with high R^2 values, negative intercepts, and slopes slightly, but not significantly, greater than 1.

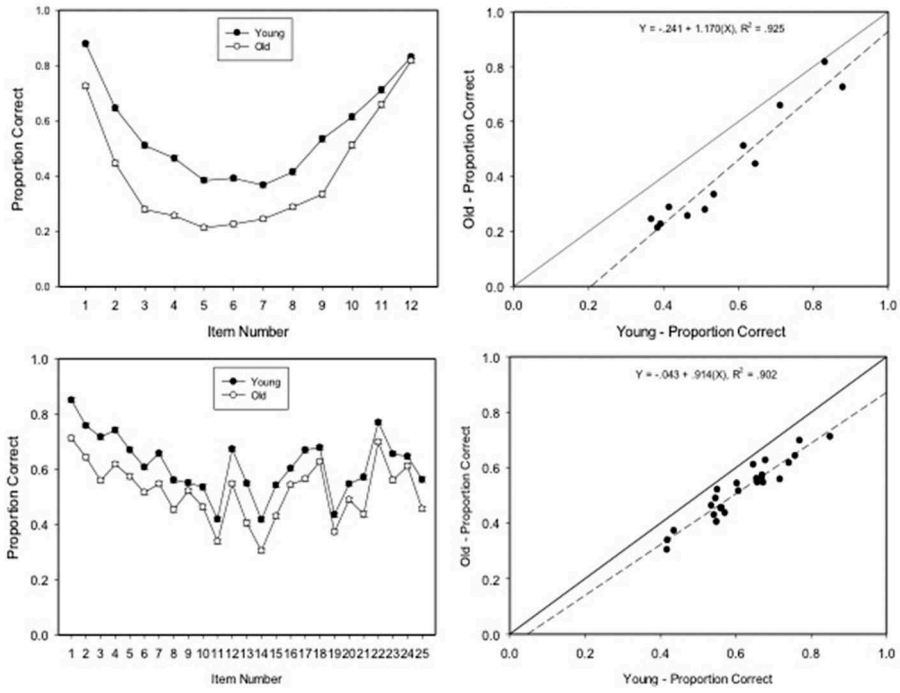

Figure 2. Serial position functions for word recall (top left) and logical memory (bottom left), and memorability functions for word recall (top right) and logical memory (bottom right) for the age contrast.

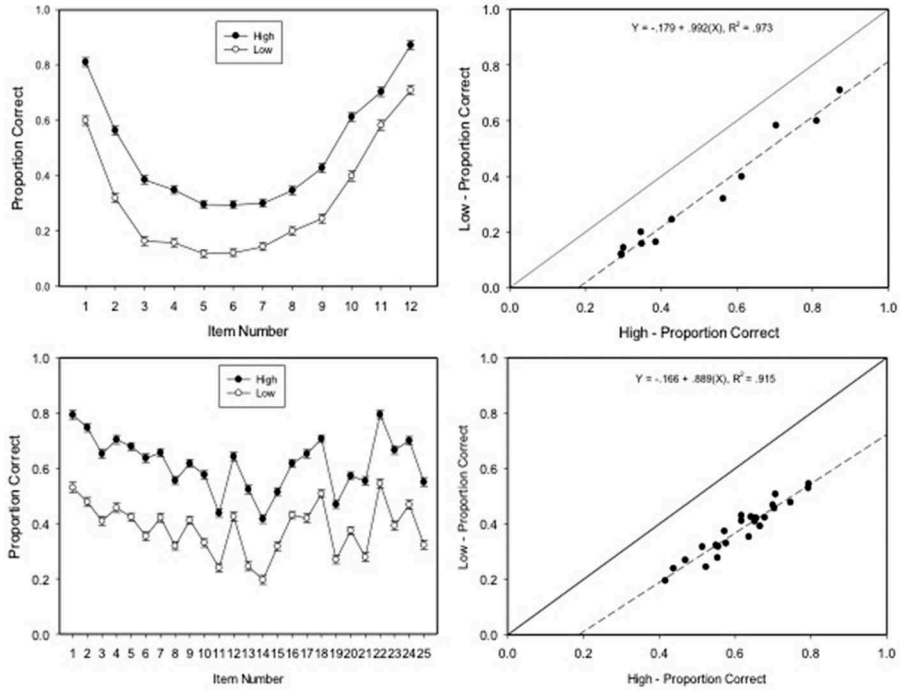

Figure 3. Serial position functions for word recall (top left) and logical memory (bottom left), and memorability functions for word recall (top right) and logical memory (bottom right) for the ability contrast.

The group-by-item interaction was also small with the logical memory measure (η^2_p = .009), and the bottom panels in Figure 2 indicate that the serial position functions were nearly parallel, and that the memorability functions were systematic. The values in Table 3 reveal high R^2 values, slightly negative intercepts, and slopes close to 1.

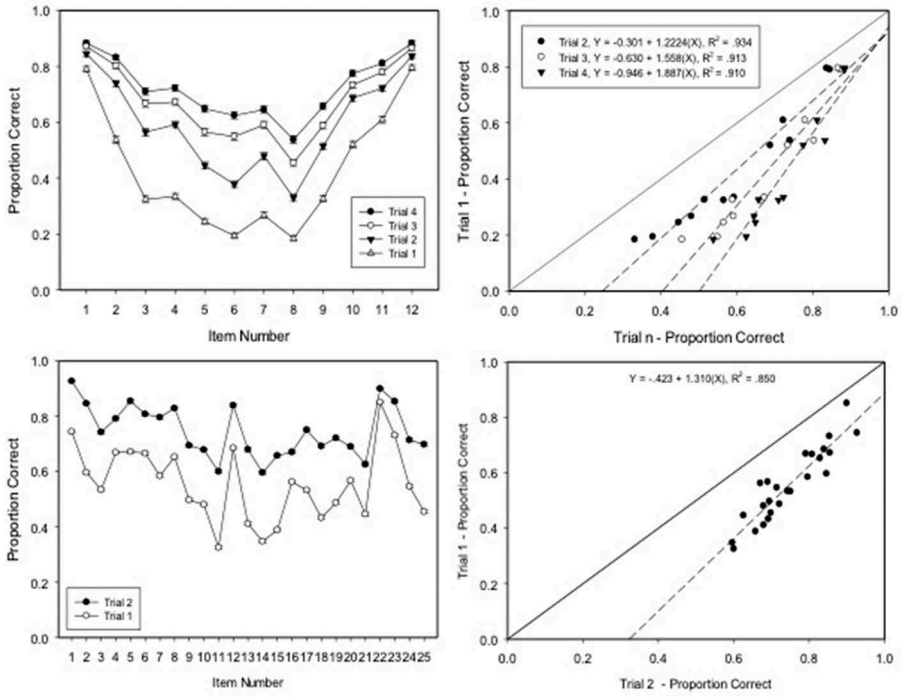

Figure 4. Serial position functions for word recall (top left) and logical memory bottom left), and memorability functions for word recall (top right) and logical memory (bottom right) for the repetition contrast.

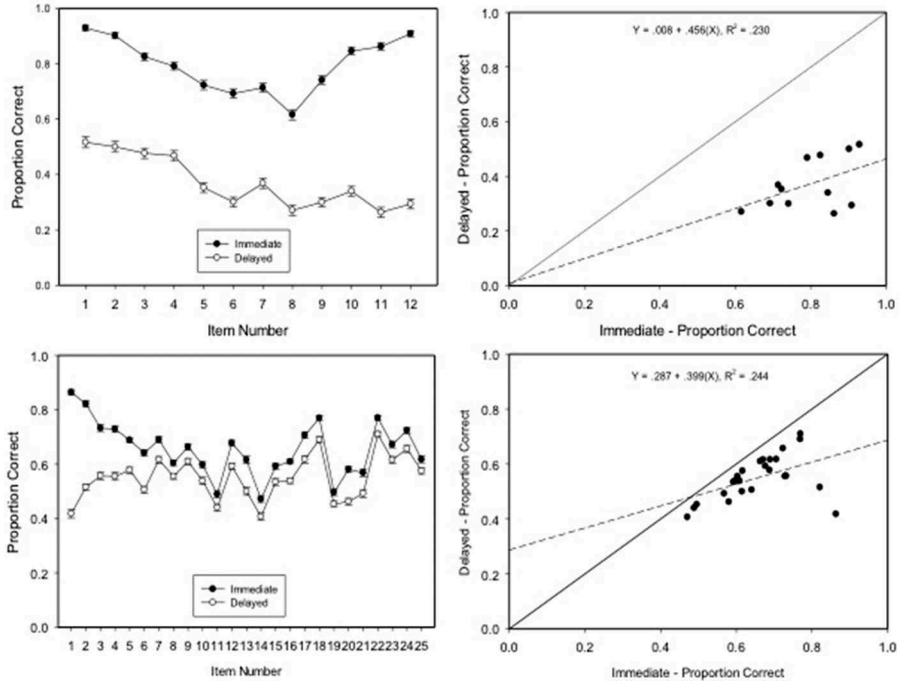

Figure 5. Serial position functions for word recall (top left) and logical memory (bottom left), and memorability functions for word recall (top right) and logical memory (bottom right) for the delay contrast.

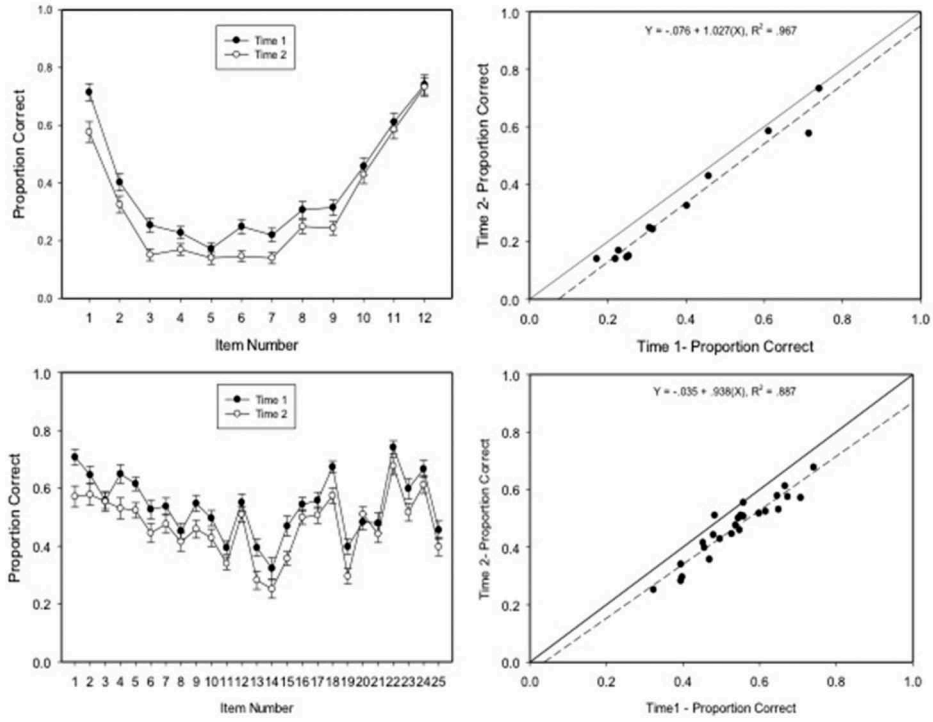

Figure 6. Serial position functions for word recall (top left) and logical memory (bottom left), and memorability functions for word recall (top right) and logical memory (bottom right) for the longitudinal contrast.

Ability

The group-by-item interaction with the recall data was small (η^2_p = .008), and the top panels in Figure 3 indicate that the serial position functions were nearly parallel with systematic memorability functions. The entries in Table 3 indicate that R^2 values were high, intercepts were slightly less than 0, and slopes were close to 1.

Very similar patterns were evident with the logical memory data, with a small group-by-item interaction (η^2_p = .007), nearly parallel serial position functions, systematic memorability functions, high R^2 values, and negative intercepts with slopes close to 1.

Repetition

In order to express the repetition memorability functions in the same format as the other comparisons, the higher performing condition (after repetition) was the reference, and the lower performing condition (the initial trial) was the contrast.

With the recall data the condition-by-item interactions were progressively larger with more repetitions, and the top panels in Figure 4 indicate that the serial position functions were less pronounced, and the memorability functions steeper, with more repetitions. The regression parameters in Table 3 revealed high R^2 values and negative intercepts with each number of repetitions, with additional repetitions associated with more positive slopes.

The interaction of condition and item was small in the logical memory data, with nearly parallel serial position functions and systematic memorability functions. The R^2 values were high, with significant negative intercepts and slopes greater than 1.

Delay

The delay contrast was between delayed performance and performance on the last exposure in the original test (i.e., Trial 4 for word recall, and Trial 1 for logical memory Story A and Trial 2 for logical memory Story B). There was a moderate condition-by-item interaction with the recall data (η^2_p = .100), and the top panels of Figure 5 indicate nearly flat serial position functions for the immediate (4th trial) and delay conditions, and unsystematic memorability functions. The regression analyses had low R^2 values, intercepts close to 0, and slopes less than 1.

The patterns were similar in the logical memory data, with a moderate interaction (partial eta^2 of .283), nearly flat serial position functions, and unsystematic memorability functions. The R^2 values in Table 3 were low, particularly with analyses based on the averages for 25 data points.

Longitudinal

The occasion-by-item interaction was not significant with the recall data, and the top panels of Figure 6 indicate that the serial position functions were nearly parallel with systematic memorability functions. The R^2 values were high, with small negative intercepts and slopes very close to 1. The logical memory data were similar, with no interaction, parallel serial position functions, and systematic memorability functions with high R^2 values, intercepts close to 0, and slopes close to 1.

Discussion

All main effects of group (condition) and item were significant in the ANOVAs, and thus the primary question in each contrast was whether the composition of total score differences varied as a function of item. Effect sizes for the condition-by-item interactions were moderately large only with the repetition and delay contrasts. Inspection of the serial position functions reveals that the interaction in the repetition contrast was associated with flattening of the serial position functions with repetition, and the interaction in the delay contrast was associated with less pronounced recency and primacy segments in the delay condition.

As indicated in the introduction, memorability analyses and the corresponding regression parameters are useful in specifying how the total score is achieved. That is, the value of R^2 is informative about the systematicity of differences across items, the intercept relative to 0 is informative about the mean difference between groups or conditions, and the slope relative to 1 is informative about possible differential relations on easy (high accuracy) or difficult (low accuracy) items.

Three different patterns were evident across the five contrasts, which were generally similar in the analyses based on all individual data points and on averages at each serial position, and in analyses with the recall and logical memory tasks. A pattern resembling that in the top left panel of Figure 1, with unsystematic memorability functions and low R^2 values, was apparent in the delay contrast. A pattern similar to the top right panel of Figure 1, with systematic memorability functions, slopes close to 1, and negative intercepts, was evident in the age, ability, and longitudinal contrasts. And finally, a pattern similar to that in the bottom right panel of Figure 1, with systematic memorability functions and slopes greater than 1, was evident in the repetition contrast.

The generally systematic relation between performance of young and old adults is consistent with earlier findings by Stine and Wingfield (1988, 1990) and their suggestion that cross-sectional age differences in memory are primarily quantitative rather than qualitative. However, this conclusion must be qualified somewhat because results in the top left panel of Figure 2 indicate that the age differences in recall were smaller in the last (recency) input positions than in other input positions.

The differences across items in the ability contrast were nearly uniform, with systematic memorability functions, and very small group-by-item interactions. This pattern suggests that the large differences in average performance associated with different levels of ability are primarily quantitative, with similar magnitude across items.

The condition-by-item interactions in the repetition contrast were large, with the greatest condition differences for items in the middle serial positions. This pattern is evident with slopes greater than 1 in the memorability functions, indicating that the condition differences were largest on items with the lowest accuracy.

The serial position functions were nearly flat in the delay contrast, with unsystematic memorability functions in both the recall and logical memory tasks. One of the reasons for the nearly flat serial position functions in the word recall task is the lack of a recency effect, corresponding to higher accuracy in later input serial positions, in the delayed condition.

The serial position functions were nearly parallel in the longitudinal contrast with systematic memorability functions and slopes close to 1, suggesting small, and primarily quantitative, changes in memory performance.

All of the group or condition effects were significant in the ANOVAs, indicating the existence of mean differences between groups or conditions. However, inclusion of item as another factor in the ANOVAs allowed interactions to be examined to determine whether the group or condition differences varied across items. In addition, the

memorability analyses indicate whether the differences are systematic and, if so, whether the differences are larger for items with relatively high, or low, levels of accuracy. Because data at the item level are available in nearly all cognitive tests, analyses such as those described in this report could be implemented without any additional data collection to provide information about the detailed nature of group or condition differences in total scores.

The application of these methods to clinical populations could be particularly informative to determine how group differences in the total score are manifested. That is, some clinical groups might differ from normal adults in a quantitative fashion, with uniformly lower performance across items, whereas other groups might differ in a more qualitative manner, with larger differences on some items than on others. Furthermore, in the latter case, identification of the characteristics of items associated with small or large differences could provide clues about the reasons for the group difference in overall performance.

Analyses such as these may also be valuable in theoretically oriented memory research. For example, the different patterns in the repetition and delay contrasts suggest that different mechanisms are likely involved in gains and losses in performance.

Four limitations of the study should be acknowledged. First, only two types of memory tasks were examined, and it is possible that the results may not generalize to other memory tasks, such as those with nonverbal material in which item effects may be less pronounced. Second, the number of longitudinal participants was relatively small, and the average interval was only 3 years. A different pattern might be obtained with a larger mean decline and more powerful analyses. Third, most of the participants were relatively high functioning, with averages of over 14 years of education and high age-adjusted scaled scores, which could limit generalizability to the broader population. However, this concern is somewhat alleviated by the finding of no qualitative differences in the ability contrast in which the low-ability group had average scaled scores substantially below the means of the nationally representative normative sample. And fourth, all of the participants were healthy adults, and thus the applicability of the methods to different types of patient groups is not known.

To summarize, scores on memory tests can be achieved in different ways, and analyses of performance at the level of individual items can be informative about how differences in the overall scores are obtained. The results of this study provided some evidence of qualitative differences in which the condition effects varied according to item, as in the repetition contrast in recall, but the differences in the other contrasts were largely quantitative, as with the cross-sectional, longitudinal, and ability contrasts.

Acknowledgements

I would like to acknowledge the valuable contributions of Alese Zeaman for conducting initial analyses on subsets of the reported data.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This research was supported by National Institute on Aging [grant number RO1AG024270]. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

References

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198. doi:10.1016/0022-3956(75)90026-6

Rubin, D. C. (1978). A unit analysis of prose memory. *Journal of Verbal Learning and Verbal Behavior*, *17*, 599–620. doi:10.1016/S0022-5371(78)90370-5

Salthouse, T. A. (2014). Correlates of cognitive change. *Journal of Experimental Psychology: General*, *143*, 1026–1048. doi:10.1037/a0034847

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*, 464–486. doi:10.1016/j.intell.2007.10.003

Stine, E. A. L., & Wingfield, A. (1988). Memorability functions as an indicator of qualitative age differences in text recall. *Psychology and Aging*, *3*, 179–183. doi:10.1037/0882-7974.3.2.179

Stine, E. A. L., & Wingfield, A. (1990). The assessment of qualitative age differences in discourse processing. In T. M. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 34–91). Amsterdam: Elsevier.

Wechsler, D. (1997). *Wechsler memory scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.