The Effect of Age and Semantic Encoding on Word Recall

Nicholas Forbes-Smith *University of Tasmania*

Aims

To determine whether the following hypotheses, outlined by the researchers, are correct:

1. For both the young and old groups, participants who used semantic encoding will recall more words than those who used verbal rehearsal but for the young group, the improvement should be larger.
2. The young people will recall more words than old group, irrespective of the level of encoding.
3. Given 1 and 2, the young group who encoded deeply should recall more words than the average of all three of the remaining groups.

Background

The data set contains measurements of 80 participants, 40 in the "Young" category and 40 in the "Old" category. The 80 participants were given a list of 35 words and asked to memorise them. Half an hour later, the number of words they could recall was recorded as the pre-treatment score. Participants were then randomly allocated to either a "Shallow" encoding group, in which they were asked to memorise another list of 35 words by rehearsing them aloud, or to a "Deep" encoding group. The Deep encoding group was taught how to use semantic encoding to help increase memory and then given the same list of words as the "Shallow" group. The post-treatment number of words recalled was then also recorded.

In order to identify differences between treatments we created a new variable which is equal to the difference between the pre and post-treatment scores for each participant (Post-score - Pre-score). This minimises the effect of individual participant's word recall abilities on the test scores, as some participants naturally have higher abilities than other participants.

The recorded variables are:

- ID - Identification number of the participant
- Age - Age category - Young or Old
- Depth - Depth of processing - Shallow for "verbal rehearsal" or Deep for "semantic encoding"
- Post - Number of words recalled post treatment
- Pre - Number of words recalled pre treatment

Findings

1. Confirming research expectations, Young participants who used who used semantic encoding were able to recall 2.0 to 5.2 more words than young participants who used verbal rehearsal *[Table 06]*. Contrary to the researchers expectations, old participants didn't show any improvement in word recall between both methods *[Table 05]*.

2. Contrary to the researchers expectations, Young participants weren't able to recall any more words than old participants, irregardless of depth of processing. *[Table 05]*

3. Young participants who used semantic encoding were able to recall 1.0 to 3.6 more words than all other participants, which confirms the researchers expectations. *[Table 06]*

Discussion

Statistical Appendix

Created Variables

Variable Name	Formula
Diff	Post-Treatment Score − Pre-Treatment Score
Cat	Age + '.' + Depth

Table 1: Created Variables table.

Computation Methods

All statistical analyses were undertaken using Jupyter lab (alpha 0.26.5) with R (v 3.4.1 on x86_64-w64-mingw32 platform) and RStudio (v 1.0.153). All were installed as part of the Anaconda3 distribution (v 4.4.0). This document was rendered using knitr, pandoc and pdflatex latex engine.

Libraries:

car(levene.test), multcomp(mcp)

Functions used during analyses:

- read.csv() was used to import CSV files into R
- anova() was used to perform two-way ANOVA
- summary() was used to print ANOVA summary tables
- plot() was used to display residual diagnostic plots and box plots
- levene.test() used to perform a Levene Test
- confint() used to generate confidence intervals
- glht() and mcp() to perform multiple comparison tests
- xtable() used to generate latex tables from summary tables

All code use for analyses and document rendering is available at:

Results

The “Cat” variable was created to allow for easy plotting and to define the contrast matrix needed for later analyses. This variable was a concatenation of the “Age” and “Depth” variables separated with a period.

The “Diff” variable was created to minimise the impact of subject-to-subject variability in our analyses. It is equal to the difference between Post-treatment score and Pre-treatment score, or the *improvement* score of treatment.

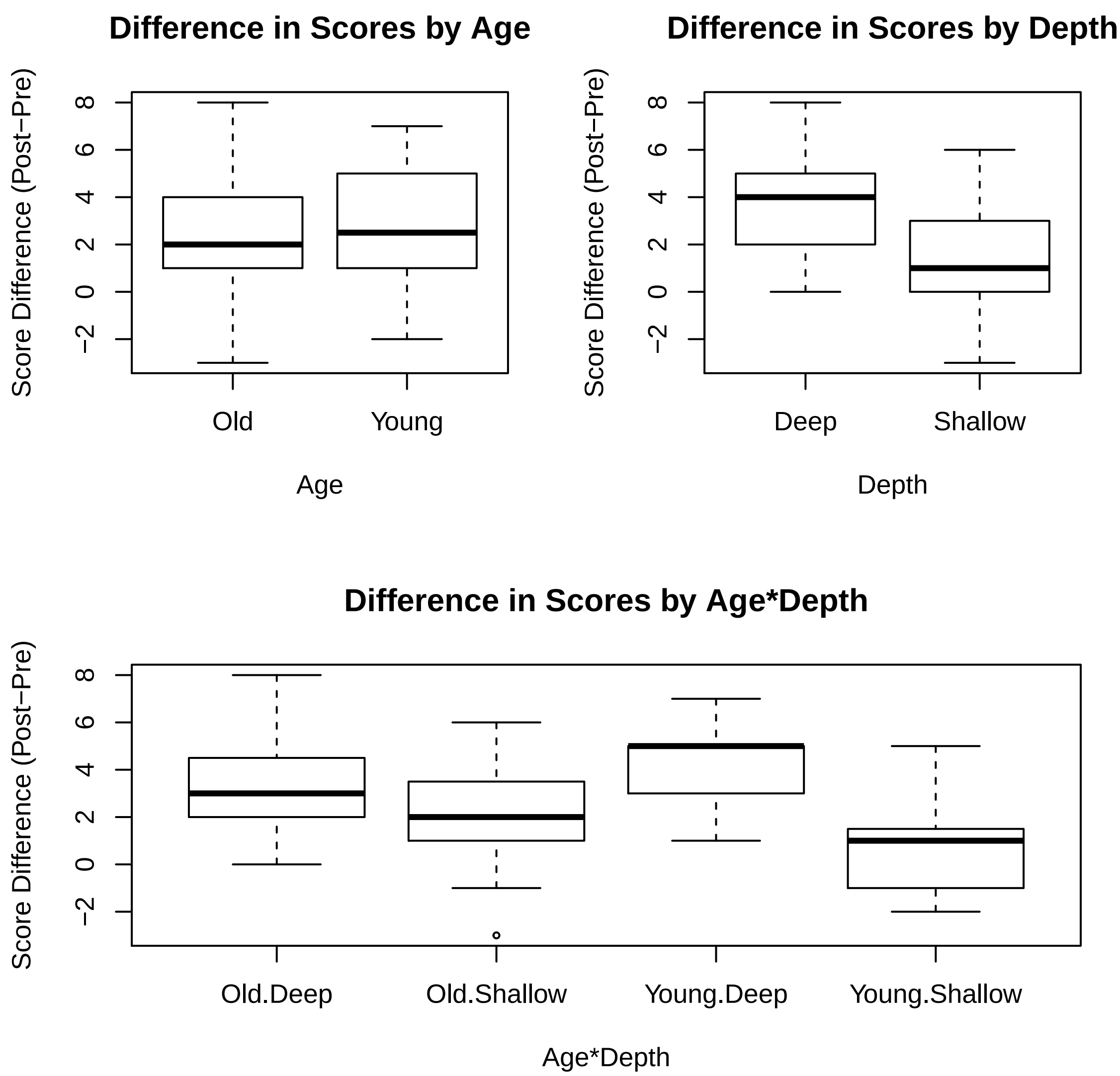

Figure 1: *The box plots show that there is more of a difference between Diff scores of Deep and Shallow treatment groups compared with Old and Young participants. The bottom boxplot also shows that Young participants who used Deep encoding have higher Diff scores than all other participants on average.*

A two-way ANOVA was then performed to determine whether there are statistically significant differences between Age and Treatment allocation and Post-Pre Treatment score (Diff variable).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Age	1	0.20	0.20	0.049	0.826
Depth	1	125.00	125.00	30.468	0.000
Age:Depth	1	24.20	24.20	5.899	0.018
Residuals	76	311.80	4.10		

Table 2: *The ANOVA summary table reports a significant difference between treatment groups (Depth) and also a significant interaction between treatment group and age of participant (Age). The large p-value for Age indicates that there is no difference between age groups, but this term can't be excluded due to the Age:Depth interaction term being significant*

The Regression Diagnostic plot shows that all assumptions required for ANOVA have been met.

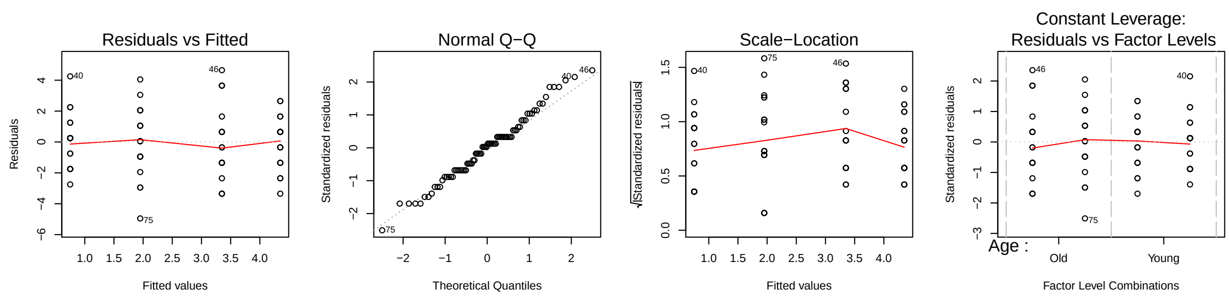

Figure 2: *The Residuals vs Fitted plot shows that there is reasonably homogeneous variance across groups and the Normal-QQ plot shows that the assumption of normally distributed residuals has also been met. There are a few outliers, namely 40, 46 and 75, which slightly break normality and have residuals which are 2 starndardised residuals away from 0. They are just within a reasonable margin, so they are probably no cause for concern.*

To confirm homoscedascitity among categories we can use a Levene Test.

	Df	F value	Pr(>F)
group	3	1.344	0.266
	76		

Table 3: *The large p-value indicates that the null hypothesis that there is homogeneous variance among groups is reasonable. Therefore, the assumption of homoscedasticity has been met.*

In order to determine the actual difference in treatment scores we will perform planned multiple comparison tests. In this case, Priori multiple comparison testing is more appropriate than Post-Hoc testing because the researchers have a predefined set a hypotheses to be tested. Post-hoc testing, such as pair-wise t-tests, will test for differences in each possible combination of factors, including but not limited to the tests outlined by the researchers, and in order to control for multiplicity the p-values must be adjusted according to the family-wise error rate. Therefore, as more tests are performed the p-values must be increasingly adjusted, which increases the change of a Type 2 error occurring.

Planned comparisons have more statistical power as less p-value adjustment is more required (due to the lower number of tests being performed). They also allow for more complicated comparisons to be made, such as the fourth test outlined below, which compares the Diff score of a single group to that of all remaining groups.

Planned comparison tests (predetermined by researchers):

Note: Shallow = "verbal rehearsal" and Deep = "semantic encoding"

1. Old subjects who used semantic encoding (Deep method) had larger Diff score than Old subjects who used verbal rehearsal (Shallow method).
2. Young subjects who used Deep method had larger Diff score than Young subjects who used Shallow method.
3. Young subjects had larger Diff scores than Old subjects
4. Young subjects who used Deep method had larger Diff scores than all other subjects.

Before we can perform said comparison tests, we must first define a matrix of contrasts.

	Old.Deep	**Old.Shallow**	**Young.Deep**	**Young.Shallow**
Old.Deep > Old.Shallow	1.00	-1.00	0.00	0.00
Young.Deep > Young.Shallow	0.00	0.00	1.00	-1.00
Young > Old	-0.50	-0.50	0.50	0.50
Young.Deep > Everything Else	-0.33	-0.33	1.00	-0.33

Table 4: *Defined Contrast matrix used for multiple comparison tests*

Now we can perform the multiple comparison tests. Note all tests are two-sided and have a Holm p-value adjustment.

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Old.Deep > Old.Shallow	1.40	0.64	2.19	0.104
Young.Deep > Young.Shallow	3.60	0.64	5.62	0.000
Young > Old	-0.10	0.45	-0.22	0.995
Young.Deep > Everything Else	2.33	0.52	4.46	0.000

Table 5: *The highly significant p-value for 'Young.Deep > Young.Shallow' indicates that there is a significant difference in Diff score between Young participants who used the 'shallow' treatment compared to Young participants who used the 'deep' treatment. Similarly, the highly significant p-value for 'Young.Deep > Everything Else' indicates that there a significant difference in Diff score between Young participants who used the 'deep' threatment compared with all other participants.*

In order to determine the magnitude of the difference between the groups we will construct simultaneous confidence intervals with a 95% family-wise error rate (using a Holm adjustment).

	Estimate	lwr	upr
Old.Deep > Old.Shallow	1.40	-0.20	3.00
Young.Deep > Young.Shallow	3.60	2.00	5.20
Young > Old	-0.10	-1.23	1.03
Young.Deep > Everything Else	2.33	1.03	3.64

Table 6: *Confirming research expectations, Young participants who used who used semantic encoding were able to recall 2 to 5 more words than young participants who used verbal rehersal and 1.0 to 3.6 more words than all other participants combined. Contrary to the researchers expectations, old participants didn't show any imrovement in word recall between both methods and young participants weren't able to recall any more words than old participants, irregardless of depth of processing.*