Pilot Study Results

Ramiro Casó - Rotterdam School of Management

2023-08-24

Table of Contents

# Intro

Hello Bram, Steven, how are you?

I’ve just wrapped up the analysis on the subjective knowledge scale, and I wanted to share something that I think might be interesting. At the end of this report are the details, but as an intro, here is a summary:

**The Analysis:**

In line with our last meeting, I’ve implemented the controls you suggested, such as subtracting the values for Time 1 (T1) and using them as a control variable. I also combined both batches of participants—with and without visual aid—and incorporated them as another control variable.

**Main Findings:**

Here’s the main finding so far:

* Understanding of Metabolism: Regardless of the text type, participants seemed to feel they understood metabolism better after reading explanatory text. This is seen in the rise in perceived understanding from T1 to T2 then its reversal after objective knowledge testing at T3.
* Metaphors Matter: An extended apt metaphor seemed to have a slight but noticeable effect compared to a non-apt metaphor. This could indicate that metaphors enhance subjective knowledge, and the aptness of the metaphor might further amplify this effect.

**Thoughts and Recommendations:**

I think our Objective Knowledge measure might need a revamp. While it did have a moderate effect, I feel it didn’t quite shatter the illusion of knowing as we hope it would. Perhaps asking participants to explain what they just read could yield a stronger effect, similar to what Fernbach does in his studies. If apt metaphors indeed increase subjective knowledge, participants in that condition may resist this illusion-shattering more robustly.

I hope these insights make sense. Obviously,your thoughts and suggestions on the analysis below are more than welcome!

**Moving Forward:**

I’ll continue to analyse the remaining variables, but I wanted to get this to you quickly, as we might still have time to gather new data.

That’s all for now. I’m looking forward to hearing your thoughts.

¡Saludos!

Ramiro

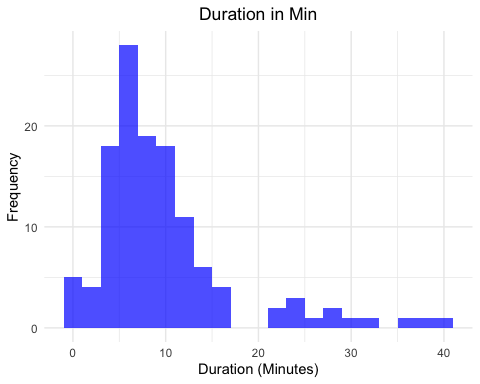
**p.s:** You probably already noticed it, but the menu in the top left side of the page helps you navigate through the report. It seemed like a lot of work to make the report, but it wasn’t, and I believe it could be a better way to share results with you hereafter. So also, let me know what you think about this “new method” of reporting ;)

# Data Cleaning

First, and just like you suggested, I’m using the two batches together and included a new variable called pilot, which differentiates the first batch (pilot\_a = with visual aid) from the second batch (pilot\_b = without visual aid)

Second, I’m removing participants that either answered too fast (below 5 minutes) or took too long (above 25 minutes).

Here is a histogram of duration\_min along with the summary statistics.

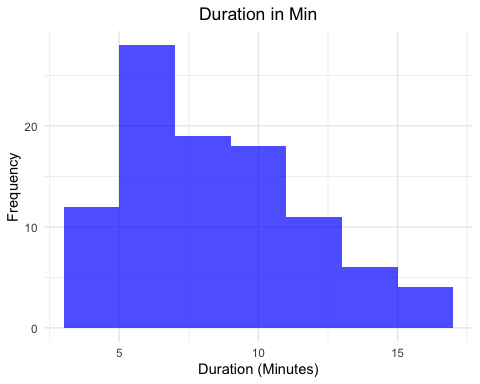


Summary of Duration (Minutes)

| Statistic | Value |
| --- | --- |
| Minimum | 0.12 |
| 1st Quartile | 5.38 |
| Median | 7.85 |
| Mean | 9.80 |
| 3rd Quartile | 11.28 |
| Maximum | 39.98 |

As can be observed, a few observations clearly did not take the survey seriously. I’m trimming all participants whose duration is below 4 minutes and above 20 minutes.

Here are the resulting histogram and descriptives of duration\_min after the removal of observations.



Summary of Duration (Minutes)

| Statistic | Value |
| --- | --- |
| Minimum | 4.12 |
| 1st Quartile | 6.12 |
| Median | 7.94 |
| Mean | 8.49 |
| 3rd Quartile | 10.18 |
| Maximum | 16.50 |

The new minimum duration is 4.12

The new maximum duration is 16.5

This reduces the data set to a total of 98 observations.

# Descriptives

With the clean data, here is a table with the mean, standard deviation, and number of observations of the main variables. .

For clarity, these are the names of the variables.

* text\_type: The main independent variable indicating the type of text that was presented, which has three levels:
  + met\_a = metaphor apt;
  + met\_na = non-apt metaphor;
  + literal = control condition.
* ave\_sk\_t1: Mean Subjective Knowledge before presenting the Text (Time 1).
* ave\_sk\_t2: Mean Subjective Knowledge after presenting the Text (Time 2).
* ave\_sk\_t3: Mean Subjective Knowledge after presenting the Objective Knowledge Scale (Time 3).
* ave\_wta: Average of the Willingness to Adopt scale.
* total\_ok: Results of Objective Knowledge Scale calculated by adding the correct
* duration\_min: Average time it took to complete the survey in minutes.

Summary Statistics for All Variables

| variable | mean | se | p10 | p25 | p50 | p75 | n |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ave\_sk\_t1 | 4.43 | 0.14 | 2.33 | 3.42 | 4.67 | 5.33 | 98 |
| ave\_sk\_t2 | 5.26 | 0.10 | 4.00 | 5.00 | 5.33 | 6.00 | 98 |
| ave\_sk\_t3 | 5.04 | 0.11 | 3.33 | 4.67 | 5.00 | 6.00 | 98 |
| ave\_wta | 4.05 | 0.07 | 3.00 | 3.67 | 4.00 | 4.67 | 98 |
| duration\_min | 8.49 | 0.32 | 4.84 | 6.12 | 7.94 | 10.18 | 98 |
| total\_ok | 7.04 | 0.18 | 5.00 | 6.00 | 7.00 | 8.75 | 98 |

Summary Statistics for Groped Variables

| text\_type | variable | mean | se | p10 | p25 | p50 | p75 | n |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| literal | ave\_sk\_t1 | 4.48 | 0.26 | 2.67 | 3.42 | 4.67 | 5.33 | 30 |
| literal | ave\_sk\_t2 | 5.23 | 0.16 | 4.30 | 4.67 | 5.00 | 6.00 | 30 |
| literal | ave\_sk\_t3 | 5.01 | 0.22 | 3.93 | 4.67 | 5.00 | 5.83 | 30 |
| literal | ave\_wta | 4.12 | 0.11 | 3.33 | 4.00 | 4.33 | 4.33 | 30 |
| literal | duration\_min | 8.35 | 0.50 | 4.88 | 6.25 | 8.11 | 10.63 | 30 |
| literal | total\_ok | 6.93 | 0.34 | 4.90 | 6.00 | 7.00 | 8.75 | 30 |
| met\_a | ave\_sk\_t1 | 4.76 | 0.22 | 3.00 | 4.00 | 4.67 | 5.67 | 33 |
| met\_a | ave\_sk\_t2 | 5.50 | 0.13 | 4.67 | 5.00 | 5.67 | 6.00 | 33 |
| met\_a | ave\_sk\_t3 | 5.25 | 0.15 | 4.37 | 5.00 | 5.33 | 5.75 | 33 |
| met\_a | ave\_wta | 3.98 | 0.14 | 3.00 | 3.67 | 4.00 | 4.67 | 33 |
| met\_a | duration\_min | 8.12 | 0.51 | 5.04 | 5.93 | 7.15 | 9.87 | 33 |
| met\_a | total\_ok | 6.91 | 0.34 | 5.00 | 6.00 | 7.00 | 8.00 | 33 |
| met\_na | ave\_sk\_t1 | 4.07 | 0.26 | 2.13 | 2.50 | 4.33 | 5.17 | 35 |
| met\_na | ave\_sk\_t2 | 5.08 | 0.20 | 3.13 | 5.00 | 5.00 | 6.00 | 35 |
| met\_na | ave\_sk\_t3 | 4.87 | 0.19 | 3.00 | 4.00 | 5.00 | 6.00 | 35 |
| met\_na | ave\_wta | 4.05 | 0.12 | 3.13 | 3.67 | 4.00 | 4.50 | 35 |
| met\_na | duration\_min | 8.96 | 0.62 | 4.79 | 5.97 | 8.20 | 11.25 | 35 |
| met\_na | total\_ok | 7.26 | 0.26 | 6.00 | 7.00 | 7.00 | 9.00 | 35 |

After the trim, we end up with more than 30 observations per condition. This, however, will change ahead, since the the variable pilot which I used to differentiate the two badges **did have an effect** on some of the analysis

The following sections explore each variable in detail.

# Subjective Knowledge

Let’s analyze it, starting with the subjective knowledge scale.

Again, for clarity, keep in mind that:

* ave\_sk\_t1: Mean Subjective Knowledge before presenting the Text (Time 1).
* ave\_sk\_t2: Mean Subjective Knowledge after presenting the Text (Time 2).
* ave\_sk\_t3: Mean Subjective Knowledge after presenting the Objective Knowledge Scale (Time 3).
* delta\_sk2\_sk1: The difference between ave\_sk\_t2 and ave\_sk\_t1
* delta\_sk3\_sk2: The difference between ave\_sk\_t3 and ave\_sk\_t2
* delta\_sk3\_sk1: The difference between ave\_sk\_t3 and ave\_sk\_t1

## SK ANOVA’s

First, let’s run a simple Repeated measured ANOVA. I’m including the variable pilotas a covariate, as you can see in the code below.

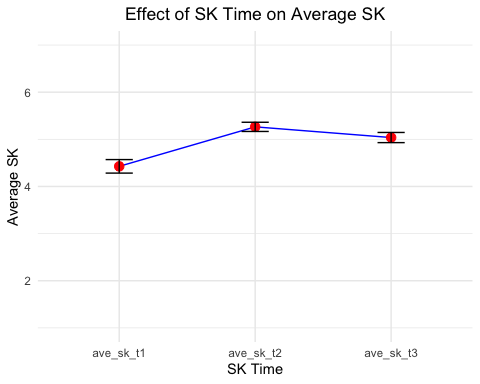
# Conduct the repeated measures ANOVA  
anova\_sk <- aov(ave\_sk ~ sk\_time + pilot + ave\_time + Error(ResponseId/sk\_time), data = sk\_df)  
#anova\_sk <- aov(ave\_sk ~ sk\_time + Error(ResponseId/sk\_time), data = sk\_df)  
  
  
# Print the summary  
summary(anova\_sk)

##   
## Error: ResponseId  
## Df Sum Sq Mean Sq F value Pr(>F)  
## sk\_time 1 3.65 3.653 1.226 0.271  
## pilot 1 2.95 2.947 0.989 0.323  
## ave\_time 1 0.01 0.007 0.002 0.961  
## Residuals 94 280.05 2.979   
##   
## Error: ResponseId:sk\_time  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sk\_time 2 35.06 17.53 30.76 2.57e-12 \*\*\*  
## Residuals 192 109.40 0.57   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

If I’m reading the output correctly, pilot didn’t have an effect on Subjective Knowledge. In that sense, we can keep the whole data set.

Below we can see a table with the means per condition, a line graph that show the results and the post-hoc analysis

| sk\_time | n\_obs | mean\_ave\_sk | se\_ave\_sk |
| --- | --- | --- | --- |
| ave\_sk\_t1 | 98 | 4.426871 | 0.1434505 |
| ave\_sk\_t2 | 98 | 5.264605 | 0.0975583 |
| ave\_sk\_t3 | 98 | 5.037801 | 0.1077886 |



## contrast estimate SE df t.ratio p.value  
## ave\_sk\_t1 - ave\_sk\_t2 -0.823 0.108 192 -7.593 <.0001  
## ave\_sk\_t1 - ave\_sk\_t3 -0.596 0.108 192 -5.501 <.0001  
## ave\_sk\_t2 - ave\_sk\_t3 0.227 0.108 192 2.092 0.0941  
##   
## Results are averaged over the levels of: pilot   
## P value adjustment: tukey method for comparing a family of 3 estimates

Looking at the plot, it is clear (as we saw) that the differences are due to the lower scores in the first measure of SK.

It seems that regardless of the experimental text, participants increased their SK after being presented with the text. There is a slight decrease in SK for time 3, that goes from 5.28 to 5.06, which seems to be moderately significant, according to the post-hoc analysis.

In other words, without accounting for the experimental stimuli, participants first increased their **Subjective Knowledge** after being presented with the text and decrease it after being asked to complete the **Objective Knowledge** scale.

Next, I conducted a mixed two-way ANOVA

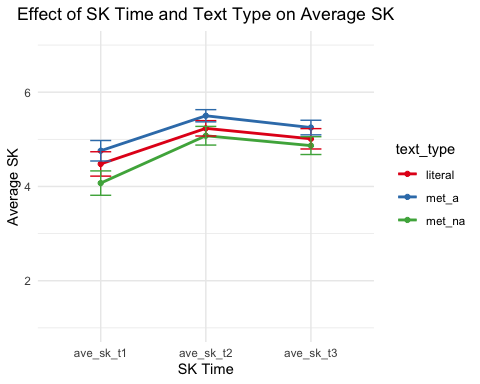
## Mixed two-way ANOVA

## Analysis of Deviance Table (Type III tests)  
##   
## Response: ave\_sk  
## Chisq Df Pr(>Chisq)   
## (Intercept) 210.5995 1 < 2.2e-16 \*\*\*  
## text\_type 6.2138 2 0.0447393 \*   
## sk\_time 15.7125 2 0.0003873 \*\*\*  
## pilot 0.7909 1 0.3738265   
## ave\_time 0.1094 1 0.7408829   
## text\_type:sk\_time 2.1997 4 0.6990809   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

the results indicate that both text\_type and sk\_time have significant main effects on the response variable ave\_sk, but their interaction is not significant. Again, the pilot variable does not have a significant effect.

Below are the table with the means, the plot and the posthoc test for the text\_type and sk\_time For ease of reading, I’m going to present the posthoc comparisons separately. —

| text\_type | sk\_time | n\_obs | mean\_ave\_sk | se\_ave\_sk |
| --- | --- | --- | --- | --- |
| literal | ave\_sk\_t1 | 30 | 4.477778 | 0.2585367 |
| literal | ave\_sk\_t2 | 30 | 5.233333 | 0.1631037 |
| literal | ave\_sk\_t3 | 30 | 5.011111 | 0.2158966 |
| met\_a | ave\_sk\_t1 | 33 | 4.757576 | 0.2172451 |
| met\_a | ave\_sk\_t2 | 32 | 5.500000 | 0.1296190 |
| met\_a | ave\_sk\_t3 | 32 | 5.250000 | 0.1548210 |
| met\_na | ave\_sk\_t1 | 35 | 4.071429 | 0.2584958 |
| met\_na | ave\_sk\_t2 | 35 | 5.076191 | 0.1980699 |
| met\_na | ave\_sk\_t3 | 35 | 4.866667 | 0.1889575 |



**Pairwise comparison just for text\_type** .

## contrast estimate SE df t.ratio p.value  
## literal - met\_a -0.223 0.251 93 -0.889 0.6486  
## literal - met\_na 0.270 0.249 93 1.088 0.5238  
## met\_a - met\_na 0.493 0.243 93 2.034 0.1098  
##   
## Results are averaged over the levels of: sk\_time, pilot   
## Degrees-of-freedom method: containment   
## P value adjustment: tukey method for comparing a family of 3 estimates

**Pairwise comparison just for sk\_time** .

## contrast estimate SE df t.ratio p.value  
## ave\_sk\_t1 - ave\_sk\_t2 -0.824 0.109 188 -7.555 <.0001  
## ave\_sk\_t1 - ave\_sk\_t3 -0.597 0.109 188 -5.472 <.0001  
## ave\_sk\_t2 - ave\_sk\_t3 0.227 0.109 188 2.082 0.0964  
##   
## Results are averaged over the levels of: text\_type, pilot   
## Degrees-of-freedom method: containment   
## P value adjustment: tukey method for comparing a family of 3 estimates

These results, to me, are indicative that the stimuli are not working as intended.

When we look at the effect of sk\_time on ave\_sk, it seems that presenting participants with an explicative text does not lead to the shattering of the illusion of knowing. Perhaps this is because participants aren’t required to make any effort to explain how metabolism works but instead are **“passively”** consuming information that differs in format. In that sense, it is possible that participants feel that they have **“learned”** something, increasing their rating ave\_sk instead of realizing they don’t know much about the subject.

This effect of sk\_time on ave\_sk is shown both in the highly significant coefficient as well as in the posthoc analysis. In the latter, all of the instances present significant difference, with the difference between T2 and T3 slightly less strong (p < 0.1).

What this indicates is that indeed participants “gained” subjective knowledge when presented with the stimuli and then lost it after they were presented with the Objective Knowledge Scale.

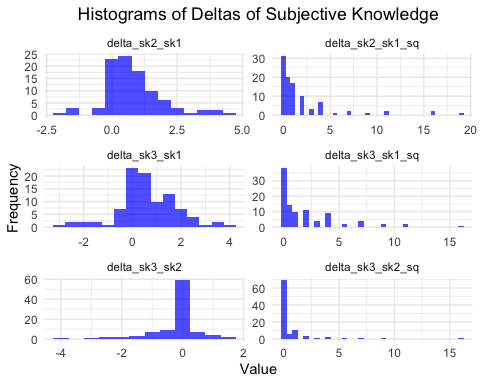
**This is VERY IMPORTANT (I believe)**

When we look at the effect of text\_type on ave\_sk, it seems that participants in the apt metaphor condition report slightly higher levels of subjective knowledge than participants in the non-apt metaphor condition, although this is not reflected in the posthoc analysis.

Now I’m going to perform the analysis using the deltas.

## SK Deltas

First, let’s take a look at the distribution of the Deltas.

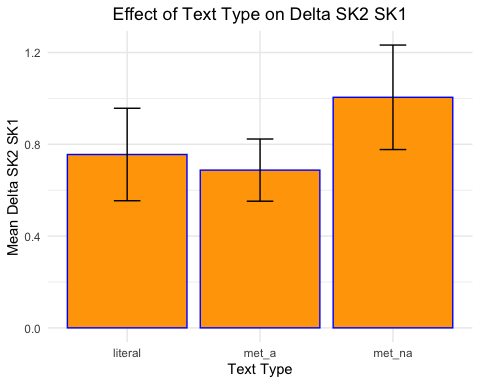


### SK Deltas without squaring

First, I want to know if there are differences in delta\_sk2\_sk1

## Df Sum Sq Mean Sq F value Pr(>F)  
## text\_type 2 1.88 0.9401 0.783 0.460  
## pilot 1 1.99 1.9886 1.657 0.201  
## ave\_time 1 2.83 2.8298 2.358 0.128  
## Residuals 92 110.40 1.2000

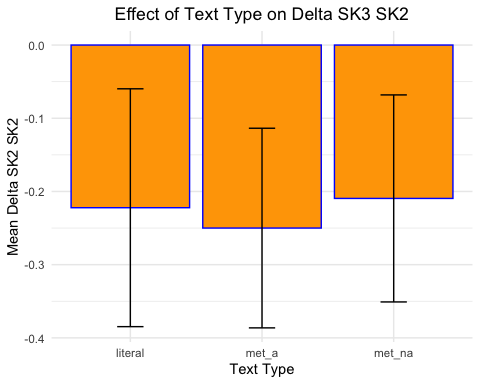
Here is the plot.



Second, I want to know if there are differences in delta\_sk3\_sk2

## Df Sum Sq Mean Sq F value Pr(>F)  
## text\_type 2 0.03 0.0141 0.020 0.980  
## pilot 1 0.56 0.5612 0.799 0.374  
## ave\_time 1 0.03 0.0272 0.039 0.844  
## Residuals 92 64.62 0.7023

Here is the plot.

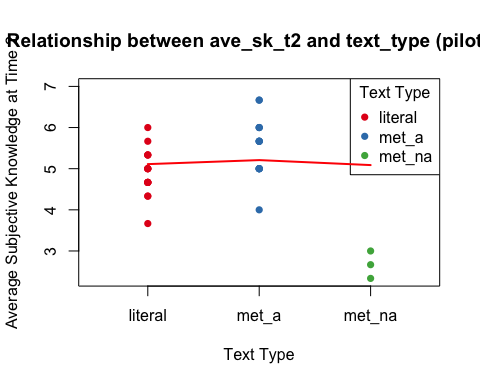
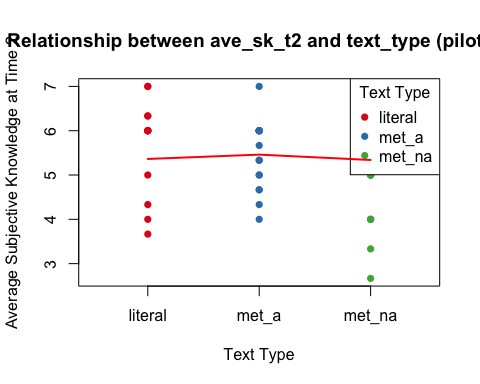


### ANOVA for av\_sk\_t2 controlling for av\_sk\_t1

Next, I want to see if there are differences in the average subjective knowledge reported in Time 2 (after the experimental stimuli was presented) controlling for the score of av\_sk\_t1 and pilot.

##   
## Call:  
## lm(formula = ave\_sk\_t2 ~ ave\_sk\_t1 + text\_type + pilot, data = sk\_df3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.03070 -0.30418 -0.00918 0.51027 2.20849   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.49420 0.29547 11.826 < 2e-16 \*\*\*  
## ave\_sk\_t1 0.42300 0.05531 7.648 1.93e-11 \*\*\*  
## text\_typemet\_a 0.09830 0.19232 0.511 0.6105   
## text\_typemet\_na -0.02302 0.18934 -0.122 0.9035   
## pilotpilot\_b -0.27348 0.15358 -1.781 0.0783 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7509 on 92 degrees of freedom  
## Multiple R-squared: 0.4207, Adjusted R-squared: 0.3955   
## F-statistic: 16.7 on 4 and 92 DF, p-value: 2.525e-10

Here are the plots



As you can see, The coefficients for the text\_type factor levels “met\_a” and “met\_na” are 0.0983 and -0.0230, respectively. Neither of these coefficients is statistically significant (p > 0.05) so there is no evidence of a difference in ave\_sk\_t2 across the text\_type levels, after controlling for ave\_sk\_t1 and pilot.

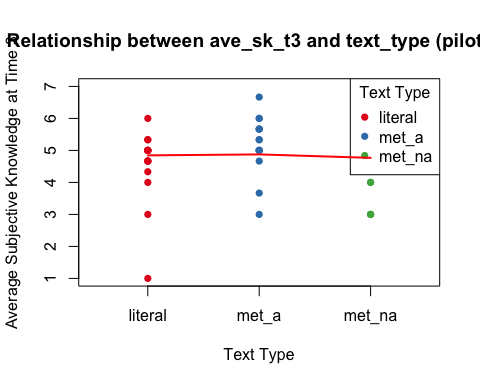
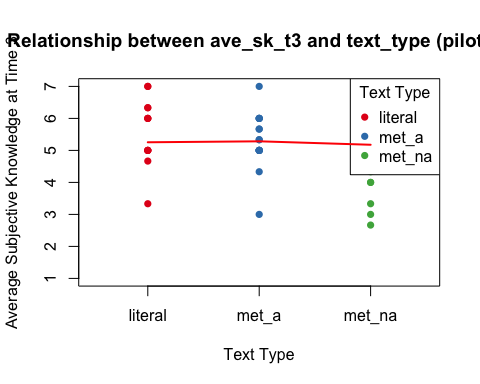
There is a mildly significant effect of pilot -0.27348 (p < 0.1).

Finally, ave\_sk\_t1 has a coefficient of 0.4230 which is highly significant (p < 0.01), which makes sense since we saw that the scores of SK are related to the scores at time 2.

### ANOVA for av\_sk\_t3 controlling for av\_sk\_t1

Lastly, I want to see if there are differences in the average subjective knowledge reported in Time 3 (after the experimental stimuli was presented) controlling for the score of av\_sk\_t1 and pilot.

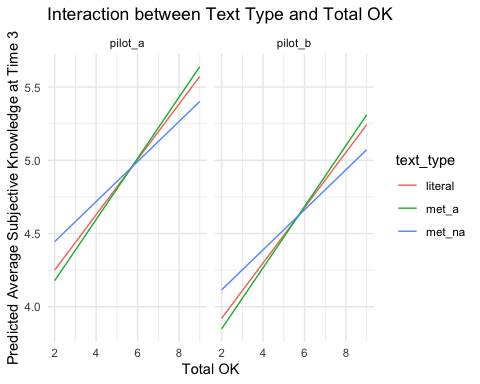
##   
## Call:  
## lm(formula = ave\_sk\_t3 ~ text\_type + pilot + ave\_sk\_t1 + total\_ok,   
## data = sk\_df3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.37606 -0.32617 0.06958 0.57743 1.56281   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.09072 0.53413 3.914 0.000175 \*\*\*  
## text\_typemet\_a 0.03060 0.21717 0.141 0.888267   
## text\_typemet\_na -0.07765 0.21390 -0.363 0.717431   
## pilotpilot\_b -0.34076 0.17521 -1.945 0.054874 .   
## ave\_sk\_t1 0.42124 0.06268 6.721 1.54e-09 \*\*\*  
## total\_ok 0.17701 0.05421 3.266 0.001541 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.847 on 91 degrees of freedom  
## Multiple R-squared: 0.4027, Adjusted R-squared: 0.3699   
## F-statistic: 12.27 on 5 and 91 DF, p-value: 4.229e-09



In the case of ave\_sk\_t3 ave\_sk\_t1and total\_ok show significant effects. The variable pilot may have a marginal effect, while the text\_type variable does not appear to have a significant relationship with IV.

Below I included total\_okas an IV just to see if there was any significant interaction, but I didn’t find one.

##   
## Call:  
## lm(formula = ave\_sk\_t3 ~ text\_type \* total\_ok + pilot + ave\_sk\_t1,   
## data = sk\_df3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.38103 -0.31405 0.06537 0.60756 1.55556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.00407 0.69930 2.866 0.00519 \*\*   
## text\_typemet\_a -0.11196 0.97068 -0.115 0.90844   
## text\_typemet\_na 0.29885 0.93353 0.320 0.74962   
## total\_ok 0.18898 0.08653 2.184 0.03159 \*   
## pilotpilot\_b -0.32943 0.17923 -1.838 0.06940 .   
## ave\_sk\_t1 0.42062 0.06334 6.641 2.39e-09 \*\*\*  
## text\_typemet\_a:total\_ok 0.01987 0.13444 0.148 0.88283   
## text\_typemet\_na:total\_ok -0.05223 0.12784 -0.409 0.68383   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8551 on 89 degrees of freedom  
## Multiple R-squared: 0.4047, Adjusted R-squared: 0.3579   
## F-statistic: 8.643 on 7 and 89 DF, p-value: 4.685e-08



# Discussion of SK Scale.

Just like before, I believe the stimuli are not working as predicted. Presenting participants with explanatory text, regardless of the type of text, seems to increase participants’ perceived understanding of how metabolism work.

When you think about it, this makes sense.

Participants have no reason to feel their SK threatened in any way because their knowledge has not been tested yet. They probably think they know more since they have just acquired additional information. This effect is shown in the significant differences in SK between T1 and T2. Also, and perhaps more importantly, the difference between T2 and T3 is also significant, but in the opposite direction, meaning that after participants were asked to answer the Objective Knowledge Scale, their perceived understanding suffered.

Finally, when the information is presented as an extended apt metaphor, it appears to have a more significant effect on SK than when presented with a non-apt extended metaphor. This effect is mild, but it could indicate that apt metaphors are better at increasing SK than non-apt metaphors.

# What to do next?

I believe we need to find a better way to shatter the illusion of knowing. Instead of using an Objective Knowledge scale, we could ask participants to explain, in their own words, what they’ve just read. Doing so will probably be difficult, especially for the literal condition. Therefore, we could expect a higher drop of SK from T2 to T3 in the literal and non-apt metaphors conditions.