

› Retrieval Practice and Learning

What is the most effective way to learn a subject? Many students focus exclusively on the *encoding* process—that is, how to get the knowledge into memory in the first place. For example, taking notes is an activity for encoding knowledge.

Retrieval, on the other hand, is the process of reconstructing that knowledge from memory. [Karpicke and Blunt](#) (2011) demonstrated that *retrieval* is more effective for learning than activities designed to promote effective encoding. They conducted an experiment in which subjects had to learn about sea otters by reading a passage. Subjects were randomly assigned to one of two conditions: some were instructed to create a [concept map](#) as they read the passage, while others were instructed to practice retrieval (i.e., read the passage, recall as much as they could, read the text again, and recall again). The two main measurements they recorded were:

1. each subject's score on a follow-up learning test one week later
2. each subject's *prediction* of how well they would do on that test

In this lab, you will analyze data from a *replication* of Karpicke and Blunt's experiment, conducted by Buttrick *et al.*

- The data file is : data.csv.
- The codebook (explaining what the variables mean) is : codebook.csv.

[] ↳ 2 cells hidden

✓ Question 1

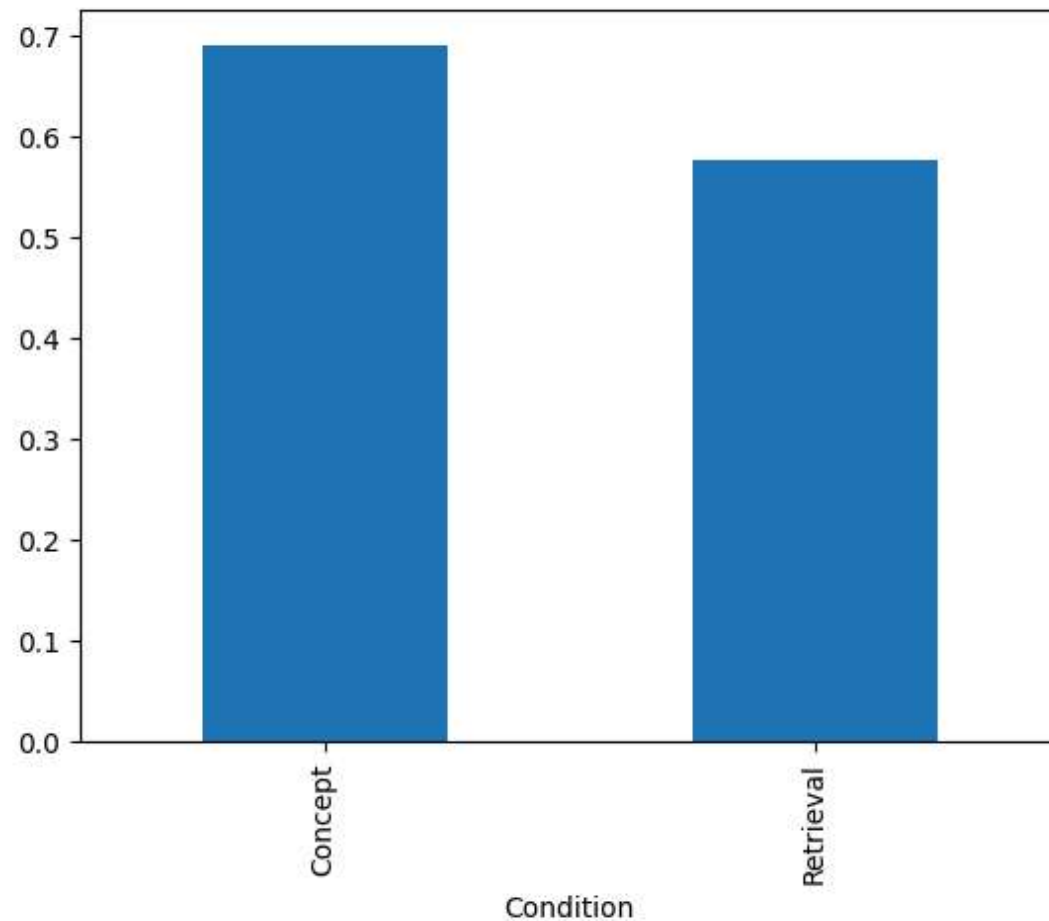
Which group felt like they learned more: the subjects who made concept maps or the ones who practiced retrieval? (Or are they about the same?) Make an appropriate visualization and explain what you see.

Hint: Use the variable `PR.2`, which contains the participants' predictions of how well they would do on a test one week later.

```
import pandas as pd
data = pd.read_csv('data.csv')
```

```
prediction_mean = data.groupby('Condition')['PR.2'].mean()
prediction_mean.plot.bar()
```

<Axes: xlabel='Condition'>



As shown by the bar chart shown above, the group that made concept maps had higher predictions than the Retrieval group.

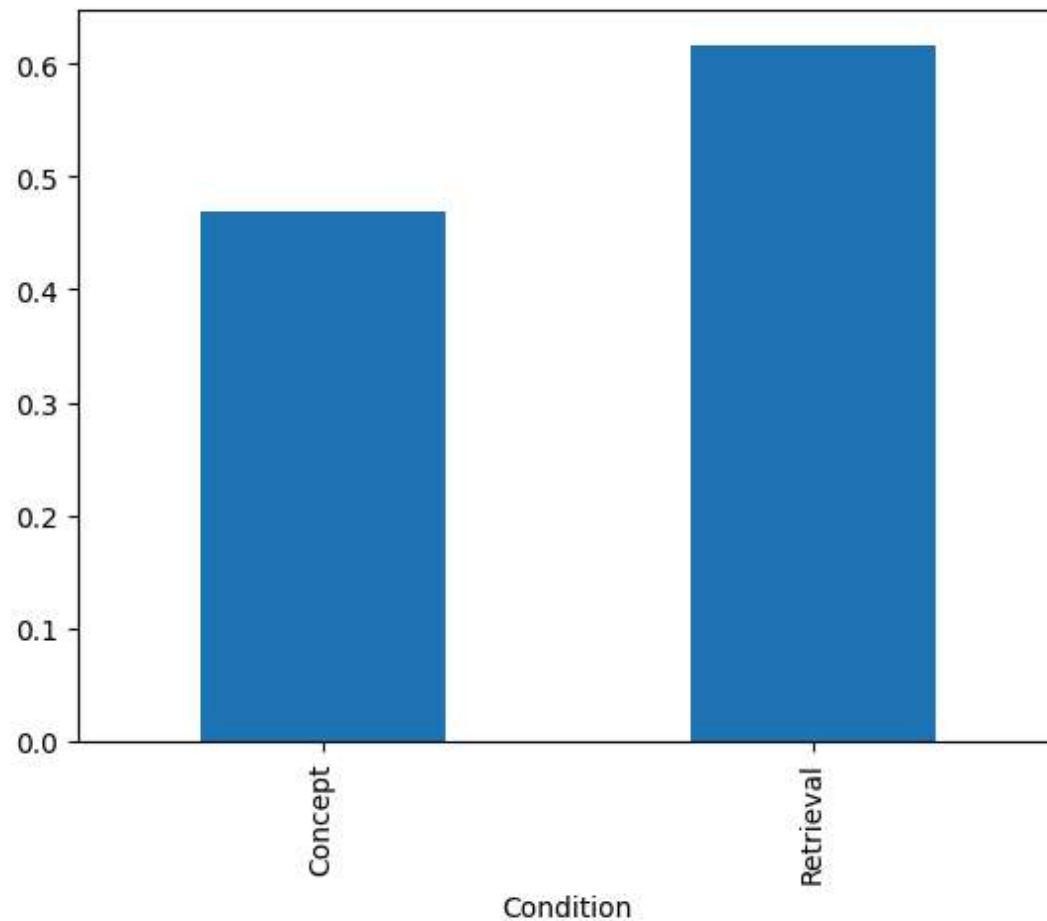
✓ Question 2

Which group actually did better on the follow-up learning test one week later? Make an appropriate visualization and explain what you see.

Hint: Don't ask which variable you should use. That is for you to figure out. Read the codebook carefully (consulting the [original paper](#) if necessary), make an informed decision, and explain your choice.

```
actual_mean = data.groupby('Condition')['TS.avg'].mean()
actual_mean.plot.bar()
```

<Axes: xlabel='Condition'>



Since the participants engaged in 2 tests a week following their predictions, it makes sense to measure the average between these two tests. The resulting bar chart shows information that contradicts what the participants predicted. The Retrieval group predicted that would perform with an average of ~60%, and the test results a week later accurately reflect this. However, the Concept group believed that would perform with an average of 70%, but in actuality performed significantly lower, with a sub-50% average.

✓ Question 3

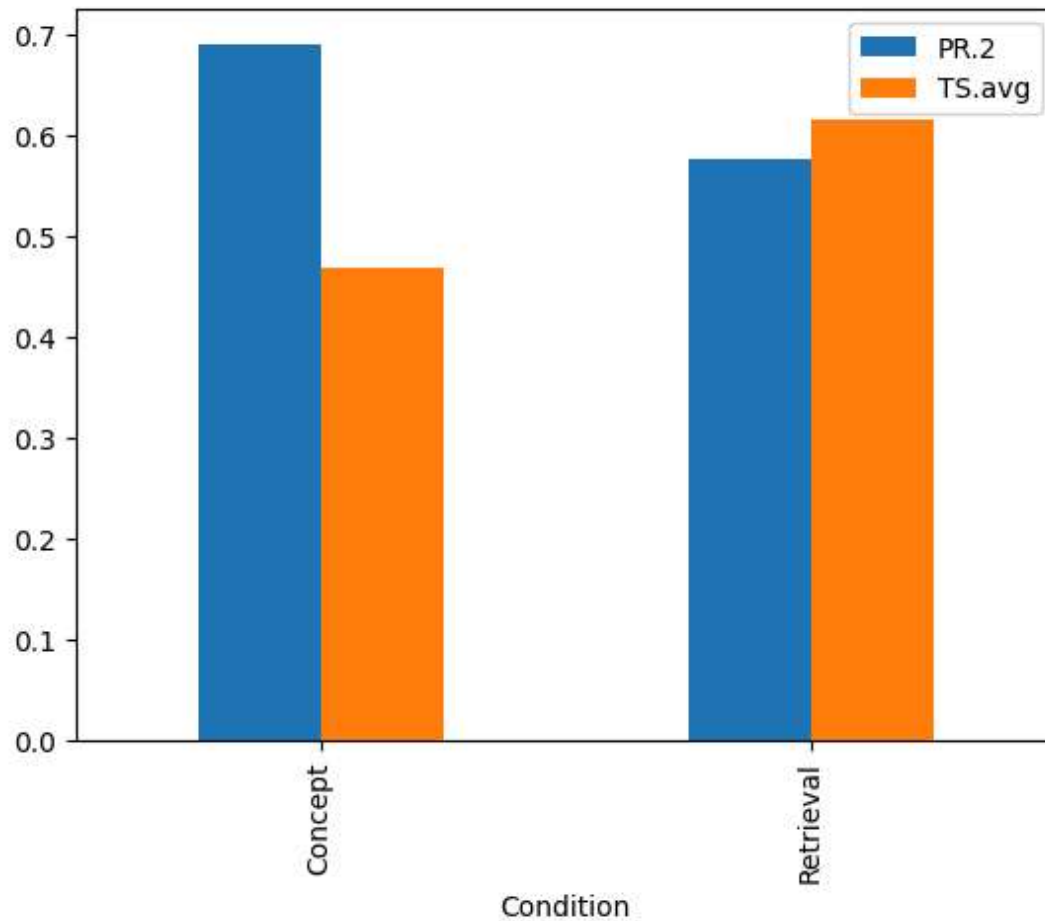
How good were subjects at predicting how well they would do on the follow-up learning test? Calculate a measure of how well subjects predicted their performance and interpret the value in context. (Include a visualization as well.)

```
pred_vs_foll = data.pivot_table(values=['TS.avg', 'PR.2'], index='Condition', aggfunc='mean')
pred_vs_foll.plot.bar()
```

```
pred_vs_foll['Percent Error'] = (pred_vs_foll['TS.avg'] - pred_vs_foll['PR.2']) / pred_vs_foll['TS.avg'] * 100
```

```
# Display the result with the percent error
print(pred_vs_foll)
```

	PR.2	TS.avg	Percent Error
Condition			
Concept	0.690385	0.468846	-47.251846
Retrieval	0.576471	0.616471	6.488550



We can see that concept test takers are more likely to overestimate their actual scores while retrieval were better at estimating their actual scores.

Concept over estimated with an error rate of 47.25% above their actual scores, while Retrieval were very close in there prediction, actually underpredicting by 6.5%.

✓ Question 4

This was a completely randomized experiment. This means that the condition that each subject was assigned to should be independent of their gender, age, and any other subject characteristics. Does that seem to be true in this case? Calculate a summary measure and/or make a visualization, and explain what you see.

From the distribution of subjects based on age and gender, we can see that retrieval had a much higher ratio of females, while concept had a much higher ratio of males. This certainly does not look random, and adds doubt to the findings. The age distribution for concept was much younger than retrieval's age distribution, which also may skew findings.

```
subj_characteristics = data.pivot_table(values=['ID'], index=['Condition', 'Age', 'Gender'], aggfunc='count').unstack(level='Gender')

subj_characteristics.plot(kind='bar', stacked=True)
```

<Axes: xlabel='Condition, Age'>

