Practice Makes Perfect: New Empirical Data and Statistical Analysis to Test Preconceived Notions

Tei Kim
Stanford Online High School
Redwood City, CA 94063, USA
teidkim@ohs.stanford.edu

Abstract

"Practice makes perfect" is passed around frequently without much evidence to support it. Although there have been experiments using mice to test if practice time and maze completion time are correlated, the results cannot directly be applied to humans. This experiment uses a custom website that randomly generates different mazes. The participants used the arrow keys to navigate through the mazes. Each participant completed four increasingly difficult variations, as the function of the arrow keys were changed for each variation. After practicing for a certain amount of time, they solved three timed mazes before resuming practice. A two sample t-test determined that participants improved with practice, but only for the most difficult variation The experiment also featured participants of ages 8 to 73 to determine whether maze completion time would increase with the participant's age. A linear regression analysis shows that younger people generally solved mazes faster than older people.

Keywords
Learning, adaptation, practice, cognition, and technology.

1. Introduction
It is commonly said that as someone practices for a longer duration of time, they become more skilled at the task. For experiments that involve simple problem solving and learning ability, researchers often have mice solve mazes. For instance, researchers created a circular maze in which the mice had to complete it multiple times. They showed that the mice learned and memorized the maze rather quickly (in a few trials), and they completed the maze quicker and more efficiently with more practice (Koopmans et al., 2002). However, these findings did not account for different ages and access to technology, and cannot be entirely applied to humans.

The experiment involved human subjects completing mazes using their computer arrow keys. Since each subject completed numerous tasks with increasing difficulty, we could determine the effects of practice time depending on the task's complexity. We could answer important questions from this information: Does performance plateau after a certain amount of practice time? If someone is already great at a task, can they still marginally improve with more practice?

Furthermore, researchers conducted an experiment to conclude that older people are slower in adapting to a change in motor variability (Sosnoff & Newell, 2008). Since the experiment would include participants from 8 to 70 years old, we could also determine how age affects performance. Factors like difficulty, practice time, and the participant's age will influence the time to complete a maze. Our prediction is that the maze completion time will increase with difficulty, decrease with practice time, and increase with the participant's age.

1.1 Objectives
We collected and analyzed data to determine how someone's maze completion time is affected by three main factors: the amount of practice time, the difficulty of the maze, and their age. By grouping the dataset based on these factors, I analyzed the trends to see if the participants became better, worse, or stayed the same.

2. Literature Review
Researchers conducted an observational study to determine the availability and use of home computers for different age groups. Their data came from the September 2001 CPS, consisting of 143,300 people (25 years or older) in 56,366

households. When the researchers took important factors like gender, education, and income into consideration, they noticed a slow linear decline in the age versus computer availability plot. They also discovered that when a computer is available, the use drops from 90% among 25-29 years old to 39% among 85-89 years old (Cutler et al., 2003). This conforms with the belief that younger people have better access to technology and use it more frequently.

This implies that younger people may perform better at computer-related tasks. Furthermore, studies show that cognitive ability sees an increase from childhood to adolescence, a gradual decrease from 25 to 45, and a rapid decrease from then onwards (Rushton & Ankney, 1996).

In addition to computer and cognitive ability related studies, researchers conducted a mice study to learn about spatial learning and memory. Four different mouse strains were introduced to a modified version of a Barnes maze: a circular arena with 12 exits evenly spaced across the wall, but only one of them leading to the home cage. The researchers determined that for all mouse strains, the time needed to find and enter the escape tunnel generally decreased with more trials. In addition, the total distance the mice traveled before finding the correct exit generally decreased with more trials. This indicates that the mice not only became quicker, but also more efficient (Koopmans et al., 2002).

3. Methods

We used Flask and JavaScript to create a website (link) that presents mazes for practice and test runs. The maze completion time was not recorded for practice runs, and the participants could reload the page to solve one of 30 different mazes. For test runs, a random maze was assigned client-side, and an internal timer began at the moment the participant pressed an arrow key (participants cannot see the timer). The timer stopped as soon as the participant reached the end of the maze. Then, they would fill out a multiple choice survey on how many minutes they practiced for the test. After pressing the submit button, the database recorded the completion time as one data entry. Figure 1 offers an example of how the completion time is recorded.


```
{
    "maze": 11,
    "name": [redacted]
    "played": 1,
    "time": 13744.800000000745,
    "variation": 0
},
```


Figure 1. How the data is stored by the computer.

The participants used their arrow keys to navigate the square from the top left corner to the bottom right corner. We created four different variations, where the functions of the arrow keys were switched with each other:

- Variation 0: none of the arrow keys are switched.
- Variation 1: the left and right arrow keys are switched.
- Variation 2: the left and right arrow keys are switched, and the up and down arrow keys are switched
- Variation 3: the arrow keys are rotated 90° clockwise. The right arrow key acts like the down arrow key, the down arrow key acts like the left arrow key, etc.

Every participant practiced and completed tests for all four variations. For each variation, the timer started when the participant began practicing the mazes. When the timer reached the 1 minute mark, the timer was paused and the participants were instructed to take a test (one test consists of three mazes that are individually timed by the computer). After completing the test, the timer was resumed when the participant began practicing again. When the timer reached the 3 minute mark, the timer was paused and the participants were instructed to take another test. This procedure was repeated when the timer reached the 5 minute and 7 minute mark.

4. Data Collection

Convenience sampling was used because there were not enough resources to perform a large and unbiased sampling. All 14 participants were relatives and friends. All of them were from middle class families and the majority of them were either teenagers or in their mid-40s and 50s. One-on-one zoom meetings were organized with each participant throughout the week. The following instructions were given (read as verbatim):

On the website, there are practice mazes and the actual tests. You will complete all four variations. The function of the arrow keys will change in each variation. When you practice in each variation, you have to reload the page after reaching the green square to get a new maze. After practicing for 1 minute, you will take a test. The test starts as soon as you press an arrow key, so I will show you an example. (Show the participant an example run). You are going to repeat these steps after 1 minute, 3 minutes, 5 minutes, and 7 minutes.

Technical support was also provided in real-time, ensuring that the practice times were recorded correctly with as little bias as possible.

5. Results and Discussion

5.1 Numerical Results

As described in the methods section, the variations became increasingly difficult with Variation 0 being the least complicated and Variation 3 being the most complicated. The two-sample t-tests were used to determine whether there is a statistically significant difference in the maze completion times for different variations.

A two-tailed test was used: a result was statistically significant if the p-value was less than the significance level of $\alpha = 0.05$. The highlighted boxes in Table 1 were the t-statistics and their corresponding p-values that were statistically significant. Both Variations 0 and 3 were significantly different from the other variations, one due to its simplicity and the latter due to its complexity. Variations 1 and 2 were not significantly different from each other. I also noted that the p-values for Variation 3 were particularly small, indicating that it was far more difficult than any other variation.

Table 1. Two-sample t-test results of different variations

T-statistics					P-value			
Variation	1	2	3		Variation	1	2	3
0	2.359	3.729	7.263		0	0.0188	0.0002	5.43E-12
1	-	1.546	5.612		1	-	0.1229	5.22E-08
2	-	-	4.273		2	-	-	2.63E-05

In terms of practice and outcome, it was expected that there would be a decrease in the maze completion time as the practice time increases. The maze completion times would be roughly correlated to the skill level of the participant, which would increase as the participant practices more.

Table 2. Two-sample t-test results of different practice times for Variation 3

T-statistics					P-value			
Practice time (in minutes)	3	5	7		Practice time (in minutes)	3	5	7

1	2.482	2.06	3.463		1	0.0135	0.0402	0.0006
3	-	0.4229	1.231		3	-	0.6726	0.2191
5	-	-	1.593		5	-	-	0.112

The t-statistics and p-values were not included for variations beside Variation 3 because they were not statistically significant. The highlighted boxes in Table 2 are statistically significant differences in the maze completion times between the practice times, namely 1 min vs. 3 min, 1 min vs. 5 min, and 1 min vs. 7 min.

5.2 Graphical Results

Figure 2 displays boxplots for the maze completion times of all participants for each variation. I observed a general upward trend in the completion time as the variation became more complex. Variation 0 has the lowest median completion time and Variation 3 has the highest. It also appears that the distribution becomes more spread out as the variation number increases. A possible explanation is that as the difficulty increases, certain participants were able to adapt easily whereas others were not able to. Outliers are represented by dots.

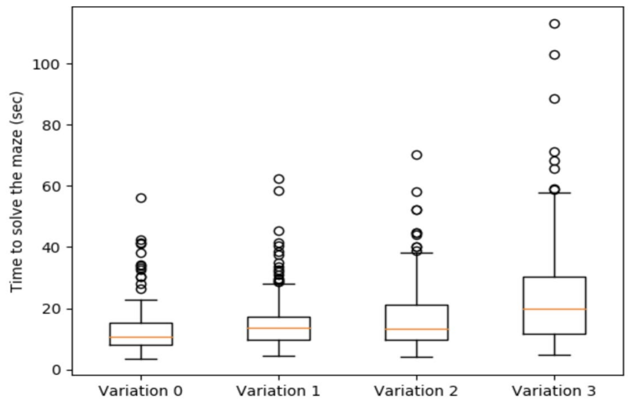

Figure 2. The boxplots for maze completion time for each variation.

As mentioned in section 5.1, there were statistically significant differences in the maze completion time between the practice times (1 min vs. 3 min, 1 min vs. 5 min, 1 min vs. 7 min). This trend is also visible in Figure 3, where the maze completion times were plotted for each practice time in Variation 3.

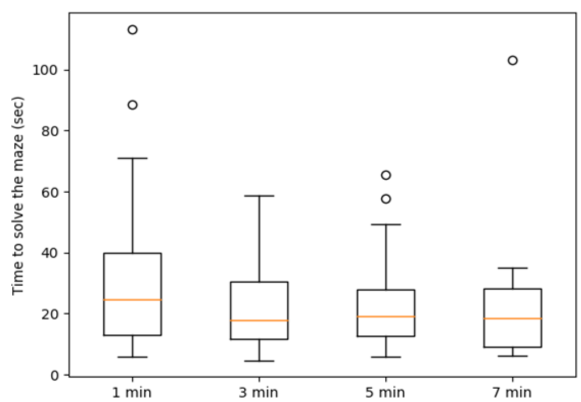

Figure 3: Maze completion time vs. practice time for

Although there was an obvious decrease in the median completion time from 1 minute to 3 minutes, there was not much change from 3 minutes to 7 minutes. There was a small decrease, but the statistical tests showed that the difference was not statistically significant (not much change).

Going back to the phrase, “Kids these days with technology,” we wanted to determine if the maze completion times were longer for older participants. In order to quantify this, the average maze completion time for each participant in Variation 3 were plotted against their corresponding ages.

When conducting regression analysis, it was first analyzed how one variable influences the other by drawing correlation relationships between the two variables: maze completion time and age. A scatterplot of the data was graphed where the x-axis represented the age in years and the y-axis represented the maze completion time. The correlation coefficient R, a measure of the strength of correlation, was calculated. The square of the correlation coefficient is $R^2 = 1 - SSresid/SSTo$, a measure of the percentage of the variation in the dependent variable that can be explained by the independent variable. SSResid represents the sum of the squared residuals, and SSTo represents the sum of the squared totals.

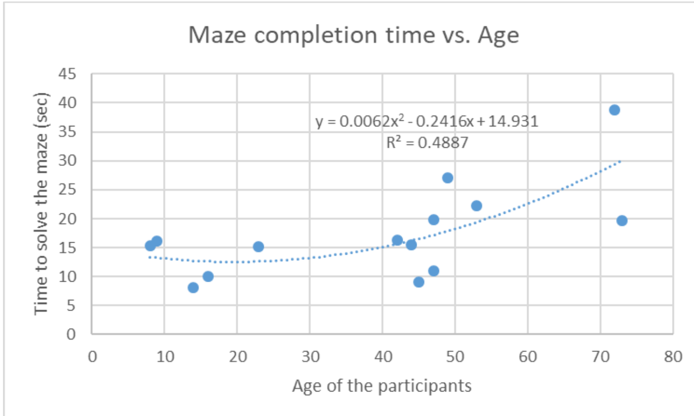

Figure 4. Second order polynomial regression for maze completion time vs. age (Variation 3)

A second order polynomial regression fits the data the best, with an R^2 value of 0.4887. From the eye test, there appears to be a general upward trend in maze completion time as the age of the participant increases. In the polynomial form $y = ax^2 + bx + c$, coefficient a determines whether the general trend would be positive or negative. The value of a was 0.0062, indicating that older people would generally complete mazes slower than younger people.

There was some noticeable variability within the same age group in Figure 4. For instance, participants in their 40s were widely spread out, as some were over the regression line and others were below it. In Figure 5, the participants were grouped by decade to get a broader overview. By doing so, there was a much better regression fit for the overall trend ($R^2 = 0.8883$).

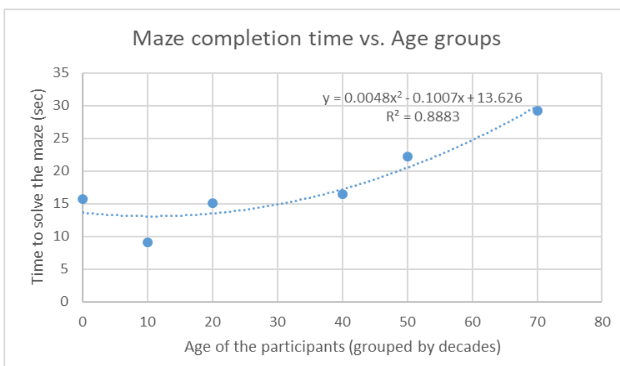

Figure 5. Second order polynomial regression for maze completion time vs. age (grouped by decade / Variation 3)

5.3 Proposed Improvements

The sample was not an accurate representation of the population, as most participants were teenagers or 40-50 years old from middle class families. If the experiment were to be repeated, future researchers should obtain a random sample of participants with diverse backgrounds (age, race, income, etc.). This would allow for a more detailed and accurate correlation analysis to gain deeper insights. Before the experiment, researchers should also conduct a survey asking participants about their access to technology: Is there an electronic device (phones, tablets, computers) in your house? How often do you use it per day?

In this experiment, participants would solve randomly-generated mazes in both practice and test runs. After compiling and analyzing the data, Figure 6 indicates that certain mazes were more difficult than others. For instance, maze 21 required the participant to navigate through a zig-zag pattern, but maze 28 involved a relatively straightforward path. Hence, because of the difficulty and increased chances of making a mistake, a participant would record a slower time if they encountered maze 21 in their test. Despite the mazes being randomly generated, the participants interestingly managed to consistently solve the mazes.

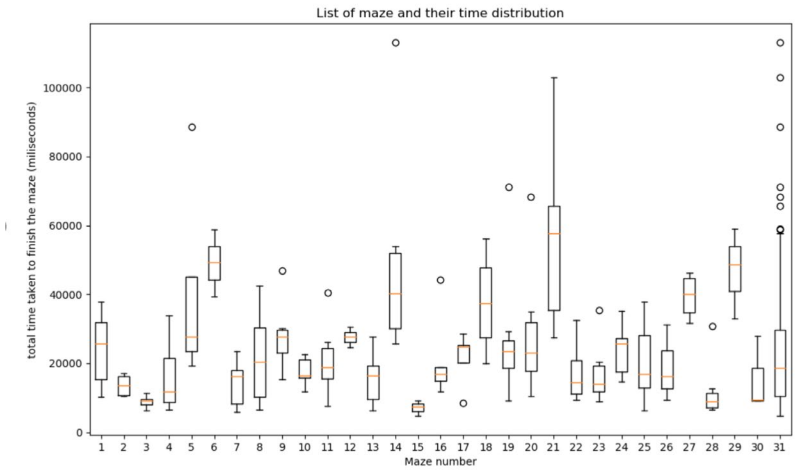

Figure 6. Possible mazes and their time distribution (strictly for Variation 3)

Future researchers can also determine if the practice timer should continue running while participants are taking the test, since they are still technically practicing. However, this change may create certain complications: slower participants may end up practicing longer than other participants, giving them an unfair advantage.

6. Conclusion

As the complexity of the variations increased, some participants continued to solve the mazes quickly while others lagged behind. This explains the increasing variability in more difficult variations as shown in Figure 2. The two sample t-tests in Table 1 show that variations 1 and 2 were not much different from each other in terms of difficulty. This supports our initial hypothesis that the difficult variations will require more time to complete than the simpler ones.

As displayed in Figure 3, the participants struggled with variation 3 after only one minute of practice. However, they quickly adjusted to it after 3 minutes of practice. From this point onwards, a further improvement in the maze completion was marginal. Even though the median maze completion time did not improve from 3 to 7 minutes of practice, the variability did decrease. This indicates that the participants were consistently performing better as a group. These results support our initial hypothesis that participants would become ‘better’ after practicing for longer periods of time. My original metric of becoming ‘better’ at solving mazes was the completion time. However, I also realized through the experiment that becoming consistent was another metric of measuring ‘better’.

In Figure 4, the regression fitting indicates that older people performed worse than younger people. This could be attributed to younger people being more exposed to technology and able to adapt more easily to environmental changes. In Figure 5, grouping participants by decade increases the R^2 value for the polynomial regression. The differences within an age group were canceled out, yielding an averaged data point that represented the overall trend better. This is especially for participants in their 40s, who had diverse occupations running from house-makers to doctors; we suspect that certain occupations prepared the person to adapt to environmental changes while others did not. These findings support the initial hypothesis that younger people are better at adapting to a change in variation than older people; this does not directly support that younger people are better at technology or adapting to changes. However, it can be inferred that there lies some underlying trends that are worth studying.

All in all, our preconceived notion that there is some negatively correlated relationship between age and adaptability to changes were, to some extent, statistically supported through this study. This study also provides support for the saying, “Practice makes perfect” when the task was difficult.

References

Cutler, S. J., Hendricks, J., & Guyer, A., Age Differences in Home Computer Availability and Use. *The Journals of Gerontology: Series B*, *58*(5), S271–S280, 2003. https://doi.org/10.1093/GERONB/58.5.S271

Ian J. Deary, Janie Corley, Alan J. Gow, Sarah E. Harris, Lorna M. Houlihan, Riccardo E. Marioni, Lars Penke, Snorri B. Rafnsson, John M. Starr, Age-associated cognitive decline, *British Medical Bulletin*, Volume 92, Issue 1, Pages 135–152, 2009. https://doi.org/10.1093/bmb/ldp033

Koopmans, G., Blokland, A., van Nieuwenhuijzen, P., & Prickaerts, J., Assessment of spatial learning abilities of mice in a new circular maze. *Physiology & Behavior*, *79*(4–5), 683–693, 2003. https://doi.org/10.1016/S0031-9384(03)00171-9

Pitts, M. W., Barnes Maze Procedure for spatial Learning and Memory in Mice. *Bio-protocol* 8(5): e2744. DOOI: 10.21768/BioProtoc.2744, 2018.

Pitts, M. W., Kremer, P. M., Hashimoto, A. C., Torres, D. J., Byrns, C. N., Williams, C. S. and Berry, M. J. , Competition between the brain and testes under selenium-compromised conditions: Insight into sex differences in Selenium metabolism and risk of neurodevelopmental disease. *J Neurosci* 25(46): 15326-38, 2015.

Roxy Peck, Chris Olsen, Jay L. Devore. Introduction to Statistics & Data Analysis. 5th Edition. Cengage Learning; ISBN: 978-1-305-11534-7, 2016.

Rushton, J.P., Ankney, C.D. Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin & Review* 3, 21–36, 1996. https://doi.org/10.3758/BF03210739

Sosnoff, J. J., & Newell, K. M., Age-Related Loss of Adaptability to Fast Time Scales in Motor Variability. *The Journals of Gerontology: Series B*, *63*(6), P344–P352., 2008. https://doi.org/10.1093/GERONB/63.6.P344

Underdahl, J., Palacio-Cayetano, J., & Stevens, R. , Assessing and Enhancing Knowledge and Problem-Solving Skills with IMMEX Software Practice Makes Perfect Assessing and Enhancing Knowledge and Problem-Solving Skills with IMMEX Software Practice Makes Perfect Feature. *Learning & Leading with Technology*, *28*, 26. www.iste.org., 2001.

Biography

Tei Kim is a rising sophomore at Stanford Online High School. He has particular interests in statistics as well as biology and chemistry. He also enjoys programming and developed the custom website for this experiment along with his older brother.