**Investigation on efficiency of extensive exercises on school-age children’s computational ability**

1/22/2019

Ruiyan Huang

FP Math L10-2

Kokming Lee

**Rationale**

Every Chinese student in local school, me included, must have gone through the process of repeatedly doing a great number of exam-oriented exercises. Even up till now, most Chinese educators are still adopting this method, because doing exercise is one of the most straightforward way to consolidate knowledge, familiarize students with exam-style questions, and meanwhile test the degree of understanding. However, over recent years, there is also a growing consensus that we should abandon this old-fashioned method, because it can discourage students’ self-motivation to learn and thus decrease the efficiency and make them unable to achieve higher score (Han 106).

As one of those who once both benefited and suffered from this exam-oriented studying method, I have also been doubting about the efficiency of this study method for a long time, so by employing what I have learned, I have aspiration to explore whether this method is efficient or not, and thus help find the more efficient way for students to study math.

**Introduction**

**Theoretical background**

When an exploration is conducted, we need to make sure if the result is meaningful (statistically significant) or it just happens by chance. A way to do this is hypothesis testing. When conducting a hypothesis test, we need to first consider a population X with unknown mean , unknown standard deviation and sample size n (n30). The value for the sample mean follows a normal distribution with mean and standard deviation according to Central Limit Theorem (Quinn 71-72). Then, we determine the hypothesis and the null distribution, which in this exploration is a T-distribution. After that, test statistic is determined by the formula

we choose either to reject or accept the null hypothesis by either calculating the p-value or determining where the test statistic will lie in the t-distribution: if it lies in the critical regions of the distribution, then we need to reject the null hypothesis (Quinn 100-114).

**Aim**

The aim of this exploration is to answer the research question: Can the mechanical training method of repeated exercises statistically significantly improve Grade 1 students’ ability of mental abacus calculation? To answer this question, an experiment will be conducted and the data will be collected and analyzed using statistics.

**Hypothesis**

In this exploration, I hypothesize that at 99% confident level, there will be a statistically significant improvement on students’ computational ability in both short time and long run after the training.

**Method**

Sampling design

The data were collected from a public primary school in Suzhou. As a public school, all the students there are admitted according to the school district, so students with a wide range of ability and family backgrounds will be admitted, including both rural and urban students. There are six grades in the school from Grade 1 to 6. In each grade, the students are randomly divided into different classes, so each class can be seen as a cluster, or a heterogeneous group.

Since Grade one, all the students in this school are required to study mental abacus calculation, which is a Chinese traditional way of calculation using abacus, so even for Grade one students, they all have the fundamental knowledge of mental abacus calculation.

In this experiment, Grade 1 students were chosen because they had just finished all the basic knowledge yet had not gone through many exercises. There are 6 classes in Grade 1. By rolling a fair die (with six sides), I conducted a clustered sampling, in which the heterogeneous groups of classes are seen as 6 clusters. I chose Class 3 as the sample for further investigation.

Data collection

First, the students in Class 3 Grade 1 were required to do a pop quiz with 40 problems, each requiring them to calculate the sum of 5 two-digit numbers using mental abacus calculation, given 10 minutes. Each correct answer will give one point, and an incorrect answer gives no points. The first set of data were thus collected.

After the quiz, a worksheet including 40 questions targeting the addition of 5 two-digit numbers are distributed to each student as homework every day for one week. The exercise should take about 15 minutes every day on average. Still, in consideration of the potential homework pressure, during those days, the students did not have any extra math homework, and thus ensure them to be in a relatively good mental and health condition.

After all the seven worksheets were send out and completed, without being informed, the students were tested again using the same questions as the first quiz but with different orders, in order to make sure that the two tests are equally difficult. The results were collected as the second set of data.

Finally, after another week, without any exercises for the whole week, just as before, the students were again tested using the same questions albeit different orders. The grades were to be the third set of data.

The three sets of data were recorded according to the points each student gets, but the results were not given until the end of the experiment in order to eliminate the stimulation from the grades, and thus reduce potential bias.

**Raw Data**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | 卜\*雯 | 蔡\*逸 | 陈\* | 程\*蕾 | 甘\*萱 | 高\*阳 | 顾\*帆 | 纪\*佳 |
| **First test** | 13 | 26 | 9 | 12 | 15 | 3 | 14 | 11 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Third test** | 18 | 30 | 17 | 14 | 14 | 5 | 11 | 17 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 李\* | 陆\*轩 | 马\*辰 | 马\*涵 | 毛\*超 | 欧\*泽 | 潘\*予 | 祈\*乐 |
| **First test** | 23 | 15 | 10 | 20 | 15 | 12 | 16 | 16 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Third test** | 23 | 16 | 15 | 22 | 16 | 11 | 23 | 14 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 祈\*涵 | 钱\*雅 | 钱\*昊 | 钱\*浩 | 盛\*妮 | 盛\*洋 | 孙\*聪 | 王\*政 |
| **First test** | 1 | 2 | 6 | 10 | 6 | 10 | 22 | 9 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Third test** | 15 | 6 | 10 | 12 | 7 | 15 | 21 | 21 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 夏\*涵 | 徐\*玥 | 杨\*燕 | 杨\*豪 | 尹\*博 | 余\*宸 | 俞\*阳 | 张\*彤 |
| **First test** | 11 | 7 | 21 | 11 | 23 | 23 | 17 | 20 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Third test** | 15 | 13 | 29 | 25 | 25 | 27 | 19 | 24 |

Table 1. Raw data collected for the three tests.

Students’ names are not written in full in order to protect the students’ privacy.

**Data processing**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | 卜\*雯 | 蔡\*逸 | 陈\* | 程\*蕾 | 甘\*萱 | 高\*阳 | 顾\*帆 | 纪\*佳 |
| **First test** | 13 | 26 | 9 | 12 | 15 | 3 | 14 | 11 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Short time** | -15 | -28 | -11 | -14 | -17 | -5 | -16 | -13 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 李\* | 陆\*轩 | 马\*辰 | 马\*涵 | 毛\*超 | 欧\*泽 | 潘\*予 | 祈\*乐 |
| **First test** | 23 | 15 | 10 | 20 | 15 | 12 | 16 | 16 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Short time** | -25 | -17 | -12 | -22 | -17 | -14 | -18 | -18 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 祈\*涵 | 钱\*雅 | 钱\*昊 | 钱\*浩 | 盛\*妮 | 盛\*洋 | 孙\*聪 | 王\*政 |
| **First test** | 1 | 2 | 6 | 10 | 6 | 10 | 22 | 9 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Short time** | -3 | -4 | -8 | -12 | -8 | -12 | -24 | -11 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 夏\*涵 | 徐\*玥 | 杨\*燕 | 杨\*豪 | 尹\*博 | 余\*宸 | 俞\*阳 | 张\*彤 |
| **First test** | 11 | 7 | 21 | 11 | 23 | 23 | 17 | 20 |
| **Second test** | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| **Short time** | -13 | -9 | -23 | -13 | -25 | -25 | -19 | -22 |

First, we consider the short time improvement. By subtracting the scores in second test and scores in the first test, we get the improvement or retrogress, which is shown in the table below.

Table 2. Data collected for the first and second time and the short time improvement.

For the first step, the mean of the short time improvement is calculated using the formula

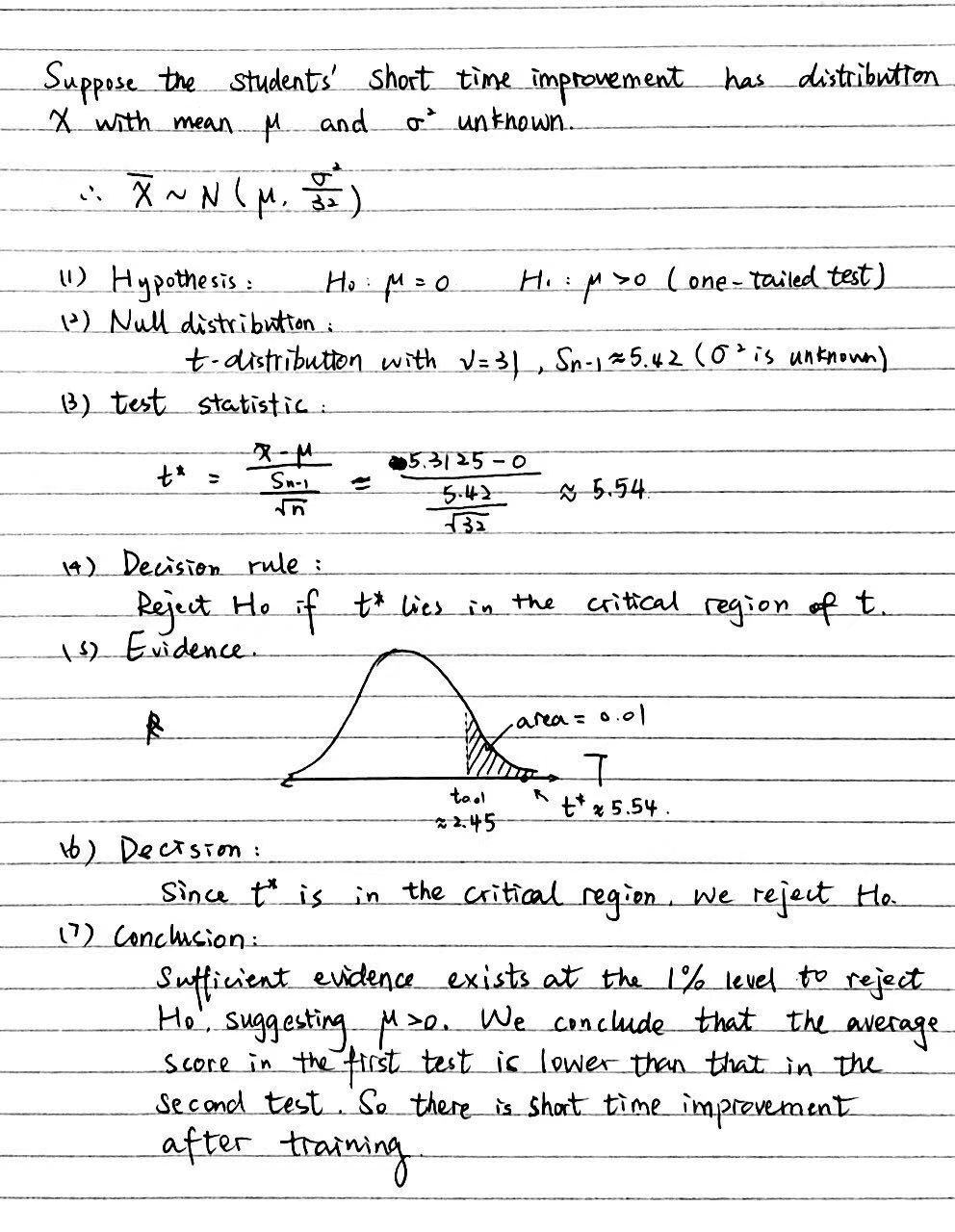
Where is the mean, xi (xi = x1, x2, x3, …, xn) is the short time improvement for each student, and n is the sample size, which is 32.

Then the variance can be calculated by

The unbiased estimator for standard deviation can thus be calculated by

The distribution of the population improvement can be seen as normally distributed, since there are many small variations in students’ grades, including luck, health and mental conditions, intelligence, preparation, etc.

Since the grades are normally distributed, in order to verify whether the improvement is statistically significant or not, we need to conduct hypothesis testing:



The probability of error can be calculated by

where type I error means to reject H0 when it is in fact true.

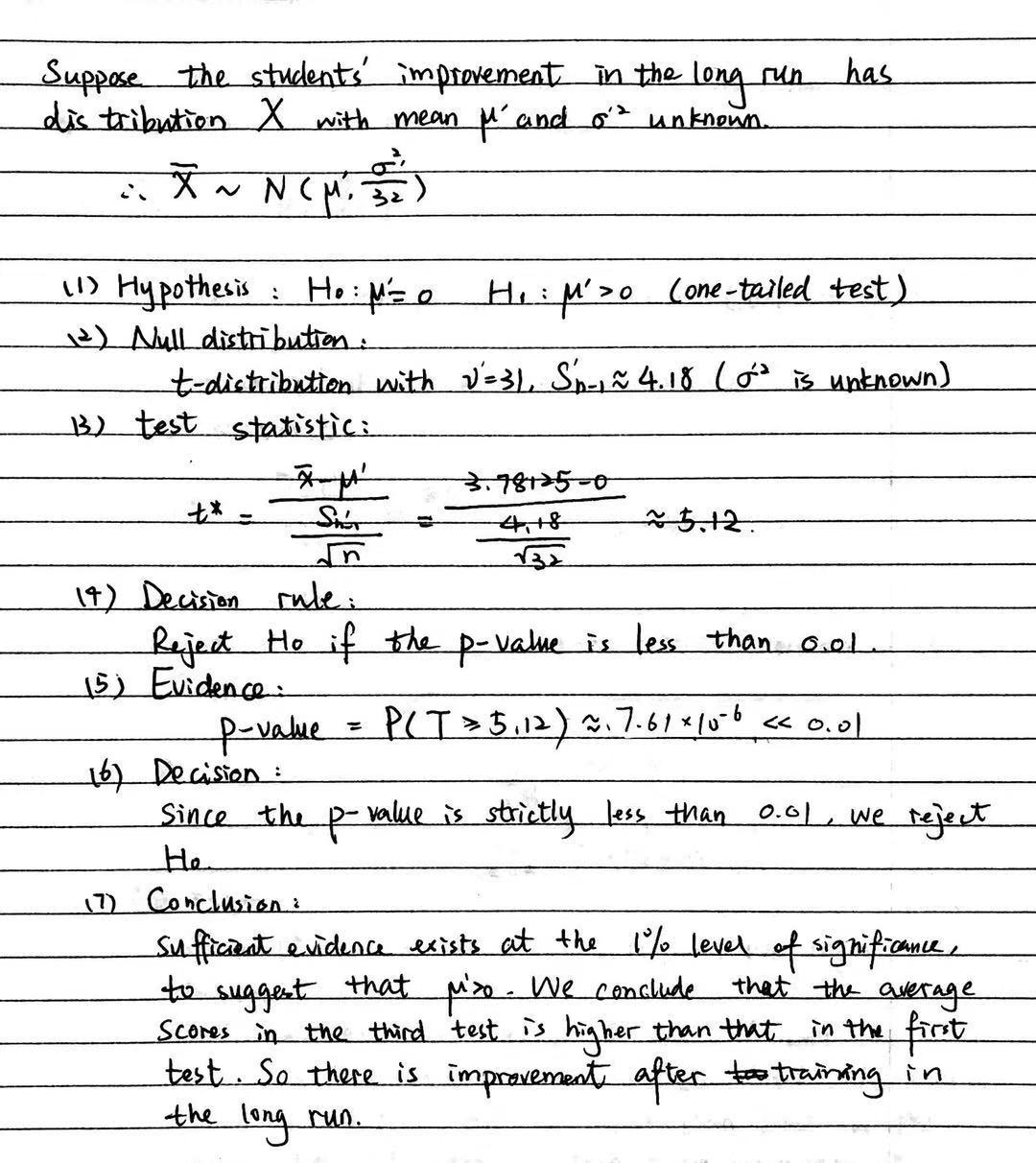
Now we need to consider the improvement in the long run, which can be computed by subtracting the scores each student gets in the third test by the scores in the second test. We get the processed data in the table below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | 卜\*雯 | 蔡\*逸 | 陈\* | 程\*蕾 | 甘\*萱 | 高\*阳 | 顾\*帆 | 纪\*佳 |
| **First test** | 13 | 26 | 9 | 12 | 15 | 3 | 14 | 11 |
| **Third test** | 18 | 30 | 17 | 14 | 14 | 5 | 11 | 17 |
| **long time** | 5 | 4 | 8 | 2 | -1 | 2 | -3 | 6 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 李\* | 陆\*轩 | 马\*辰 | 马\*涵 | 毛\*超 | 欧\*泽 | 潘\*予 | 祈\*乐 |
| **First test** | 23 | 15 | 10 | 20 | 15 | 12 | 16 | 16 |
| **Third test** | 23 | 16 | 15 | 22 | 16 | 11 | 23 | 14 |
| **long time** | 0 | 1 | 5 | 2 | 1 | -1 | 7 | -2 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 祈\*涵 | 钱\*雅 | 钱\*昊 | 钱\*浩 | 盛\*妮 | 盛\*洋 | 孙\*聪 | 王\*政 |
| **First test** | 1 | 2 | 6 | 10 | 6 | 10 | 22 | 9 |
| **Third test** | 15 | 6 | 10 | 12 | 7 | 15 | 21 | 21 |
| **long time** | 14 | 4 | 4 | 2 | 1 | 5 | -1 | 12 |
|  |  |  |  |  |  |  |  |  |
| **Name** | 夏\*涵 | 徐\*玥 | 杨\*燕 | 杨\*豪 | 尹\*博 | 余\*宸 | 俞\*阳 | 张\*彤 |
| **First test** | 11 | 7 | 21 | 11 | 23 | 23 | 17 | 20 |
| **Third test** | 15 | 13 | 29 | 25 | 25 | 27 | 19 | 24 |
| **long time** | 4 | 6 | 8 | 14 | 2 | 4 | 2 | 4 |

Table 3. Data collected for the first and third test and the improvement in the long run.

Likewise, using the same method as we are calculating the statistics for the short time improvement, we can calculate that for the improvement in the long run and get

Since the grades, as before, are normally distributed, in order to determine whether the improvement in the long run is statistically significant or not, we also need to conduct hypothesis testing:



The probability of error can be calculated by

where type I error means to reject H0 when it is in fact true.

**Conclusion and evaluation**

We can now verify the hypothesis that, at 99% confident level, there is a statistically significant improvement on students’ computational ability in both short time and long run after the training.

However, limitations exist in this exploration.

First, clustered sampling and small sample size can lead to disproportional or unrepresentative samples. For example, in my experiment, there are more boys than girls because of clustered sampling as well as relatively small sample size. It will be better if I have chosen a larger sample size and conducted the method of stratified sampling by classifying students into different stratifies and select students according to the real proportion.

Second, the control of third variables is imperfect. Since the seven exercises are assigned as homework, the students finished them without any supervision, so I could not make sure the quality of the homework is the same for different people. It will be better if they had done it at school with my or other teachers’ supervision and grade them afterwards. Also, I cannot make sure that every sample is under equally health and mental conditions, which can also affect the results, but that can be diminished because of the randomization of the sample.

Thirdly, some assumptions are made during this exploration. Even though it is very likely that the students’ grades are normally distributed because it is an outcome for the combination of many small variations in nature, it is still possible that the real distribution is not perfectly normally distributed.

Finally, the results are not generalizable because of the following reasons:

First of all, in the sampling design, the primary school we chose for data collection was restricted in only one district, so we cannot assume that there is still improvement in other places. To solve this, we can select students from different regions all around China but that will be too time-consuming and costly.

Also, the sample was only selected from Grade one, because it will be hard to quantify the improvement if the students have different mathematical levels. Since Grade one might be a special period for the students as they are relatively young, we cannot assume that students of all ages have equal amount of improvement. A possible solution is to repeat the experiment in different grades. Still, that will be fairly time-consuming and I did not have enough time or energy for that.

What’s more, the time given to prove the improvement in the long run is not enough. We can see a decline in the improvement after seven days, but we cannot make sure how the decline of improvement will be like afterwards. It is possible that the improvement will keep declining or it will remain at almost the same level. If time permits, a better solution is to test the students more frequently and examine the whole trend of their improvement of mathematical computational ability, which will also require a control group that stops doing the exercises and an experiment group that keeps doing the exercises, in order to diminish the effect of third variables, like the use of other knowledges they learn during usual math courses.

Lastly, there is no comparison between different training methods, meaning that I cannot make sure this training method of repeatedly exercising is a more efficient way compared with others. But we can still see there is a huge improvement at a very low level of possibility of error, which is only 0.01.

**Conclusion**

In conclusion, by using hypothesis testing and calculating the possibility of errors, we can prove that, at level 1% significance, there is improvement in students’ mathematical computational ability in both short time and long run after the training method of repeatedly exercising.

**Works cited**

Catherine, Quinn, et al. “E. Distribution of the Sample Mean and the Central Limit Theorem.” Mathematics for the International Student: Mathematics HL (Option): Statistics and Probability, first ed., Haese Publication, 2013, pp. 71-72.

Catherine, Quinn, et al. “H. Significance and Hypothesis Testing.” Mathematics for the International Student: Mathematics HL (Option): Statistics and Probability, first ed., Haese Publication, 2013, pp. 100–114.

Han, Jinmei. “浅析"题海战术"对小学数学的影响.” Future Talent, vol. 11, 2016, p. 106.