MONEY SAVVY U INTERMEDIATE CURRICULUM

Washington State 2021-2022

Evaluative Report

HAGEDORN

Evaluation Services

Background

In the 2021-2022 year pre- and post-tests were collected from schools all over the state of Washington. The sample we collected of pre- and post-tests (N = 1003) came from 20 Washington schools and 21 individual teachers who participated in the program.

To investigate the effectiveness of this program, the 10 question multiple choice test, called "The Money Savvy U™ Personal Finance Curriculum test" (MSUPFT) was used. A portion of the questions on this instrument were drawn from the JumpStart Coalition Personal Finance Literacy test (items 1, and 5 through 10) and a portion were provided by the curriculum developers (items 2 through 4).

Because we had some 175 fewer post-tests than pre-tests, we chose to use our whole sample of 1003 pre-and post-tests and analyze them with independent samples inferential statistics. instead of matched scores with paired samples inferential statistics. This is an acceptable alternative to paired samples, particularly when you have unequal numbers of pre- and post-tests.

Since these 1003 tests came from middle school students, for whom this curriculum was designed, Table 1 shows how many students of each grade are represented in the sample.

Table 1. Grades represented by total number of pre- and post-tests

Grade listed	Frequency	Percentage
6	374	37.3%
7	442	44.1%
8	187	18.6%
Total	1003	100.0%

Results

Test: As a whole

Taken as a whole, the mean raw test scores from pre to post changed from nearly an average of 3.4 items correct on the pre-test and 5.5 items correct on the post-test. This change, of 2.1 more items (p < .001) had an effect size of 0.43 , using $r = \frac{z}{\sqrt{N}}$, where N is the number of total observations 1003 and Z was 13.7 (Rosenthal, 1994). The correct rule of thumb for interpretation of this effect size is: 0.1 = small effect, 0.3 = medium effect, 0.5 = large effect. As a whole, the improvement was statistically significant (did not occur by chance) with a medium effect size.

The two figures below are of exactly the same scale, both vertically and horizontally. This allows one to visually inspect how the numbers of students' correct responses changed because of participation in the program.

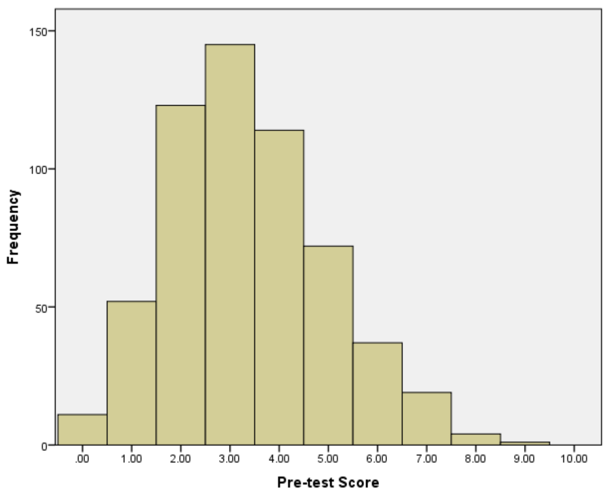

Figure 1. Distribution of raw pre-test scores scores Mean = 3.37, SD = 1.63, N = 578

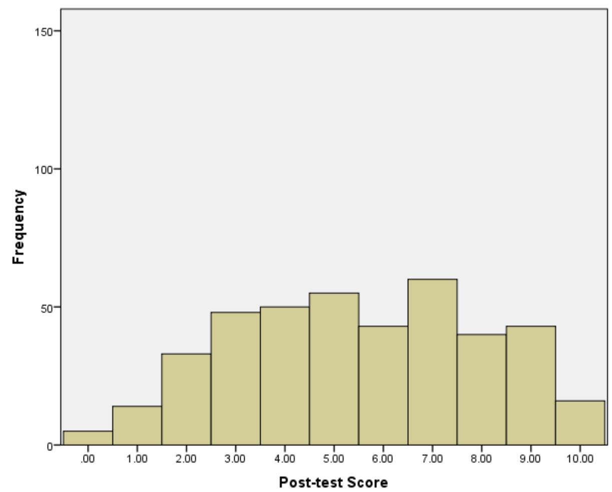

Figure 2. Distribution of raw post-test Mean = 5.51, SD = 2.49, N = 407

Compared to last year's very small sample (N=56), this year's more complete sample of students started off with slightly more prior knowledge (3.37, up from last year's mean of 2.44 correct on the pre-test). This is the equivalent of one more correct response, on average. The post-test mean, of the 407 students who completed the post-test, was 5.51, slightly up from last year's score of 4.78. While a mean improvement in raw score of 2.14 is slightly down from last year's 2.34 items, looking at a "normal" pre-pandemic year (2017-2018) we see that 2.14 is actually more than the 2.09 improvement in 2018.

Test: Item by item

The results from this year's independent sample of pre- and post-tests (N = 1003) showed a statistically significant improvement on all 10 items. Table 3 provides the percentage of the students who got each item correct on the pre-test and the percentage of students who got it right on the post-test. The change in the percentage of students getting each item right on the pre-test and on the post-test is also provided.

Table 4 provides these changes again, but with the chi square calculated for these changes in percentages of correct responses. The appropriate non-parametric inferential test for independent samples is the Fisher's exact test (Siegel, 1956, pp. 96-104). The statistical significance of each of these Fisher's exact tests is included in the table as well. The appropriate effect size for changes in dichotomous (0 = wrong, 1 = right) scores from independent samples, is the simple odds ratio: namely, what are the odds of improving your response to an item on the post-test *after* taking instruction. If the odds ratio calculated for an item were 3.5, this implies, that students who participated in the program had 3.5 times greater likelihood of getting the answer correct when they took the post-test after instruction. To interpret these effect sizes, Chen, Cohen & Chen (2010) suggest that odds ratios higher than 1.68 indicate a small effect (association), odds ratios higher than 3.47 indicate a moderate (or medium) effect, and odds ratios higher than 6.71, indicating a strong effect.

Table 3. Percent of Correct Responses for Each Item on the Pre and Post-tests

Item	Item topic	Pre-test % correct	Post-test % correct	Change
1	power of compounding	41	73	32%
2	smart money strategy	38	50	12%
3	budget benefits	57	70	13%
4	rule of 72	23	54	31%
5	credit card finance charges	27	51	24%
6	loan type with highest interest	36	64	28%
7	bad use of credit	28	54	26%
8	appropriate definition of investing	33	52	19%
9	reasons for salary increase	59	71	16%
10	best investment over 18 years	9	25	16%

Table 4. Inferential Statistics and Effect Sizes for Changes in Percentages of Correct Responses

Item		Change	χ^2	p value	Odds	Interp.
1	power of compounding	32%	100.5	.000	3.89	medium
2	smart money strategy	12%	15.0	.000	1.65	small
3	budget benefits	13%	17.5	.000	1.76	small
4	rule of 72	31%	105.6	.000	4.04	medium
5	credit card finance charges	24%	59.3	.000	2.79	medium
6	loan type with highest interest	28%	93.1	.000	3.87	medium
7	bad use of credit	26%	70.0	.000	3.03	medium
8	appropriate definition of investing	19%	35.9	.000	2.18	small
9	reasons for salary increase	16%	13.0	.000	1.63	small
10	best investment over 18 years	16%	51.7	.000	3.62	medium

All of the changes from pre- to post-test were positive with statistical significance. Six items (#'s 1, 4,5,6,7 and 10) have medium effect sizes, which implies the odds of students getting these items correctly after instruction are at least 2.5 times better (with the best odds pushing 4 times more likely). Four items (#'s 2, 3, 8 and 9) have small effect sizes, which implies that the odds of students getting these items correctly after instruction are at least 1.5 times better.

Table 5 shows the mean post-test scores by grade level. On average, 6th grade students got 6.3 questions correct on the post-test compared to 7th graders who got 5 questions correct and 8th graders who got 4.6 questions right.

Table 5. Mean post-test scores by grade level

Grade	N	Post-test mean	Standard Deviation
6	178	6.3	2.41
7	164	5.0	2.55
8	72	4.6	2.00

To more rigorously investigate these differences discernable by grade, a Kruskal-Wallis test was used on both the mean pre-tests and post-tests. This is the non-parametric equivalent of an ANOVA. The results indicated that the mean change scores did not vary between grades with statistical significance on the pre-test ($p = 0.805$) but they did on the post-test ($p = 0.000$), with the interesting result that 6th grade students who took the post-test did better (with statistical significance on a Scheffe post-hoc test) than both the 7th and 8th grade students. While the Kruskal-Wallis test showed that this statistically significant difference existed in general, it did not show the exact results for all possible comparisons (6 vs.7, 6 vs.8, and 7 vs.8). In parametric statistics, the appropriate test would be ANOVA with post-hoc tests. The reason non-parametric tests are used is because our data are not normally distributed, so using a parametric test violates the normality assumption. Nevertheless, the Scheffe post-hoc used here is the most appropriate, conservative test otherwise, and it showed the 6th graders scored higher, with statistical significance of less than 0.001, than either the 7th or 8th grade students.

Conclusions

This year's sample of tests (N = 1003) indicate that the Money Savvy U™ Curriculum in 2021-2022 had a statistically significant impact ($p<.001$) on the learning of these middle school students as measured with the MSUPFT. On average, the change in mean raw scores changed from a bit more than 3 (3.4) responses out of ten correct to very nearly 6 (5.5). The overall effect size for this gain in test score is medium. Changes in the percentages of participants getting individual items correct before and after instruction clearly indicate improvement on all items as evidenced by statistically significant results ($p<.000$) using the Fisher exact inferential technique.

Student responses to Item #10 are intriguing because on both the pre-test (only 9%) and the post-test (only 25%) chose the correct answer: the stock market. This was the lowest percentage correct – less than half the percentage correct of *every* other set of item responses. What does this reflect of student perceptions of the stock market? What does it indicate about how the teacher explains this topic? Are middle school students averse to the stock market despite instruction and the likelihood that their understanding of it is simplistic? The improvement on this item has been low for many years of these observations. Does this imply a tsmall may be needed to the curriculum on the topic of stocks as a long-term investment strategy?

The normalized learning gain observed in 2021-2022 is 32%. In 2017-2018, the last pre-pandemic study with a matched sample of 734 students, had a normalized gain of 35%. It seems the students and teachers are getting back to "normal" with regards to this program.

Again, all the empirical evidence suggests that the Money Savvy U™ Curriculum, still has a positive, statistically significant impact on middle students' understanding of personal finance.

References

Chen, H., Cohen, P. and Chen, S. (2010). How big is a big odds ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation. 39,* 4 (pp. 860-864).

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), The Handbook of Research Synthesis. (pp. 231-244). New York: Russell Sage Foundation.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. (pp. 96-104). New York: McGraw-Hill.