



**Evaluative Report  
Department of Financial Institutions Program  
Washington State**

**August 2012**

Eric A. Hagedorn, Ph.D.

Hagedorn Evaluation Services  
El Paso, TX

## **Introduction**

The purpose of this study is to evaluate the effectiveness of the Money Savvy U<sup>®</sup> Intermediate Personal Finance Curriculum on pupils in schools in Washington State that used this curriculum.

To investigate the effectiveness of this program a 10 question multiple choice test, called “The Money Savvy U<sup>®</sup> Personal Finance Curriculum test” was used. A portion of the questions were drawn from the JumpStart Coalition Personal Finance Literacy test (items 5, 9, and 10) while the remaining 7 items were provided by the curriculum developers. This study also evaluates the effectiveness of this test in terms of classical and IRT item analysis.

## **Conclusions**

These data indicate that the Money Savvy U<sup>®</sup> Curriculum had a statistically significant impact on the learning of these school children as measured with the test used. The overall effect size for this gain in test score is medium, close to 65% of one full standard deviation of improvement. Changes in the percentages of participants getting individual items correct before and after instruction clearly indicate statistically significant improvement on all items except Item #10 (which is best college savings plan over 18 years: U.S. savings bond, savings account, checking account, stocks?).

While there is clearly room for student improvement (as indicated by getting roughly only 5 out of 10 answers correct, on average, and a Hake Index of 21%), the item facilities indicate a number of difficult items on the test. This is not a bad thing: these items allow this test to assess the learning of stronger students (in addition to average students).

These data from 2011-2012 indicate a program that successfully impacts learning as measured with this instrument.

## **Methodology**

### **Mean Raw Score Analysis**

There were 138 matched pre- and post-tests. These included the participating students’, teachers’ (5 teachers), and school’s names (4). This allowed for matching individual pre and post-tests. Once matched and recorded, either a paired-samples t-test or the non-parametric Wilcoxon Signed Ranks test could be performed on the mean raw scores (out of 10) to determine if student responses changed from pre to post in a statistically significant manner.

Any statistically significant change from pre to post will be identified and interpreted. The effect size of any significant change will also be calculated. The effect size is essentially the ratio of the change to the standard deviation of the change score (Cohen, 1992).

The “Hake Index” is a measure of how much of an improvement has been made from pre to post in terms of the amount the students would have to improve from the mean pre-score to everyone getting 100% correct on the post-test (2001). For instance, a Hake Index of 50% implies that the students improved half as much as they could have.

In addition to this overall analysis of raw score improvement, the percentages of students choosing the correct responses and incorrect responses to each item are provided for the 138

matched tests. The percentage change in students getting the item correct on the post-test who had gotten it wrong on the pre-test are also provided.

### Classical Item Analysis

Classical item parameters (facility and point biserial correlation coefficients) can tell you a great deal about how easy or difficult particular items were to the students who answered these items. They can also tell you about the relationship between students' total scores as compared to whether they got a particular item correct.

Facility, the percentage of students who correctly answered an item, is essentially a measure of item "easiness" and "difficulty." These item difficulties for both the pre and post-tests will be sorted according to rank. The point biserial Correlation coefficient is a special type of correlation between a dichotomous item score (either right or wrong, indicated as "1" or "0," respectively) and the total score. More or less, a good, difficult item would be gotten right by a student with a high total score (implying a student with higher ability) and gotten wrong by a student with a lower total score (a student with lesser ability), thus each of these student's item and total scores would contribute to a higher correlation coefficient (varying as usual from 0 to 1). A problematic item would be one that if a student got it correct, they were likely to have a low total score, or on the contrary, if a student got this item wrong, they nevertheless had a higher total score. This kind of item would give a lower point biserial correlation. The rule of thumb for point biserial correlation coefficients is that values less than 0.200 are problematic, items between .200 and .400 are acceptable and items better than .400 are good. Note that the values of the point biserial on the post-test are probably more meaningful than those on the pre-test: students may have had no idea how to answer these questions before instruction.

## Results

### Grade Demographics

One hundred and thirty eight students from 4 schools and 5 teachers completed tests before and after instruction. For these matched tests, the participants were from the grades indicated in Table 1. The predominant number of students (93.5%) are in the 6<sup>th</sup> and 7<sup>th</sup> grades.

**Table 1.** Student Grade Level

	Frequency	Percent
6	102	73.9
7	27	19.6
8	9	6.5
Total	138	100

### Comparing Matched Tests: Mean Raw Scores and Item Percentages Correct

Table 2 shows the mean raw score (total number correct) for the entire group of pre-tested students compared to the mean raw scores for the same group of post-tested students.

**Table 2.** Descriptive Statistics for Raw Scores on Pre and Post-tests

	N	Mean	Std. Deviation
Pre	138	3.32	1.696
Post	138	4.75	1.999

As these data were not normally distributed, a Wilcoxon Signed Ranks test (the non-parametric equivalent of the paired samples t-test) was used to determine whether the mean increase of 1.43 more questions correct was likely to have occurred by chance. The Wilcoxon Z value was 6.646 which indicates that there was less than one chance in 1000 that this increase occurred by random chance ( $p = .000$ ).

The effect size for this improvement is 0.65, which indicates that the improvement was approximately 65% of a pooled standard deviation. Cohen considered a “medium” effect size (1992). The Hake Index is 21%. This indicates that the improvement from pre to post was essentially 21% of the total improvement that could have been made if everyone got a perfect score on the post-test. Table 3 shows the percentage of students getting each item correct or not on the pre and post-tests.

**Table 3.** Percent of Total Wrong “0” and Correct “1” both Pre and Post

<b>Item 1</b>		Pretest %	Posttest %	% students improving
Valid	0	59.4	24.6	34.8
	1	40.6	75.4	
	Total	100.0	100.0	
<b>Item 2</b>		Pretest %	Posttest %	
Valid	0	51.4	41.3	10.1
	1	48.6	58.7	
	Total	100.0	100.0	
<b>Item 3</b>		Pretest %	Posttest %	
Valid	0	50.7	42.0	8.7
	1	49.3	58.0	
	Total	100.0	100.0	
<b>Item 4</b>		Pretest %	Posttest %	
Valid	0	79.7	63.8	15.9
	1	20.3	36.2	
	Total	100.0	100.0	
<b>Item 5</b>		Pretest %	Posttest %	
Valid	0	74.6	61.6	13.0
	1	25.4	38.4	
	Total	100.0	100.0	
<b>Item 6</b>		Pretest %	Posttest %	
Valid	0	86.2	53.6	32.6
	1	13.8	46.4	
	Total	100.0	100.0	
<b>Item 7</b>		Pretest %	Posttest %	
Valid	0	71.7	53.6	18.1
	1	28.3	46.4	
	Total	100.0	100.0	
<b>Item 8</b>		Pretest %	Posttest %	
Valid	0	71.7	70.3	1.4
	1	28.3	29.7	
	Total	100.0	100.0	

Item 9		Pretest %	Posttest %	
Valid	0	34.8	30.4	4.4
	1	65.2	69.6	
	Total	100.0	100.0	
Item 10		Pretest %	Posttest %	
Valid	0	86.2	86.2	0.0
	1	13.8	13.8	
	Total	100.0	100.0	

### Item Analyses

The percentage of students getting each item correct (facility) as well as the point biserial correlation coefficients for each item, are given in Table 4 for both the pre and post-tests.

**Table 4.** Classical Test Theory Item Parameters

Item	Pre		Post	
	Facility	P.Bis	Facility	P.Bis
1	0.406	0.411	0.754	0.352
2	0.486	0.534	0.587	0.446
3	0.493	0.437	0.580	0.560
4	0.203	0.304	0.362	0.339
5	0.254	0.365	0.384	0.423
6	0.138	0.181	0.464	0.384
7	0.283	0.492	0.464	0.493
8	0.283	0.388	0.297	0.488
9	0.652	0.465	0.696	0.509
10	0.138	0.119	0.138	0.192

Thus, the easiest item on the pre-“The Money Savvy U<sup>®</sup> Personal Finance Curriculum test” was item 9: 65% of the students who completed this item got it correct. On the post-test 70% of the students got it correct. The easiest item on the post-test was number 1: 75% of the students got it correct. On the pre-test, items 6 and 10 were the most difficult with only 14% of students getting it correct. On the post-test, item 10 was the most difficult with the same number of students getting it correct

On the pre-test, two items (#6 and #10) had low point biserial correlation coefficients, i.e. less than 0.200. This indicates that a student’s getting these items correct, does not correlate with their overall score. On the post-test, only Item #10 kept a low point biserial correlation coefficient. It is worth noting that this was the hardest item on both the pre and post-tests.

**Table 5.** Items sorted from easiest to hardest.

Item	Pre		Post	
	Facility	Easiest	Item	Facility
9	0.652		1	0.754
3	0.493		9	0.696
2	0.486		2	0.587
1	0.406		3	0.580
7	0.283		6	0.464
8	0.283		7	0.464
5	0.254		5	0.384
4	0.203		4	0.362
6	0.138		8	0.297
10	0.138	Hardest	10	0.138

### References

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Hake, R. R. (2001). Suggestions for administering and reporting pre/post diagnostic tests. Retrieved May 5, 2004, from <http://www.physics.indiana.edu/~hake/TestingSuggestions051801.pdf>