Evaluative Report
Department of Financial Institutions Program
Washington State

July 27, 2014

Eric A. Hagedorn, Ph.D.

Hagedorn Evaluation Services
El Paso, TX

Money Savvy GENERATION®
HELPING KIDS GET SMART ABOUT MONEY™

Introduction

The purpose of this study is to evaluate the effectiveness of the Money Savvy U 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 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 item analysis.

Conclusions

These data indicate that the Money Savvy U 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 large, almost two full standard deviations (1.98) of improvement. Changes in the percentages of participants getting individual items correct before and after instruction clearly indicate improvement on all items.

While there is clearly room for student improvement (as indicated by getting roughly only 6 out of 10 answers correct, on average, and a Hake Index of 51%), 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 2013-2014 indicate a program that successfully impacts learning as measured with this instrument, and seems to impacts it more so each year. In 2011-2012, the effect size was only 65% of a standard deviation of improvement and the Hake Index was 21%. In 2012-2013, the effect size was 88% of a standard deviation of improvement, and the Hake Index was 42%.

Methodology

Mean Raw Score Analysis

There were 245 matched pre- and post-tests. These included the participating students', teachers' (4 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 557 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 students 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

Five hundred and fifty seven students from 10 schools and 21 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^{th} and 7^{th} grades.

Table 1. Student Grade Level

Grade	Frequency	Percent
6	198	80.8
7	47	19.2
Total	245	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	245	3.88	1.60
Post	245	5.95	2.18

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 2.07 more questions correct was likely to have occurred by chance. The Wilcoxon Z value was 10.94 with a significance of $p < .001$, which indicates that there was less than one chance in 1000 that this increase occurred by random chance.

The effect size for this improvement is 1.98, which indicates that the improvement was very nearly 2 pooled standard deviations. Cohen considered a "large" effect size (1992). The Hake Index is 51%. This indicates that the improvement from pre to post was essentially 51% 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 both tests.

Table 3. Percent of Total Wrong "0" and Correct "1" both Pre and Post

Item 1		Pretest %	Posttest %	% students improving
Valid	Wrong	53.1	16.7	
	Right	46.9	83.3	36.4
	Total	100.0	100.0	
Item 2		Pretest %	Posttest %	
Valid	Wrong	40.8	38.4	
	Right	59.2	61.6	2.4
	Total	100.0	100.0	
Item 3		Pretest %	Posttest %	
Valid	Wrong	45.3	31.4	
	Right	54.7	68.6	13.9
	Total	100.0	100.0	
Item 4		Pretest %	Posttest %	
Valid	Wrong	70.6	36.3	
	Right	29.4	63.7	34.3
	Total	100.0	100.0	
Item 5		Pretest %	Posttest %	
Valid	Wrong	78.4	54.7	
	Right	21.6	45.3	23.7
	Total	100.0	100.0	
Item 6		Pretest %	Posttest %	

Valid	Wrong	81.2	39.2	
	Right	18.8	60.8	42
	Total	100.0	100.0	
Item 7		Pretest %	Posttest %	
Valid	Wrong	71.4	42.4	
	Right	28.6	57.6	29
	Total	100.0	100.0	
Item 8		Pretest %	Posttest %	
Valid	Wrong	59.6	56.7	
	Right	40.4	43.3	2.9
	Total	100.0	100.0	
Item 9		Pretest %	Posttest %	
Valid	Wrong	56.7	15.9	
	Right	43.3	84.1	40.8
	Total	100.0	100.0	
Item 10		Pretest %	Posttest %	
Valid	Wrong	90.2	73.1	
	Right	9.8	26.9	17.1
	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

	Pre		Post	
Item	Facility	P.Bis	Facility	P.Bis
1	0.469	0.356	0.833	0.261
2	0.592	0.457	0.616	0.471
3	0.547	0.409	0.686	0.441
4	0.294	0.305	0.637	0.532
5	0.216	0.441	0.453	0.562
6	0.188	0.186	0.608	0.484
7	0.286	0.363	0.576	0.458
8	0.404	0.445	0.433	0.594
9	0.788	0.348	0.841	0.461
10	0.098	0.239	0.269	0.402

Thus, the easiest item on the pre-"The Money Savvy U Personal Finance Curriculum test" was item 9: 79% of the students who completed this item got it correct. On the post-test 84% of the students got it correct and it remained the easiest item. On the pre-test, item 10 was the most difficult with only 10% of students getting it correct. On the post-test, item 10 remained

the most difficult with 27% of students getting it correct. On the pre-test, one item (#6) had a low point biserial correlation coefficient, i.e. less than 0.200. This indicates that a student's getting these items correct, does not correlate with their overall score. On a pre-test, this is not of great concern. On the post-test, the items all had good point biserial correlation coefficients.

Table 5. Items sorted from easiest to hardest.

	Pre		Post	
Item	Facility	Easiest	Item	Facility
9	0.788		9	0.841
2	0.592		1	0.833
3	0.547		3	0.686
1	0.469		4	0.637
8	0.404		2	0.616
4	0.294		6	0.608
7	0.286		7	0.576
5	0.216		5	0.453
6	0.188		8	0.433
10	0.098	Hardest	10	0.269

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