MONEY SAVVY KIDS™
Basic Personal Finance Curriculum

Evaluative Report
Department of Financial Institutions Program
Washington State

September, 2005

Eric A. Hagedorn, Ph.D.

Hagedorn Evaluation Services
El Paso, TX

Money Savvy Generation®

Introduction

The purposes of this study are to: 1) measure the effectiveness of a program called Money Savvy Kids™ on the attitudes and knowledge of young children in public schools in Washington State; and 2) to begin to assess if the program is differentially effective with children of varied demographic variables (e.g. lower socio-economic status, ethnicity, school location, etc).

Money Savvy Kids™ is curriculum developed by Money Savvy Generation of Lake Bluff, Illinois.. The curriculum includes eight lessons:

- The History of Money
- Where Does Money Come From?
- Kids Can Earn Money Too!
- Saving Money and Bank Field Trip
- Spending Money
- Donating Money
- Investing Money
- Family Money Press Conference

An important part of Money Savvy Kids™ curriculum is the Money Savvy Pig™. This is a four slot piggy bank. It provides teachers and parents with a fun and interesting way to introduce children to saving, spending, investing, and donating. Each child participating in the program receives a Money Savvy Pig™. During the academic year 2004-2005, 43 elementary school teachers, representing 50 classrooms received Money Savvy Kids™ materials and curriculum training. Training was provided either in person, via telephone or via self-study materials. They were asked to implement the program in their classrooms and to use a pre-and post test with the students. Usable pre and post tests were received back from 14 classrooms, representing two hundred and eighty-two second graders and 19 third graders with pre- and post-tests that could be matched (by identifiable names)

To investigate the effectiveness of this program, Dr. Mark Schug of the Center for Economics Education at the University of Wisconsin – Milwaukee, developed survey (see Appendix A) measuring student beliefs about savings habits, handling money, the role of business, etc. This survey has been used in each subsequent evaluation study since the first such study at the end of the 2003-2004 school year. This study was featured in the academic journal *The Social Studies* in Spring 2005 (Schug & Hagedorn, 2005). This survey was given to the Washington students before receiving their Money Savvy Pigs and after they had completed their training. This report presents the analysis and interpretation of the results of those surveys.

Conclusions

Overall, the aggregate data indicate that the Money Savvy Kids™ program is effective in positively affecting students' attitudes and knowledge about spending, saving and investing money. The paired samples data indicates statistically significant improvements on seven out of ten items. While the effect sizes were small, the pre-test scores on Items 1, 2, 4, 6, 7 & 8 were already leaning the correct way. The small increases implied an even greater majority choosing the correct response, and it is very unlikely this occurred by chance.

None of the significant changes, for the entire sample, indicated inappropriate understandings. The non-significant changes on Items 3, 5 and 9 do not indicate a problem, however. Item 3 refers to having things when I want them – a position opposed to saving. Students disagreed with this on average on the pre-test and even more so on the post-test, but not with statistical significance. Item 5, which states: "The thing I enjoy most about earning money is getting to spend it later on." This statement is somewhat confusing: it implies saving with "later on" but refers explicitly to spending. The mean on the pre-test was 2.548 (SD=.7540) and on the post-test 2.568 (SD=.7536). These mean scores are between agree and uncertain, moving slightly towards agreeing. Item 9 states: "It is important for families to keep money in real banks." The pre-test mean was 2.688 (SD=.7039) and the post-test mean was 2.714 (SD=.6414). Both these scores indicate fairly strong agreement, with a change too small to be statistically significant. This item may be common knowledge for most students, even before the pre-test.

When the data were analyzed by school, some statistically significant differences occurred between some schools. When the scores at these schools were linked with the demographic variables of percentage of free or reduced lunch or percentage of non-Caucasians, no identifiable pattern emerged. Some low SES schools did extremely well and others did poorly and the same may be said for the higher SES schools. Some schools with high percentages of minorities did better than those with high percentages of Caucasians and vice versa. These differences in school performance are likely attributable to teacher characteristics, including pedagogical approach.

In this evaluator's professional opinion, these data indicate that the Money Savvy Kids™ curriculum worked effectively for students of varied backgrounds in Washington State. This is consistent with results found in a previous study where participating students in affluent Chicago suburbs learned comparably to urban Chicago public school students, as measured with the Money Savvy Kids Assessment. These two studies suggest the generalizability of the statement: the Money Savvy Kids™ curriculum is effective across a wide variety of English reading students.

Methodology

The Money Savvy Kids™ Assessment is a 10 item, Likert scale instrument. A three point response format was used: a smiley face for agree (with a value of 3), a straight mouth face for don't know or unsure (with a value of 2) and a frown face for disagree (with a value of 1). Dr. Schug had a literacy expert check the questions for roughly a second grade reading level.

The completed pre- and post-tests were to include the participating student's name. This would allow 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 would be performed on the data to determine if student responses changed from pre to post in a statistically significant manner. The paired samples t-test is appropriately used if the data did not differ significantly from a normal distribution. Normality is determined using the Kolmogorov-Smirnov test of normality (with Lillefiors correction) and the Shapiro-Wilk test. If the data do differ significantly from the normal distribution, one uses the Wilcoxon Signed Ranks test.

(Test and survey data can often deviate from the normal distribution due to floor effects on pre-tests and ceiling effects on post-tests. Another factor which can cause deviations from normality are outliers – test scores that are very low or very high. There are two general approaches to dealing with non-normality: data cleaning and transforming and using non-parametric statistics. Data cleaning includes removing outliers. Data transformations involve mathematical transformations of data, such as taking the logarithms of the data, and if this generates a normal distribution, doing statistical tests on the transformed scores. This evaluator prefers to accept the data as they are and use the appropriate non-parametric tests as needed.)

Any statistically significant changes from pre- to post- would be identified and interpreted. A statistically significant difference in means from pre- to post- indicates the likelihood that such a difference in mean in the population would occur by chance. For instance, an increase of mean score on item 3 of .31 (on a scale of 1 to 5) occurs by chance only once in a thousand, as indicated by a p value equal to .001. While this information implies statistical significance (likelihood of occurring by chance), it says nothing about "how big" or "how important" a change of .31 is. To begin to understand these issues, one calculates effect sizes. The effect size is essentially the ratio of the change to the standard deviation of the change scores. If the standard deviation of the change scores for Item 3 were around .3, the effect size would be about 1, indicating the change was roughly one whole standard deviation. In the literature, such an effect size is considered "large" (Kirk, 1995). If the standard deviation of the change scores was around 3 (indicate great variability in student responses to Item 3), the effect size would only be .10 – representing a change of about $1/10^{th}$ of a standard deviation. This effect size is considered "small," even though the likelihood that such a change occurred by chance is very unlikely.

Analysis of variance (ANOVA) with post-hoc tests was proposed to determine if students of a particular socio-economic status, ethnicity, or primary language preference perform differently on the Money Savvy Kids Assessment. ANOVA, however are only appropriate when the data are normally distributed and the assumption of equal variances is met. If these assumptions are not met the Brown-Forsythe robust test of equality of means and the non-parametric Kruskal-Wallis test would be used. While each of these tests would determine if statistically significant

differences in means occurred between different groups of students, they would not indicate which specific groups differed. To determine this (and begin to interpret exactly how the groups differed – by SES or language, etc.) post-hoc tests need to be interpreted. If the normality and equal variances assumptions are not met, such post-hoc tests are not available in standard statistical software. If this were the case, ANOVA results would be compared to the Brown-Forsythe and Kruskal-Wallis results. If they were similar, a conservative interpretation of the post-hoc tests would be used to interpret any significant differences. In the event that the ANOVA (or Brown-Forsythe and Kruskal-Wallis) tests did not indicate significant differences at all, one could interpret that the assessment scores did not vary by differing groups of students.

Results

Entire Sample: Mean Item Changes

301 students could be identified by name and completed the pre- and post-tests. The average scores and standard deviations for each item are given in Table 1. Post-test items marked with an asterisk indicate a statistically significant improvement in average student response from pre to post.

Table 1. Item response averages and standard deviations for paired samples data.

	Pre	**SD**	**Post**	**SD**
Item 1	2.676	0.5907	2.854***	0.4226
Item 2	1.590	0.8136	1.334***	0.6511
Item 3	1.399	0.7586	1.342	0.6825
Item 4	2.606	0.7167	2.716*	0.5801
Item 5	2.548	0.7540	2.568	0.7356
Item 6	2.548	0.7747	2.103***	0.9306
Item 7	2.196	0.7680	1.912***	0.8464
Item 8	2.610	0.6767	2.761**	0.5910
Item 9	2.688	0.7039	2.714	0.6414
Item 10	2.075	0.8805	1.944*	0.8719

Table 2. Significantly changed item response averages and effect size of changes.

Item	Z value	Exact 2-tailed significance	Effect size
1. I know a lot about how to handle my money.	-4.743	.000	0.27
2. I believe that people act selfishly when they save money.	-4.777	.000	0.28
4. I believe it is important to save money for the things that I want to buy in the future.	-2.327	.020	0.13
6. It is best to put the money you save in your room at home.	-5.923	.000	0.37
7. When I invest in stocks, I will always make money.	-4.844	.000	0.28
8. Business people help others by providing them with goods and services.	-2.973	.003	0.17
10. I believe saving money helps me but not help anyone else.	-2.121	.034	0.12

What Tables 1 and 2 tell us about student responses to individual items. The average response of the students to item 1 changed from 2.676, leaning towards agreeing, to 2.854, which leans even more towards strongly agreeing. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance, less than 1 in1000 times. The .27 effect size indicates that this improvement is roughly one quarter of an average standard deviation in size. Cohen considers this a "small effect."

The average response of the students to item 2 changed from 1.590, on the disagreeing side of uncertain, to 1.334, which is more strongly disagreeing. This indicates an improvement in student understanding, because it is appropriate for students to disagree with this item. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance, less than 1 in1000 times. The .28 effect size indicates that this improvement is slightly more than one quarter of an average standard deviation in size. Cohen considers this a "small effect."

The average response of the students to item 4 changed from 2.606, leaning towards agreeing, to 2.716, which leans even more towards strongly agreeing. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance, less than 2 times in 100. The .13 effect size indicates that this improvement is 13% of an average standard deviation in size. Cohen considers this a "small effect."

The average response of the students to item 6 changed from 2.548, leaning towards agreeing, to 2.103, which is very close to uncertain. This indicates an improvement in student understanding, because even though the average post-test score is uncertain, this average decreased from the pre-test because more students disagreed with this item, which was the learning objective. The exact two-tailed significance implies that this change in average score could only have occurred

by chance less than 1 out of 1000 times. The .37 effect size indicates that this decrease in score is roughly 37% of an average standard deviation in size. Cohen considers this a "small effect."

The average response of the students to item 7 changed from 2.196, very close to uncertain, to 1.192, which leans more towards disagreeing. This indicates an improvement in student learning because it is more appropriate for students to disagree with this item. The two-tailed exact significance implies that this change in average score could only have occurred by chance, less than 1out of 1000 times. The .28 effect size indicates that this improvement is roughly 28% of an average standard deviation in size. Cohen considers this a "small effect."

The average response of the students to item 8 changed from 2.610, leaning towards agreeing, to 2.761, which leans more towards agreeing. This indicates an improvement in student learning because it is more appropriate for students to agree with this item. The two-tailed exact significance implies that this change in average score could only have occurred by chance, 1 out of 1000 times. The .17 effect size indicates that this improvement is roughly 17% of an average standard deviation in size. Cohen considers this a "small effect."

The average response of the students to item 10 changed from 2.075, very close to uncertain, to 1.944, which leans towards disagreeing. Even though the average post-test score is still very close to uncertain, this average decreased from the pre-test because more students disagreed with this item, which was the learning objective. The two-tailed exact significance implies that this change in average score could only have occurred by chance, 17 out of 1000 times. The .12 effect size indicates that this improvement is roughly 12% of an average standard deviation in size. Cohen considers this a "small effect."

Analyses by Demographic Groupings

Change Scores. Means and standard deviations for each of the item change scores (post-score minus pre-score) are listed in Table 3. Note that items 2, 3, 6, 7, and 10 are items that students should disagree with, therefore, negative change scores indicate a good thing – less students agreeing after than before.

Table 3. Descriptive statistics for change scores

	N	Mean	Std. Deviation
ch1	301	.1777	.66267
ch2	301	-.2558	.91525
ch3	301	-.0565	.81556
ch4	301	.1096	.84731
ch5	301	.0199	.92445
ch6	301	-.4452	1.19142
ch7	301	-.2841	1.00906
ch8	301	.1512	.88010
ch9	301	.0266	.85593
ch10	301	-.1312	1.08331
Valid N (listwise)	301		

These data all differed from the normal distribution - the Kolmogorov-Smirnov (with Lillefiors correction) and the Shapiro-Wilk tests of normality indicated that there was less than one chance in a thousand that these data could have come from a normally distributed population. Because of this, traditional ANOVA analyses (see Table 4) were supplemented with Welch and Brown-Forsythe robust tests of equality of means (see Table 5) and non-parametric Kruskal-Wallis tests (see Table 6). Significant differences are indicated with asterisks and yellow shading (for electronic version).

Table 4. Analyses of variance for each change score by school groupings.

		Sum of Squares	df	Mean Square	F	Sig.
ch1	Between Groups	5.098	11	.463	1.058	.396
	Within Groups	126.643	289	.438		
	Total	131.741	300			
ch2	Between Groups	26.147	11	2.377	3.051	.001*
	Within Groups	225.155	289	.779		
	Total	251.302	300			
ch3	Between Groups	7.907	11	.719	1.084	.374
	Within Groups	191.633	289	.663		
	Total	199.540	300			
ch4	Between Groups	10.686	11	.971	1.372	.186
	Within	204.696	289	.708		

	Groups					
	Total	215.382	300			
ch5	Between Groups	5.372	11	.488	.562	.859
	Within Groups	251.009	289	.869		
	Total	256.380	300			
ch6	Between Groups	62.273	11	5.661	4.500	.000*
	Within Groups	363.573	289	1.258		
	Total	425.846	300			
ch7	Between Groups	30.292	11	2.754	2.892	.001*
	Within Groups	275.171	289	.952		
	Total	305.463	300			
ch8	Between Groups	17.842	11	1.622	2.185	.015*
	Within Groups	214.530	289	.742		
	Total	232.372	300			
ch9	Between Groups	8.157	11	.742	1.013	.435
	Within Groups	211.630	289	.732		
	Total	219.787	300			
ch10	Between Groups	26.323	11	2.393	2.123	.019*
	Within Groups	325.743	289	1.127		
	Total	352.066	300			

You will note that Items 2, 6, 7, 8 and 10 are significantly different on each of this tests.

Table 5. Welch and Brown Forsythe Robust Tests of Equality of Means.

		Statistic(a)	df1	df2	Sig.
ch1	Welch	1.205	11	94.207	.294
	Brown-Forsythe	.966	11	161.124	.480
ch2	Welch	2.576	11	93.658	.007*
	Brown-Forsythe	2.891	11	181.520	.002*
ch3	Welch	1.486	11	92.450	.150
	Brown-Forsythe	1.191	11	199.369	.295
ch4	Welch	1.537	11	94.251	.131
	Brown-Forsythe	1.334	11	179.519	.209
ch5	Welch	.476	11	95.746	.914
	Brown-Forsythe	.533	11	149.056	.878
ch6	Welch	4.304	11	94.638	.000*
	Brown-Forsythe	4.369	11	192.175	.000*
ch7	Welch	2.966	11	94.765	.002*
	Brown-Forsythe	2.912	11	219.415	.001*
ch8	Welch	2.230	11	97.505	.018*
	Brown-Forsythe	2.307	11	160.109	.012*
ch9	Welch	1.187	11	97.122	.306
	Brown-Forsythe	1.080	11	221.998	.378
ch10	Welch	2.114	11	93.829	.026*
	Brown-Forsythe	2.009	11	197.393	.029*

a Asymptotically F distributed.

Table 6. Kruskal-Wallis test results.

	ch1	ch2	ch3	ch4	ch5
Chi-Square	14.235	33.990	13.870	15.212	6.259
df	11	11	11	11	11
Asymp. Sig.	.220	.000*	.240	.173	.856

Table 6 (continued). Kruskal-Wallis test results.

	ch6	ch7	ch8	ch9	ch10
Chi-Square	45.052	32.880	29.986	13.163	25.214
df	11	11	11	11	11
Asymp. Sig.	.000*	.001*	.002*	.283	.008*

The significant differences indicated by each of these tests for Items 2, 6, 7, 8 and 10 tell us that at least 2 of the 12 groups had mean scores significantly different from one another on each of these items. These tests do not tell us which 2 (or more) groups had such differences. To determine which groups are significantly different, one typically uses post-hoc tests. Post-hoc tests typically require normal data and homogeneity of variance. Because the data are definitely not normal, the post-hoc tests performed will be followed up with non-parametric Mann-Whitney tests. In addition any items violating homogeneity of variance will be tested with a post-hoc test appropriate for this situation – the Tamhane post-hoc test, rather than the Scheffe test..

Table 7. Test of homogeneity of variances.

	Levene Statistic	df1	df2	Sig.
ch2	2.455	11	289	.006*
ch6	1.877	11	289	.042
ch7	1.593	11	289	.100
ch8	3.524	11	289	.000*
ch10	1.225	11	289	.269

The Tamhane post-hoc tests indicate significant differences between groups 3 and 8 for Item 2 (p=.008) and between groups 1 and 11 for item 8 (p=.009). Item 2 refers to saving money being greedy. Group 8 represents a school with a fairly low percentage of free and reduced lunch students (14%). Their change score of .1563 is in the wrong direction – more of them believed saving money was greedy after participating in the program. Group 3 represents a school with a fairly high percentage of free and reduced lunch students (89%). They, on the other hand, had the greatest possible improvement with respect to item 2: - .6875. Double checking the difference between Group 8 and Group 3 with a Mann-Whitney test indicated a U of 371.5 (p= .000).

Item 8 states that business people help others by providing goods and services. Group 1 had the highest change in the correct direction: .5714. This school has 64% of its students participating in the free and reduced lunch program. Group 11, which had the greatest change in the wrong direction: -.4118, has 31% of its students eating free and reduced price lunches. The difference between Groups 1 and 8 gave a Mann-Whitney U of 125.5 (p=.000).

The Scheffe post-hoc tests on Items 6, 7, and 10 indicated significant differences between groups on only Item 6, between Groups 4 and 1, and 4 and 11. Item 6 refers to it being best to save your money in your room at home. This is an item where accurate learning leads to a negative change score. Group 4 had the largest change in the right direction: -1.375. Group 4 represents a school with a moderate percentage of free and reduced lunch students (49%). Group 1 had the smallest change in the right direction: -.0571. Group 1 is 64% free and reduced lunch. The difference between Groups 4 and 1 has a Mann-Whitney U of 189.0, which is significant at the p = .000 level. Group 11 was the only group which moved on average the wrong direction on Item 6 – a change score of .4706. As such, it differed quite a bit with Group 4 – having a Mann-Whitney U of 78.5, which is significant at the p=.000 level.

A final test to determine if demographic variables were related to scoring well or less well, on items with significant improvements, was to calculate both the parametric Pearson product moment (see Table 8) and the nonparametric Spearman correlation coefficients (see Table 9). While the sample size for these analyses was 11 (12 schools, minus one for which no demographic data were supplied), no statistically significant correlations of either type were found. This is consistent with the various analyses of variance. Note that the percentage of free and reduced lunches correlated strongly and significantly with percent minorities. Despite the small sample size, this well documented social statistic appeared unambiguously.

Table 8. Pearson correlation coefficients between demographic variables and significant change scores.

		freered	Permin
freered	Pearson Correlation	1	.735(**)
	Sig. (2-tailed)	.	.010
permin	Pearson Correlation	.735(**)	1
	Sig. (2-tailed)	.010	.
ch2ave	Pearson Correlation	-.079	-.139
	Sig. (2-tailed)	.817	.684
ch6ave	Pearson Correlation	-.311	-.070
	Sig. (2-tailed)	.351	.839
ch7ave	Pearson Correlation	-.021	.004
	Sig. (2-tailed)	.952	.990
ch8ave	Pearson Correlation	.243	.426
	Sig. (2-tailed)	.472	.191
ch10ave	Pearson Correlation	.305	-.063
	Sig. (2-tailed)	.362	.855
	N	11	11

** Correlation is significant at the 0.01 level (2-tailed).

Table 9. Spearman correlation coefficients between demographic variables and significant change scores.

		freered	permin
freered	Correlation Coefficient	1.000	.773(**)
	Sig. (2-tailed)	.	.005
permin	Correlation Coefficient	.773(**)	1.000
	Sig. (2-tailed)	.005	.
ch2ave	Correlation Coefficient	-.191	-.045
	Sig. (2-tailed)	.574	.894
ch6ave	Correlation Coefficient	-.527	-.264
	Sig. (2-tailed)	.096	.433
ch7ave	Correlation Coefficient	-.064	-.200
	Sig. (2-tailed)	.853	.555
ch8ave	Correlation Coefficient	.136	.427
	Sig. (2-tailed)	.689	.190
ch10ave	Correlation Coefficient	.255	-.164
	Sig. (2-tailed)	.450	.631
	N	11	11

** Correlation is significant at the 0.01 level (2-tailed).

Appendix A: Money Savvy Kids™ Assessment

Directions: Teachers, please read each of the following 10 sentences together in class. Explain the following directions to the children: If you ***agree*** with the statement, use your pencil to circle the ***face with the smile***. If you ***don't know*** or are ***unsure*** about the statement, circle the ***face with the straight mouth***. If you ***disagree*** with the statement, circle the ***face the frown***. Please circle only one face for each question.

1.	I believe I know a lot about how to handle my money.	
2.	I believe that people act selfishly when they save money.	
3.	I believe it is important to have the things I want when I want them.	
4.	I believe it is important to save money for the things that I want to buy in the future.	
5.	The thing I enjoy most about earning money is getting to spend it right away.	
6.	It is best to save your money in a secret place in your bedroom.	
7.	I believe that some places to put my savings - - like putting money in banks - - are safer than others.	
8.	I believe business people help others by providing them with goods and services to buy.	
9.	It is important for families to keep money in real banks.	
10.	I believe saving money helps me but not help anyone else.	

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (Third ed.). Pacific Grove: Brooks/Cole Publishing Company.

Schug, M. C., & Hagedorn, E. A. (2005). The Money Savvy Pig™ goes to the big city: Testing the effectiveness of an economics curriculum for young children. *The Social Studies, 96*(2).