



**Evaluative Report
North Dakota Department of Securities**

October, 2005

Eric A. Hagedorn, Ph.D.

Hagedorn Evaluation Services
El Paso, TX

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 North Dakota.

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, 9 elementary school teachers in 8 schools, representing 9 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.

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 North Dakota 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 eight out of ten items. While the effect sizes were small (with the exception of item 8, which had a medium effect size), the pre-test scores on Items 1, 2, 3, 4, & 10 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. On items 6 and 7, the mean scores went from the agreeing side of uncertain to the disagreeing side (as appropriate).

None of the significant changes, for the entire sample, indicated inappropriate understandings (for the entire group as a whole). The non-significant changes on Items 5 and 9 do not indicate a problem, however. Item 5, which states: “The thing I enjoy most about earning money is getting to spend it later on.” As noted in previous studies, this statement is somewhat confusing: it implies saving with “later on” but refers explicitly to spending. Item 9 refers to families keeping money in real banks. Students already agreed with this rather strongly on the pre-test (2.80) and while an increase in agreement to 2.86 is not statistically significant, it does indicate even more agreement..

When the data were analyzed by school, some statistically significant differences occurred between some schools. This was useful in identifying schools where the mean change scores on some items went the wrong way from pre to post. This may identify teachers who either misunderstood a concept and communicated this misconception effectively or teachers who understood a concept and did not succeed in communicating it to enough of their students.

In this evaluator’s professional opinion, these data indicate that the Money Savvy Kids™ curriculum worked effectively for the students who participated in this study from North Dakota. These results are consistent with those found in previous studies in both urban and suburban Chicago; in Cleveland, Ohio, and in Washington State. These 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 Lilliefors 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 issue, 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 1/10th of a standard deviation. This effect size is considered “small,” even though the likelihood that such a change occurred by chance is very unlikely.

Finally, analyses of variance (both parametric and non-parametric) and post-hoc tests were used to identify schools which had statistically significant changes in mean scores in the wrong direction.

Results

Entire Sample: Mean Item Changes

159 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	2.68	.532	2.87*	.340
Item 2	1.53	.780	1.31*	.607
Item 3	1.46	.753	1.30*	.642
Item 4	2.71	.546	2.88*	.399
Item 5	2.42	.782	2.45	.752
Item 6	2.30	.853	1.83*	.886
Item 7	2.01	.755	1.76*	.811
Item 8	2.78	.503	2.87*	.378
Item 9	2.80	.489	2.86	.425
Item 10	1.88	.787	1.71*	.849

* Statistically significant change from pre- to post.

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.	-4.328	.000	0.36
2. I believe that people act selfishly when they save money.	-3.183	.001	-0.27
3. I believe it is important to have the things I want when I want them.	-2.391	.017	-0.20
4. I believe it is important to save money for the things that I want to buy in the future.	-3.33	.001	0.27
6. It is best to put the money you save in your room at home.	-5.826	.000	-0.53*
7. When I invest in stocks, I will always make money.	-3.149	.002	-0.25
8. Business people help others by providing them with goods and services.	-2.016	.054 (.044)	0.16
10. I believe saving money helps me but not help anyone else.	-2.105	.036	-0.18

*Medium effect size. All others are small.

What Tables 1 and 2 tell us about student responses to individual items.

In general these two tables show that there were statistically significant improvements in student understanding on 8 out of 10 items on the assessment.

The average response of the students to item 1 changed from 2.68, leaning towards agreeing, to 2.87, which leans even more towards strongly agreeing. More students, on average believe that they know how to handle their money. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance, less than 1 in 1000 times. The .36 effect size indicates that this improvement is one third of an average standard deviation in size. Cohen considers this a “small effect.”

The average response of the students to item 2 changed from 1.53, on the disagreeing side of uncertain, to 1.31, which is more strongly disagreeing. This indicates an improvement in student understanding, because it is appropriate for students to disagree with the notion that saving money is selfish. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance only 1 in 1000 times. The -.27 effect size indicates that this improvement is roughly 27% an average standard deviation in size. Cohen considers this a “small effect.” The minus sign indicates that the average score decreased from pre to post (which is appropriate for this item).

The average response of the students to item 3 changed from 1.46, on the disagreeing side of uncertain, to 1.30, which is more strongly disagreeing. This indicates an improvement in student understanding, because it is appropriate for students to disagree with the notion that one must have what one wants when one wants it. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance only 17 in 100 times. The -.20 effect size indicates that this improvement is roughly 20% an average standard deviation in size. Cohen considers this a “small effect.” The minus sign indicates that the average score decreased from pre to post (which is appropriate for this item).

The average response of the students to item 4 changed from 2.71, leaning towards agreeing, to 2.88, which leans even more towards strongly agreeing. This is an improvement of student understanding because it is appropriate for students to believe that saving money for future purchases is important. The two-tailed exact significance implies that this improvement in average score could only have occurred by chance, less than 1 times in 1000. The .27 effect size indicates that this improvement is 17% of an average standard deviation in size. Cohen considers this a “small effect.”

The average response of the students to item 6 changed from 2.30, uncertain, but leaning towards agreeing, to 1.83, which is much closer 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 the idea that saving money in your room is the best method of saving. 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 -.53 effect size indicates that this decrease in score is roughly 53% of an average standard deviation in size. Cohen considers this a “medium effect.”

The average response of the students to item 7 changed from 2.01, very close to uncertain, to 1.76, which leans more towards disagreeing. This indicates an improvement in student learning because it is more appropriate for students to disagree with the idea that investing in the stock market always pays off. The two-tailed exact significance implies that this change in average score could only have occurred by chance 2 out of 1000 times. The -.25 effect size indicates that this improvement is roughly 25% of an average standard deviation in size. Cohen considers this a “small effect.”

The average response of the students to item 8 changed from 2.78, leaning towards agreeing, to 2.87, which leans more towards agreeing. This indicates an improvement in student understanding because it is appropriate for students to believe that business people help others by providing goods and services. The exact two-tailed significance of .054 indicates that this change should be rejected as not significant, however the asymptotic two-tailed significance 0.044 indicates that this change could only have occurred by chance 44 times in a 1000. In light of the closeness of the exact probability to .05, this result is considered significant. The .16 effect size indicates that this improvement is 16% of an average standard deviation in size. Cohen considers this a “small effect.”

The average response of the students to item 10 changed from 1.88, on the disagreeing side of uncertain, to 1.71, which leans towards disagreeing. This indicates an improvement in student understanding because more students should disagree that saving money helps only the saver. The two-tailed exact significance implies that this change in average score could only have occurred by chance, 36 out of 1000 times. The -.18 effect size indicates that this improvement is roughly 18% of an average standard deviation in size. Cohen considers this a “small effect.”

Analyses by Schools/Classrooms

These data all differed from the normal distribution - the Kolmogorov-Smirnov (with Lilliefors 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 non-parametric Kruskal-Wallis tests (see Table 5). Significant differences are indicated with asterisks and yellow shading (for electronic version).

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	159	.1887	.51766
ch2	158	-.2152	.80872
ch3	159	-.1635	.83343
ch4	157	.1688	.62418
ch5	159	.0314	.83758
ch6	156	-.4679	.87562
ch7	159	-.2547	1.03402
ch8	156	.0897	.54916
ch9	158	.0665	.57303
ch10	157	-.1720	.97519
Group	159	4.74	2.265

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

		Sum of Squares	df	Mean Square	F	Sig.
ch1	Between	3.830	8	.479	1.865	.070
	Within	38.510	150	.257		
	Total	42.340	158			
ch2	Between	7.564	8	.946	1.481	.169
	Within	95.119	149	.638		
	Total	102.684	157			
ch3	Between	8.347	8	1.043	1.543	.147
	Within	101.402	150	.676		
	Total	109.748	158			
ch4	Between	6.971	8	.871	2.397	.018*
	Within	53.806	148	.364		
	Total	60.777	156			
ch5	Between	2.794	8	.349	.485	.866
	Within	108.049	150	.720		
	Total	110.843	158			
ch6	Between	9.580	8	1.197	1.611	.126
	Within	109.260	147	.743		
	Total	118.840	155			
ch7	Between	41.361	8	5.170	6.079	.000**
	Within	127.573	150	.850		
	Total	168.934	158			
ch8	Between	9.350	8	1.169	4.595	.000**
	Within	37.393	147	.254		
	Total	46.744	155			
ch9	Between	2.045	8	.256	.769	.630
	Within	49.508	149	.332		
	Total	51.552	157			
ch10	Between	6.838	8	.855	.894	.523
	Within	141.519	148	.956		
	Total	148.357	156			

Table 5. Kruskal-Wallis test results.

	ch1	ch2	ch3	ch4	ch5
Chi-Square	16.030	9.637	13.367	14.430	5.256
Df	8	8	8	8	8
Asymp. Sig.	.042*	.291	.100	.071	.730

	ch6	ch7	ch8	ch9	ch10
Chi-Square	11.878	35.555	31.532	6.126	8.611
Df	8	8	8	8	8
Asymp. Sig.	.157	.000**	.000**	.633	.376

The ANOVA indicate significant differences for Items 4, 7, and 8. This tells us that at least 2 of the 9 groups had mean scores significantly different from one another on each of these items. The Kruskal-Wallis tests indicate significant differences for Items 1, 7 & 8. Again this tells us that at least 2 of the 9 groups had mean scores significantly different from one another. None of these tests 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. 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. The purpose of these tests is to identify schools which fared significantly more poorly than the others.

Table 6 indicates which change scores violated the assumption of homogeneity of variance, which determines whether it is appropriate to use Scheffe or Tamhane post-hoc tests. Item 7 was thus analyzed using the Scheffe post-hoc test and Items 1, 4 and 8 were analyzed with the Tamhane post-hoc test.

Table 6. Test of homogeneity of variances.

	Levene Statistic	df1	df2	Sig.
ch1	3.072	8	150	.003
ch4	2.029	8	148	.047
ch7	1.406	8	150	.198
ch8	4.672	8	147	.000

The Tamhane post-hoc test on Item 1 indicated no significant difference (which is consistent with the non-significant F), nevertheless, inspection of the means plot shows zero average change from pre to post for groups 2 & 3.































The Scheffe post-hoc test on Item 4 (that it is important to save for things that you want in the future), did indicate significant differences between Groups 3 and 5. The problematic school is in Group 3. Their mean score on this item was a positive $-.1034$ ($SD = .61788$). This implies that more children at Group 3 disagreed with this statement after instruction. The discerning reader may note that a Scheffe test is not appropriate for nonhomogenous variances, so a Mann-Whitney U test was conducted to verify the difference between groups 3 and 5. The results confirm a significant difference, at the .004 level.

The Scheffe post-hoc tests for Item 7 (investing in stocks always leads to profit) indicate significant differences between three groups (3, 5 and 8 with inappropriate response means (positive, in this case) and group 7 which had the largest, correct, negative mean change score.

The Tamhane post-hoc tests for Item 8 (business people help others by providing them with goods and services) indicate 3 significant differences between groups 3, 6 and 8 with group 7. Groups 3, 6 and 8, had incorrect negative mean change scores and group 7 had the largest correct (positive) change score.

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 put the money you save in your room at home. |  |  |  |
| 7. | When I invest in stocks, I will always make money. |  |  |  |
| 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).