

Wings for Kids Evaluation
Academic year 2005/2006
Academic Achievement Outcomes

Zorana Ivcevic

Susan Rivers

and

Marc A. Brackett

Health, Emotion, & Behavior Laboratory

Department of Psychology

Yale University

***Wings for Kids* Evaluation Study: Academic Achievement Outcomes**

This report describes the continuation of the evaluation of *Wings for Kids* after-school program for elementary school children. In this part of the evaluation, report cards were obtained from all students in the Memminger elementary school at the end of the 2005/2006 academic year, along with the information about which students participated in *Wings*. The goal was to compare measures of academic grades, absenteeism, tardiness, and social development between students who attended *Wings* and those who did not attend *Wings*. Furthermore, students' performance was examined in longitudinal analyses to address whether *Wings* students are more likely to maintain high academic achievement or improve than are *non-Wings* students.

The report includes three sections:

- (1) description and discussion of measures and procedures employed;
- (2) report and interpretation of results; and
- (3) conclusions and suggestions for future research.

I. Measures and procedure

Information on academic achievement was obtained from report cards at the end of the 2005/2006 academic year. Report cards contained information on performance in different academic subjects, school attendance/absence, tardiness, and social and personal development. Students in first and second grades were evaluated on a 3-point scale. For each academic subject, students were graded as 'consistently demonstrating', 'sometimes demonstrating', or 'rarely demonstrating' a skill or knowledge set. Students in third through sixth grade were evaluated on a percentile grade-point scale (55-100 points). Analyses were performed separately for students receiving three-point grades and numeric grade points.

Nine academic subjects were analyzed: English language arts, reading, social studies, science, math, health, art, music, and physical education. All analyses were performed on final grades for the academic year to increase reliability of the grade measures. In some cases, report cards did not contain final year-end grade and the final grade was computed based on all available evaluation components across all grading periods. Report cards are not uniform across different grade levels and students in different grades do not get evaluated on the identical set of academic subjects. Because of this, sample sizes for different analyses are not identical (sample sizes are reported with results of each statistical analysis). Furthermore, criteria for social and personal development were not uniform across grade levels. These different criteria were pooled together and one composite score was computed as a proxy for grade appropriate measures of social and personal development.

Table 1 gives an overview of the number of *Wings* and *non-Wings* students in each grade level. Note that every student did not have a grader for every subject; therefore the exact number of students in each statistical analysis deviates from the total number of students for whom report cards were available.

Table 1

Overview of the number of *Wings* and *non-Wings* students in each grade level

Grade	<i>non-Wings</i>	<i>Wings</i>
1	26	16
2	26	16
3	24	14
4	22	13
5	27	16
6	34	7
Total	159	82

II. Results of statistical analyses

In this section we present the results of statistical analyses comparing academic achievement of *Wings* and *non-Wings* students. The analyses address the following sets of questions:

- (1) Is there a difference in academic outcomes for students who attended *Wings* and those who did not attend *Wings* during academic year 2005/2006? Which areas of academic and social and personal development show greatest differences between *Wings* and *non-Wings* students?
- (2) Does academic achievement change between successive academic years? Are there differences in the rate of change for students who attended *Wings* in 2005/2006 and those who did not attend *Wings* during this academic year? Is there a significant change in academic achievement for students with different patterns of involvement in *Wings* during 2004/2005 and 2005/2006 (e.g., attending *Wings* during one academic year and not the other)?

1. Differences between *Wings* and *non-Wings* students during 2005/2006 academic year

Grades at the end of the 2005/2006 academic year were compared for *Wings* and *non-Wings* students in three sets of analyses: (1) The number of students receiving each letter grade (for students graded on the three-point scale), (2) Differences in mean grade-points, and (3) Differences in grades for students in different grade levels.

The first set of analyses examined differences between *Wings* and *non-Wings* students who were graded on a three-point qualitative scale. Students in first and second grades were evaluated to indicate if their performance ‘rarely’, ‘sometimes’, or ‘consistently’ showed the

targeted skills. Most students in both *non-Wings* and *Wings* groups received two grades indicating partial or full success in meeting academic challenges (see Table 2). There were no significant differences in student performance in the two groups. It should be noted that the three-point scale may not be sensitive to differences between students. For example, we do not know whether a student who ‘sometimes’ shows a skill does that 60, 70, or 80% of the time. A stronger test of the difference in academic performance would employ a more sensitive grading scale. Students in grades three through six were graded on a percentile point scale which offers an opportunity for such analysis.

Comparing the mean grade points for *non-Wings* and *Wings* students in grades three through six, statistically significant differences were found between the two groups in art and physical education (see Table 3). For both subjects, *Wings* students received higher grades. For art, *non-Wings* students received an average grade of 93.54 and *Wings* students received an average grade of 95.70. For physical education, *non-Wings* students received an average grade of 94.21 and *Wings* students received an average grade of 94.82. While these differences in mean grade-points were statistically significant, they are not practically significant. Mean grades for both groups of students indicate a performance in the same high-level category; on a traditional five-point letter scale, grades for both groups would be As. Statistically significant differences were not obtained for any other academic subjects.

Information on the number of days students were absent or tardy was available for students in all grade levels. A statistically significant difference between *non-Wings* and *Wings* students was obtained for the number of days students were absent from school. As shown in Table 3, *non-Wings* students were on average absent 8.02 days and *Wings* students were on average absent 4.23 days. In other words, *Wings* students on average spend approximately four

days more in school than *non-Wings* students, which may be a difference that has practical significance.

Table 2

Number of *non-Wings* and *Wings* students with different grades in different academic subjects: three-point grading scale

		<i>non-Wings</i>		<i>Wings</i>		Total		
		Count	Expected	Count	Expected	Count	Expected	
Reading	Rarely	3	1.8	0	1.2	3	3	$\chi^2(2) = 1.994, p = 0.369$
	Sometimes	16	16.5	11	10.5	27	27	
	Consistently	31	31.7	21	20.3	52	52	
	Total	50	50	32	32	82	82	
Language arts	Rarely	3	2.5	1	1.5	4	4	$\chi^2(2) = 0.292, p = 0.589$
	Sometimes	22	23.3	16	14.7	38	38	
	Consistently	26	25.2	15	15.8	41	41	
	Total	51	51	32	32	83	83	
Social studies	Sometimes	12	11.7	7	7.3	19	19	$\chi^2(1) = 0.030, p = 0.861$
	Consistently	39	39.3	25	24.7	64	64	
	Total	51	51	32	32	83	83	
Science	Sometimes	12	12.9	9	8.1	21	21	$\chi^2(1) = 0.220, p = 0.639$
	Consistently	39	38.1	23	23.9	62	62	
	Total	51	51	32	32	83	83	
Math	Rarely	3	2.5	1	1.5	4	4	$\chi^2(2) = 0.338, p = 0.845$
	Sometimes	18	18.4	12	11.6	30	30	
	Consistently	30	30.1	19	18.9	49	49	
	Total	51	51	32	32	83	83	
Health	Consistently	24	24	15	15	39	39	

		<i>non-Wings</i>		<i>Wings</i>		Total		
		Count	Expected	Count	Expected	Count	Expected	
Art	Consistently	35	35	20	20	55	55	
Music	Sometimes	11	9.5	4	5.5	15	15	$\chi^2 (1) = 0.838, p = 0.360$
	Consistently	24	25.5	16	14.5	40	40	
	Total	35	35	20	20	55	55	
Physical ed	Consistently	35	35	20	20	35	20	

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two

groups. A close match between Count and Expected would indicate a non-significant finding.

Table 3

Differences between *non-Wings* and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Reading	N	106	50	0.19	154	0.850
	M	83.92	83.68			
	SD	7.55	7.46			
Language arts	N	113	57	-0.02	168	0.984
	M	85.12	85.14			
	SD	7.64	8.16			
Social studies	N	114	57	-0.86	169	0.390
	M	85.43	86.47			
	SD	7.76	6.81			
Science	N	114	57	-1.58	169	0.117
	M	87.71	87.49			
	SD	7.33	6.14			
Math	N	113	57	-0.63	168	0.525
	M	83.71	84.49			
	SD	7.61	7.51			
Health	N	99	49	-0.90	146	0.368
	M	87.04	88.31			
	SD	8.15	7.73			
Art	N	99	47	-1.79	144	0.075
	M	93.54	95.70			
	SD	6.96	6.40			
Music	N	99	47	-0.49	144	0.627
	M	87.63	88.34			
	SD	8.15	8.58			
Physical ed	N	106	50	-2.73	121.24	0.007
	M	94.21	94.82			
	SD	2.22	0.44			

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Tardiness	N	158	82	-0.52	238	0.604
	M	11.32	10.33			
	SD	14.51	13.17			
Absence	N	158	82	5.58	237.311	<0.001
	M	8.02	4.23			
	SD	6.79	3.72			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Another set of analyses examined differences between *non-Wings* and *Wings* students in different grades. It is possible that participation in Wings has differential effect on students in different grades and that significant differences between *non-Wings* and *Wings* students can therefore be obtained for only some grades, but not others.

Tables 4 to 11 show performance of *non-Wings* and *Wings* students on criteria for academic success in first through sixth graders. Grade-specific differences between students in *non-Wings* and *Wings* groups were observed for music and physical education. Fifth grade *Wings* students had higher grades in music than *non-Wings* students. On average, *non-Wings* students received 88.95 grade-points and *Wings* students received 93.46 grade-points in music. On a five-point scale, *non-Wings* students received an average B grade and *Wings* students received an average A grade in music. Sixth grade *Wings* students had significantly higher grades in physical education than *non-Wings* students. *Wings* students received 94.57 points, while *non-Wings* students received 93.36 points in physical education. This difference is statistically significant, but very small and not practically important. Both groups of students received an equivalent of an average A grade. Overall, significant grade level differences were small and focused in subjects that are somewhat secondary to school success.

Significant differences were found between *non-Wings* and *Wings* students on variables of school attendance for several grades. First grade *Wings* students were significantly less tardy than *non-Wings* students. On average, *Wings* students were tardy 5.12 days and *non-Wings* students were tardy 12.56 days. Furthermore, significant differences in the number of absences between *non-Wings* and *Wings* students were found for first, fourth, and sixth graders. *Wings* students were present in school for average of three to seven days more than *non-Wings* students. These differences may be of practical significance.

The final set of statistical analyses compared *non-Wings* and *Wings* students on their grades for social and personal development. Table 12 shows the number of students who received each grade. Most students received the two highest grades, indicating that they show target skills at least some of the time. Overall, there was no statistically significant difference between *non-Wings* and *Wings* students; 94.4% of *non-Wings* students received the two highest grades and 97.3% of *Wings* students received the two highest grades. When comparing *non-Wings* and *Wings* students at each grade level, a statistically significant difference was found for second grade students. The highest grade was received by 42.3% of *non-Wings* students and by 81.2% of *Wings* students. There was no difference between the two groups in any other grade levels. In order to make reliable conclusions about the reason for these differences, this difference should be replicated in a different school or during a different academic year. If this difference could be replicated, it is possible that the *Wings* curriculum contributes to development of skills that are particularly important for second grade students or that *Wings* uses teaching methods that are particularly appealing or effective with second grade students.

Table 4

Number of **first grade** *non-Wings* and *Wings* students with different grades in different academic subjects: three-point grading scale

		<i>non-Wings</i>		<i>Wings</i>		Total		
		Count	Expected	Count	Expected	Count	Expected	
Reading	Sometimes	8	7.9	5	5.1	13	13	$\chi^2(1) = 0.003, p = 0.960$
	Consistently	17	17.1	11	10.9	28	28	
	Total	25	25	16	16	41	41	
Language arts	Sometimes	12	12.4	8	7.6	20	20	$\chi^2(1) = 0.059, p = 0.808$
	Consistently	14	13.6	8	8.4	22	22	
	Total	26	26	16	16	42	42	
Social studies	Sometimes	1	1.2	1	0.8	2	2	$\chi^2(1) = 0.126, p = 0.722$
	Consistently	25	24.8	15	15.2	40	40	
	Total	26	26	16	16	42	42	
Science	Sometimes	1	1.9	2	1.1	3	3	$\chi^2(1) = 1.118, p = 0.290$
	Consistently	25	24.1	14	14.9	39	39	
	Total	26	26	16	16	42	42	
Math	Sometimes	7	7.4	5	4.6	12	12	$\chi^2(1) = 0.091, p = 0.763$
	Consistently	19	18.6	11	11.4	30	30	
	Total	26	26	16	16	42	42	
Health	Consistently	24	24	15	15	39	39	
Art	Consistently	10	4	10	4	10	4	

		<i>non-Wings</i>		<i>Wings</i>		Total	
		Count	Expected	Count	Expected	Count	Expected
Music	Consistently	10	4	10	4	10	4
Physical ed	Consistently	10	4	10	4	10	4

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two

groups. A close match between Count and Expected would indicate a non-significant finding.

Table 5

Differences between **first grade** *non-Wings* and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Tardiness	N	25	16	2.79	31.63	0.009
	M	12.56	5.12			
	SD	12.30	4.13			
Absence	N	25	16	4.25	37.76	<0.001
	M	8.80	3.56			
	SD	5.16	2.68			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Table 6

Number of **second grade** *non-Wings* and *Wings* students with different grades in different academic subjects: three-point grading scale

		<i>non-Wings</i>		<i>Wings</i>		Total		
		Count	Expected	Count	Expected	Count	Expected	
Reading	Not Yet	3	1.8	0	1.2	3	3	$\chi^2(2) = 2.077, p = 0.354$
	In Progress	8	8.5	6	5.5	14	14	
	Proficient	14	14.6	10	9.4	24	24	
	Total	25	25	16	16	41	41	
Language arts	Not Yet	3	2.4	1	1.6	4	4	$\chi^2(2) = 0.591, p = 0.744$
	In Progress	10	11.0	8	7.0	18	18	
	Proficient	12	11.6	7	7.4	19	19	
	Total	25	25	16	16	41	41	
Social studies	In Progress	11	10.4	6	6.6	17	17	$\chi^2(1) = 0.170, p = 0.680$
	Proficient	14	14.6	10	9.4	24	24	
	Total	25	25	16	16	41	41	
Science	In Progress	11	11.0	7	7.0	18	18	$\chi^2(1) = 0.000, p = 0.987$
	Proficient	14	14.0	9	9.0	23	23	
	Total	25	25	16	16	41	41	
Math	Not Yet	3	2.4	1	1.6	4	4	$\chi^2(2) = 0.407, p = 0.816$
	In Progress	11	11.0	7	7.0	18	18	
	Proficient	11	11.6	8	7.4	19	19	
	Total	25	25	16	16	41	41	

Art	Proficient	25	16	25	16	25	16	
		<i>non-Wings</i>		<i>Wings</i>		Total		
		Count	Expected	Count	Expected	Count	Expected	
Music	In Progress	11	9.1	4	5.9	15	15	$\chi^2(1) = 1.518, p = 0.218$
	Proficient	14	15.9	12	10.1	26	26	
Physical ed	Proficient	25	16	25	16	25	16	

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two

groups. A close match between Count and Expected would indicate a non-significant finding.

Table 7

Differences between **second grade non-Wings** and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Tardiness	N	26	16	-1.54	40	0.130
	M	9.50	17.12			
	SD	14.21	17.50			
Absence	N	26	16	1.60	40	0.117
	M	7.42	4.62			
	SD	5.99	4.57			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Table 8

Differences between **third grade non-Wings** and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Reading	N	24	14	-0.01	36	0.996
	M	86.42	86.43			
	SD	7.18	4.52			
Language arts	N	24	14	0.24	36	0.815
	M	86.17	85.64			
	SD	7.70	4.01			
Social studies	N	24	14	-0.10	36	0.919
	M	90.50	90.71			
	SD	6.92	4.87			
Science	N	24	14	-0.46	36	0.652
	M	89.96	91.00			
	SD	7.48	5.41			
Math	N	24	14	-0.80	35.97	0.431
	M	88.67	90.07			
	SD	6.81	4.05			
Health	N	23	14	-0.15	35	0.881
	M	90.61	90.93			
	SD	6.91	4.92			
Art	N	24	14	-0.01	36	0.994
	M	98.21	98.21			
	SD	2.52	2.42			
Music	N	24	14	0.62	36	0.538
	M	88.83	86.93			
	SD	8.03	10.73			
Physical ed	N	24	14	-0.06	36	0.952
	M	94.83	94.86			
	SD	1.43	0.36			

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Tardiness	N	24	14	1.32	36	0.196
	M	7.42	4.07			
	SD	8.72	4.83			
Absence	N	24	14	1.60	36	0.117
	M	5.83	4.40			
	SD	3.64	3.36			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Table 9

Differences between **fourth grade** *non-Wings* and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Reading	N	22	13	0.32	33	0.754
	M	80.14	79.15			
	SD	8.54	9.43			
Language arts	N	22	13	0.68	33	0.504
	M	83.82	81.54			
	SD	9.24	10.35			
Social studies	N	22	13	0.35	33	0.731
	M	84.23	83.38			
	SD	7.37	6.17			
Science	N	22	13	-0.48	33	0.633
	M	83.64	84.85			
	SD	7.57	6.43			
Math	N	22	13	0.33	33	0.743
	M	80.95	79.92			
	SD	8.88	8.94			
Health	N	22	13	0.25	33	0.802
	M	82.64	82.00			
	SD	7.15	7.32			
Art	N	22	13	0.14	33	0.891
	M	96.41	96.15			
	SD	6.06	3.53			
Music	N	22	13	0.50	33	0.620
	M	85.73	84.23			
	SD	9.56	6.42			
Physical ed	N	22	13	-0.62	33	0.542
	M	94.59	94.92			
	SD	1.92	0.28			

		non-Wings	Wings	t	df	p
Tardiness	N	22	13	0.47	33	0.644
	M	12.09	9.77			
	SD	15.71	11.14			
Absence	N	22	13	1.83	33	0.076
	M	6.95	3.62			
	SD	6.03	3.36			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Table 10

Differences between **fifth grade** *non-Wings* and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Reading	N	27	16	0.16	41	0.874
	M	84.04	83.69			
	SD	6.73	7.35			
Language arts	N	27	16	-0.56	41	0.576
	M	86.37	87.75			
	SD	6.37	9.70			
Social studies	N	27	16	-0.31	41	0.759
	M	85.33	86.06			
	SD	7.30	7.77			
Science	N	27	16	-1.51	41	0.140
	M	86.96	89.50			
	SD	5.48	5.08			
Math	N	27	16	-1.07	41	0.290
	M	82.78	84.94			
	SD	5.94	7.09			
Health	N	21	15	-0.74	34	0.465
	M	94.67	95.00			
	SD	1.02	1.69			
Art	N	19	13	-1.49	30	0.146
	M	96.26	99.08			
	SD	6.19	3.33			
Music	N	19	13	-1.80	30	0.082
	M	88.95	93.46			
	SD	6.99	6.92			
Physical ed	N	27	16	-0.78	41	0.441
	M	94.37	94.81			
	SD	2.24	0.40			

		non-Wings	Wings	t	df	p
Tardiness	N	27	16	-0.83	41	0.409
	M	12.70	16.81			
	SD	14.91	16.76			
Absence	N	27	16	1.01	41	0.314
	M	8.67	8.98			
	SD	6.25	3.82			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Table 11

Differences between **sixth grade non-Wings** and *Wings* students: mean grade points

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Reading	N	33	7	-0.71	38	0.479
	M	84.54	86.57			
	SD	7.10	5.06			
Language arts	N	33	7	-0.58	38	0.563
	M	85.15	86.86			
	SD	7.34	5.01			
Social studies	N	34	7	-0.32	39	0.751
	M	82.20	83.28			
	SD	8.04	8.60			
Science	N	34	7	0.73	39	0.468
	M	82.62	80.43			
	SD	7.52	5.00			
Math	N	33	7	0.85	38	0.399
	M	82.91	80.28			
	SD	7.62	5.99			
Health	N	33	7	0.74	38	0.462
	M	82.64	80.43			
	SD	7.48	5.00			
Art	N	34	7	1.54	39	0.132
	M	86.88	83.57			
	SD	4.90	6.55			
Music	N	34	7	-0.64	39	0.529
	M	87.26	89.28			
	SD	7.92	6.10			
Physical ed	N	33	7	-2.20	33.43	0.035
	M	93.36	94.57			
	SD	2.66	0.79			

		<i>non-Wings</i>	<i>Wings</i>	t	df	p
Tardiness	N	34	7	1.05	39	0.300
	M	12.97	5.43			
	SD	18.28	10.24			
Absence	N	34	7	2.25	39	0.030
	M	9.62	2.57			
	SD	8.04	3.91			

Note. N – number of students in each group for whom data was available; M – mean grade

points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance

of mean differences in scores between two groups; df – degrees of freedom for the statistical test;

p – probability that statistical test indicates no significant difference between groups

Table 12

Number of *non-Wings* and *Wings* students with different grades in social and personal development

		non-Wings		Wings		Total		
		Count	Expected	Count	Expected	Count	Expected	
Total sample	Rarely	7	5.6	2	3.4	9	9	$\chi^2(2) = 1.152, p = 0.562$
	Sometimes	59	58.1	34	34.9	93	93	
	Consistent	59	61.3	39	36.8	98	98	
	Total	125	125	75	75	200	200	
Grade 1	Sometimes	9	10.5	8	6.5	17	17	$\chi^2(1) = 0.973, p = 0.324$
	Consistent	17	15.5	8	9.5	25	25	
	Total	26	26	16	16	42	42	
Grade 2	Sometimes	15	11.1	3	6.9	18	18	$\chi^2(1) = 6.133, p = 0.013$
	Consistent	11	14.9	13	9.1	24	24	
	Total	26	26	16	16	42	42	
Grade 3	Rarely	2	1.3	0	0.7	2	2	$\chi^2(2) = 1.331, p = 0.514$
	Sometimes	9	8.8	5	5.2	14	14	
	Consistent	13	13.9	9	8.1	22	22	
	Total	24	24	14	14	38	38	
Grade 4	Rarely	4	3.8	2	2.2	6	6	$\chi^2(2) = 0.391, p = 0.822$
	Sometimes	13	13.8	9	8.2	22	22	
	Consistent	5	4.4	2	2.6	7	7	
	Total	22	22	13	13	35	35	

		non-Wings		Wings		Total		
		Count	Expected	Count	Expected	Count	Expected	
Grade 5	Rarely	1	0.6	0	0.4	1	1	$\chi^2(2) = 0.763, p = 0.683$
	Sometimes	13	13.8	9	8.2	22	22	
	Consistent	13	12.6	7	7.4	20	20	
	Total	27	27	16	16	43	43	

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two groups. A close match between Count and Expected would indicate a non-significant finding.

2. Longitudinal analyses

Longitudinal analyses examined changes in academic success for students in *Wings* and *non-Wings* groups between academic years 2004/05 and 2005/06. Three sets of analyses were performed: (1) Examining grade-point changes in students who were not in *Wings* and those who were in *Wings* in 2005/06; (2) Comparing patterns of performance between students who were in *Wings* and those who were not in *Wings* in 2005/06; and (3) Examining performance in 2005/06 for students with different patterns of participation in *Wings* during two successive years.

2.1. Change in grade-points

This set of analyses asked whether *non-Wings* and *Wings* students show the same changes in grades between 2004/05 and 2005/06. Analyses were performed separately for students who attended *Wings* during 2005/06 and who did not attend *Wings* during 2005/06. These analyses were performed on students for whom information were available for both years¹. Table 13 shows the average grade-points for *non-Wings* and *Wings* students in 2004/05 and 2005/06 and presents results of statistical tests comparing significance of mean differences in grade-points.

Significant differences in average grade-points were found for several academic subjects and attendance variables, including science, health, art, music, and tardiness. In all cases and for both *non-Wings* and *Wings* students, performance declined from 2004/05 to 2005/06. Also, in most cases the observed declined was similar in *non-Wings* and *Wings* groups.

¹ Note that the sample sizes will be smaller for these analyses than when comparing *non-Wings* and *Wings* students during one academic year. For example, students who were in first or second grade during 2004/05 would not be included in these analyses because they were graded on a different scale, and students who were in sixth grade in 2004/05 would not be included in these analyses because they would have moved to middle school in 2005/06 and their data would not be available.

Non-Wings students had significantly higher scores in science in 2004/05 (average 88.72 grade points) than in 2005/06 (average 85.16 grade-points). Both mean grades indicate B-level performance and it is not clear whether this difference in grade-points has substantial practical significance. The difference in performance of *Wings* students was not statistically significant, although the mean scores were similar to those of *non-Wings* students (87.34 for 2004/05 and 85.90 for 2005/06). Similar results were obtained for art. *Non-Wings* students had significantly higher scores in art for academic year 2004/05 than for 2005/06. Although the difference in scores was statistically significant, scores for both academic years indicated A-level grades (94.95 for 2004/05 and 92.69 for 2005/06) and therefore the difference should not be considered practically significant. *Wings* students had similarly high scores, but the difference between the two academic years was not statistically significant (average 94.48 grade-points for 2004/05 and 93.15 for 2005/06).

There was a significant change in students' mean grade-points between 2004/05 and 2005/06 for health, music, and tardiness. Both *non-Wings* and *Wings* groups were significantly more successful in 2004/05, suggesting that the changes in grades were not due to the effects of the intervention program but to some other variable such as developmental changes in student motivation and behavior. Grade point change in health and music indicated a decline in performance from A-level to B-level. For health, *non-Wings* students had an average grade of 92.28 in 2004/05 and 87.00 in 2005/06 and *Wings* students had an average grade of 91.48 in 2004/05 and 86.41 in 2005/06. Similarly, for music, *non-Wings* students had an average grade of 92.26 in 2004/05 and 89.10 in 2005/06 and *Wings* students had an average grade of 92.20 in 2004/05 and 88.85 in 2005/06. The average number of days students were tardy increased significantly between the two academic years. *Non-Wings* students were on average tardy 7.88

days in 2004/05 and 10.38 days in 2005/06 and *Wings* students were on average tardy 6.95 days in 2004/05 and 10.60 days in 2005/06.

Table 13

Mean differences in grade points between academic years 2004/2005 and 2005/2006

		<i>Non-Wings</i>					<i>Wings</i>				
		2004/05	2005/06	t	df	p	2004/05	2005/06	t	df	p
Reading	N	56	56	1.55	55	0.127	29	29	1.20	28	0.240
	M	82.73	84.16				81.28	83.10			
	SD	7.69	6.42				8.99	8.13			
Language arts	N	56	56	1.18	55	0.243	29	29	0.81	28	0.424
	M	85.41	86.41				83.31	84.52			
	SD	8.48	6.58				9.80	10.24			
Social studies	N	57	57	-1.25	56	0.216	29	29	-0.25	28	0.805
	M	85.81	84.79				85.07	84.76			
	SD	7.26	7.49				7.28	7.63			
Science	N	57	57	-3.29	56	0.002	29	29	-0.94	28	0.357
	M	88.72	85.16				87.34	85.90			
	SD	6.61	6.27				6.22	6.63			
Math	N	56	56	0.20	55	0.841	29	29	-0.79	28	0.425
	M	82.89	83.07				83.07	81.86			
	SD	8.62	6.49				8.35	8.46			
Health	N	57	57	-3.67	56	0.001	29	29	-2.71	28	0.011
	M	92.28	87.00				91.48	86.41			
	SD	6.66	7.45				6.48	8.56			

		<i>Non-Wings</i>					<i>Wings</i>				
		2004/05	2005/06	t	df	p	2004/05	2005/06	t	df	p
Art	N	39	39	-1.91	38	0.064	20	20	-0.97	19	0.342
	M	94.95	92.69				94.95	93.15			
	SD	0.32	7.47				0.22	8.37			
Music	N	39	39	-2.83	38	0.007	20	20	-2.18	19	0.042
	M	92.26	89.10				92.20	88.85			
	SD	2.35	7.26				2.35	7.62			
Physical ed	N	48	48	-1.29	47	0.204	23	23	0.19	22	0.852
	M	94.50	94.06				94.70	94.74			
	SD	1.40	2.22				1.10	0.54			
Tardiness	N	95	95	2.07	94	0.041	55	55	2.35	54	0.022
	M	7.88	10.38				6.95	10.69			
	SD	9.98	13.56				7.61	14.36			
Absence	N	40	40	1.25	39	0.219	26	26	-1.11	25	0.276
	M	4.98	5.78				5.15	4.12			
	SD	3.87	4.43				5.72	4.08			

Note. N – number of students in each group for whom data was available; M – mean grade points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance of mean differences in scores between two groups; df – degrees of freedom for the statistical test; p – probability that statistical test indicates no significant difference between groups

2.1. Change in patterns of academic success

How many *non-Wings* and *Wings* students show stable underperformance, decline, improvement, and stable success between 2004/05 and 2005/06? We defined success as receiving a B grade or higher and underperformance as receiving a C grade or lower. Comparing student performance in the two academic years, we defined the patterns of academic success as follows: (1) stable underperformance: receiving a C or lower for both years; (2) decline: receiving a B grade or higher in 2004/05 and C or lower in 2005/06; (3) improvement: receiving a C or lower in 2004/05 and B or higher in 2005/06; and (4) stable success: receiving a B grade or higher for both academic years. Stable underperformance and decline categories can be more generally considered to indicate failure, while improvement and stable success categories can be considered to indicate academic success. This set of analyses can be very informative as it compares differences in grades that have immediate practical importance.

Patterns of student success differed across academic subjects (see Tables 14 through 19). For reading, *non-Wings* students were relatively evenly distributed among the four categories of academic performance (25% showed stable success and 23.3% improved; overall 48.2% of successful students). The percentage of successful students was similar in *Wings* student: 31.1% of students showed stable success and 10.3% improved (overall of 41.4% of students in successful categories). However, most *Wings* students showed stable underperformance (51.7% students).

Most *non-Wings* students were successful in English language arts: 50% showed stable success and 14.3% improved (overall 64.3% of students in successful categories). On the other hand, *Wings* students showed a pattern of either stable underperformance (48.3% of students) or stable success (44.8% of students). Only 3.4% of *Wings* students improved between the two

academic years (overall, 48.2% of successful students). A similar pattern was observed for grades in social studies. Most students in the *non-Wings* group showed stable success (45.6% of students), while most *Wings* students showed either stable underperformance (31.0% of students) or stable success (37.8% of students). However, the overall number of successful students in the two groups was similar (52.6% of *non-Wings* students and 51.7% of *Wings* students).

Most students in both groups were successful in science; 50.9% of *non-Wings* students showed stable success and 7.0% improved (57.9% success), and 41.4% of *Wings* students showed stable success and 17.2% improved (58.6% success). While the two groups do not appear to differ on the overall rate of success, more *Wings* students showed improvement and more *non-Wings* students showed stable success. On the other hand, most students in both groups showed underperformance in math; 41.1% of *non-Wings* students and 41.4% of *Wings* students performed at C or lower level in both academic years. Less than half of the students in both groups were successful; among *non-Wings* students only 28.6% showed stable success and 14.3% improved (42.9% success) and among *Wings* students only 27.6% showed stable success and 10.3% improved (37.9% students).

Most students in both *non-Wings* and *Wings* groups showed stable success in performance in health; 50.9% of *non-Wings* students and 48.3% of *Wings* students received B grade or higher in both academic years. An additional 8.8% of *non-Wings* students improved (59.6% success) and 6.9% of *Wings* students improved (55.2% success). The second largest category of students showed decline; 35.1% of *non-Wings* students and 31.0% of *Wings* students received lower grades in 2005/06 than in 2004/05.

Grades in art, music, and physical education showed minimal or no variation between the two academic years. Both groups of students showed stable success or improvement in these

subjects (only one student in the *non-Wings* group showed stable underperformance in music). Grades in these subjects are generally quite skewed towards high grades and not very sensitive to differences among students.

Students' social and personal development was evaluated on a qualitative three-point scale. Table 20 presents number of students in *non-Wings* and *Wings* groups who received different grades during 2004/05 and 2005/06. Examination of student grades shows most prominently that few students are judged to be failing in social and personal development. In the *non-Wings* group, only 6.6% of students rarely showed target skills in 2004/05 and only 5.3% of students rarely showed target skills in 2005/06. In the *Wings* group, 4.2% of students rarely showed target skills in 2004/05 and 12.5% students rarely showed target skills in 2005/06. Also, most students in both groups – 34.2% of *non-Wings* and 37.5% of *Wings* students – showed stable high level of social and personal development skills (i.e., consistently showing desired behaviors). An additional 19.7% of *non-Wings* and 16.7% of *Wings* students showed the target skills some times during both academic years.

A difference between the two groups emerged when comparing the number of students whose grades declined or improved. Decline was defined as decrease in grades for at least one category (e.g., from consistently showing social and personal development skills in 2004/05 to showing them sometimes in 2005/06) and improvement was defined as an increase in grades for at least one category (e.g., from rarely showing social and personal development skills in 2004/05 to showing them sometimes in 2005/06). More *Wings* than *non-Wings* students showed improvement (25.0% of *Wings* versus 13.2% of *non-Wings* students). Also, fewer *Wings* than *non-Wings* students showed decline (20.8% *Wings* versus 30.3% *non-Wings* students). Of note, it was not possible to conduct a statistic test of these differences because there were too few

students in most categories. This informal analysis offers suggestions of the effects Wings can have on students' social and personal development. These effects primarily consist of buffering students from decline in their skills and behavior and helping children improve from the level at which they entered the program.

Table 14

Change in reading grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06

		2004/05						
		<i>Non-Wings</i>			<i>Wings</i>			
		C or lower	B or higher	Total	C or lower	B or higher	Total	
2005/06	C or lower	Count	19 ^a	10 ^b	29	15 ^a	2 ^b	17
		Expected	16.6	12.4	29	10.6	6.4	17
	B or higher	Count	13 ^c	14 ^d	27	3 ^c	9 ^d	12
		Expected	15.4	11.6	27	7.4	4.6	12
	Total	Count	32	24	56	18	11	29
		Expected	32	24	56	18	11	29

$$\chi^2(1) = 1.72, p = 0.189$$

$$\chi^2(1) = 11.95, p = 0.001$$

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline; ^c – improvement; ^d – stable success

Table 15

Change in English language arts grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06

		2004/05						
		<i>Non-Wings</i>			<i>Wings</i>			
		C or lower	B or higher	Total	C or lower	B or higher	Total	
2005/06	C or lower	Count	14 ^a	6 ^b	20	14 ^a	1 ^b	15
		Expected	7.9	12.1	20	7.8	7.2	15
	B or higher	Count	8 ^c	28 ^d	36	1 ^c	13 ^d	14
		Expected	14.1	21.9	36	7.2	6.8	14
	Total	Count	22	34	56	15	14	29
		Expected	22	34	56	15	14	29

$$\chi^2(1) = 12.30, p < 0.001$$

$$\chi^2(1) = 21.54, p < 0.001$$

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline; ^c – improvement; ^d – stable success

Table 16

Change in social studies grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06

		2004/05						
		<i>Non-Wings</i>			<i>Wings</i>			
		C or lower	B or higher	Total	C or lower	B or higher	Total	
2005/06	C or lower	Count	17 ^a	10 ^b	27	9 ^a	5 ^b	14
		Expected	9.9	17.1	27	6.3	7.7	14
	B or higher	Count	4 ^c	26 ^d	30	4 ^c	11 ^d	15
		Expected	11.1	18.9	30	6.7	8.3	15
	Total	Count	21	36	57	13	16	29
		Expected	21	36	57	13	16	29

$$\chi^2(1) = 15.04, p < 0.001$$

$$\chi^2(1) = 4.14, p = 0.042$$

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline; ^c – improvement; ^d – stable success

Table 17

Change in science grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06

		2004/05						
		<i>Non-Wings</i>			<i>Wings</i>			
		C or lower	B or higher	Total	C or lower	B or higher	Total	
2005/06	C or lower	Count	7 ^a	17 ^b	24	5 ^a	7 ^b	12
		Expected	4.6	19.4	24	4.1	7.9	12
	B or higher	Count	4 ^c	29 ^d	33	5 ^c	12 ^d	17
		Expected	6.4	26.6	33	5.9	11.1	17
	Total	Count	11	46	57	10	19	29
		Expected	11	46	57	10	19	29

$$\chi^2(1) = 2.59, p = 0.107$$

$$\chi^2(1) = 0.47, p = 0.494$$

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline; ^c – improvement; ^d – stable success

Table 18

Change in math grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06

		2004/05						
		<i>Non-Wings</i>			<i>Wings</i>			
		C or lower	B or higher	Total	C or lower	B or higher	Total	
2005/06	C or lower	Count	23 ^a	9 ^b	32	12 ^a	6 ^b	18
		Expected	17.7	14.3	32	9.3	8.7	18
	B or higher	Count	8 ^c	16 ^d	24	3 ^c	8 ^d	11
		Expected	13.3	10.7	24	5.7	5.3	11
	Total	Count	31	25	56	15	14	29
		Expected	31	25	56	15	14	29

$$\chi^2(1) = 8.24, p = 0.004$$

$$\chi^2(1) = 4.24, p = 0.039$$

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline; ^c – improvement; ^d – stable success

Table 19

Change in health grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06²

		2004/05						
		<i>Non-Wings</i>			<i>Wings</i>			
		C or lower	B or higher	Total	C or lower	B or higher	Total	
2005/06	C or lower	Count	3 ^a	20 ^b	23	4 ^a	9 ^b	13
		Expected	3.2	19.8	23	2.7	10.3	13
	B or higher	Count	5 ^c	29 ^d	34	2 ^c	14 ^d	16
		Expected	4.8	29.2	34	3.3	12.7	16
	Total	Count	8	49	57	6	23	29
		Expected	8	49	57	6	23	29

$$\chi^2(1) = 0.03, p = 0.859$$

$$\chi^2(1) = 1.46, p = 0.227$$

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two

groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline;

^c – improvement; ^d – stable success

² Statistical test could not be computed to test differences in performance between 2003/04 and 2004/05. Test is meaningful only when the expected frequencies for each category are at least 1 and when no more than 20% of the categories have expected frequencies of less than 5. To maximize the usefulness of the available data and reach at least tentative conclusions, we were liberal about these criteria and present the statistical tests even when they are somewhat violated. When more than 50% of categories have expected frequencies less than 5, the statistical test is not presented.

Table 20

Change in social and personal development grades for *Wings* and *non-Wings* students from 2004/05 to 2005/06

		2004/05								
		<i>Non-Wings</i>				<i>Wings</i>				
		Rarely	Sometimes	Consistently	Total	Rarely	Sometimes	Consistently	Total	
2005/06	Rarely	Count	2 ^a	1 ^b	2 ^b	5	0 ^a	1 ^b	1 ^b	2
		Expected	0.3	1.6	3.2	5	0.3	0.6	1.1	2
	Sometimes	Count	1 ^c	15	20 ^b	36	3 ^c	8	8 ^b	19
		Expected	1.9	11.4	22.7	36	2.4	5.9	10.7	19
	Consistently	Count	1 ^c	8 ^c	26 ^d	35	3 ^c	6 ^c	18 ^d	27
		Expected	1.8	11.1	22.1	35	3.4	8.4	15.2	27
L	Total	Count	4	24	48	76	6	15	27	48
		Expected	4	24	48	76	6	15	27	48
$\chi^2(4) = 15.93, p = 0.003$					$\chi^2(1) = 1.46, p = 0.227$					

Note. Count – number of students in each category; Expected – takes into consideration the difference in the sample sizes of the two

groups. A close match between Count and Expected would indicate a non-significant finding; ^a – stable underperformance; ^b – decline;

^c – improvement; ^d – stable success

2.3. Performance in groups of students with different patterns of participation in Wings

Some students attend Wings during one academic year and stop attending Wings the following year. Other students do not attend Wings one academic year and attend it the following year. Still other students attend Wings for several consecutive years and some never attend Wings. It is possible that academic outcomes could differ for these groups of students. For example, Wings might be effective for students who participate in the program during a current year of assessment even though they did not attend it previously. Alternatively, Wings might be effective only if students attend it for more than one consecutive year. The following set of statistical analyses compares performance of students with four different patterns of participation in Wings: (1) students who attended Wings both in 2004/05 and 2005/06; (2) students who attended Wings in 2004/05, but not in 2005/06; (3) students who did not attend Wings in 2004/05, but attended Wings in 2005/06; and (4) students who did not attend Wings either in 2004/05 or 2005/06.

Table 21 shows mean scores in different academic subjects and attendance variables for students in the four different patterns of Wings participation. There was no statistically significant difference among the four groups on academic grades. However, the two variables of school attendance differed among the groups. Comparisons of all possible pairs of the four groups showed that groups of students who participated in Wings for both academic years and those who participated in Wings only in 2005/06 differed in the number of days they were tardy for school. Students who participated in Wings during both academic years were significantly less tardy (6.53 days) than students who participated in Wings only during 2005/06 (17.43 days). Contrary to expectations, students who attended Wings for one academic year did not show significant difference in relation to students who did not attend Wings. Also, students who

participated in Wings for both academic years were significantly less absent from school (3.38 days) than students who did not attend Wings (7.22 days). Students who participated in Wings for one academic year did not significantly differ from students who participated in Wings during both years or from students who never participated in Wings. These results do not support the hypothesis that Wings contributes to academic success, but suggest that Wings might be related to school participation. The present results suggest that the best outcome is achieved when students participate in Wings for two years.

Table 21

Differences between groups of students with different patterns of participation in Wings during 2004/05 and 2005/06

		<i>Wings</i> 04/05 <i>Wings</i> 05/06	<i>Wings</i> 04/05 <i>Non-Wings</i> 05/06	<i>Non-Wings</i> 04/05 <i>Wings</i> 05/06	<i>Non-Wings</i> 04/05 <i>Non-Wings</i> 05/06	F	df	p
Reading	N	27	18	14	56	0.34	3, 111	0.797
	M	83.48	85.39	85.07	84.64			
	SD	7.55	4.67	7.28	6.98			
Language arts	N	29	20	18	60	0.65	3, 123	0.587
	M	83.90	85.95	85.89	86.22			
	SD	9.05	6.56	8.29	6.77			
Social studies	N	29	20	18	61	0.57	3, 124	0.636
	M	86.31	88.05	87.17	85.75			
	SD	7.50	7.58	6.09	7.39			
Science	N	29	20	18	61	1.00	3, 124	0.394
	M	86.21	86.75	89.22	86.34			
	SD	6.84	6.70	4.77	6.76			
Math	N	29	20	18	60	0.33	3, 123	0.801
	M	83.76	85.65	84.94	84.13			
	SD	8.93	5.66	6.25	7.09			
Health	N	27	18	14	57	0.94	3, 112	0.422
	M	86.11	88.06	90.28	87.81			
	SD	8.65	7.78	5.66	7.46			

		<i>Wings</i> 04/05	<i>Wings</i> 04/05	<i>Non-Wings</i> 04/05	<i>Non-Wings</i> 04/05	F	df	p
		<i>Wings</i> 05/06	<i>Non-Wings</i> 05/06	<i>Wings</i> 05/06	<i>Non-Wings</i> 05/06			
Art	N	26	16	12	51	0.42	3, 101	0.740
	M	94.58	93.50	96.33	94.98			
	SD	7.04	7.14	6.17	6.71			
Music	N	26	16	12	51	0.21	3, 101	0.892
	M	87.62	89.06	86.67	87.82			
	SD	7.92	7.79	10.52	7.88			
Physical ed	N	27	18	14	57	0.63	3, 112	0.595
	M	94.78	94.33	94.93	94.44			
	SD	0.51	1.97	0.27	1.97			
Tardiness	N	34	23	21	71	2.80	3, 145	0.042
	M	6.53	11.91	17.43	10.03			
	SD	8.23	13.60	19.18	14.18			
Absence	N	34	23	21	71	4.00	3, 145	0.009
	M	3.38	6.74	5.90	7.22			
	SD	3.41	4.57	4.62	6.51			

Note. N – number of students in each group for whom data was available; M – mean grade points; SD – standard deviation in grade points; t – value for a t-test statistic testing significance of mean differences in scores between two groups; df – degrees of freedom for the statistical test; p – probability that statistical test indicates no significant difference between groups

III. Conclusions and suggestions for future research

The results of statistical analyses comparing academic performance of *non-Wings* and *Wings* students did not offer support of program effectiveness in relation to academic achievement. Few significant differences were found between the two groups and they focused in areas not central to academic success, such as music or physical education. While these subjects are valuable to overall student development, they are not likely to determine students' success in schooling or level of educational attainment.

On the other hand, statistical analyses suggested that participation in *Wings* might have an effect on student attendance (i.e., absences and tardiness) and development of social and personal skills. There was a trend for *Wings* students to show less tardiness, fewer absences, and greater improvement in social and personal skills than *non-Wings* students. However, it is important to note that the reporting of student absences was not reliable. For example, when adding the number of days students were absent and the number of days students were present, the sum often differs. While we do not have a reason to believe that there was a systematic bias that would favor either *Wings* or *non-Wings* students, it is important to note the unreliability of the data. This stresses the need to replicate the findings presented in this report when making conclusions about effectiveness of *Wings*. We consider the present findings as tentative/preliminary rather than conclusive evidence.

It would be important to continue *Wings* evaluation research. In the following section we offer several suggestions concerning the selection of study participants and measures to be used in future research.

The validity of any evaluation study largely depends on adequate selection of study participants. In an ideal case, participants get selected randomly out of the total pool of students

at a certain school or school district and they get randomly assigned to either the experimental (i.e., *Wings*) or control group (i.e., *non-Wings*). Any pre-existing differences in students' family background or personal skills and behaviors get effectively cancelled out. Any group differences can then be attributed to the intervention program. However, random assignment of students is most often not practically possible. Parents voluntarily sign up their children to participate in *Wings*, which eliminates a chance for random assignment. An alternative approach is to assess *Wings* students before starting the program and select a group of control students with similar key attributes. This procedure is called matching and should be done on basic demographic variables of age and gender and any additional variables that are believed to be important for achieving desired goals. If students in the two groups are not matched, it is not possible to discern whether the observed change in criterion variables was due to initial differences between groups or they are due to the effects of the program. For example, if one desired effect of the program is to improve students' academic performance, it would be important to match *Wings* and *non-Wings* students on their prior academic achievement, intellectual abilities, parents' education, and parents' beliefs about education. While information on students' academic achievement and intellectual abilities can be obtained from official school records, information about parents could be collected when they sign up for *Wings* or during a parent-teacher night. The parent survey can be as short as asking for their education, occupation, and one question on how much they believe that doing well in school is important for success in life.

If the primary goal of *Wings* is to influence students' social and emotional skills, it would be important to test students on these skills before the beginning of the intervention. Target skills should be specifically defined and assessed according to these definitions. For example, one of the stated goals for *Wings* is to improve children's understanding of emotions.

Understanding emotions can be defined in different ways, from an ability to appropriately define emotion words, to an ability to describe circumstances that evoke different emotions, to being able to identify events in one's own life or the lives of significant others when they felt a certain emotion. Which emotions do children understand before starting the program? Which emotions should children be able to understand at different grade levels? Answers to these specific questions will determine appropriate assessments. For some skills, assessments may be age or grade specific (e.g., first graders should understand the meaning of basic emotions of happiness, sadness, fear, anger, surprise, and disgust, while sixth graders should also understand the meaning of disappointment, empathy, and alienation). Other skills could be assessed in an age-neutral way (e.g., whether students disrupt class or push other children).

While many positive outcomes in social and emotional development are correlated and therefore the intervention can have a significant effect on many student attributes, it is advisable to focus at least initial stages of program evaluation on a relatively small number of very specific goals. This procedure would ensure that students could be matched on important characteristics. Based on the results presented in this report where statistically significant differences between *Wings* and *non-Wings* students were observed primarily for variables of tardiness and school absences, it can be hypothesized that *Wings* can have an effect on children's attitudes toward school and motivation for school. The Appendix contains one possible self-report instrument measuring academic motivation. This measure is based on the expectancy/value theory of achievement motivation (Eccles et al., 1983; Eccles & Wigfield, 2002; Midgley, Feldlaufer, & Eccles, 1989; Wigfield, et al., 1997).

The questionnaire was designed to assess motivation in specific academic areas (e.g., math, reading, social studies), but it can be adapted to measure motivation in school more

generally. The survey measures two components of academic motivation: competence beliefs and subjective task value. Competence items ask children how good they are at each subject (e.g., math or reading), how good they are in these activities relative to the other things they do, how good they are relative to other children, how well they expect to do in the future, and how good they think they would be at learning something new in each domain. Subjective task value concerns perceived importance, interest, utility, and cost of working on different academic subjects. These items ask children how interesting or fun each activity is, how important they think being good at the activity is, how useful they think each activity is, and how much putting effort into each academic activity keeps them from other activities.

The leaders of Wings noted that students should not be asked to fill long self-report instruments. They were concerned about validity of these assessments because of fatigue of loss of interests that they observed in children during previous assessments. An alternative form of assessment to self-reports that is often times used in elementary school children is sociometric measurement. In this approach, students are asked to nominate one of several students in their class who have a certain attribute. The attributes can be chosen based on the goals of the program. For example, students can be asked to nominate who is most willing to share a toy or a game, who is most likely to fight or push other kids on the playground, or who is most likely to answer a teacher's question in class. The advantage of this approach is that it can be easily administered and that children readily understand it and can enjoy it. Examination of students' answers can show whether *Wings* students are more likely to be nominated for positive behaviors than *non-Wings* students.

We noticed that the teacher reports of school attendance were not reliable. The total number of days in an academic year should be the same for all children. However, when adding

the number of days students were reported to be present and number of days they were reported to be absent, the sum often times differs. Teachers might not notice this difference when dealing with a large number of students and these discrepancies might not be of great importance on an individual level. On the other hand, these discrepancies can have an important effect on the validity of research results. For example, the results in this report should not be interpreting as indicating that *non-Wings* students are on average absent four days more than *Wings* students because we cannot have confidence in the teacher reports of student attendance. We propose to supplement teacher reports with independent observations. Research assistants blind to student participation in *Wings* can sit in classrooms for one week at several points during a school year. These research assistants can be trained to record student absences, tardiness, student participation (e.g., answering teacher's questions), and any disciplinary actions for disruptive behavior. This approach would offer information based on in-class observations of behavior, which is a more important measure than student self-reports or even teacher evaluations. Furthermore, these observations would not be biased by observer's knowledge of whether a student participates in *Wings* and they would offer an opportunity to follow changes in students' behavior throughout the academic year.

One final suggestion concerns monitoring of students' academic achievement. We believe that research should not substantially rely on academic grades. The analyses presented in this report show that grades are severely skewed. Average grades were rarely below 85 points and they were never below 80. This presents a twofold problem. First, it is much harder to detect any differences between groups because uniformly high grades do not leave enough variation in scores. Second, most students appear to be succeeding, which is not supported by standardized test scores that were available for previous academic years. Because of this, we

suggest to complement examination of student grades with an analysis of standardized test scores and advise that future evaluation research places greater emphasis on test performance than grades. Other measures could include homework completion, class participation, and involvement in extracurricular activities.

In conclusion, we suggest another wave of evaluation studies to examine program effectiveness for Wings. This study should focus on several changes in comparison to previous data collections: (1) have a more careful selection of study participants, (2) select a small number of very specific criteria for program success in areas of social and emotional development and academic achievement (i.e., list most important end results of a successful program), and (3) collect data for at least two years. It would be advisable to do the next set of evaluation studies in a school other than Memminger. While this school has supported the program for the longest period of time and therefore can be expected to be most successful in its implementation, many students would have been involved in Wings prior to the beginning of the evaluation research, which could bias the results.

References

- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spense (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco: W. H. Freeman.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109-132.
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes toward mathematics before and after the transition to junior high school. *Child Development*, *60*, 981-992.
- Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*, 451-469.

Appendix

Questionnaire of student motivation in different academic areas

Children's Motivation for Math, Reading, and Social Science

1. Competence beliefs (repeated for each subject)

- How good in math are you? (not at all good - very good)
- If you were to list all the students in your class from the worst to the best in math where would you put yourself? (one of the worst - one of the best)
- Some kids are better in one subject than in another. For example, you might be better in math than in reading. Compared to most of your other school subjects, how good are you in math? (a lot worse in math than in other subjects - a lot better in math than in other subjects)
- How well do you think you will do on your next math test? (not at all good - very good)
- Compared to other students, how well do you expect to do in math this year? (will be one of the worst – will be one of the best)
- How good would you be at learning something new in math? (not at all good - very good)

2. Task value:

a. Importance

- How important is it to you to get good grades in math? (not at all important - very important)
- I feel that being good at solving problems that involve math is (not at all important - very important)
- It is important to me to learn the course material in this class. (not at all important - very important)
- Understanding subject matter in this course is very important to me. (not at all important - very important)

b. intrinsic interest

- In general, I find working on math assignments (very boring – very interesting)
- How much do you like doing math? (not at all – a lot)

- Do you spend as much time as you do in math (because you have to in order to finish the work OR because you just like doing math)
- Would you take more math if you didn't have to? (a. I very definitely would take more math; b. I probably would take more math; c. Maybe I would take more math; d. I'm not sure; e. Maybe, but not that likely; f. I probably would not take any more math; g. I very definitely would not take any more math)

c. utility value (usefulness)

- In general, how useful is what you learn in math? (not at all useful - very useful)
- How useful is what you learn in math for your daily life outside school? (not at all useful – very useful)
- How useful do you think the math you are learning will be for what you want to do after you graduate and go to work? (not at all useful - very useful)
- Compared to most of your other activities, how useful is what you learn in math? (not at all useful – very useful)

d. cost (academic/activity, social)

- Is the amount of effort it will take to do well in your math course this year worthwhile to you? (not at all worthwhile - very worthwhile)
- How much does the amount of time you spend on math keep you from doing other things you would like to do? (not at all– a lot)

Each question is answered on a 7-point scale except when a dichotomous (either/or) response format is noted. Questions can be reformulated to better suit the purposes of the researcher. For example, instead of asking about specific academic areas, items can be worded to ask about general attitudes about school. Also, it is possible to pick either some components of motivation (e.g., competence beliefs) or only a smaller number of items from each component.