Report on an Outcome Evaluation of the Arrowsmith Program For Treating Learning Disabled Students

Prepared by William J. Lancee, Ph.D.

November 20, 2005

Executive Summary: Empirical Outcome Study of the Arrowsmith Program for Treating Learning Disabled Students at Arrowsmith School

A three year outcome study of the work of the Arrowsmith Program treating 79 children with learning disabilities at Arrowsmith School was generously funded by the Donner Canadian Foundation.

The results were informative and encouraging. The amount of improvement was slightly dependent on intake severity level (the number of performance problem areas on intake). The rate of improvement varied from one year to three years, and was dependent on initial severity. The amount and rate of improvement were not dependent on other baseline characteristics such as age, gender or IQ. Furthermore, the rate of improvement was not dependent on the type of impairment at intake. All deficit areas identified by the Arrowsmith Program improved as a result of the application of Arrowsmith Program cognitive exercises. A specificity of effect was found suggesting that the cognitive exercises could be directly linked to performance improvement. Moreover, students who through specific cognitive exercises improved with respect to AP cognitive functions also improved on related achievement tests.

In the study sample, the cognitive deficits tended to be multi-dimensional, and there was no clear pattern of combinations of deficits. In other words, a given AP student was likely to have more than one deficit and his or her combination tended to be specific to the student.

This study, combined with previous research of the program, strongly supports the effectiveness of the Arrowsmith Program for a wide spectrum of learning problems. These results provide hope for parents and teachers, and open up opportunities for children struggling with learning difficulties.

Outcome Evaluation of the Arrowsmith Program For Treating Learning Disabled Students.

(November 20, 2005, W.J. Lancee)

Introduction:

Learning Disabilities (LD) seriously affect academic and emotional development and are unlikely to remit without specialized intervention. Students with learning disabilities tend to fall farther and farther behind their peers in academic performance and subsequently tend to have a low sense of self-worth. Klein and Mannuza (2000)[1] followed 104 children with LD who initially did not have emotional difficulties. Sixteen years later, these children, when compared to 124 controls, had a much lower status occupational level and continued to struggle with a high prevalence of psychiatric and addiction disorders.

Various special education programs have been developed to address learning disabilities. The approach of the Arrowsmith Program is first to distinguish finely between elemental cognitive impairments and then to implement an individualized task-oriented program that challenges the identified deficit. It is thought that these highly targeted cognitive exercises create ways for the brain to provide the necessary functionality for encoding and decoding spoken and written discourse, and for storing, organizing, processing, and integrating knowledge. If this is successful, the child can rejoin his or her peers in normal academic progress. It should be understood that successful graduates of the Arrowsmith Program will require some time to make up for the learning time that was lost due to the original impairment. The developer of the Arrowsmith Program has high expectations for the successful graduates and believes that they will become academically and occupationally competitive.

Objective:

In 2001, the Donner Canadian Foundation funded a 3 year study that was designed to follow a sample of students attending the Arrowsmith School in Toronto. When data collection was completed (in the fall of 2004), the project had attained extensive performance and skills data over three years for each of 79 students. This sample exceeded the initial target of 40 students.

It was hypothesized that each student would have his or her own specific pathway through the program. First, specific neurocognitive functional deficits would be identified, followed by a tailored program of cognitive exercises. It was postulated that a student who adhered to the program would increase his or her skill in doing the assigned program tasks and have corresponding improvements in academic performance scores.

[1]Klein, R.G. and Mannuza, S. (2000). Children with complicated reading disorders grown up. In L.L. Greenhill (Ed.), *Learning disabilities: Implications for psychiatric treatment.* Washington: American Psychiatric Press.

It should be noted that during the collection of data for the Donner study, a second distinct and non-overlapping one-year study was carried out. The design limitations and small sample size, the study results strongly support the Arrowsmith Program as instrumental in changing the developmental course of the majority of children with LD in this sample".

While this earlier study was supportive of the effectiveness of the program, it was limited by having a one-year duration. The Donner study provides a detailed view of students in the Arrowsmith Program over three years. For example: What happens to students who struggle in the first year of the program? What is the expected rate of improvement for various deficits? Moreover, since the students were at the Arrowsmith School, the data collected on each student was more comprehensive and more complete and the students received more time doing AP cognitive exercises each day.

Study Sample:

All students enrolled in the Arrowsmith School participated in the study. Since study assessments and measures are already routinely carried out as part of the standard program, no additional consent was required. Of the 72 students that enrolled in the AP program in 2001-2003, three were over 20 years old and three were only mildly impaired. These 6 students were excluded from the study sample. There were 13 students who were already in the AP when the study started. These 13 students were also entered into the study (using data from their first year in the school as time 1 data in the study). The total number of students followed was 79 (72-6+13), with a male to female ratio of 2 to 1 (53 male, 26 female). All students in the study sample completed at least one year. The majority of students were in the study for two years (62%), and 23 students (29%) were in the study for 3 years.

When a student graduated from the program he or she exited the study. For comparison across students at the follow-up time points (end of years 2 and 3), a student's last available data point was used (forward propagation of last available data). This strategy makes the (not unreasonable) assumption that students who graduate from the program do not spontaneously lose their improvements. Some students were not finished their program at the completion of the study and continued in the program in the subsequent year. Only data collected during the three years of the study was analyzed.

The majority of the students (81%; 64 students) received six 40 minute periods per day of AP cognitive exercises and one 40 minute period each of English and mathematics instruction while 19% (15 students) received eight 40 minute periods per day of AP cognitive exercises and no academic instruction.

Figures 1 to 2 provide a breakdown of the sample by age and grade.

Figure 1 Distribution by Age

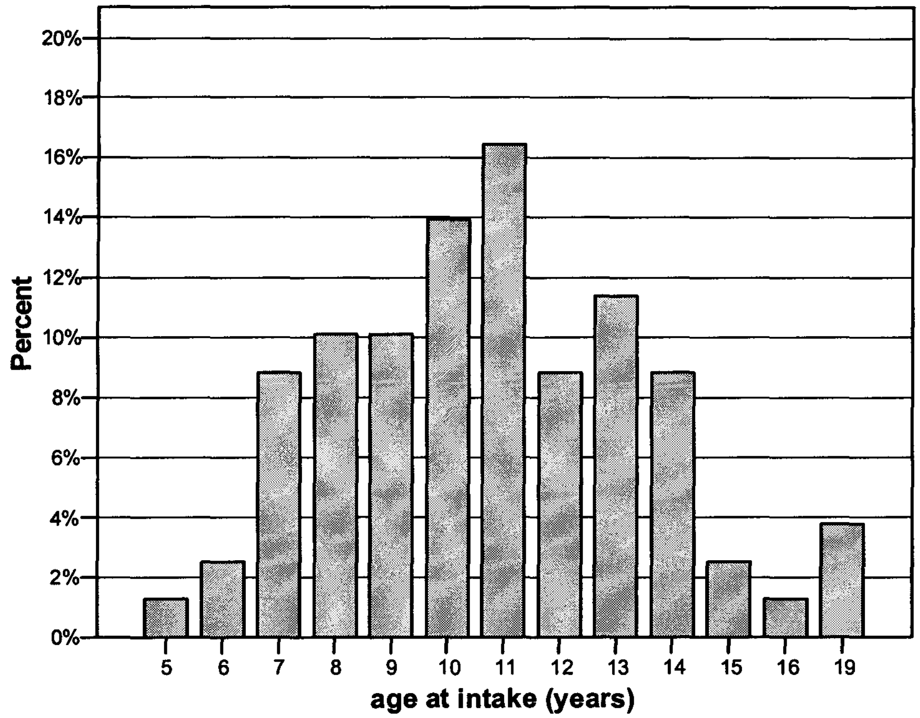


Figure 2 Distribution by Grade

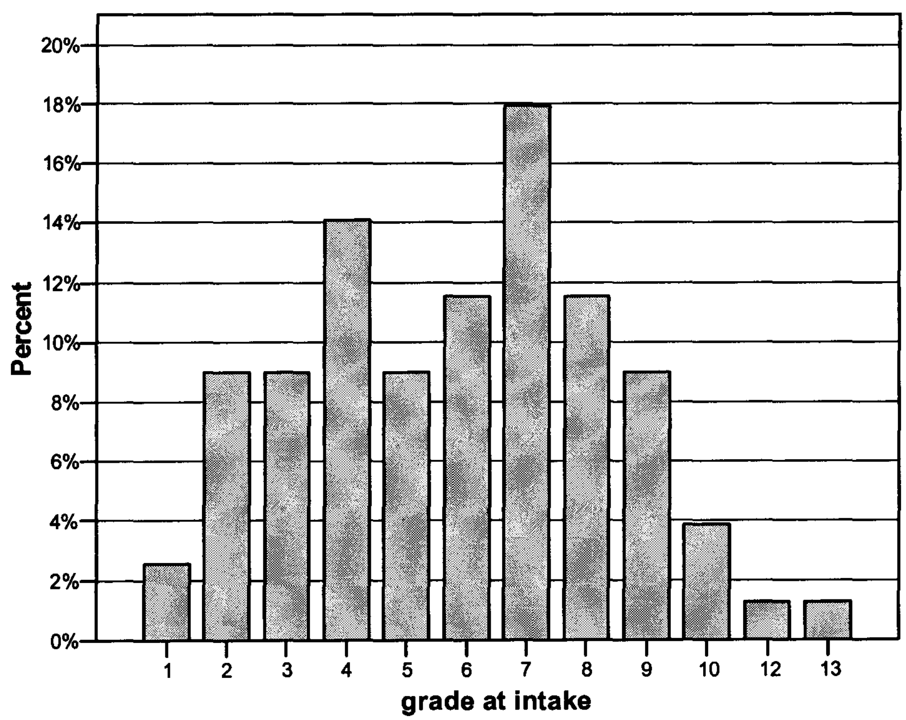


Figure 3 shows the relationship between grade and age in the study sample. As can be seen, for grades 1 to 9, the mean age per grade was as expected in a non-LD sample. Students starting the AP in grades 10 to 13 were older than expected for their grade.

Figure 3 Mean Age by Grade

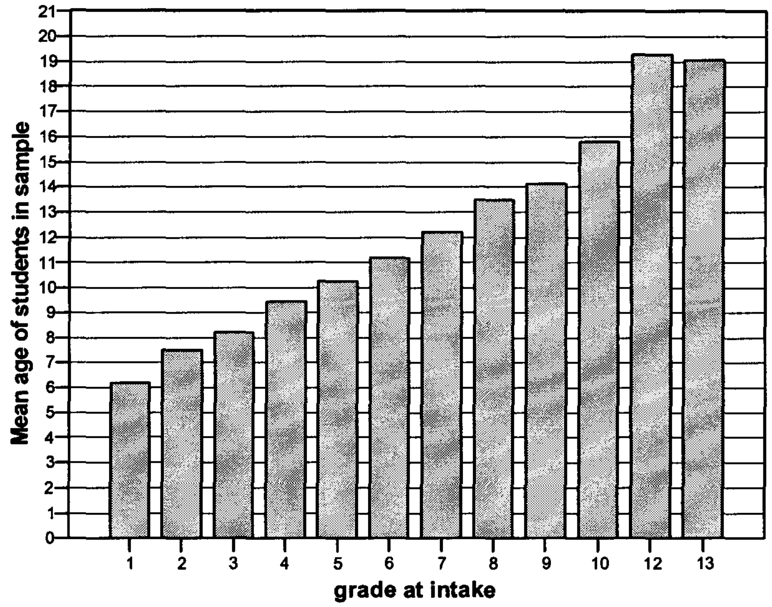


Figure 4 Cumulative distribution of IQ at intake

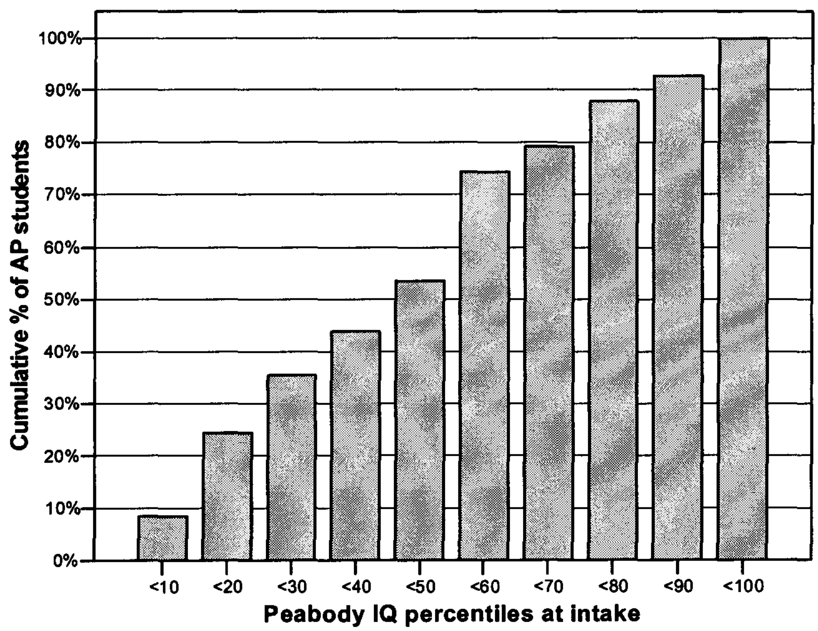


Figure 4 shows that the distribution of IQ in the AP sample is not different from expected, and comparable to non-LD students. Starting IQ scores did not vary significantly with age or gender. Mean IQ score was 97.1 (±14.5).

Study Results

Sample description

Severity at intake was defined in terms of performance on the following standard achievement tests:

(1) Monroe-Sherman Achievement Test - Copying text
(2) Monroe-Sherman Achievement Test - Crossing out letters
(3) Monroe-Sherman Achievement Test - Vocabulary
(4) Monroe-Sherman Achievement Test - Auditory letter memory
(5) Monroe-Sherman Achievement Test - Visual letter memory

(6) Wide Range Achievement Test - Word recognition
(7) Wide Range Achievement Test - Word recognition with phonics
(8) Wide Range Achievement Test - Spelling
(9) Wide Range Achievement Test – Arithmetic

(10) Test of Written Language 3rd version (age 7 and older) - Overall writing
(Students under 7 years old completed the Test of Early Written Language version 2)
(11) Test of Written Language – Contrived writing
(12) Test of Written Language – Spontaneous writing

(13) Woodcock Reading Mastery - Word Attack
(14) Woodcock Reading Mastery - Word Comprehension
(15) Woodcock Reading Mastery - Passage Comprehension

These 15 tests were chosen because:

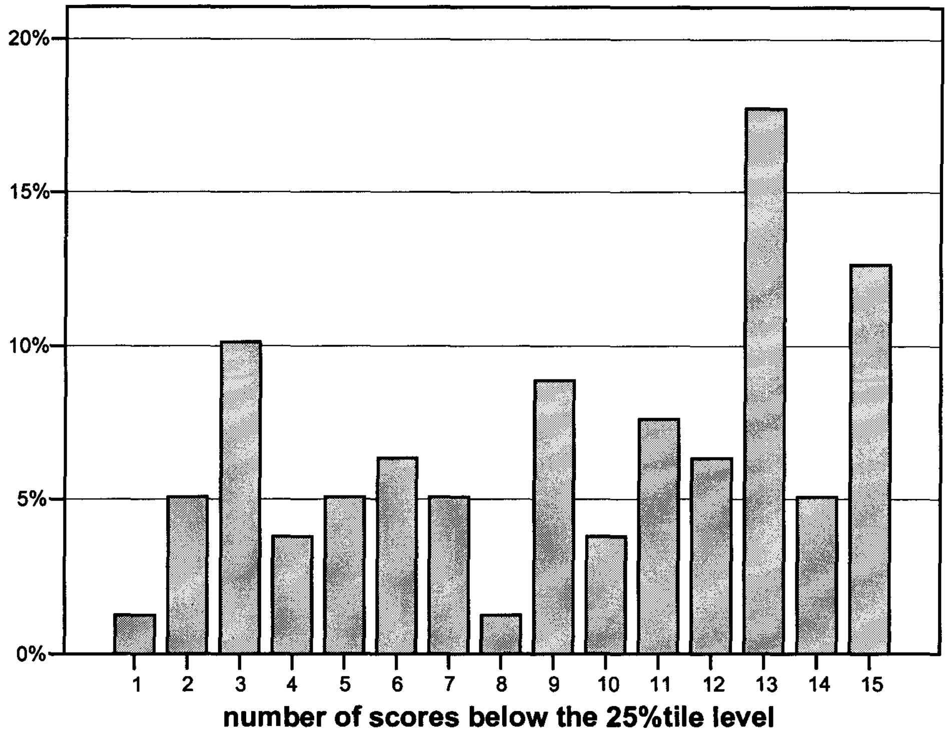
(a) They are part of the standard Arrowsmith Program test battery
(b) They are commonly accepted correlates of academic achievement
(b) Percentile scores are available
- AP students can be compared to their peers
- composite severity scores can be computed

The study inclusion criterion was:
"Having at least one of the 15 percentile scores under the 25%tile level"

Figure 5 shows the distribution of number of percentile scores under the 25%tile level. This number ranges between 1 (inclusion criterion) to 15 (maximum possible).

The sample can be divided into three distinct levels of severity based on the number of percentile scores under the 25%tile level at the start of the program. The most severe were those students who had all tests below the 25%tile level. There were 10 students (13%) who scored in this severe category (severity level 3).

Figure 5 Distribution of Severity at Intake



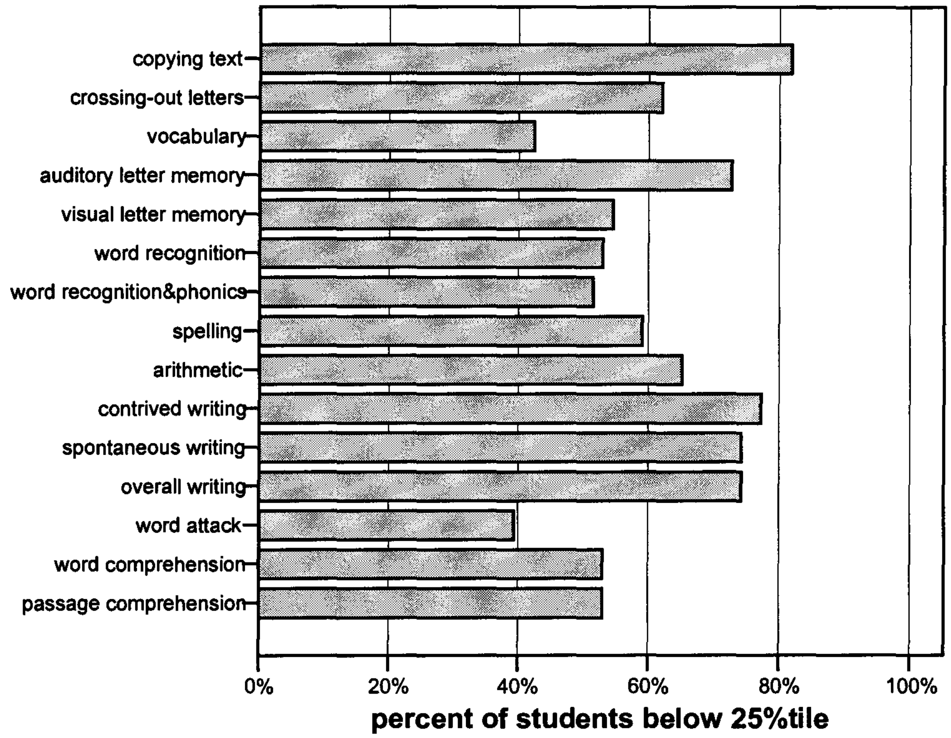
The least severe students were those who had less than eight (i.e. 1 to 7) of the fifteen tests below the 25%tile level. There were 29 students (37%) who were in this category (severity level 1).

The largest group fell between these two levels (severity level 2). There were 40 (50%) who had from eight to fourteen tests below the 25%tile level.

Severity at intake was not related to age, starting grade, or gender (respectively, F(12,66]=.884; F[11,66]=.999; F[1,77]=1.309; all statistically non-significant).

Figure 6 shows the profile of achievement problems at intake across the whole sample. That is, each bar represents the proportion of the sample that is performing ***below*** the 25%tile level on each test respectively. For example, the first bar shows that 83% of the sample performed poorly on the Monroe-Sherman "copying text" task. This graph also illustrates that the AP student population is highly heterogeneous with respect to the 15 achievement domains.

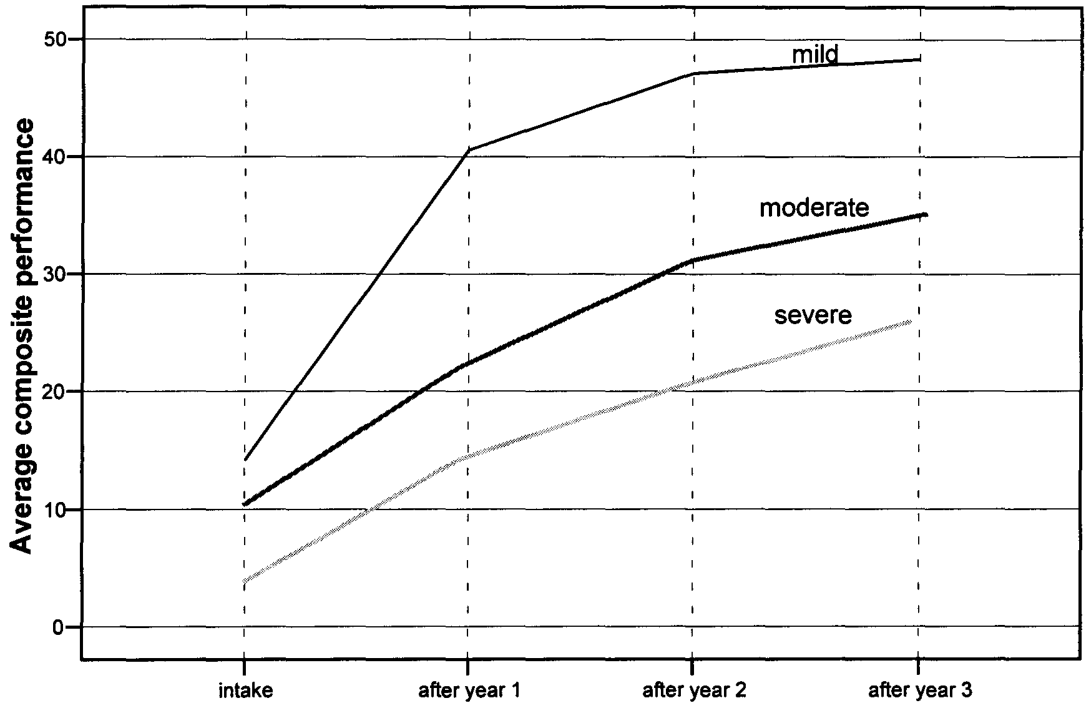
Figure 6 Achievement Problems at Intake



Improvement over time

Figure 7 shows the relationship between intake severity level and rate of improvement. The y-axis represents a mean composite performance score. This score is computed for each student as the average %tile across that student's problem domain. For example if, at intake, a student (call her Mary) is below the 25%tile level on (1) copying text and (2) crossing out letters and (3) word recognition, then Mary's problem domain would be (1), (2) , and (3). For Mary, the change in average performance over (1), (2) and (3), would represent her overall improvement in academic achievement.

Figure 7 Intake Severity Level and Rate of Improvement



For students with a milder level of severity, most of the improvement occurs in the first year. These students improve from a mean of 14%tile to a mean of 41%tile in year one, 47%tile at end of year 2, and 48%tile at end of year 3.

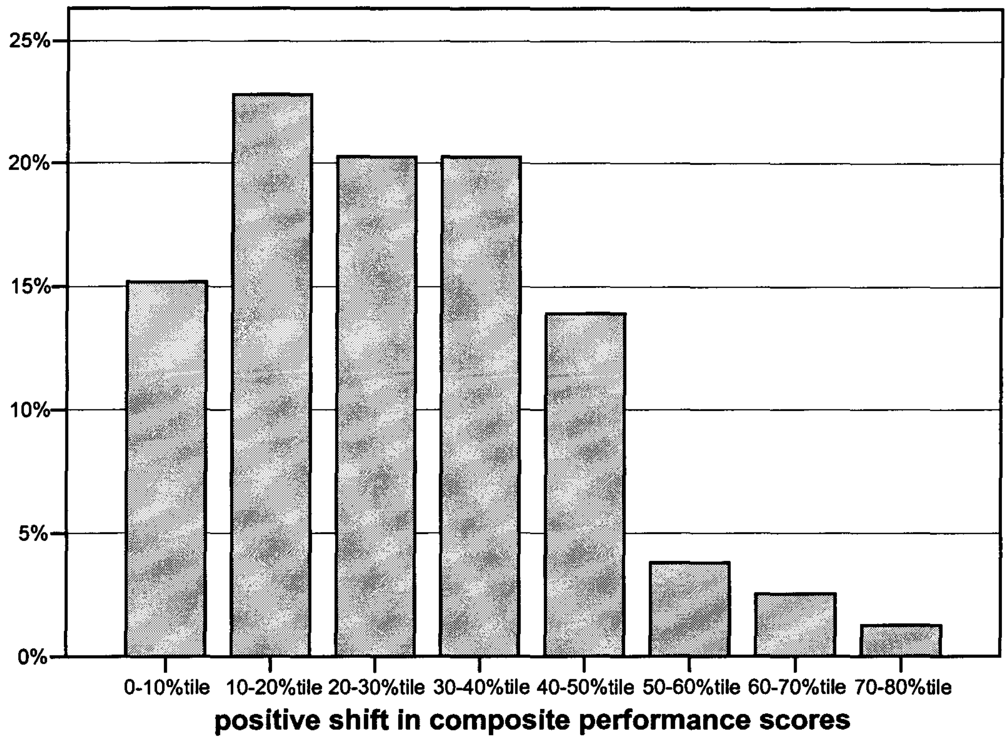
For students with moderate intake severity the rate of improvement is slightly slower. Improvements range from a mean of 11%tile to a mean of 23%tile in year one, 31%tile at end of year 2, and 35%tile after year 3.

For students with 15 or more achievement scores below the 25%tile level at intake, improvement is almost linear, with students making continuous progress from a mean of 6%tile to a mean of 15%tile in year one, 21%tile at end of year 2, and 27%tile at end of year 3. Two of these students did not improve.

A score at the 25%tile level and above is considered to be in the normal range. On the average, at study completion, AP students have moved 6 achievement test scores into the normal range. One AP student deteriorated (the number of achievement test scores to below the 25%tile level increased).

Figure 9 shows the improvement in terms of mean change in composite achievement score.

Figure 9 Distribution overall improvement (from intake to study completion)



Amount of improvement was not related to age, starting grade, or gender (respectively, F(12,66]=.732; F[11,66]=1.134; F[1,77]=1.256; all statistically non-significant).

The amount of improvement was slightly related to intake severity level. Table 1 breaks down the mean improvement from intake to study completion by intake severity (F[2,76]=4.822, p<05).

Table 1 Improvement by Intake Severity

Intake Severity Level	Mean Change in %tile	N	Std. Deviation
1.00 1-7 < 25%tile	34.1	29	15.8
2.00 8-12 < 25%tile	24.4	40	15.0
3.00 15 < 25%tile	20.5	10	13.9
Total	27.3	79	15.9

After controlling for severity at intake, improvements were not correlated with baseline IQ scores (Pearson r=0.112, not sig.) This result suggests that scores on achievement tests at intake are more predictive than baseline IQ scores.

Support for the Validity of the Arrowsmith Program for Treating Learning Disabled Students.

AP cognitive exercises were assigned based on identified need. All students with a deficit in X were assigned X-cognitive exercises. Students without a deficit in X were not assigned X- cognitive exercises. Because exercise assignment is not random, the present study is not designed to test whether improvement could occur without AP cognitive exercises.

Recall that the approach of the Arrowsmith Program is first to distinguish finely between elemental cognitive impairments and then to implement an individualized task-oriented program that challenges the identified deficit. These highly targeted cognitive exercises are thought to create ways for the brain to provide the necessary functionality for encoding and decoding spoken and written discourse, and for storing, organizing, processing and integrating knowledge. If these ideas are valid one would expect that reading, writing, and arithmetic would have common as well as distinct requirements with respect to elemental cognitive functions.

For example, whereas skill in arithmetic operations requires the ability to relate one symbol to another in non-linear ways, understanding ***symbol relations*** is not a critical requirement for reading and writing. For reading and writing, ***symbol recognition*** and motor skills for rapidly following and producing symbol sequences (***motor symbol sequencing***) are more critical.

One approach to explore the relationship between AP cognitive functions and performance tests is to apply a Factor Analysis to the pool of variables comprising the 15 performance tests and 7 common AP cognitive functions. For this analysis we exclude the 10 students who have deficits on all performance tests. These students have poor scores in all tests and AP cognitive functions and therefore can not used in an analysis of divergent relationships.

Based on the criterion of Eigenvalues>1, a 6 factor solution was found in 7 iterations using the principal component method. Varimax rotation with Kaiser normalization was applied to separate the factors. Table 2 shows the significant loadings (greater than .40) on the 6 factors.

Table 2 Six Rotated Factors

	1	2	3	4	5	6
visual letter memory	.885					
word recognition	.844					
word recognition & phonics	.884					
Spelling	.850					
word attack	.777					
AP-symbol recognition	.845					
overall writing (TOWL)		.840				
spontaneous writing (TOWL)		.778				
contrived writing (TOWL)		.770				
passage comprehension			.810			
word comprehension			.736			
Vocabulary			.688			
AP-artifactual thinking			.429			
AP-symbolic thinking			.463			
crossing-out letters				.831		
copying text				.793		
AP-motor symbol sequencing				.728		
timed arithmetic					.747	
AP-supplementary motor					.755	
AP-symbol relations					.732	
auditory letter memory						.648
AP-memory for information						.800

As can be seen, the AP cognitive functions (bold) associated with each cluster of performance tests are consistent with the theoretical framework of the Arrowsmith Program (see Appendix A).

For example, a deficit in *Symbol Recognition,* is manifested as the difficulty in visually recognizing and remembering a symbol or word that has been seen before, and results in performance problems with word recognition, reading, spelling, and remembering symbol patterns as demonstrated in Table 2.

Note that the intake severity of AP functions are not strongly correlated with the Test of Written Language (TOWL), possibly because the TOWL involves a broad range of mutually compensating functions. Therefore the TOWL should not be used to identify specific deficits. However, changes in AP cognitive functions are correlated with changes in the TOWL, reflecting the generalized consequence of improved performance on specific functions. Table 3 shows the correlation between ***changes*** in AP cognitive functions and ***changes*** in performance over two years of the AP program. In each row, only students who had performance problems in that area are included in calculating the correlation. A two year period is used to ensure sufficient variance among students. A one year period doesn't capture enough changes in some of the performance areas, and at the end of three years almost all students have improved significantly. At the end of year two a proportion of students have not yet improved and we can therefore test the link between **improvement or no improvement** and **changes** in each AP cognitive function. Only significant correlations are shown ($p<.01$). [Note that the two students who did not improve at all were excluded from this analysis.]

It can be seen that for the most part the pattern of correlations is as expected based on the AP theoretical framework. It is important to note the specificity of observed effects. For example, improvement in performance tasks that require the sequencing of symbols is correlated with improvement in AP-motor symbol sequencing task performance. Improvement in performance tasks that require recognition of symbols is correlated with improvement in AP-symbol recognition task performance. Improvement in arithmetic performance is correlated with improvement in AP-supplementary motor task performance. Improvement in AP-symbol relations task performance is correlated with improvement on a number of tests as this function relates to understanding and processing information including test instructions. Improvement in AP-artifactual thinking performance leads to improved attentional capacity which may explain its relation to improved performance on a number of tests. There are some additional significant correlations which are not predicted by the framework, but these are likely due to the multiplicity of deficits in this sample of students. Problems with AP motor symbol sequencing, symbol relations, and supplementary motor functions tend to co-occur in the sample.

Table 3 Correlations between Changes in Performance Problems
And Changes in Severity of Cognitive Impairment

Changes in Performance Test Scores	# of students <25%tile on intake	AP motor symbol sequencing	AP symbol relations	AP memory for information	AP symbolic thinking	AP symbol recognition	AP artifactual thinking	AP supple-mentary motor
copying text	65	.38						
crossing-out letters	51	.48	.34			.39	.32	
Vocabulary	37					.30		
auditory letter memory	59						.35	
visual letter memory	48	.39	.46		.35	.83	.42	
word recognition	45		.31			.30	.35	
word recognition & phonics	44		.37				.34	
Spelling	49					.42	.37	
timed arithmetic	50		.35	.30				.42
contrived writing	59						.35	
spontaneous writing	57	.31	.40					
overall writing	57		.42				.34	
word attack	33					.67		
word comprehension	43		.39					
passage comprehension	44		.38					

Summary and Conclusion

Previous research on the Arrowsmith Program has supported its effectiveness in broad terms. The present study funded by the Donner Canadian Foundation provides specific answers to important questions about why and how the AP cognitive exercises are effective.

1. For whom is the Arrowsmith Program most effective?

    ANSWER: All but 2 out of the 79 students did improve. The 2 students who did not improve were not distinguishable from other students in their pattern of test results at intake

    a. Are some impairments irreparable?
    ANSWER: all AP cognitive functions improved through AP cognitive exercises

    b. Does it depend on student characteristics (age, gender, IQ)?
    ANSWER: age, gender, and IQ (controlling for intake severity) were not related to improvement.

2. What affects the rate of improvement?

    ANSWER: rate of improvement depended primarily on the extent of impairment on intake – the number of performance problem areas

    a. Is it related to the type of impairment at intake?
    ANSWER: rate of improvement was not related to any <u>specific</u> functional deficit

    b. Is it related to other student characteristics (age, gender, IQ)?
    ANSWER: *rate* of improvement was not related to age, gender or starting IQ

3. Are performance problems identifiable in terms of AP cognitive functions?

    ANSWER: Since a specific cognitive impairment can effect more than one academic performance area, identification requires testing for specific impairments and cannot be based on performance on achievement tests

4. Can exercises be directly linked to performance improvement?

    ANSWER: There is a specificity of effect. Students who, through specific exercises, improve with respect to AP cognitive functions also improve on related achievement tests.

Appendix A: AP 19 Dysfunctions and Learning Outcomes

Learning Dysfunction (AP Cognitive Function)	Description	Common Features	Learning Outcomes
Motor Symbol Sequencing	Difficulty learning and producing a written sequence of symbols.	Messy handwriting, miscopying, irregular spelling, speech rambling, careless written errors in mathematics, poor written performance.	Develop/improve fine motor skills, sequential motor memory and motor planning in writing, improve handwriting; reduce careless errors in written work; capacity for hand-eye coordination in writing and eye tracking in reading.
Symbol Relations	Difficulty understanding the relationships among two or more ideas or concepts.	Reversals of b-d, p-q, difficulty reading a clock, needing to reread material to comprehend it, problem understanding cause and effect, trouble with logic and mathematical reasoning.	Develop/improve capacity necessary for understanding relationships between concepts necessary for logical and mathematical reasoning and conceptual understanding of material either written or spoken.
Memory for Instructions	Difficulty remembering chunks of oral information.	Trouble remembering and following oral instructions, difficulty following lectures or extend-ed conversations, problem acquiring general information through listening.	Develop/improve auditory memory and the capacity to remember and follow oral instructions and retain oral information for learning; improve the capacity to remember chunks of information in the correct order.
Predicative Speech	Difficulty seeing how words and numbers interconnect sequentially into fluent sentences and procedures.	Problem putting information into one's own words, speaking in incomplete sentences, difficulty using internal speech to work out consequences, trouble following long sentences, breakdown of steps in mathematical procedures.	Develop/improve the capacity to remember a sentence of increasing difficulty and length; improve the ability to put information into own words; develop the capacity for the sense of how symbols (words and numbers) interconnect sequentially; improve the ability to follow procedures in mathematics; develop the ability to write and speak in complete sentences.
Broca's Speech Pronunciation	Difficulty learning to pronounce syllables and then integrate them into the stable and consistent pronunciation of a word.	Mispronouncing words, avoiding using words because of uncertainty of pronunciation, difficulty thinking and talking at the same time, flat and monotone speech with lack of rhythm and intonation, limited ability to learn and use phonics.	Develop/improve the capacity for sound-symbol correspondence; develop the phonemic memory necessary for the phonetic aspect of reading; develop the ability to pronounce multi-syllabic words correctly; develop the ability to read with greater oral expression

Symbolic Thinking	Difficulty developing and maintaining plans and strategies through the use of language.	Problem being self-directed and self organized in learning, limited mental initiative, difficulty keeping attention focused on task to completion, trouble seeing the main point.	Develop/improve the ability to grasp the main point of written or orally presented material; develop the ability to state the main idea of a selection using one's own words; develop the ability to generate and maintain plans and strategies for problem solving; develop the capacity to express ideas more clearly in writing; develop the capacity to self-direct, to develop initiative and to remain focused on tasks to completion.
Symbol Recognition	Difficulty in visually recognizing and remembering a symbol or word that has been seen before.	Poor word recognition, slow reading, difficulty spelling, trouble remembering symbol patterns such as mathematical or chemical equations.	Develop/improve the capacity to visually recognize and remember words or symbols necessary for reading, spelling and to a lesser extent mathematics.
Lexical Memory	Difficulty remembering several unrelated words.	Problems with paired associative memory, trouble learning how to read due to difficulty associating the word with its sound, trouble following oral information.	Develop/improve auditory memory for words and vocabulary development.
Artifactual Thinking	Difficulty registering and interpreting non verbal information.	Problems interpreting non-verbal information such as body language, facial expression and voice tone, difficulty registering one's own emotions, weak social skills. Is also related to attention to detail.	Develop/improve the capacity for non-verbal thinking and problem-solving; develop the ability to interpret body language, facial expression and voice tone and to respond appropriately in interpersonal interactions; develop ability to interpret and modulate one's emotions. Improved attentional capacity.
Supplementary Motor	Difficulty carrying out internal sequential mental operations, such as mental mathematics.	Finger counting, trouble retaining numbers in one's head, difficulty making change, problem learning math facts, poor sense of time management.	Develop/improve the capacity for number sense; develop the capacity for carrying out internal mental computation of addition and subtraction; develop the ability to use time effectively through scheduling and organization; develop an understanding of quantification related to money, time, space.

Study Consultant

Dr. William Lancee is Head of Research in the Department of Psychiatry at Mount Sinai Hospital and Associate Professor, Department of Psychiatry, University of Toronto. He has a Bachelor in Mathematics from the University of Waterloo. In the first part of his career, he was a statistical consultant to major pharmaceutical companies. Later he received his Ph.D. from the Institute of Medical Science at University of Toronto, and has designed and carried out more than 20 peer reviewed collaborative studies, including large-scale epidemiological studies as well as psychotherapy intervention studies. He is author of three published psychological measurement tools: The Nurse Observed Behaviour Scale; the Staff Patient Interaction Response Scale; and the Perceived Family Burden Scale. He has supervised 15 graduate students. His special interest is in the mathematical modeling of complex systems using cellular automata, genetic programming, and neural networks. He has published 37 papers on a wide variety of topics. Recent papers related to family issues and child development are:

Maunder R., **Lancee, W.J.**, Greenberg, G., Hunter, J., Fernandes, B. Insecure attachment in a subgroup of ulcerative colitis defined by ANCA status. Dig Dis Sci, 45, 2000, 2127-2132.

Hazelton, R. **Lancee, W.J.,** O'Neil, M.K. (1998) The controversial long-term effects of parental divorce: the role of early attachment. Journal of Divorce and Remarriage, 29(1)1-18.

Beitchman, J.H., Brownlie, E.B., Inglis, A., Wild, J., Ferguson, B., Schachter, D., **Lancee, W.J.**, Matthews, R., Wilson, B. (1996) Seven-Year Follow-up of Speech/Language Impaired and Control Children: Psychiatric Outcome. Journal of Child Psychology and Psychiatry 37(8)961-970.

Levene, J.E., **Lancee, W.J.**, Seeman, M.V. (1996) The Perceived Family Burden Scale: Measurement and Validation. Schizophrenia Research, 22, 151-157.

Beiser, M., **Lancee, W.J.**, Gotowiec, A, Sack, W., & Redshirt, R. (1993). Measuring Self-Perceived Role Competence Among First Nations and non-Native Children. Canadian Journal of Psychiatry. Vol 38, 412-419.