




Annexure O


← → <https://onesearch.library.utoronto.ca/faq/are-books-peer-reviewed> Are books peer reviewed? | ... X


File Edit View Favorites Tools Help


 UNIVERSITY OF TORONTO LIBRARIES





 Research


 Services


 Visit

 Ask

 About

 My Account

 UofT

 Search

All FAQs

Contacts

- Library information by email or 416-978-8450
- Report a problem with an online resource
- Staff and departments
- Faculty liaison librarians
- Computer help

Research help

- Chat
- Drop-in
- Consultations
- Email
- Telephone

Are books peer reviewed?

"Peer review" is the editorial process that scholarly articles go through before they are published in a journal. Since books go through a different editorial process before publication, they aren't peer reviewed. They can be still be good scholarly sources, though.

More information:

- What is peer review?
- Does a book count as an academic source?

See other FAQs about:

- Books
- Peer review
- Research help

Can't find what you're looking for? [Contact us.](#)

Library links

- [Library Home](#)
- [Renew items and pay fines](#)
- [Library hours](#)
- [Ask](#)
- [Site map](#)
- [Portal](#)

Libraries

- [Robarts](#)
- [Gerstein](#)
- [OISE](#)
- [Engineering](#)
- [UT Mississauga Library](#)
- [UT Scarborough Library](#)
- [Information Commons](#)
- [All libraries](#)

Contact

University of Toronto Libraries
130 St. George St., Toronto, ON, M5S 1A5
library.info@utoronto.ca
416-978-8450




[Map](#)

[About web accessibility.](#) Tell us about a web accessibility problem.


[About online privacy and data collection.](#)

© University of Toronto. All rights reserved.

Connect with us

-  [Twitter](#)
-  [YouTube](#)
-  [Instagram](#)

[more social media](#)



Annexure P

Comments on an Arrowsmith Program Evaluation Report (8 Month Review for a 3 Year Pilot):

Completed for the Vancouver School Board (VSB) by Linda Siegel - School year 2002-2003

Abstract

In 2003 the Vancouver School Board (VSB) received an 8 month Arrowsmith Program Evaluation Report from Linda Siegel, a Professor at the University of British Columbia in Educational and Counselling Psychology and Special Education. Two programs for elementary students with learning disabilities were to be compared over a three year period. The Arrowsmith Program (designed to improve cognitive functioning over a three to four year period) and an Extended Learning Assistance Class (ELAC) (which focused on reading and writing) were to be compared on measures of students' cognitive ability and academic achievement. The Arrowsmith Program was to be piloted for three years (the length of time it takes to complete the program) and the VSB received funding through the Vancouver Foundation (a non-profit organization). After the first eight months of the pilot, tests of cognitive ability and academic achievement were given to all students involved in the study and an evaluation was written by Siegel (only achievement measures were reported). Siegel stated in her conclusion of her 8 month evaluation that, "On all but the Comprehension and Spelling measures, ELAC performed at higher levels than Arrowsmith, often by a relatively large amount." Results from this review of Siegel's data indicate her statements regarding the ELAC group are incorrect and do a disservice to those who could benefit from the Arrowsmith Program. There are a variety of serious methodological, calculation and statistical problems associated with the Siegel evaluation, and even with those problems there was no statistical support for the above conclusion. Analysis by two independent statisticians shows that the only statistically significant findings supported by the test scores presented in Siegel's evaluation are that the Arrowsmith students showed greater gains in Comprehension and Spelling subtest measures as compared to the ELAC group. No subtests showed statistical significance in favor of the ELAC group. The Arrowsmith Program was discontinued at the VSB and Siegel's eight month evaluation was noted as being influential in this decision. Additionally, this study has been cited by Siegel, media and professionals involved in education as evidence against the effectiveness of the Arrowsmith Program.

On April 12th, 2013 a copy of Siegel's 2003 evaluation of the Arrowsmith Program was sent to me by email for review. Linda Siegel is the Dorothy C. Lam Chair in Special Education and a Professor at the University of British Columbia in Educational and Counselling Psychology and Special Education. I had heard that a study existed and had been quoted over the last 10 years (2003 to 2013) by Siegel as a critical evaluation of the effectiveness of the Arrowsmith Program. The original evaluation report was given to Val Overgaard, Associate Superintendent, Learning Services, Vancouver School Board (VSB) who is now retired. A copy of Siegel's evaluation was given to me by a parent who received it directly from Siegel. The copy I received was confirmed as being identical to the one sent to the VSB through a direct written confirmation from Val Overgaard on May 2nd, 2013.

Siegel is highly critical of the Arrowsmith Program. She has been quoted as stating that the Arrowsmith Program is "a fraud" (The Passionate Eye: Fixing My Brain, 2008). Siegel's name and this study have been noted in a recent Scientific American Mind magazine (Wickelgren, 2013). The article notes, "Indeed, one

small eight – month investigation of the school's curriculum led by educational psychologist Siegel of the University of British Columbia failed to show that it significantly improved students' scores on a battery of cognitive and achievement tests" (Wickelgren, 2013, p. 45). With this in mind, I felt it necessary to closely review Siegel's research and evaluation and provide public comment. Two academics with extensive backgrounds in statistics and research methodology were also consulted. The first is Dr. Darren Irwin, Associate Professor, Biodiversity Research Centre and Department of Zoology, University of British Columbia. The second is Dr. William J. Lancee, Associate Professor, University of Toronto, Toronto, Canada.

Siegel's evaluation was not peer-reviewed or published in an academic journal. Rather, it was used to assess the effectiveness of the Arrowsmith Program at Cunningham Elementary School as compared to a learning assistance classroom at another elementary school in Vancouver, British Columbia, Canada (Nootka Elementary School's Extended Learning Assistance Class – ELAC). The ELAC classroom focused on teaching reading and writing skills to children with severe learning disabilities.

The following points for comment are made with the notion that any research/evaluation should be conducted with the highest research standards possible. For example, when studying children with learning disabilities, experimental design groups should be as closely matched with each other as possible in terms of type of learning disability, age, socioeconomic background and educational and remediation experience. The researcher should also have a full awareness of the theory and curriculum of the programs being evaluated. This assists in the proper design and length of research/evaluation. This also assists in making conclusions regarding the data being presented to the researcher. In addition, the researcher must do an in-depth statistical analysis of the data to test for an effect.

The Arrowsmith Program is designed to improve cognitive functions underlying a variety of learning disabilities. Children with significant behaviour difficulties are not recommended for the Arrowsmith Program due to the amount of focus required on the cognitive exercises. The Arrowsmith Program is not an academic achievement program. It does not teach reading, writing, spelling and math directly through academic curriculum. Nevertheless, progress has been observed at Arrowsmith Program sites in achievement areas without specific academic instruction; this is not always the case, and depends on the severity of the child's cognitive weaknesses. Therefore, it is recommended that students working on the Arrowsmith Program also receive daily Math and Language (reading, writing and spelling) instruction during the implementation of the program.

The Arrowsmith Program also recognizes that there are additional cognitive functions that need to be addressed outside of those affecting specific academic achievement skills for children with learning disabilities. These cognitive functions are related to executive functioning, social perception, language processing, visual memory, auditory memory and processing, and reasoning. Thus, a child's Arrowsmith Program is designed not for one specific learning disability, such as Dyslexia, but rather to address a multitude of learning challenges.

Any study into the effectiveness of the Arrowsmith Program should look at both cognitive functioning and academic achievement. This is due to the theory proposed by the Arrowsmith Program that, as cognitive functions improve, so does the ability of the child to improve academic achievement skills. A study that looks at academic achievement alone, without investigating improvements in cognitive functioning, would not be a proper analysis of the effectiveness of the Arrowsmith Program.

The following points for comment have been made after careful reading of Siegel's study. The study can be made available to anyone interested in reading it. It is important to note that this re-analysis of the statistical data was conducted independently by both Dr. Darren Irwin and Dr. William J. Lancee who have extensive backgrounds in statistics and research methods. I felt it important to compare two independent academic reviews of the data. It should be noted that several sections in the "Points for Comment" section below were written primarily by Drs. Irwin and/or Lancee and then edited by Howard Eaton.

Siegel Evaluation: Points For Comment

1. Failure to Report Drop-Out Effects and Impact on Statistical Results

The failure to use 'intention-to-treat' analysis or discuss attrition effects makes results impossible to interpret. There was a 100% completion rate for Arrowsmith Program students and approximately 47% for the ELAC sample. The final ELAC sample reported has an N of 7, out of a class of 15. This is at least half the sample for whom the report does not provide Pre values and for whom the analysis does not try to compensate.

A common reason for anyone to drop out of a program is that it isn't working for them. That means that improvement seen in program completers is most likely over-estimated. A conservative approach is to include drop-outs in a study with Post scores based on the expected change if there was no intervention. For some measures the expected change is zero points, but the expected change value for other measures may be negative (i.e., dropping further behind peers).

If intention-to-treat analysis were done, most of the improvements seen in the ELAC sample would most likely be reduced. In short, losing more than half of the study group could cause serious bias in the results if the included individuals are a non-random sample of those initially in the study group.

2. Arrowsmith Student Selection/Comparison Group

Arrowsmith Program Students

Ten students (N=10) were selected by the Vancouver School Board at Cunningham Elementary School in Vancouver, British Columbia for the Arrowsmith Program experimental group. This was their first year in the Arrowsmith Program. From these 10 students, 3 were identified as having a Learning Disability. Seven out of ten, or 70% did not have a classification of a Learning Disability. It is not stated in the report if this 70% had a classification of any sort. In other words, 70% of the Arrowsmith Program students had not received psycho-educational assessment and proper special needs identification. Siegel also noted in her study that 33% of the students experienced fatigue and hunger due to not receiving breakfast at home. It is also evident that a number of the ten students spoke English as a second language (ESL). In addition, Siegel notes that "some children in the Arrowsmith Program did have problems with attention and distractibility" (Siegel, 2003). All ten students in the Arrowsmith Program completed the 2002-2003 academic year (100% retention). The Arrowsmith Program students did not receive any math instruction over the course of the 2002-2003 academic year, and some did not receive reading instruction. This is against the recommendation of the Arrowsmith Program.

Comparison Group – Nootka Elementary School Extended Learning Assistance Class (ELAC)

As stated above, the data for seven students (N=7) was reported in Siegel's study. It is noted that a class of 15 started the school year with seven students leaving (47% retention rate of the original starting class in the learning assistance centre; although data was presented only for 7 of the remaining 8). The seven students whose data was reported had all been diagnosed with severe learning disabilities. There is no mention of these students' cognitive or intellectual abilities. The Woodcock Johnson III Cognitive battery was reportedly administered to the children, but the report from Siegel makes no mention of the results of these assessments. This is the same for the Arrowsmith Program students. Thus, it is not possible to do a comparison of the two groups' cognitive and intellectual abilities prior to the start of the study based on the data reported. Given that the Arrowsmith Program focuses on improving cognitive abilities, this would have been useful information or data to compare. The focus in the ELAC group was improving reading and writing skills. Siegel writes, "He [the teacher] feels that the primary role of the Extended Learning Assistance program is to improve the reading and writing of students with severe learning disabilities" (Siegel, 2003). Finally, many of the seven students whose data was reported had been in the Extended Learning Assistance Class for more than one year (three of the seven students were in year two of the program; two of the seven students were in year three of the program; and two of the seven students had no data reported related to time in the ELAC program), and so for the majority of these students, this was not their first year in the program designed to improve reading and writing. For all of the Arrowsmith Program students, this was their first year in the program.

Note:

In a letter to the Learning Disabilities Association of Ontario's (LDAO) Executive Director, dated October 23, 2001, Siegel commented on the Arrowsmith School and Toronto Catholic District School Board Research Project. This research was being conducted prior to her own evaluation for the VSB. Siegel raised several issues with this design in her letter to the LDAO. Two key points worth noting (as related to Siegel's Vancouver School Board research/evaluation of the Arrowsmith Program) were her comments regarding knowing whether or not the student being studied had a learning disability and the effective use of comparison groups.

1) Siegel writes, "We have no way of knowing whether or not the children are actually learning disabled. Although they may be labeled that way by the school, we need independent verification of the disability through test scores. Keogh, Kavale and others have written about the lack of validity of school system identification of learning disabilities. The test scores necessary to define a child as learning disabled should be outlined. We cannot call a child learning disabled unless there is a low test score on at least one achievement measure; this is the minimum expectation."

2) Siegel writes, "A comparison group is absolutely critical. Furthermore, the comparison group must have the same type of learning disabilities and be the same age and grade level as the target group. They also must receive the same Pre- and Post- tests in the same interval. Without this comparison group, it is impossible to conclude anything about the efficacy of the treatment."

These are important insights into how to properly conduct research in the field of Learning Disabilities. It would be fair to assume that in any study conducted by Siegel a similar standard of research conduct would be undertaken.

3. Data Analysis:

Re-analysis of data and changes to results:

This re-analysis of data was done independently by Dr. Darren Irwin and Dr. William J. Lancee, although, we are confident that any researcher experienced with statistics can access Siegel's study, extract the data, and conduct the same statistical tests that we have. We can provide data files in a variety of formats including SPSS (Statistical Package for Social Sciences), JMP, or Excel.

The data Siegel reported in her evaluation was analyzed using statistical software (i.e., SPSS). It is apparent that the data presented for two students in the ELAC group is internally inconsistent due to unknown reasons. We provide details below.

ELAC student #1's Comprehension subtest Pre-Post is 79 to 77 (a change of -2 pts), but most of the subsequent statistical analysis is based on a Post value of 97 with a change of +18 pts (however, in the "Overall Findings" table, the SD is calculated using a Post value of 77).

[The reason for the error/modification is unknown, but the preceding data cell 'Word Attack' has a Post value of 97.]

ELAC student #5's Letter Word subtest Pre-Post is 76 to 52 (a change of -24) but all statistics are based on a Post value of 76 with a change of 0 pts.

[The reason for the error/modification is also unknown, but sometimes investigators deem a measurement to be too far from expected to be correct - in this case a drop of 24 points. Setting the 'missing/incorrect' value to the Pre score is not conservative and is unacceptable since it results in an inflated improvement score for one of the comparison groups (the ELAC program)]

Because of the very small sample size, these errors overestimate the ELAC improvements dramatically

Based on original values reported for individual students:

Correct averages for ELAC for the Letter Word test should be Post=79.29, change=-3.42

Correct averages for ELAC for the Comprehension test should be Post=73.71, change=-3.28

Re-analysis of the data, using the original values from individual students:

There are three students with no severe problems (defined for this analysis as standard scores < 80) in any of the areas measured (two AP students and one ELAC student). Most students in both samples have problems in only a few of the areas measured.

In order to determine how a student with an actual weakness in an area of achievement performed in the two groups, another statistical analysis was conducted. For each student, an overall index of critical improvements was calculated as follows:

- A. If the Pre-score for a given measure is under 80 then the change from Pre to Post is designated as 'critical'.
- B. Critical changes are averaged for each student (negative as well as positive changes).

The Y-axis in the graph below represents mean improvement scores for the two programs.

The mean critical change for the AP sample is 11.79 (SD=7.83).

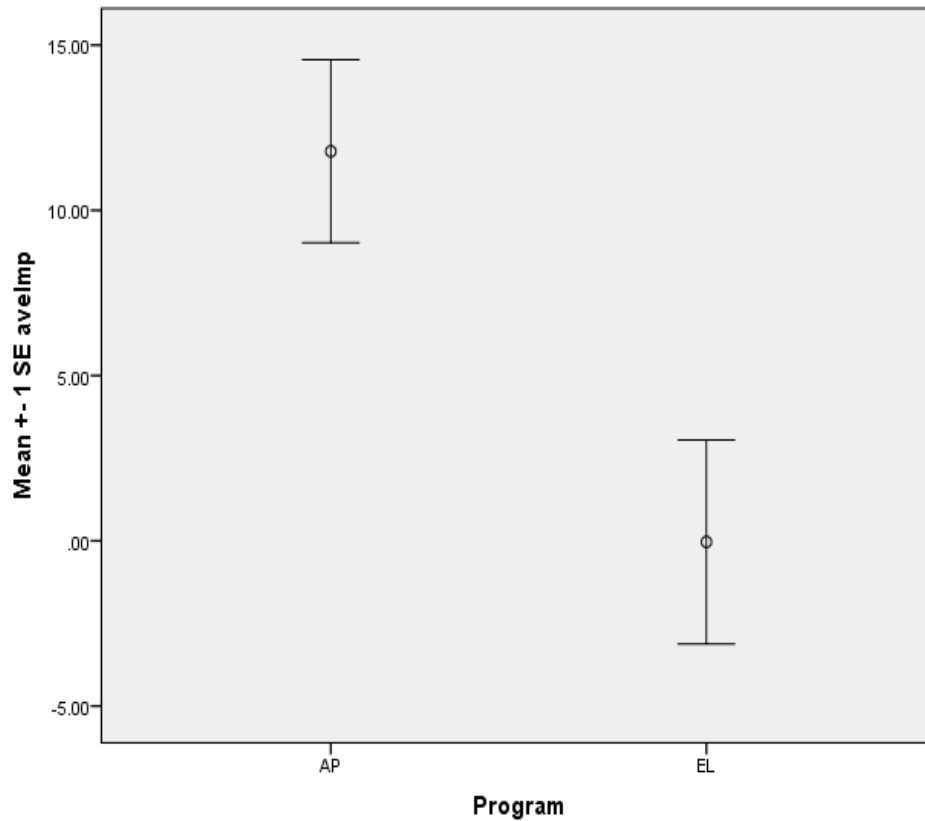
The mean critical change for the ELAC sample is -0.03 (SD=7.56).

The difference is statistically significant: $F(1, 12) = 8.049$; $p = .015$.

This technique recognizes the fact that different students have different problems and need to improve on the weaknesses they have, rather than on the weaknesses they don't have. The above analysis excludes the three students who do not have any scores below 80.

The mean score for the AP sample is significantly greater than zero. The 95% Confidence Interval is +5.24 to +18.33. The 95% Confidence Interval for the ELAC sample is -7.97 to +7.97.

The following graph outlines the improvement of the Arrowsmith Program students over the ELAC students during this eight month study.



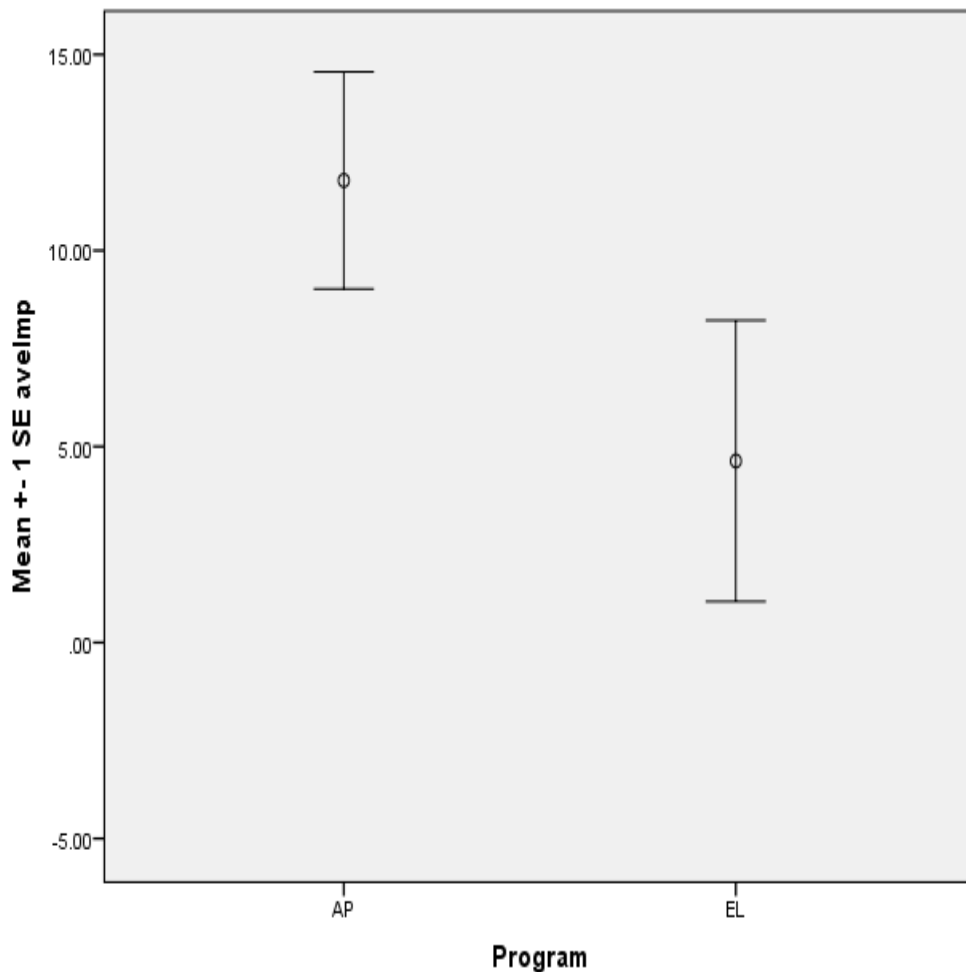
The following is the result when the two discrepant Post values (based on reported Diff – Pre rather than the reported Post) are put in for the two ELAC students (#1: 97 instead of 76; #5: 76 instead of 52).

The pattern is still there but now the difference is NOT significant.

The mean critical change for the AP sample is 11.79 (SD=7.83). [THIS IS THE SAME AS BEFORE]

The mean critical change for the ELAC sample is now 4.63 (SD=8.79). [HIGHER THAN BEFORE]

The difference is not statistically significant: $F(1, 12) = 2.585, p = .134$.



The following table shows how the average critical improvement values are distributed in the sample. They are arranged from the lowest to the highest. With this chart, one can see the number of Arrowsmith Program (AP) students with larger scores for average critical improvement.

Critical Average Improvement	Program	ID number	gender	grade
Not Applicable (no scores < 80)	AP	3	M	5
Not Applicable (no scores < 80)	AP	8	M	4
Not Applicable (no scores < 80)	ELAC	3	M	.
-9.67	ELAC	5	F	6
-6.00	ELAC	6	M	6
-2.00	ELAC	1	F	6
1.50	ELAC	2	F	6
1.80	AP	9	M	2
3.00	AP	7	M	6
4.63	ELAC	4	M	7
5.00	AP	1	M	3
11.33	ELAC	7	M	6
11.50	AP	10	M	4
15.00	AP	4	M	4
16.00	AP	2	F	7
19.00	AP	6	M	4
23.00	AP	5	M	5
n=14	n=17	n=17	n=17	n=16

A note about demonstrating lack of effect - the risk of "type II error":

It is a fundamental principle in the Applied Sciences that, to show that an effect is small or nonexistent, one must first make a serious and meaningful effort to try to prove the effect. It is only when such effort fails that one can conclude that no effect existed in the first place. Failure to make a serious and meaningful effort to prove effectiveness makes it impossible to conclude ineffectiveness. The power of the proof is directly related to the quality of the effort. A small study sample, insensitive measures, and liberal data manipulation all undermine the quality of the effort and cannot be used to argue that an intervention does not work.

Using the Grade 5-7 group comparison and concluding remarks by Siegel

The 2003 report divides the Arrowsmith results into grades 2-4 and 5-7, so that the grade 5-7 Arrowsmith group can then be compared with the ELAC group (which was also grade 5-7). Even though the report did not indicate which of the original ten Arrowsmith students became the four individuals included in the grade 5-7 group, we were able to determine this by closely examining the means and comparing them to the scores of individuals. This reveals that Arrowsmith cases #2, #3, #5, and #7 are included in the grades 5-7 group. These four Arrowsmith Program students result in the same means and SD's presented in the Mean Improvements table for that age group. For the ELAC data in the same table, the same values were arrived at, except for Letter Word ID and Comprehension, due to the discrepancy between reported Pre-Post and Diff scores.

Siegel's report ends the statistical analysis there, without doing a statistical test of the differences in the grade 5-7 groups. Rather, Siegel concludes, without doing a statistical test, that the ELAC group showed "relative success" (2003). In order to analyze the data, we performed a series of ANOVAs (Analysis of Variance), none of which resulted in P-values below 0.05. Thus, there is not sufficient evidence in the data that the Diffs differed between the two programs in any of the scores. Surprisingly, however, the 2003 report concludes from the table (without presenting any statistical test of differences in the grade 5-7 groups), that, "What we see from the table . . . is fairly strong evidence for ELAC's success. On all but the Comprehension and Spelling measures, ELAC performed at higher levels than Arrowsmith, often by a relatively large amount. This provides support for the relative success of ELAC" (Siegel, 2003). The differences are small compared to the variation within groups, and (when done by us) statistical tests show no statistically significant differences. Yet Siegel makes a very strong statement that the evidence provides support for ELAC over Arrowsmith. Siegel seems to base this purely on a count of how many mean improvements are greater in ELAC than Arrowsmith—in their table, 5 test mean improvements are greater (although not significantly) in ELAC, 2 are greater in Arrowsmith, and in 1 the means are equal. If Siegel had wanted to test this, she could have used a binomial test. We did so, and it comes out to $P = 0.45$, far from surprising under a null hypothesis that the programs are equal. Hence the verbal argument from Siegel in the 2003 report has no statistical merit. There is in fact no evidence in the "Overall Comparison" section of the 2003 report that ELAC performs better than Arrowsmith.

Extending the analysis by properly accounting for correlations between Pre and Post

The "Overall Comparison" section of the 2003 report has two major flaws that may have prevented an accurate conclusion. First, it was apparently based on two incorrect Diff values (Pre to Post calculations with ELAC student #1 and #5). Second, it did not properly control for variation in "Pre" scores (such variations can be due to differences in age, grade, or other factors) and the expected relationship between Pre and Post scores. Thus, an analysis was done that controls for Pre scores, as follows:

For each test, a regression line was calculated that best describes the relationship between Pre (the explanatory variable) and Post (the response variable) for all students in the study (Arrowsmith and