Sub-group Analysis of AIE Data
1 Dec 08

For the sub-group analysis eight groups were formed based on their 0506 assessment score and number of years the child participated in the program. Assessment scores were sorted from lowest to highest and divided into four categories. Then, each category was divided further into two groups based on the number of years the child participated in AIE.

Two groups representing participation was formed. Those who participated 0 or 1 year are referred to as the "comparison" group and those who participated 2 or 3 years are referred to as the "program" group. This grouping was used because previous analysis indicated that this is where the major break is with respect to effect of participation on assessment scores. That is, participation for one year is no different from not participating at all; and there were little differences between participating two years or participating three years. Grouping two years together was also the best way to ensure that enough cases were retained for analysis.

The two categories used in the analyses are defined as:

Category 1 = Lowest scorers: participated 0 or once (comparison) vs. participated 2 or 3 times (program)

Category 2 = Below average: participated 0 or once (comparison) vs. participated 2 or 3 times (program)

These were the only two categories analyzed because it is less likely to see improvement for children scoring above average, and because there were no children in category 4 (highest scores) who participated in AIE 2 or 3 years.

The analysis began with testing whether the two groups were different on the 0506 measure. If they were then an effort was made to make the groups more similar by dropping the lowest and highest scores from each respective group. This is very important because if the groups are different at the beginning then the basis for arguing that AIE caused them to be different at the end is seriously undermined, especially if the margin of difference is the same at the end. For example, for the RND analysis, the category 1 comparison group was significantly different from the program group on 0506 TerraNova scores. Six of the very lowest scores in the category 1 comparison group were dropped to make the group more similar at baseline to the program group.

In order to compensate for the very small sample, the probability level for determining statistical significance was set at 0.10, instead of the usual 0.05. The minimal sample size for analysis is 4 per group (there must be one more case than the number of outcomes (three years). While this analysis is useful in determining the

impact of participation for this group of children, it lacks sufficient sample size for making generalizations to a larger population or for making policy decisions.

In brief, this analysis should be presented as descriptive rather than conclusive because:

    (1) the sample size is very small, and

    (2) there is insufficient data to take into account which school the child is from (i.e., nesting is ignored). It is quite possible that differences are due to the differences between schools rather than participation in the program. There is insufficient data to test this competing explanation. Ignoring nesting also means the statistical tests are biased in favor of finding significance when none exists.

RESULTS

For the following graphs the primary interest is in testing whether the slopes of the lines are flat (indicating change over time), and whether the lines have any meaningful separation between them (difference in means).

Summary:

Overall, the most promising results are for the lowest scoring students on RND, LNG, and SCI. The program effect for students who score below the mean at baseline is less promising, however. The small sample size places limitations on the analysis that can be conducted and the stability of the results. Results should be used only for descriptive purposes and not to make inferences to a larger population or policy related decisions.

RND Category 1: Lowest scorers – significant differences found

The program groups shows significant improvement over the years (both lines have significant slope) where the comparison group does not show improvement over the years (its line is flat). However the difference between the groups (i.e., the distance between the lines) is significant for AY0506 ($p=.06$) and AY0708 ($p=.06$). Even though cases were removed in order to make the means equal at baseline (AY0506) the groups remained significantly different. Given the magnitude of difference at 0708, it would be difficult to not attribute this improvement to participation in AIE.

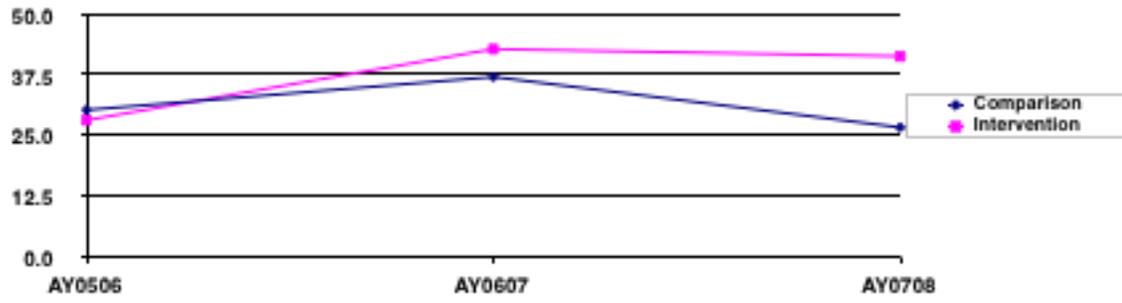| Category 1: Lowest scorers	N	AY0506	AY0607	AY0708
Comparison	4	12.0	15.5	14.7
Program	4	17.2	28.5	47.5
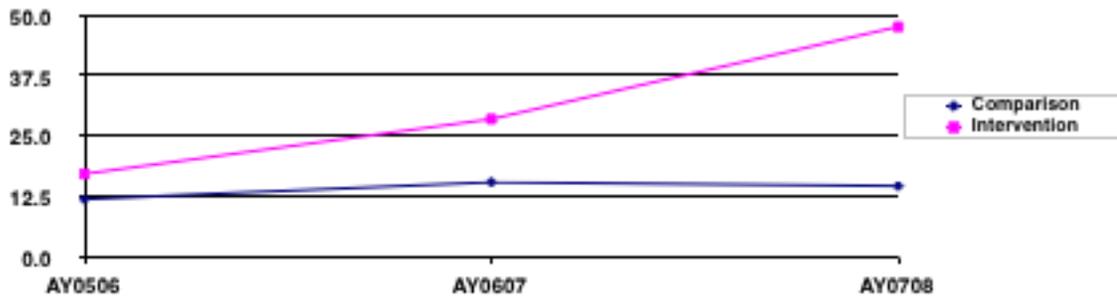
RND Category 2: Below average scorers - No significant differences found

Neither the comparison nor Program show improvement over the years, nor is there difference between them at any time point. In other words, both the lines are flat, and have no meaningful separation between them.

Category 2: Below average scorers	N	0506	0607	0708
Comparison	6	30.2	37.0	26.6
Program	7	28.1	42.7	41.2

LNG Category 1: Lowest scorers – significant differences found

The comparison and program groups show significant improvement over the years (both lines have significant slope). There is also significant difference between the groups for AY0607 only (p=.08); but the groups are not significantly different for AY0506 or AY0708. Even though the 0708 difference is larger than 0607, there is more variation in 0708.

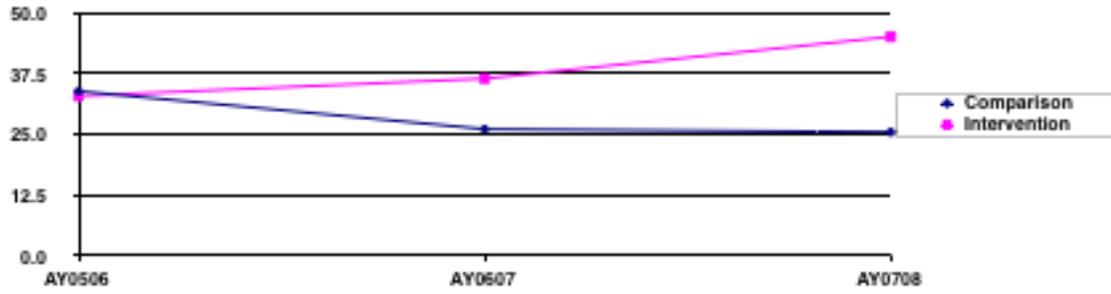| Category 1: Lowest scorers	N	AY0506	AY0607	AY0708
Comparison	5	12.6	21.8	27.0
Program	4	16.2	32.0	39.2
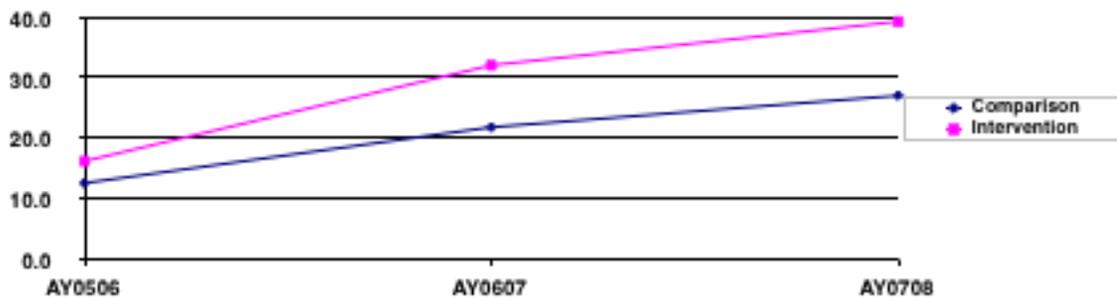
LNG Category 2: Below average scorers – no significant differences found

Again, neither the comparison nor Program shows significant improvement over the years (both lines are essentially flat). There are no differences in mean for any years.

Category 2: Below average scorers	N	AY0506	AY0607	AY0708
Comparison	5	33.8	26.0	25.4
Program	9	32.8	36.4	45.0

MTH Category 1: Lowest scorers – significant differences found

There is significant upward trend for the comparison and Program showing significant improvement over the years (both lines have significant slope). However there are no significant differences between means for any years.

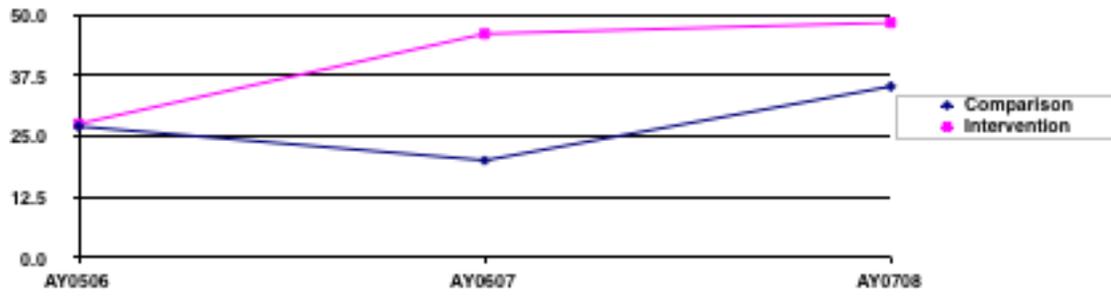| Category 1: Lowest scorers	N	AY0506	AY0607	AY0708
Comparison	5	12.8	15.8	43.6
Program	5	15.4	36.8	37.0
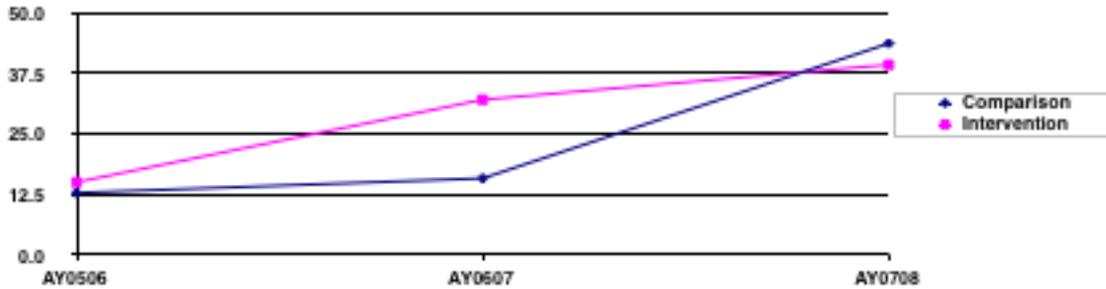
MTH Category 2: Below average scorers – significant differences found

There is significant upward trend for the Program participants showing significant improvement over the years (both lines have significant slope). There is a significant difference between means for year AY0607 only.

Category 2: Below average scorers	N	AY0506	AY0607	AY0708
Comparison	4	27.0	20.0	35.2
Program	6	27.5	46.0	48.3

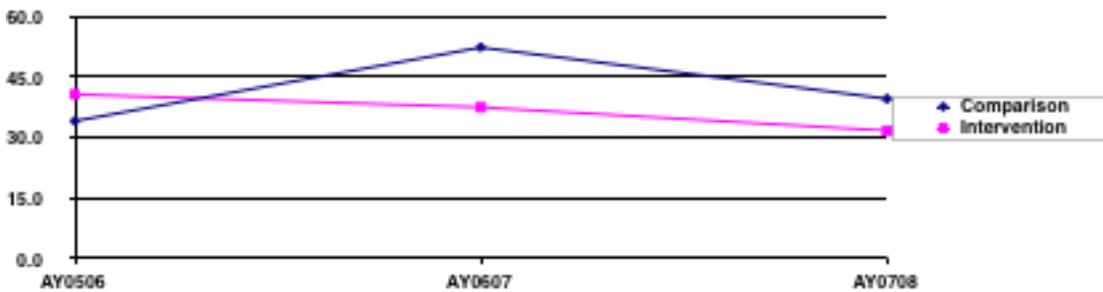==**SCI Category 1: Lowest scorers – significant differences found**==

==There is significant upward trend for both the comparison and program showing significant improvement over the years (both lines have significant slope). And there are significant differences between means for AY 0607 and 0708.==

Category 1: Lowest scorers	N	AY0506	AY0607	AY0708
Comparison	3	7.0	12.6	24.3
Program	4	13.0	37.7	48.2
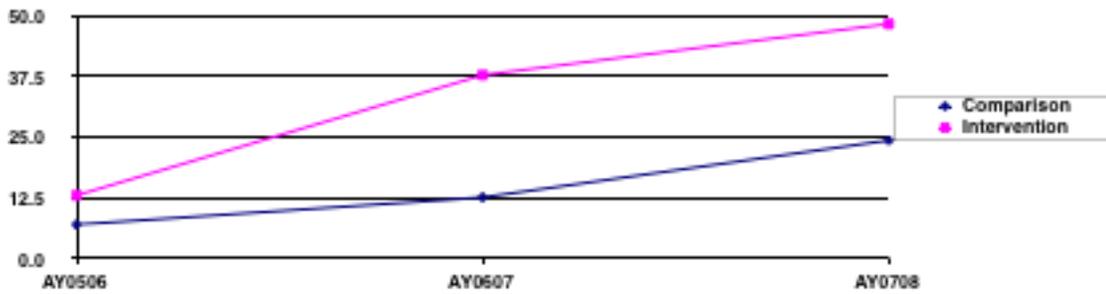
SCI Category 2: Below average scorers – no significant differences found

Category 2: Below average scorers	N	AY0506	AY0607	AY0708
Comparison	9	34.0	52.2	39.5
Program	6	40.6	37.4	31.6

SOC Category 1: Lowest scorers – no significant differences found

Category 1: Lowest scorers	N	AY0506	AY0607	AY0708
Comparison	5	18.2	19.8	34.2
Program	2	25.5	35.0	22.0
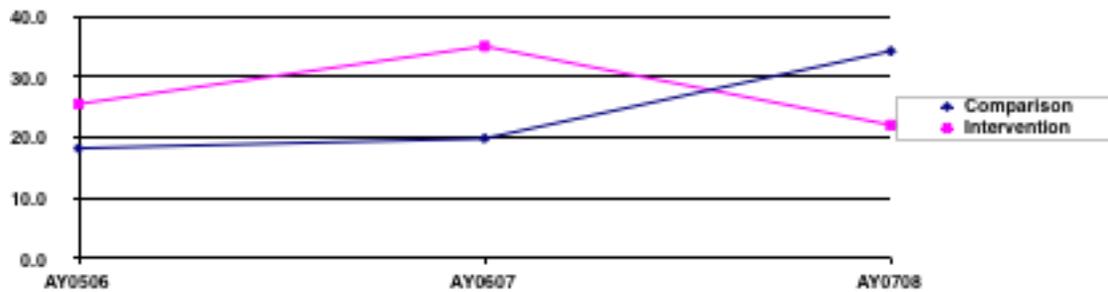
SOC Category 2: Below average scorers – INSUFFECIENT DATA FOR ANALYSIS