

Project WILD Evaluation Learner Knowledge and Attitude Gains

The Environmental Program Evaluation Working Group
The Ohio State University
2000-2001

Joe E. Heimlich, Ph.D.
Diane Cantrell, Ph.D.
Hongxia Duan, MS

Summary:

This report summarizes the findings of a highly controlled study using a variety of pre- and post-measures with treatment and control groups to measure the impact of Project WILD on student learning and attitudes. One unit from the Project WILD framework was selected and participating teachers from eight classrooms conducted five activities from that unit.

The evaluation instrument consisted of three parts: a knowledge test of twenty-three items, an attitude measure of twenty-nine items, and five demographic characteristics. Teachers were asked to complete a class demographic profile for the purpose of matching the treatment and control groups.

Findings indicated that Project WILD does have an impact on knowledge about wildlife and attitudes toward wildlife and constructs of wildlife in general. It was determined that no single activity can teach all the concepts of the framework, but when a series of activities are used within a unit of the framework, the desired learning about these concepts is achieved.

There was a clear pattern of knowledge gain through Project WILD. Students in the control groups tended to have consistent or negative gain scores in pre- and/or post-measures, while students in the treatment groups had positive gain scores overall. In general, there were no significant differences between treatment and control group scores in pre-measure comparisons, which suggests that any gains could be attributed to the use of Project WILD.

Findings also indicated that using Project WILD does have an impact on student attitudes toward wildlife and related constructs of wildlife. As attitudes are slow to change, the immediacy of the measure serves as an indicator for directional attitudinal shift: students participating in Project WILD activities had a slightly higher attitude measure, but with a lower standard deviation, indicating a consistency in the response patterns of the students. This study supports previous studies of the impact of Project WILD participation on attitudes toward wildlife and constructs of wildlife. The verification of this controlled study suggest that other findings are likely generalizable to larger populations.

Background

This evaluation effort fits within the “Impact Studies” component of the Project WILD evaluation plan designed by the Institute for Learning Innovation. The evaluation questions addressed by this study fall under the a broader one, “How does Project WILD affect behavior, knowledge, skills, and attitude of users?”

As recommended by the ILI design, the study follows a quasi-experimental design with treatment and control groups and used a Solomon-Four design. The subquestions include “how are activity use and implementation correlated with learning,” and “how do students intend to behave (toward the environment) and why?”

The evaluation was designed to measure knowledge gain and affect shift. The goal is to measure gain in knowledge through exposure to Project WILD and to ascertain the effect of this exposure on student attitudes. One of the benefits of this type of study is to be able to make claims regarding the use of the materials. Therefore, the study was specific and restricted. Teachers involved in the study treatment were required to use at least five of the activities from the first unit, Ecological Principles, of the new Project WILD activity guide. This was done to ensure that students had the same opportunities for information acquisition, and enough repetition of affect to determine if there is a short-term shift in attitude orientation.

There were eight classrooms (n=224) involved in the preliminary study. This equates to two classes at each of four levels of treatment. The treatment consisted of teaching at least 5 of 7 WILD activities that had been preselected as a teaching unit. The first treatment level involves a pre-test, the treatment, and a post-test. The second level is pre-test and post-test with no treatment. The third level uses treatment and a post-test and the fourth level is post-test only. In this manner, it is possible to measure across groups to minimize test and external influences. All posttests were conducted at the same time by classrooms paired by control and treatment. Posttests were also conducted in the paired classrooms at the same time.

Instrumentation

The evaluation instrument consisted of three parts: a knowledge test of twenty three items, an attitude measure of twenty-nine items, and a five demographic characteristics. Teachers were asked to complete a class demographic profile for the purpose of matching the treatment and control groups.

The knowledge test was constructed based on the “Ecological Knowledge” section of the “Conceptual Framework” that is outlined in the Project WILD guide (pp. 461-465). The concepts in the framework serve as the foundation for the ecological activities in the first section of the book (p. 2-176). The questions were all multiple choice with one correct answer and three realistic distractors. The items were reviewed to ensure that the concepts were included in at least two of the activities in the teaching unit and then reviewed by a panel of experts to determine age appropriateness and accuracy. The test was also reviewed by fifth grade teachers (inner-city school teachers who use Project WILD) to verify that the items would be appropriate

and answerable by fifth grade students having experienced the WILD activities. The test was constructed to be difficult and to cover the entire ecological knowledge framework, even though five activities from the unit would not cover all constructs. In this way, gain scores could be maximized. By creating a core of questions that assume vocabulary without testing vocabulary, and assume requisite knowledge rather than testing basic knowledge, the instrument is more likely to have low pre-measure scores and thus reveal gain on a more sensitive level. In addition, the more difficult instrument would create a more consistent response pattern in “guessing” and so should minimize scores due to guessing. In the implementation, all items had response patterns with standard deviations of .5 or less and Bartlett’s Test of Sphericity revealed a significance of .000.

In the item rotation analysis (IRT), there were seven items that could be challenging the structure of the instrument. When additional data are gathered (see note in next steps), questions 1, 14 and 19 (needs of living organisms; changes in habitats; and transition zones) will be analyzed using IRT to gauge the level to which the items are/are not present in the activities. These items all had individual potential to raise the overall scale integrity by .06. Several other items, 2, 6, 8, and 9 (carrying capacity; niche; energy cycle; and predator/prey relationships) could individually raise the integrity by .05. Item analysis with additional data will clarify the strength and value of maintaining these items in the instrument.

The attitude measure had two dimensions. Some items were general in nature and, therefore, applicable to any of the Project WILD activities. Others measured attitude and intent related specifically to the teaching unit. After a review of all studies of Project WILD, plus an exploration of other attitude measures of youth’s attitudes toward nature and wildlife, it was determined that the scale developed by Lyn Fleming (1982) had both the strongest reliability and appropriateness for the study. Several items were modified slightly (due primarily to language issues and time/cultural change in society) and some items were dropped that did not have clear direction for attitude. In this study, the alpha was .6327 for the standardized item reliability coefficient. The Chi Square on the application of the instrument for the total scale revealed a significance of .000. There were no items that on an orthogonal rotation would raise the alpha by any significance.

Instruments were developed and tested through March 2001. Due to personnel changes in the state coordinator’s office for Project WILD, there were delays in obtaining names and contact information for teachers. This delay forced the study to be conducted during the final weeks of the school year. Several of the teachers’ addresses were incorrect and they received their packages of materials too late to participate in the study. For this reason, a replication study is planned for autumn 2001 using the twelve classrooms that had agreed to participate but were unable to due to the timing.

Findings

Significance

There were no statistically significant differences in scores between any of the pretest measures (pre- treatment; pre-control), and between any of the pre-measures and the post control measure with the exception of one pretest control group with the post control group (.038). This suggests that a test effect may be an influence on groups are not exposed to the Project WILD activities and as the standard deviation was .236 for the pre/post group and .350 for the post only, the significance is more likely.

Although there were important gains in knowledge between control and treatment groups, there were no significant relationships between pre and post attitude or knowledge scores for either the treatment or control groups. These findings will be discussed below.

Significant Relationships

	O O	X O	O
O X O	Not significant	Not significant	Not significant
O O		Not significant	* p= .038
X O			Not significant

Knowledge

On the pre/post measures, there was a gain in the learning of .910 (which would translate to approximately 5% score increase) in the treatment group. There was a negative gain of .375 (or approximately 4% score decrease) in the control group. The median score on the treatment group went from a 5.5 to a 7.0, and the mode increased from 2.59 to 6.0. For the control group, the median and mode scores on both the pre-and the posttests were 8.0.

This suggests that there was a discrepancy in the prior knowledge between the treatment and the control groups, though not in a pattern of statistical significance. This is supported by the only significant difference being between the pre/post control group and the post-only control group. The gain in both median and mode scores for the treatment group with only a minor gain in the mean score suggests that the outlying high scores that raised the mean were regressed to the mean so that the posttest score reflects a more consistent knowledge across the students. The consistent median and mean scores of the control groups coupled with the reduction in the mean score suggest that there were a number of students in the control group who were consistent in their higher scores (greater preexisting knowledge) and the regression to the mean of other students likely caused a lowering of the overall mean of the control group on the posttest. These observations are supported also by the consistent standard deviations of 2.391 to 2.219 for the control group pre/post; and 2.799 for the treatment group.

The treatment only group had a mean score of 7.082 with a median score of 7.0 and an mode of 8.0 with a standard deviation of 2.362. The control group, post-test only, had a mean score of 6.54 with a median score of 6.0 and a mode of 5.0 with standard deviation of 2.95. The lower standard deviation, though slight, indicates that along with the higher score, the treatment group post score was a result of the treatment. That the control group, posttest only score was not significantly different from the pretest scores of both treatment groups supports this observation.

Overall, the post-test comparisons between treatment and control groups reveal a clear pattern of both a higher mean score (7.085 versus 6.85) and a lower deviation (2.54 versus 2.71). The standard error mean was also reduced in the treatment group signifying a more consistent pattern of correct responses across the treatment groups.

Knowledge Gain Scores and Post Means

	Gain Score Changes	Post-only Mean Scores
Treatment	.910	7.082
Control	- .375	6.540

Attitude

The treatment groups, both pre/post and post only, had slightly higher mean scores on attitude outcome measures (3.180 versus 3.117) with a slightly lower standard deviation (.311 versus .314) which again suggests that the treatment of the WILD activities normed attitude measures at a slightly higher level than those who did not participate in the WILD activities. As attitudes are slow to shift, the individual measures of treatment and control reveal that the gain is, in part, the narrowing of the deviation between the pre/post treatment and control groups.

There were no significant differences in any of the pre-test measures or any of the pretest attitude measures with the post-only control group. There was a statistically significant difference between the pretest measures and the treatment post-test group ($\alpha = .038$). It is interesting that the only statistically significant difference occurred with the post-only treatment group which could suggest that the attitude of students using WILD are more likely affected when there is not the “framing” of the pre-measure.

Teacher and Student characteristics

To control for external influences, teachers were paired for the level of treatment on several factors: rural/urban/suburban school; public/private; number of students, average income of students’ families; years of experience; academic degree areas of study; and subject(s) taught. For the pairs, the only variation was in number of years teaching, though one pair did have a discrepancy in average income of families (less than \$25,000 versus \$25,000-50,000).

Students were asked a few questions regarding prior involvement in natural settings, and memberships in youth groups that have an environmental focus. Of all the youth characteristics (sex, age, memberships, prior experiences), only two had significance:

- 1) The treatment groups revealed a significance in attitude scores for those students who belonged to “Campfire” (.029) in the posttest. Campfire membership did not have a statistically significant relationship in the post measure overall, however.

- 2) There was also a significant relationship (.046) of membership in Boy Scouts in the pre-test of attitude for the control and treatment groups. There was no such relationship revealed in posttest measures.

These findings suggest that student memberships, prior experiences, sex, and age were not influences on the measures.

Summary Observations:

- 👉 Statistically, the four groups (pre/post treatment; pre/post control; post only treatment; and post only control) were not different. This allows for comparisons among the groups. This also allows for generalization to other fifth grade classrooms that have similar profiles.
- 👉 There is a gain in knowledge through the use of the WILD activities as constructed for this study. The higher mean, median and mode scores of the control group on the pretests suggest that, because it was the end of the school year, the control group classes had a higher level of requisite knowledge. That the gain was positive for the treatment group and the standard deviation narrowed. Coupled with the especially strong gain score in the mode score from 2.59 to 6.0 and the negative gain for the control group suggests that the WILD activities did teach constructs within the framework that were components of the knowledge test.
- 👉 The connections between the framework and the activities are not necessarily explicit. While the seven preselected activities contained concepts from the framework, these concepts may not have been central or explicit to the learners who participated in the activities. This creates an opportunity for Project WILD to demonstrate how activities may be used to support other lessons and to do so, the teacher should make these ties explicit. Without such connection, the framework concepts will not be consistently measurable at a level of significance.
- 👉 Project WILD does have an influence on attitudes of students toward wildlife and constructs of wildlife. Prior studies reveal that Project WILD creates a solid affective foundation and increases support over time.
- 👉 Challenges to significance: the increased knowledge scores on the treatment groups were mitigated by initially higher scores on the control groups. Gain scores were positive on the treatment groups and negative on control, and the median and mode scores of the control group suggest that the higher mean on the pretest could be accounted for through regression to the mean. The increase in the treatment group is more consistent as mode and median scores increased in proportion to the mean scores. However, the gain score between the treatment and control groups in knowledge was reduced in significance due to the mean being used for the significance measure.

Next Steps

- 👉 Due to timing constraints, several teachers willing to participate in the study were unable to do so for this cycle. Data will be collected from these teachers and added to the existing database. The increase of classes at each level of treatment or control may solidify the gain score patterns and with the increased numbers, may make these gains statistically significant.
- 👉 This autumn, an Item Response analysis will be conducted on the knowledge component of the instrument. This is being done to alter the knowledge questions used for analysis. Expert review identified that the activities do include concepts from the framework, but as mentioned above, these may not be central to the activity. The Item Response analysis will reveal those items that may be skewing scores. Elimination of those items in a statistical analysis may more likely reveal cognitive gain and significance as suggested by the narrowing of standard deviation of scores in the treatment groups.
- 👉 There is a planned replication of the study as a parallel Solomon-Four design for a masters thesis in Massachusetts. These data can be compiled with existing data and create a stronger base from which to draw conclusions.