

Hydroville Curriculum Project

Interim Evaluation Report:
Mysterious Illness Outbreak (MIO)
Implementation 2002-2003

Donald G. MacGregor, Ph.D.

MacGregor-Bates, Inc.

1010 Villard Ave.

PO Box 276

Cottage Grove, OR 97424

Tel: (541) 942-5727

Email: donaldm@epud.net



Background

Evaluation Design – Overview: The evaluation design for MIO was essentially the same as for Pesticide Spill. The most significant changes were:

- Reduction in the number of evaluation items on the pretest and the posttest instruments (from 56 items to 31 items); and,
- Placement of all Background Activity Knowledge Items in a single set and administered at the end of the Background Activities as a battery rather than after each activity.

The reduction in the number of evaluation items resulted in the loss of one of the measurement scales. Some of the other scales were shortened.

Evaluation Design – Problems and Challenges: The problems and challenges are the same as the previous year.

Key Results: Teacher Evaluations

This aspect of the evaluation is currently under review. A review of the teacher evaluations indicated some inconsistencies in teachers' responses to the questions in the instrument concerning how much of each of the Background Activities they used (e.g., All, Part, None), and their reasons for modifying the Activities if they did so.

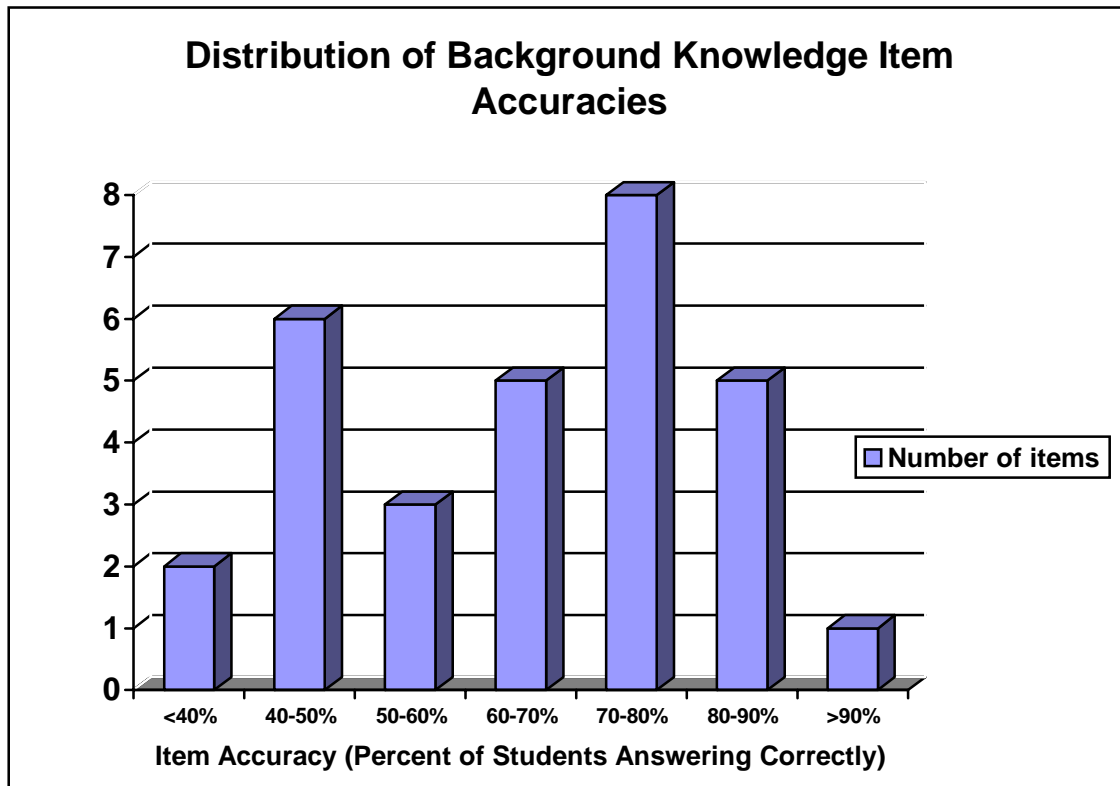
It appears from a first pass through the data that the curriculum was very unevenly implemented. While over 90% of the teachers used the "John Snow" module as provided, utilization of the other modules ranged widely from approximately 70% of teachers using a module to less than 30%. The protocols also suggest that some teachers were quite consistent in their utilization of modules (e.g., used all or most modules) while other were much less consistent and may have used only a small number of modules (e.g., two or three). Pending confirmation of this, we may want to re-analyze some of the student results based on the *degree* to which the curriculum was implemented (e.g., students receiving 8 or more modules vs. students receiving 4 or less). Sue is currently reviewing the teacher evaluation data with these issues in mind.

Key Results: Student Evaluations

Background Activity Knowledge Items. The MIO evaluation contained a battery of 30 Background Activity Knowledge Items administered after all of the Background Activities had been completed. We analyzed these results in terms of (a) individual item accuracies, and (b) student performance across all of the items in the battery.

Individual Item Accuracies. The 30 items varied considerably in terms of difficulty. Figure 1 shows a distribution of the background knowledge item accuracies.

Figure 1. Distribution of Background Knowledge Items by Percent Correct.



The results in Figure 1 are for all students who completed the Background Activity Knowledge Items. The figure shows the number of items falling into each of seven accuracy groupings, where each grouping represents the percentage of all students who got the item correct. As an example, for two of the items the percentage of students who got those items correct was less than 40%. Only one item was answered correctly by more than 90% of the students.

The two *most difficult* items in terms of percent correct were:

First step in investigation of reported cases of respiratory illness (35.8% correct)

Order signal words on chemical labels from most toxic to least toxic (38.9% correct)

The two *least difficult* items were:

Which county had highest number of case in 2002? (94.2% correct)

Avg. length of lettuce seed radicle if a 30% NaCl solution was tested (83.2% correct)

Further analyses on these data are possible and we will want to review these results before proceeding.

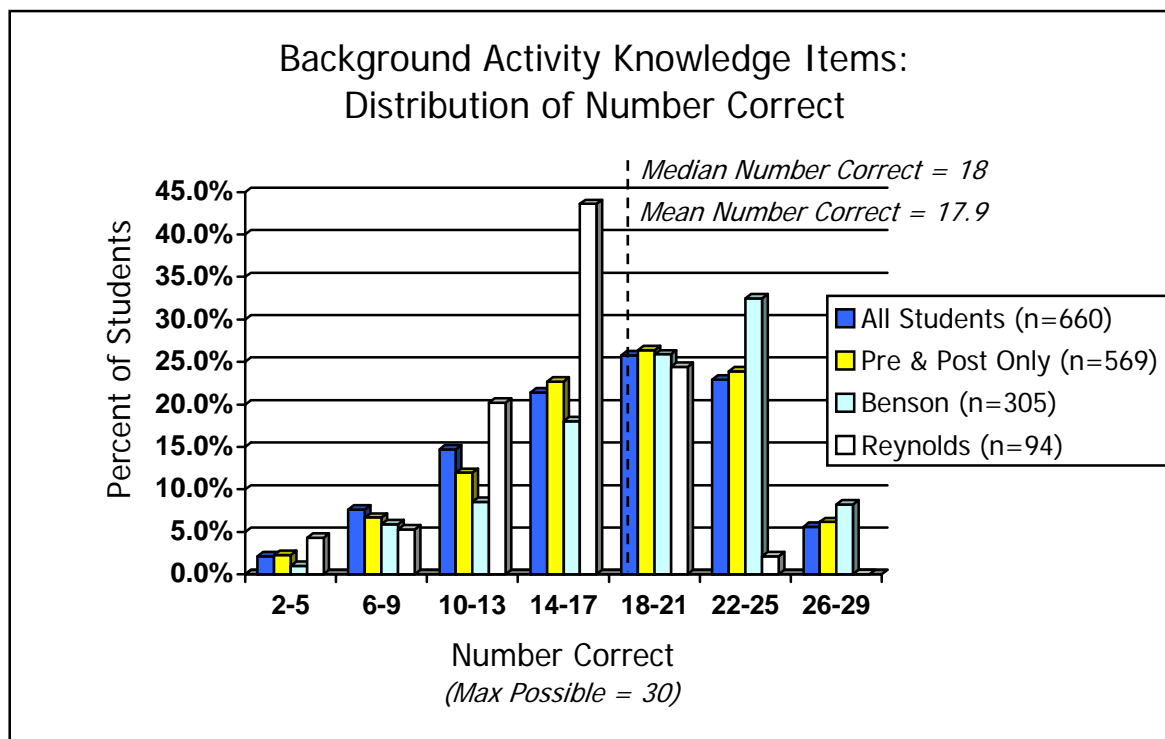
Student Performance Across Items. Looking at the set of Background Activity Knowledge Items as a 30-item test, we can evaluate how well students performed. Table 1 show the average percent correct that students achieved overall and for each of the seven sites in the program. The overall percent correct was 59.6%. However, there was considerable variance by site. Benson, Westview and Crescent Valley had relatively high student performance on the battery, with performance exceeding 60% on average. AIM, Reynolds and LaGrande were much lower with performance in the 30% to 50% range. In some cases, particularly large sample sizes (e.g., Benson, Reynolds) or particularly small sample sizes (e.g., LaGrande, AIM) make between school comparisons difficult.

Table 1. Background Knowledge Items: Average percent correct across 30 items by school and overall.

School	Number of Cases	Average % Correct
Benson	305	63.6%
Gateway	46	57.8
Reynolds	94	49.4
Westview	52	64.7
AIM	13	47.7
Crescent Valley	42	62.1
La Grande Alt. HS	17	32.3
<i>Overall</i>	<i>569</i>	<i>59.6%</i>

The summary statistics in Table 1 can also be looked at in terms of their distributional characteristics. Figure 2 shows distributions of the number correct for various student and site groups. In the figure, number of items correct is shown categorically on the horizontal axis, while the percentage of students achieving that number correct is shown on the vertical axis.

Figure 2. Distribution of Background Activity Knowledge Number Correct for all students (n = 660), students completing both Pretest and Posttest (n = 569), Benson students (n=305), and Reynolds students (n=94).



The median number of items correct across all students was 18 items (out of a possible 30). The “All Students” group are those students who answered the Background Knowledge Items regardless of whether they also completed both the pretest and posttest evaluations. The “Pre & Post Only” group is that subset of students who completed the BA Knowledge Items and completed both the pretest and posttest evaluations. Also shown in the figure are separate distributions for the Benson and Reynolds sites.

The “Pre & Post Only” group did slightly better than “All Students” on the knowledge items with higher percentages of students getting a higher number of items correct. The most pronounced outcome in the figure is the strong performance of Benson students both when compared to all students and when compared to the Reynolds group. The Reynolds site tended to underperform the average and underperform Benson. Our preliminary review of the teacher evaluation results suggests that the implementation of the curriculum may have been most complete at Benson, and that a subset of the Benson teachers may have implemented most of the modules. This factor could account for their high level of student performance on the BA Knowledge Items.

We may want to review these results in light their potential meaning for analyzing some aspects of the pretest/posttest evaluation.

Problem Solving Pretest and Post Test: The Problem Solving Pretest and Post Test consisted of a 31-item evaluation instrument that was composed of three separate modules and seven distinct measurement scales or concepts:

Table 2 shows each of the seven measurement scales in terms of (a) scale name, (b) scale code, (c) items comprising each scale as keyed to the original instrument, and (d) a description of the scale in terms of the relationship between numerical and conceptual values. The scales are also summarized below:

- *Importance of Explanations:* This 7-item scale assesses the importance students gave to seven possible explanations for differences in cancer rates between two communities. Higher scores reflect higher importance assigned to explanations. The purpose of the scale is to assess how well students are able to discriminate between various explanations for a health effect.
- *Information Seeking:* This 6-item scale assesses students' judgments of the informativeness of a number of pieces of information in the context of determining the possible cause(s) of difference in cancer incidence rates between two communities. Higher scores reflect greater informativeness of information items
- *Personal Involvement:* This 4-item scale assesses students' level of comfort with personal involvement in a number of tasks associated with working on a community committee to help identify the cause(s) of differential cancer rates between two communities. Higher scores reflect greater comfort with personal involvement.
- *Intuitive Toxicology:* This 2-item scale assesses students' perceptions of the relationship between chemical exposure and health effects or outcomes. Lower scores reflect greater agreement with "expert" views about toxicology.
- *Self-Protection:* This 2-item scale assesses students' perceived ability to protect themselves from chemical hazards. Higher scale values reflect greater perceived self-protection ability.
- *Sensitivity to Chemical Risks:* This 3-item scale assesses students' attitudes and perceptions of chemical risks in the home and in the environment. Higher scores reflect greater sensitivity to chemical risks.
- *Value for Science:* This 11-item scale assesses students' perceptions and attitudes toward science in society and as part of their daily life. Lower scores reflect a lower value for science.

Table 2. Hydroville Evaluation Scales: 2002-2003 Implementation Year
(Mysterious Illness Outbreak)

Scale Name	Scale Code	Items	Description
Importance of Explanations		mv1 through mv7	7-item scale. Higher score reflects greater importance assigned to explanations.
Intuitive Toxicology		cp23, cp26	2-item scale. Lower score reflects greater agreement with “expert” views.
Information Seeking		mv8 through mv13	6-item scale. Higher score reflects greater informativeness of items.
Sensitivity to Chemical Risks		cp20, cp21, cp22,	3-item scale. Higher score reflects greater sensitivity to chemical risks.
Self-Protection		Cp24, cp25	2-Item scale. Higher score reflects greater perceived ability for self protection.
Personal Involvement		mv14 through mv17	4-item scale. Higher score reflects greater comfort with personal involvement.
Value for Science	<i>ValueSci</i>	cp27, sc28 through sc32	6-item scale. Higher score reflects higher (more positive) value for science.

These are essentially the same scales as for the Pesticide Spill evaluation with two notable exceptions. First, the reduction in the number of items on the evaluation instrument resulted in the loss of the *ImmedAction* scale. Second, the original *ChemRisk* scale has been divided into two scales: a *ChemRisk* scale that measures risk attitudes toward chemical hazards, and a *SelfPro* scale that measures students' perceived ability to self-protect against chemical risks. Some of the other scales have a reduced number of items compared to the number of items in the same scale in the Pesticide Spill evaluation.

Regarding the split of *ChemRisk* into two separate scales, we plan to redo the Pesticide Spill analysis with respect to these two scales. This will yield a more comparable analysis of the evaluation data for the 2001 and 2002 implementations.

Scale Means by Site and Total Sample. Table 3 shows the scale means by site and for the total sample. There were considerable differences in the statistical power available at the different sites: sample sizes ranged from a high of 313 (Benson) to a low of 19 (LaGrande). This variance in statistical power suggests that greater concern needs to be given to controlling Type II error than might otherwise be the case, and the results shown in Table 3 are reported down to a significance level of $p < .10$. This difficulty was experienced as well for the Pesticide Spill evaluation.

Examining the contents of Table 3 we can see that there are some fairly consistent effects across schools for some of the scales, and less consistent effects for others:

- A moderately strong scale result occurred on the *IntuiTox* scale: of the seven sites, five showed statistically significant changes from pretest to post test on this scale as did the total sample ($p < .001$). The direction of change from pretest to posttest reflected greater agreement with “expert” views about toxicology and the relationship of chemical exposure to health outcomes.
- A fairly strong pattern of significant results was also obtained on the *SelfPro* scale: three of the seven sites reached significance, as did the total sample ($p < .001$). The direction of change from pre to post indicates that students became more confident in their ability to self-protect against chemical hazards.
- A moderately significance effect was obtained on the *InfoSeek* scale: three of the seven sites reached significance, as did the total sample ($p < .01$). The direction of change from pretest to posttest indicates that students found the information items to be *less* informative. We can speculate that the content of the curriculum may have made students more aware of the complexities of epidemiology and, therefore, caused them to see relatively simple information gathering activities as less informative than they previously believed. Other potential hypotheses could be explored to account for this effect.

Table 3. Scale pretest and posttest means, and mean difference scores by school and for total sample.¹

Site	<i>ChemRisk</i>	<i>SelfPro</i>	<i>Explanation</i>	<i>InfoSeek</i>	<i>IntuiTox</i>	<i>PersInvolv</i>	<i>ValueSci</i>
Benson (n=313)							
Pre	2.81	2.88	2.76	3.22	2.86	3.16	2.92
Post	2.78	3.03	2.81	3.18	2.67	3.18	2.89
Diff	-.03	0.15**	.04	-.04	-.19***	.02	-.03
Gateway (n=52)							
Pre	2.72	2.81	2.83	3.39	2.59	3.09	2.82
Post	2.74	2.88	2.84	2.28	2.83	3.22	2.92
Diff	.01	0.08	.01	-.10	.24*	.11	.10
AIM (n=47)							
Pre	2.89	2.86	2.63	3.31	2.64	3.17	2.69
Post	2.89	2.95	2.87	3.23	2.33	3.07	3.68
Diff	.00	0.09	.25**	-.08	-.31*	-.10	-.01
Reynolds (n=101)							
Pre	2.71	2.80	2.75	3.27	2.79	3.05	2.97
Post	2.77	3.10	2.86	3.17	2.58	3.08	2.90
Diff	.06	0.30**	.11*	-.19*	-.21*	.04	-.07
Westview (n=52)							
Pre	2.66	2.82	2.64	3.25	2.69	3.09	2.94
Post	2.66	3.01	2.57	3.08	2.54	3.07	2.89
Diff	.00	0.19 ^a	-.07	-.17**	-.14	-.02	-.05
C. Valley (n=48)							
Pre	2.78	3.06	2.89	3.42	2.74	3.13	3.03
Post	3.91	3.12	2.79	3.32	2.98	3.21	2.80
Diff	.13	0.06	-.10	-.10	.24 ^a	.08	-.22**
La Grande (n=19)							
Pre	2.82	2.68	2.36	2.89	2.82	2.41	2.61
Post	2.83	3.05	2.51	3.09	2.64	2.85	2.86
Diff	.01	0.36	.14	.20*	-.18	.44**	.25**
TOTAL (n=632)							
Pre	2.78	2.86	2.75	3.26	2.78	3.11	2.90
Post	2.79	3.03	2.79	3.19	2.65	3.14	2.87
Diff	.01	0.16***	.04*	-.07**	-.13***	.03	-.03

^a p<.10

* p<.05

** p<.01

*** p<.001

****p<.0001

¹ Mean Difference Scores are computed across subjects by taking the average difference of the posttest minus the pretest.

- A weak but significant effect was obtained on the Explanation scale: two of the seven sites reached statistical significance as did the overall sample ($p < .05$). The direction of change from pretest to posttest indicates that students found the explanations for the differential cancer rates in the Mayville scenario as more important. The scale effect is, however, largely attributable to two items that called upon students to assess the importance of “differences in the use of household chemicals” and “differences in ages of people” who live in the two comparison communities. With respect to the item concerning age, the content of the “John Snow” module was heavily oriented toward demographic risk factors in accounting for disease rates, and age was among the factors presented. The “John Snow” module was the first module in the curriculum and was the module most completely implemented.

Table 4 below summarizes the results in Table 3, showing by site the particular scales reaching significance at each site.

Table 4. Summary of sites with significant pre/post change scores.

Site	Scales with Significant Pre/Post Change Scores
Benson HS	<i>IntuitTox</i> ***, <i>SelfPro</i> **
AIM	<i>IntuiTox</i> *, <i>Explanation</i> **
Gateway LC	<i>IntuitTox</i> *
Reynolds HS	<i>IntuiTox</i> *, <i>SelfPro</i> **, <i>InfoSeek</i> *, <i>Explanation</i> *, <i>PersInvolv</i> *
Westview HS	<i>SelfPro</i> ^a , <i>InfoSeek</i> **
Crescent Valley HS	<i>IntuiTox</i> ^a , <i>ValueSci</i> **
La Grande Alt. HS	<i>InfoSeek</i> *, <i>PersInvolv</i> **, <i>ValueSci</i> **
<i>Overall</i>	<i>Intuitox</i> ***, <i>SelfPro</i> ***, <i>Infoseek</i> **, <i>Explanation</i> *

^a $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Individual Scales: Selected Item Results. Selected item results for the scales *IntuiTox*, *SelfPro*, *InfoSeek*, and *Explanation* are shown in Tables 5, 6, 7, and 8 respectively. For the *IntuiTox* (Table 5) and the *SelfPro* (Table 6) scales, all scale items are shown for each of the seven sites and overall. For the *InfoSeek* scale (Table 7) the three scale items contributing most to the significant outcome are shown for all sites combined only. For the *Explanation* scale (Table 8) the two scale items contributing most to the significant outcome are shown by site and overall. I suggest reviewing these four tables, and we can discuss how we might want to interpret these results in light of how the curriculum was implemented.

Summary Notes

The following are some general notes and thoughts on these results:

- The outcome of the evaluation for MIO was surprising to me. I was not expecting as much of a pre-post effect as was obtained, particularly on the *IntuiTox* scale.
- The results on the *SelfPro* scale are interesting and appear to parallel a similar set of results we obtained on Pesticide Spill: from pretest to posttest students seem to gain confidence in their ability to protect themselves against chemical risks in the home and confidence in their ability to read and understand warning labels. The implication(s) of this result is something we should discuss. For example, is this change in confidence appropriate?
- I am concerned about the very high level of variability in how the curriculum was implemented. We need to look more closely at any relationship this might have to BA Knowledge performance, and to pre/post changes in the student evaluations.
- Essentially, we have only one solid site in this evaluation and that is Benson. Of students who completed both pretest and posttest, over half are Benson students. In addition, it appears that the curriculum may have been implemented more (or at least as) completely at Benson compared to all other sites. This strongly reinforces the Benson results.
- Collapsing the analysis across sites is an issue to discuss. It only makes sense to collapse across sites if (a) the sites have similar teaching approaches, and (b) the curriculum was implemented to a similar degree and in a similar way across sites. The former is more easily determined than the latter.
- The Reynolds site performed surprisingly poorly on the BA Knowledge Items. Again, this may be related to how the curriculum was implemented.
- I will be thinking about the implications of all of this for the Air Quality implementation and evaluation. We should schedule a meeting to review these preliminary results.

Don MacGregor

Table 5. Pretest mean, posttest mean, and mean difference scores for two items comprising the *IntuiTox* scale (Lower score equals stronger agreement with expert model of toxicology).

<i>“If a chemical pesticide is released into the environment, then everyone in that environment is exposed to the chemical.”</i>	N	Pretest	Posttest	Difference
Benson	313	3.01	2.75	-0.24 ^{***}
Gateway	52	2.65	2.79	0.20
Reynolds	101	2.86	2.60	-0.31 ^{**}
Westview	52	2.59	2.57	0.02
AIM	47	2.74	2.38	-0.43 ^a
Crescent Valley	48	2.80	2.98	0.23
La Grande	19	2.92	2.76	-0.08
<i>Overall</i>	<i>632</i>	<i>2.88</i>	<i>2.71</i>	<i>-0.16^{***}</i>

<i>“If a person is exposed to a chemical pesticide that can cause cancer in humans, then that person will probably get cancer some day.”</i>	N	Pretest	Posttest	Difference
Benson	313	2.65	2.57	-0.09
Gateway	52	2.45	2.67	0.17
Reynolds	101	2.65	2.53	-0.19 ^a
Westview	52	2.69	2.50	-0.30 ^a
AIM	47	2.51	2.33	-0.19
Crescent Valley	48	2.75	2.76	-0.04
La Grande	19	2.85	2.50	-0.40
<i>Overall</i>	<i>632</i>	<i>2.64</i>	<i>2.56</i>	<i>-0.11^{**}</i>

^a p<.10

* p<.05

** p<.01

*** p<.001

**** p<.0001

Table 6. Pretest mean, posttest mean, and mean difference scores for two items comprising the *SelfPro* scale (Higher score equals great perceived self-protection ability.)

<i>“For most of the chemicals I am exposed to in daily life, including chemical pesticides, I feel I know how to protect my health and safety.”</i>				
	N	Pretest	Posttest	Difference
Benson	313	2.75	2.84	0.07
Gateway	52	2.61	2.85	0.15
Reynolds	101	2.71	2.91	0.23*
Westview	52	2.62	2.94	0.38**
AIM	47	2.74	2.79	0.06
Crescent Valley	48	2.89	2.82	-0.09
La Grande	19	2.75	3.11	0.33
<i>Overall</i>	<i>632</i>	<i>2.73</i>	<i>2.87</i>	<i>0.12**</i>

<i>“I can read and understand most of the warning labels on chemical pesticides in my home.”</i>				
	N	Pretest	Posttest	Difference
Benson	313	2.96	3.21	0.24***
Gateway	52	3.02	3.08	0.02
Reynolds	101	2.91	3.16	0.31**
Westview	52	3.02	3.06	0.00
AIM	47	3.26	3.28	0.05
Crescent Valley	48	3.10	3.44	0.37*
La Grande	19	2.67	2.89	0.29
<i>Overall</i>	<i>632</i>	<i>2.99</i>	<i>3.19</i>	<i>0.21***</i>

^a p<.10

* p<.05

** p<.01

*** p<.001

**** p<.0001

Table 7. Pretest mean, posttest mean, and mean difference scores for three items from the *InfoSeek* scale (Higher score equals greater information value).

	Pretest	Posttest	Difference
“Have a soil test done to determine what chemicals may be in the soil.”	3.65	3.54	-0.11**
“Check local library for newspaper reports about the plant.”	3.01	2.95	-0.06 ^a
“Check the Internet for information about causes of cancer.”	3.17	3.03	-0.16***
<i>Overall Scale</i>	<i>3.26</i>	<i>3.19</i>	<i>-0.07**</i>

^a p<.10

* p<.05

** p<.01

*** p<.001

****p<.0001

Table 8. Pretest mean, posttest mean, and mean difference scores for two items from the *Explanation Scale*. (Higher score equals greater information value.)

<i>“Differences in the use of household chemicals”</i>				
	N	Pretest	Posttest	Difference
Benson	313	2.90	3.13	0.23****
Gateway	52	2.88	3.02	0.13
Reynolds	101	2.85	3.18	0.30**
Westview	52	2.55	2.69	0.19
AIM	47	2.62	2.93	0.30 ^a
Crescent Valley	48	3.17	3.21	0.04
La Grande	19	2.57	2.83	0.29
<i>Overall</i>	<i>632</i>	<i>2.85</i>	<i>3.07</i>	<i>0.22****</i>

<i>“Differences in ages of people”</i>				
	N	Pretest	Posttest	Difference
Benson	313	2.28	2.53	0.24***
Gateway	52	2.34	2.67	0.34*
Reynolds	101	2.30	2.70	0.45****
Westview	52	2.20	2.42	0.28
AIM	47	2.20	2.83	0.69***
Crescent Valley	48	2.74	2.64	-0.13
La Grande	19	1.58	2.00	0.33
<i>Overall</i>	<i>632</i>	<i>2.30</i>	<i>2.57</i>	<i>0.29****</i>

^a p<.10

* p<.05

** p<.01

*** p<.001

****p<.0001