Hydroville Curriculum Project

Interim Evaluation Report:
Pesticide Spill (PS)
Implementation 2001-2002

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

NIEHS

Oregon State University | Environmental Health Sciences Center

Hydroville, USA

Background

Evaluation Design – Overview: The evaluation design is a quasi-experimental design that uses a pretest and posttest administration of an evaluation instrument, and a single-administration control group. The pretest administration occurred during the fall of the year, approximately two to three weeks after the beginning of the term and before the curriculum implementation. The posttest administration occurred after the curriculum implementation. The control group administration occurred near the beginning of the fall term.

The evaluation design has an exploratory focus, particularly with respect to measurement development. Previous experience with the SMILE program indicated that for some categories of students, significant changes in attitudes and perceptions about science and the environment could take place as the result of activities-based environmental science instruction. The present evaluation design implemented a broad range of measurement concepts with an emphasis on perception and attitude change. Key measurement concepts included (a) reasoning about chemical risks and toxicology, (b) value for science in daily life, (c) teamwork attitudes, and (d) information skills.

Evaluation Design – Problems and Challenges: The application of quasi-experimental evaluation designs in the context of a real-world classroom setting pose some challenges and compromises. These include (but are not limited to):

- <u>Non-uniform curriculum implementation.</u> Not all background activities were utilized, some were partially utilized and some were adapted to meet teachers needs (e.g., Risk assessment, Decision analysis, Landfill). Consequently, sites and classroom differed in terms of the curriculum (treatment) actually received by students. This effects the evaluation design because in some settings only part of the curriculum may have been implemented, thereby reducing the size of the possible treatment effect. Unless statistical power can be increased by other means (e.g., increasing sample size), the evaluation may tend show that treatment effects do NOT exist when in fact they do.

- <u>Limited ability to monitor teachers' administration of evaluation materials.</u> It was the intention of the evaluation design to control the administration of evaluation materials such that the pretest and posttest closely bracketed the curriculum implementation period. Although teachers made every effort to accommodate the evaluation as part of their classroom activities, some variance in the timing of pretest and posttest administrations occurred. In addition, teachers were not able to closely monitor students' evaluation work, resulting in relatively largely amount of missing or unusable data. Data losses resulted in limited statistical power at some sites due to small sample sizes.

- <u>Non-homogeneous sites (schools).</u> The set of sites at which the curriculum was implemented varied greatly in terms of a number of relevant variables, including the type school program (e.g., integrated vs. non-integrated curricula) and prior educational background of students. The lack of site homogeneity suggests collapsing across all sites to obtain and overall evaluation result may be in appropriate, and that the evaluation should properly be framed as a set of independent implementations. This, however, leaves open the relevance of the control group results since it is unclear with which site they should be compared. In the present design, each student serves as their own control and the value of a separate control group resides in the ability to assess the potential sensitizing effect of the pretest on the assimilation of concepts during curriculum implementation.

Key Results: Teacher Evaluations

Background Activity Evaluations. After each of the 10 background activities, teachers completed a brief evaluation form. Table 1 provides a partial summary of teacher responses. In general, teachers utilized most of the activities as they were given in the curriculum. However, background activities that occurred near the end were more likely to be partially used or not used at all (i.e., Risk Asses., D/A, Landfill). The time budget for the background activities in terms of both teacher preparation and classroom time was in almost every case exceeded, sometimes by a substantial margin. Though this was less the case for later background activities, these tended to be only partially utilized and it appears that teachers may have limited these activities to stay within time constraints.

Expert Groups and Overall Scenario Evaluations. Teacher evaluation forms were also administered for each of the five expert groups as well as for the overall scenario. Table 2 provides a partial summary of teacher responses to these evaluations. With respect to the expert groups, students assigned to a particular expert group tended to complete the prerequisite background activities, except for the Regulatory Compliance expert group. Though on the balance teachers indicated that the background activities prepared students for their involvement in the expert groups, this was more strongly the case for the Mechanical Engineer and Soil Scientist groups than for the other groups.

However, it appears from the evaluation results that students had difficulty accessing and/or utilizing outside experts to help with their expert group. Likewise, teachers were generally not able to involve outside facilitators, mentors and volunteers to assist with the expert groups.

In general, teachers appeared to have implemented the majority of the curriculum as presented to them. However, later parts of the curriculum, particularly modules having to do with risk assessment and decision making, may have been implemented only partially. From teachers' indications of the amount of time they spent in preparation and in the classroom, the curriculum required more extensive preparation and more classroom time than originally budgeted. Though both students and teachers experienced difficulty obtaining outside assistance with the expert groups, the background activities were judged by teachers to adequately prepare students for their expert group work in most cases.

Table 1. Pesticide Spill Background Activity Teacher Evaluations

Activity Use, Prep. Time and Classroom Time	Pump It Up!	Super Soaker	Circum. & Diameter	Chroma-tography	Soil Class.	Soil Perm.	Toxic. Testing	Risk Assess	D/A	Landfill
Used all of activity	13	11	8	11	10	11	11	6	4	4
Used part of activity	1	2	3	1	1	1	0	3	4	3
Didn't use activity	0	1	2	0	1	0	1	3	4	5
Total responses	14	14	13	12	12	12	12	12	12	12
Percent Used All or Part	*100.0%*	*92.9%*	*84.6%*	*100.0%*	*91.7%*	*100.0%*	*91.7%*	*75.0%*	*66.7%*	*58.3%*
Percent Used All	*92.9%*	*78.6%*	*61.5%*	*91.7%*	*83.3%*	*91.7%*	*91.7%*	*50.0%*	*33.3%*	*33.3%*
Target Prep. Time (min.)	25	10	10	20	25	35	60	10	10	10
Average Prep. Time	70	44	28	64	40	40	73	16	17	21
Percent deviation	*180.0%*	*340.0%*	*180.0%*	*220.0%*	*60.0%*	*14.3%*	*21.7%*	*60.0%*	*70.0%*	*110.0%*
Target Class Time (min.)	45	70	45	50	55	55	145	50	55	55
Average Class Time	82	109	73	87	76	81	138	69	60	89
Percent deviation	*82.2%*	*55.7%*	*62.2%*	*74.0%*	*38.2%*	*47.3%*	*-4.8%*	*38.0%*	*9.1%*	*61.8%*

Table 2. Teacher evaluations of expert groups and overall scenario.

	Overall Scenario	Reg. Comp.	Env. Tox.	Chemist	Mech. Eng.	Soil Sci.
Overall Scenario						
Problem Exposition:						
Preparation Time (min).	86					
Class Time (min).	210					
Problem Presentation:						
Preparation Time (min).	28					
Class Time (min).	235					
Expert Groups						
Use of Activity:						
Used All		3	9	9	9	9
Used Some		6	0	0	0	0
Preparation Time (min).		50	90	55	74	82
Class Time (min.)		95	264	226	232	278
"Did students assigned to this group complete all of the prerequisite background activities relevant for this expert group?" (yes/no)	n/a	2/1	6/2	8/1	9/0	7/1
"Did the background activities adequately prepare students assigned to this expert group?" (yes/no)	n/a	1/2	5/3	5/4	7/2	6/2
"Did students use outside experts/scientists to help with this expert group?" (yes/no)	n/a	1/2	2/6	1/8	1/8	0/7
"Were students able to successfully complete the Team Report for this expert group?" (yes/no)	n/a	3/0	7/1	9/0	9/0	8/0
"Were you able to involve outside facilitators, mentors, and volunteers to assist with the scenario in this expert group?" (yes/no)	n/a	1/2	2/6	3/6	3/6	1/7

Key Results: Student Evaluations

Background Theme Evaluations. At the end of each Background Theme, students completed a theme evaluation. A key element of this evaluation included a set of knowledge items relevant to each Background Theme. Table 3 summarizes accuracy scores (percent correct) for each of the eight Background Themes.

Table 3. Accuracy scores for background activity theme knowledge items.

Background Activity Theme	Item #1	Item #2	Item #3	Item #4	Item #5	Item #6	Average Accuracy
Pump It Up!	50.9%	68.7%	38.8%	--	--	--	52.8%
Graphs	78.7	82.7	50.7	--	--	--	70.7
Chromatography	72.6	83.7	60.1	--	--	--	72.1
Soils	75.1	66.2	70.3	45.6%	48.7%	43.7%	58.3
Risk Assess.	72.4	71.5	88.6	--	--	--	77.5
Tox. Testing	66.2	59.8	50.4	46.9	--	--	55.8
Decision Analysis	39.7	62.1	35.9	--	--	--	45.9
Landfill	25.2	63.6	25.9				38.2
TOTAL							*58.9%*

Of the eight theme evaluations, six had three knowledge items each, one had four knowledge items (Tox. Testing), and one had six knowledge items, for a total of 28 knowledge items altogether. All but two of the knowledge items were four-alternative, forced-choice responses. Table 3 shows the accuracy score for each knowledge item by theme evaluation, as well as the average accuracy for each theme. Most of the accuracy scores were above 50%, with only two items (Item #1 and Item #3 on Landfill) scoring at the level of guessing (i.e., 25% for a four-alternative response mode). Average accuracy was quite high, particularly for the Graphs, Chromatography, and Risk Assessment themes. Accuracy was lowest for the Landfill theme. Not shown in Table 3 are the numbers of respondents for each of the accuracy scores shown. These varied greatly depending upon theme, as not all classes completed all of the themes. For the earlier themes in the curriculum (e.g., Pump It Up!), valid responses were on the order of 250 to 280 students, while for later themes (e.g., Decision Analysis, Landfill), valid responses were on the order of 110 students.

Problem Solving Pretest and Post Test: The Problem Solving Pretest and Post Test consisted of a 56-item evaluation instrument that was composed of three separate modules and seven distinct measurement scales or concepts:

- "Mayville." The first module was developed around a fictitious newspaper story reporting on an elevated cancer incidence in the community of "Mayville" as compared to other communities in the state. A battery of 33 items had students rate (a) the importance of various explanations as causes of the cancer rate difference, (b) the priority they would give to each of a number of actions they might take as part of a community committee, (c) the informativeness of each of a number of tests and the like that their committee might undertake, and (d) their personal level of comfort taking on roles in working to solve the problem. These four groups of items corresponded to each of four measurement concepts: (a) *Importance of Explanations*, (b) *Need for Immediate Action*, (c) *Information Seeking*, and (d) *Personal Involvement*.

- "Chemical Pesticides in the Environment." The second module contained 14 items to assess (a) students' attitudes and perceptions about chemical pesticides and their potential risks and impacts on the environment, and (b) their intuitive notions about toxicology and the relationship of chemical exposure to health outcomes. These 14 items assessed two measurement concepts: (a) *Sensitivity to Chemical Risks*, and (b) *Intuitive Toxicology*.

- "Science and Daily Life." The third module contained 9 items pertaining to students' attitudes about science in general and as applied to the problems of daily life. This set of items assessed a single measurement concept: *Value for Science*.

Table 4 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.

- *Importance of Explanations*: This 7-item scale assesses the importance students gave to seven possible explanations for differences in cancer rates between two communities. Lower scores reflect higher importance ratings. The purpose of the scale is to assess how well students are able to discriminate between various explanations for a health effect.

- *Need for Immediate Action*: This 5-item scale assesses students' perceived need to take immediate action in the face of a health effect report. The purpose of the scale is to assess a possible change in student focus from action to analysis as a result of exposure to the curriculum. Lower scores reflect higher priority for taking immediate action.

Table 4. Hydroville Evaluation Scales: 2001-2002 Implementation Year

Scale Name	Scale Code	Items	Description
Importance of Explanations	Explanation	mv1 through mv7	7-item scale. Lower score reflects greater importance assigned to explanations.
Need for Immediate Action	ImmedAction	mv8 through mv12	5-item scale. Lower score reflect higher priority for immediate actions.
Intuitive Toxicology	IntuitTox	cp37, cp40, cp41, cp43, cp44	5-item scale. Lower score reflects greater agreement with "expert" views.
Information Seeking	InfoSeek	mv13 through mv19	7-item scale. Lower score reflects greater informativeness of items.
Sensitivity to Chemical Risks	ChemRisk	cp34, cp35, cp36, cp38, cp39, cp42, c45	7-item scale. Lower score reflects greater sensitivity to chemical risks.
Personal Involvement	PersInvolv	mv20 through mv27	8-item scale. Lower score reflects greater comfort with personal involvement.
Value for Science	ValueSci	cp46, cp47, sc48 through sc56	11-item scale. Lower score reflects lower value for science.

- *Information Seeking*: This 7-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. Lower scores reflect greater informativeness of information items

- *Personal Involvement*: This 8-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. Lower scores reflect greater comfort with personal involvement.

- *Intuitive Toxicology*: This 5-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.

- *Sensitivity to Chemical Risks*: This 7-item scale assesses students' attitudes and perceptions of chemical risks in the home and in the environment. Lower 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.

Scale Means by Site and Total Sample. Table 5 shows the scale means by site and for the total sample.[1]

As can be seen in Table 5, there were considerable differences in the statistical power available at the different sites: sample sizes ranged from a high of 219 (Reynolds) to a low of 32 (Gateway). 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 5 are reported down to a significance level of $p<.10$.

Examining the contents of Table 5 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 five analyzed sites, four showed statistically significant changes from pretest to post test on this scale. 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. As an example, students at posttest were significantly more likely to *disagree* that "If a chemical pesticide is released into the environment, then everyone in that environment is exposed to the chemical."

A fairly strong pattern of significant results was also obtained on the PersInvolv scale: two of the five sites reached significance, as did the total sample. The direction of change from pre to post indicates that students became more comfortable with personal involvement in activities relating to environmental decision making.

A somewhat surprising result is the lack of significant change in students' attitudes about science (ValueSci scale). Though low statistical power may offer some explanation, and examination of mean pretest and posttest scores in Table 5 shows near equality in most cases. We speculate that for students in general positive views about science have become the norm as schools across the

[1] The Nestucca site was not included in the analysis because only 10 valid responses were available.

state have emphasized science and math skills in their curricula for a number of years. Thus, this scale may refect a "ceiling effect" whereby attitudes are already so positive that only a great deal of treatment effect can lead to a positive increase.

Table 5. Scale means by school and for total sample.

Site		Chemrisk	Explanation	ImmedAction	InfoSeek	IntuitTox	PersInvolv	ValueSci
Benson (n=115)								
	Pre	2.82	2.62	2.85	2.99	2.71	2.96	2.79
	Post	2.69	2.58	2.92	2.95	2.56	2.92	2.73
	Diff	-.12	-.04	-.07	-.05	-.14*	-.04	-.05
Corvallis (n=136)								
	Pre	2.83	2.65	2.98	3.03	2.60	2.82	2.81
	Post	2.74	2.52	2.90	2.90	2.69	2.86	2.84
	Diff	-.11	-.13**	-.09[a]	-.13**	.09[a]	.04	.02
Gateway (n=32)								
	Pre	3.00	2.67	2.95	3.13	2.91	3.07	2.49
	Post	2.91	2.80	3.09	3.22	2.62	3.13	3.02
	Diff	-.08	.15	.14	.08	-.26*	.02	.08
Reynolds (n=219)								
	Pre	2.91	2.74	3.12	3.13	2.65	3.02	2.95
	Post	2.81	2.76	3.08	3.15	2.63	3.12	2.96
	Diff	-.10	.02	-.04	.01	-.02	.10**	.02
Westview (n=34)								
	Pre	2.84	2.48	3.09	3.02	2.83	2.94	3.00
	Post	2.68	2.69	3.00	3.01	2.55	3.16	2.87
	Diff	-.13	.21*	-.09	-.01	-.22[a]	.22*	-.15[a]
TOTAL (n=536)								
	Pre	2.88	2.67	3.01	3.07	2.67	2.95	2.88
	Post	2.77	2.66	2.99	3.04	2.63	3.01	2.88
	Diff	-.11	-.01	-.02	-.03	-.04	.06*	.00

[a] p<.10
* p<.05
** p<.01
*** p<.001
****p<.0001

Table 5 can also be evaluated by looking at individual schools/sites. Viewed this way, the Corvallis site exhibited significant changes on four of the seven scales, as did the Westview site. The Benson and Reynolds sites each had significant changes on two scales, and the Gateway site on one scale. None of the individual sites went without at least one significant scale effect from pretest to posttest.

Table 6 below summarizes these results by showing by site the particular scales reaching significance.

Table 6. Summary of sites with significant pre/post change scores.

Site	Scales with Significant Pre/Post Change Scores
Benson HS	IntuitTox*
Corvallis HS	IntuitTox[a], Explanation**, ImmedAction[a], InfoSeek**
Gateway	IntuitTox*
Reynolds HS	PersInvolv**
Westview HS	IntuitTox[a], Explanation*, PersInvolv*, ValueSci[a]
Overall	**PersInvolv****

[a]p<.10, *p<.05, **p<.01, ***p<.001

Of the five sites, three (Benson, Corvallis, & Westview) showed a pattern of significant change from pretest to posttest for the IntuitTox scale, suggesting an effect of the curriculum on students' intuitions about toxicology in the direction of a more scientific perspective.

A potentially interesting pattern of results occurred for the ChemRisk scale. As can be seen in Table 5, this scale did not show significant changes from pretest to posttest in students' perceptions of chemical risks. It was hypothesized at the outside of the project that providing instruction in chemical risk assessment may have a sensitizing effect on student attitudes about chemicals in the environment. However, a more complex outcome may have occurred. This is best explained by examining Table 7, which contains a summary of responses to the seven items comprising the ChemRisk scale.

Table 7. Pretest and posttest responses to items comprising the ChemRisk scale.

Item		Disagree	Agree	DK	Mean[a]
"In general, chemical pesticides make life safer." (#cp34)					
	Pretest	52.7%	26.2%	20.9%	2.21
	Posttest	47.1	35.2	17.6	2.35**
"Overall, chemical pesticides used in agriculture do more good than harm." (#cp35)					
	Pretest	38.5%	34.4%	26.0%	2.39
	Posttest	32.6	50.9	16.1	2.61****
"Chemical pesticides sold for use by consumers in their homes are not a source of environmental health risk." (#cp36)					
	Pretest	55.9%	22.7%	20.5%	2.15
	Posttest	60.6%	25.7	13.4	2.19

Item		Disagree	Agree	DK	Mean[a]
"For most chemicals I am exposed to, I know how to protect my health." (#cp38)					
	Pretest	28.6%	53.3%	17.0%	2.72
	Posttest	28.4	59.2	11.7	2.74
"I have read and can understand most of the warning labels on chemical pesticides in my home." (#cp39)					
	Pretest	17.8%	74.2%	6.2%	3.11
	Posttest	24.0	70.7	4.8	2.96**
"I can read and understand most of the warning labels on chemical pesticides commonly used in the home." (#cp42)					
	Pretest	13.9%	79.5%	5.7%	3.14
	Posttest	16.7	78.6	4.6	3.06*
"We should reduce or eliminate our reliance on the use of animals in research to test the safety of chemicals used in consumer products." (#cp45)					
	Pretest	28.2%	52.9%	17.0%	2.81
	Posttest	32.2	51.7	16.1	2.68**

[a]Asterisked means indicate a significant difference between matched pretest and posttest responses (*t*-test for paired comparisons).
* p<.05
** p<.01
***p<.001
****p<.0001

The set of 7 items contains (a) three relating to views about chemical pesticides in general and in particular chemical pesticides in homes (#'s cp34, 35, & 36), (b) three items relating to students' perceptions of their ability to protect themselves against possible harm from chemicals (#'s cp38, 39, & 42), and (c) one item relating to the use of animals for chemical toxicity testing (#cp45). All items were responded to on a one to four-category scale ranging from 1= "strongly disagree" to 4 = "strongly agree." With the exception of Item #cp45 (which was reverse scored), higher levels of "agreement" which corresponds to higher numerical values, reflects a more positive view of chemicals and a more confident attitude about personal protection against chemicals. Likewise, lower scale values reflect greater sensitivity to chemical risks.

Taken on the face, the results in Table 7 reveal a somewhat more complex picture. On the one hand, students' perceptions of chemicals became *more positive* from pretest to posttest: at posttest they were more likely to perceive that "chemical pesticides make life safer" and that "chemical pesticides used in agriculture do more good than harm." On the other hand, they were less likely at posttest to agree that they can "read and understand" pesticide warning labels, suggesting a more risk averse posture.

It is conceivable that the effect of the curriculum on students' attitudes about chemical risks can be understood in terms of two general themes: on the one hand, students learn that they do not know and understand as much about chemicals as they might have thought (thereby reducing their confidence in what they know about chemicals in the home), but on the other hand develop a more positive view about chemicals and their management through the curriculum and Scenario experience of effectively managing the cleanup of a pesticide spill. In other words, students may have learned simultaneously that their own knowledge about chemical risks is lower than they expected, but that chemical accidents don't turn out as badly as they might have thought previously and such events can be mastered with their involvement in away that leads to a positive outcome. We should discuss this and other possible interpretations.

Taking Stock

On the whole, the evaluation demonstrated a modest though statistically reliable effect of the curriculum on students' attitudes, perceptions, and reasoning about chemicals in the environment and environmental toxicology. Given the difficulties associated with implementing a pretest/post test design in a set of working classrooms distributed across a wide range of instructional settings and with somewhat dissimilar curriculum formats (i.e., integrated vs. non-integrated classrooms), the results obtained are all the more significant. It appears from these results that students who participated in the Pesticide Spill Curriculum adopted a more scientific perspective on the meaning of chemical exposure and how the risks of chemical should be managed.