

Backyard Bugs Summary of Research Results

Overview

Kinder Magic Software (KM) conducted field tests over a range of elementary school classrooms in five states, involving the assessment of 401 students in grades Kindergarten through fifth. The composition of this group reflected diversity from the perspectives of age, gender, ethnicity, socio-economic status and geographic representation. The Level 1 data (representing students in Kindergarten and first grade) consists of 100 complete sets of pretest, posttest, and science journal scores. The Level 2 data (representing students in third, fourth and fifth grade) consists of 175 records. The data were comprised of 1) teachers' responses to a survey designed to assess their perceptions of the materials, 2) students' responses to Level 1 and 2 pretests and posttests, scored by a classroom teacher, and 3) students' responses to science journal prompts (requiring a variety of response types), graded by two additional classroom teachers. The scoring was conducted according to rubrics developed by KM, detailing the possible answers for each student instrument in the assessment. The three teachers who scored the pretests, posttests, and science journal pages were trained in the procedures for scoring the instruments by KM project staff.

Since the role of the classroom teacher is essential in maximizing the impact of the *Backyard Bugs* materials, it was important to gather feedback from the 22 teachers who participated in the field tests, to assess their perceptions of the quality and utility of these materials. Hence, a teacher survey was administered as part of the evaluation process. The results of this survey were consistently positive.

Both Level 1 and Level 2 teachers awarded high scores to eight dimensions of the *Backyard Bugs* materials, with the highest scores being associated with the relevance of the curriculum and the interest level of the students. Since the science journal pages were graded by two classroom teachers who were not part of each level's classroom evaluation, it was imperative to assess the degree of consistency among these teachers in deriving scores that accurately described the level of academic performance of each pupil. For this reason, the science journal scores were analyzed for inter-rater reliability. Remarkably high levels of inter-rater reliability were obtained.

Analysis of the pretest and posttest data for students in Level 1 and Level 2 showed that students in both groups made significant gains [$t(191) = 8.139, p < 0.0001$].]

Specific details related to all data analyses are covered below in this section of the report. Copies of the pretest, posttest and science journal and teacher survey forms for both Level 1 and Level 2 may be found here:

Level 1 [Pretest](#) and [Posttest](#)

Level 1 Science Journals:

[Close-up Observation of Dragonfly](#)

[Monarch Butterfly Food](#)

[Monarch Caterpillar Food](#)

[Whirligig](#)

[Complete Metamorphosis](#)

[Teacher survey form](#)

Level 2 [Pretest](#) and [Posttest](#)

Level 2 Science Journals:

[Biodiversity Activity 1](#)

[Biodiversity Activity 2](#)

[Biodiversity Activity 3](#)

[Monarch Survival](#)

[Mimicry](#)

[Parasite & Host](#)

[Science Journal Pages Rubrics](#)

Teacher Survey Results

Teacher surveys were administered to twelve Level 1 and eleven Level 2 teachers who participated in the field test. The table below lists the mean scores for each teacher group pertaining to their responses to eight scaled questions contained in the survey. These are Likert scale items, with scores ranging from 1 = “strongly disagree” to 5 = “strongly agree.” (The survey also included 16 open-response items.)

Mean Scores for Teacher Survey Items

SURVEY ITEMS	LEVEL 1	LEVEL 2	TOTAL
The lesson plans were helpful.	4.42	3.45	3.96
2. The software was easy to use with my students.	4.33	3.45	3.91
3. The meaning of each item on the screen was understandable.	4.25	3.36	3.83
4. My students liked the program.	4.58	4.55	4.57
5. All of the help I needed was easily available.	4.33	3.64	4.00
6. The content level was appropriate for my students.	3.25	3.18	4.00
7. The science Journal pages were appropriate for my students.	3.17	2.82	3.00
8. This was a worthwhile experience for my students.	4.17	3.91	4.04

As can be seen from the above table, teachers expressed consistently high regard for the materials. Since a score value higher than 3.00 reflects a favorable statement, one

can see that all eight scaled items earned favorable mean scores from the 22 participating teachers. It should be noted that while the ratings were consistently high, the highest ratings were obtained from those teachers who instructed Kindergarten and first grade students. This is noteworthy because it demonstrates successful application of learning activities and tools in a classroom environment for very young learners. These learning activities required the students to make generalizations from their learning experiences and apply their discoveries to new scenarios and novel data. This type of learning is consistent with the higher levels of Bloom's Taxonomy of Learning Objectives.

Selected quotes from teachers in response to the open-response items follow:

"Excellent quality. The videos were awesome... Being able to watch the metamorphosis is much better than just describing it."

"Very appropriate. Also good to have the habitats for each sections...illustrates where different animals live."

"We loved it! We were oohing and ahing a lot. The bugs were so cool and interesting."

"Liked how up-close they are. We loved the microscope section as well. Wonderful!"

"I loved it all - I haven't seen anything like it. It was wonderful to experience. The children had the best time! We had an Author Study on Eric Carle and they were asking questions about the whirligig beetle and the butterfly and cricket to his stories about those bugs."

"I liked interacting with the different habitats. "

"Absolutely outstanding! We could never see it really happen in nature! Easy to access"

"Interesting - I like being able to stop and study a frame and then move on."

"Insects are age-appropriate and integrate with our 3rd grade science standards."

"Students loved the fact that these bugs could be found near them."

"Math - especially graph exercises provide practice of skills outside "math class""

"Insects are studied via lab, text, and research, so this is a perfect extension."

"Many of the students in this class receive Title I services. The students thoroughly enjoyed the software. In fact they have asked me for more bugs! This program has really helped them understand the frailness of our environment. An extra bonus is...I learned so much too!"

"They loved the biodiversity lesson, because they had a field guide to match up."

"Monarch survival simulation - they loved how it was a game."

"The software grabbed the attention of my students. It was great for my English Language Learners that need a visual."

Inter-Rater Reliability of Science Journal Scores

The teachers were given rubrics to evaluate the students' products for each of the science journal pages. The field test teachers were told that they could score the science journal pages for their own use, but that KM would be scoring these materials for the purpose of the evaluation. The two classroom teachers who scored the science journal pages were provided training in the procedures for scoring the instruments by Kinder Magic project staff. In order to insure consistency across a level's science journal page evaluation, one teacher scored all the Level 1 science journal pages, while the second teacher scored all the level 2 science journal pages. Since there is always an issue of subjectivity in assessment of student performance, even with a well-developed scoring rubric, it raises the question of inter-rater reliability among the two teachers who scored the science journal pages for the evaluation of *Backyard Bugs*. For this reason, a controlled assessment of inter-rater reliability was conducted. The two teachers who scored the science journal pages for the evaluation independently scored ten students' science journal pages. The teachers traded levels, so that the teacher who originally scored the level 1 science journal pages scored the subset of level 2 science journal pages, and vice versa. Science journal pages for the controlled assessment of inter-rater reliability were selected to represent all grade levels and participating teachers. The inter-rater reliability coefficients for the Level 1 and Level 2 science journal pages are shown in the tables below:

Inter-rater Reliability for Level 1 Science Journal Pages

Level 1 Science Journal Lessons	Inter-rater Reliability (r)	p
Close-up Observation of Dragonfly	0.908	< 0.0001
Monarch Butterfly Food	1.000	< 0.0001
Monarch Caterpillar Food	1.000	< 0.0001
Whirligig	0.745	0.007
Complete Metamorphosis	0.905	< 0.0001
Total Science Journal Page Score	0.913	< 0.0001

Inter-rater Reliability for Level 2 Science Journal Pages

Level 2 Science Journal Lessons	Inter-rater Reliability (r)	p
Biodiversity Activity 1	0.997	< 0.0001
Biodiversity Activity 2	0.944	< 0.0001
Biodiversity Activity 3	0.972	< 0.0001
Monarch Survival	0.998	< 0.0001
Mimicry	0.975	0.014
Parasite & Host	0.727	0.013
Total Science Journal Page Score	1.000	< 0.0001

As can be seen from the data in the above two tables, remarkably high levels of inter-rater reliability were obtained. Ary, Jacobs, and Razavieh (2002), in discussing "acceptable" and "optimal" properties of psychometric materials, recommend that

objective assessment instruments should demonstrate reliability levels at or above 0.90 to be acceptable, with 0.95 being recommended as optimal. As can be seen from the above two tables, the reliability figures derived generally ranged at the “optimal” level. This speaks to the clarity of the *Backyard Bugs* rubrics as well as instructions for its use.

Demographics of the Field Test Samples

As was previously mentioned, the demographic profile of the student field test groups met the proposed project objective by including a sample that reflected diversity from the perspectives of age, gender, ethnicity, socio-economic status, and geographic representation. Those 100 students in Level 1 represented seven different schools and 12 different teachers. There were 28 kindergarten and 72 first grade students. Their ages ranged from 5 – 8, with a mean age of 6.60 years. There were 39 males and 61 females. The states in which these children attended school were California, Colorado, New Jersey and Virginia. Teachers’ reported descriptions of the students’ ethnic identities can be found in the [Inclusion Enrollment Table](#).

The 175 students in Level 2 represented six different schools and nine different teachers. There were 75 third grade, 74 fourth grade and 26 fifth grade students. Their ages ranged from 8 – 12, with a mean age of 9.51 years. There were 87 males and 88 females. The states in which these children attended school were Alaska, California, Colorado, New Jersey, and Virginia. Teachers’ reported descriptions of the students’ ethnic identities can be found in the [Inclusion Enrollment Table](#).

The fact that several dimensions of diversity were represented in the field test samples has strong implications for adaptation in school curriculum. These materials demonstrated their utility across a broad spectrum of elementary school-aged learners.

Descriptive Statistics; Overall Student Performance

The score profile for Level 1 students is shown in the following table:

Measures	Mean Raw Score	% of Highest Attained Score
Pretest	19.81	76.19
Posttest	20.81	80.03
Science Journal Page Scores		
Close-up Observation of Dragonfly	7.41	61.75
Monarch Butterfly Food	3.80	95.00
Monarch Caterpillar Food	3.87	96.75
Whirligig	1.95	48.75
Complete Metamorphosis	10.54	47.91
Total Science Journal Score	27.57	67.24

The score profile for Level 2 students is shown in the following table:

Measures	Mean Raw Score	% of Highest Attained Score
Pretest	8.26	55.07
Posttest	11.04	61.33
Science Journal Scores		
Biodiversity Activity 1	50.98	80.92
Biodiversity Activity 2	20.83	83.32
Biodiversity Activity 3	21.54	86.16
Monarch Survival	24.51	66.24
Mimicry	20.73	60.97
Parasite & Host	5.17	57.44
Total Science Journal Score	117.22	79.74

As was previously mentioned, analysis of the pretest and posttest data for students in Level 1 and Level 2 showed that students made significant gains from their pretest to posttest performance [$t(191) = 8.139, p < 0.0001$]. As can be seen from the above two tables, Level 1 students increased their performance level by an average of 4% and Level 2 students increased their performance level by an average of 6%. This suggests the potential utility and impact of the *Backyard Bugs* materials to increase students' knowledge of the topic, as well as generalize to areas of science beyond those specifically covered in the materials. An additional noteworthy statistic is the strong correlation between scores on these two measures ($r = .859, df = 190, p < 0.0001$). This speaks to the psychometric equivalence of the pretest and posttest instruments.

Performance Scores Related to Student Demographics

Additional pertinent analyses of the above four performance scores were conducted. These analyses examined score differences related to gender, age, and teacher description of students' ethnicity and "English Language Learner" status. The following findings were obtained from these analyses:

There were no differences in performance regarding gender or age of students. The fact that no differences were found among these 275 students related to gender and age suggests the appropriateness and applicability of the *Backyard Bugs* materials for the broad population of young learners. Regarding data on ethnicity, subgroup sizes were too small and unequal to allow meaningful analysis of the data. In order to further explore the issue of ethnicity related to performance the "ethnicity" data for Level 1 and Level 2, students was recoded as follows:

Non-Minority	(n = 195)
Minority	(n = 80)

An analysis of score differences on the four performance measures revealed some noteworthy findings. While the pretest scores of the minority students averaged almost

four points below those of the non-minority students, they were able to decrease this performance gap by more than 50 % on the posttest. The minority students averaged a 3.75 point pre-post gain, while the non-minority students averaged only a 1.57 point pre-post gain on the same materials. This was a significant difference in gain [$t(190) = 3.769, p < 0.0001$]. Another finding of note is that there was no significant difference between the minority students and the non-minority students on their total science journal page scores. These findings suggest some powerful implications. The minority students, while reflecting lower baseline scores, were able to “close the gap,” making significantly greater gains, as a result of participating in the *Backyard Bugs* learning experience. It would appear that the materials contributed to “leveling the playing field” between these two groups. This was true for the “hands-on” science journal page performance measure as well as the more traditional posttest assessment. In fact, the minority students averaged a higher score on the science journal pages than the non-minority students. The table below shows mean scores for both groups across these four measures.

Performance Comparisons Between Minority and Non-minority Students

GROUPS	PRETEST	POSTTEST	GAIN	SCIENCE JOURNAL
MINORITY	10.47	14.24	3.75	37.64
NON-MINORITY	14.23	16.25	1.57	34.05

The same profile of score differences was observed when comparing those students identified by their teachers as “Hispanic” to the non-Hispanic students on these four measures. While the Hispanic students trailed their counterparts by an average of three points on the pretest, they made significantly greater pre-post gains [$t(188) = 3.255, p = .001$]. As a result, they closed the gap between the two groups on posttest performance. Additionally, there were no significant differences between these two groups on their average science journal page scores. The table below shows mean scores for both groups across these four measures.

Performance Comparisons Between Hispanic and Non-Hispanic Students

GROUPS	PRETEST	POSTTEST	GAIN	SCIENCE JOURNAL
HISPANIC	10.68	15.25	4.61	26.80
NON-HISPANIC	13.66	15.68	1.92	35.80

An analysis of the four performance score differences between individual classrooms, specific school sites, and specific states was not possible due to the small and unequal subgroup sizes. This was true for both Level 1 and Level 2 groups. If however, the *Backyard Bugs* materials were utilized on a consistent and universal basis throughout a school site, district, or state, then it seems plausible that an examination of these materials would enable teachers and administrators to discern systematic differences across groups. This could be of utility in assessing how materials and curriculum might

be implemented to tailor learning experiences and maximize individual learner outcomes.

Summary

As can be seen from the preceding discussion a large and diverse sample of students, teachers, and learning environments was incorporated into the field tests. Data collection and analyses examined “hands-on” learning outcomes and their relationship to a variety of demographic factors.

The feedback teachers provided on the teacher survey instrument yielded consistent and positive ratings regarding the value, appropriateness, interest level and novelty of the *Backyard Bugs* materials. Since the teacher’s role as the facilitator of the learning environment is critical, the high regard for the materials that these teachers expressed is a very valuable finding.

The diversity of the student sample, as manifested by ethnicity, age, gender, socio-economic status, and geographic representation, speaks to the potential of *Backyard Bugs* to be utilized on a broad basis. Related to ethnic diversity, the findings of statistical comparisons between minority and non-minority students as well as between Hispanic and non-Hispanic students were of particular interest and value. While significant pre-post test gains were observed for the entire sample, the minority students evidenced a level of gain that was significantly greater than non-minority students. Further, the minority students’ performance on the science journal pages was equal to their non-minority counterparts. The same trend of findings was observed when comparing the Hispanic to the non-Hispanic students. This shows the potential of the *Backyard Bugs* materials to “level the playing field” for students who are often the most challenged by traditional learning environments and teaching materials.

Reference

Ary, D., Jacobs, L.C., & Razavieh, A. (2002) *Introduction to Research in Education* (6th Ed.) Belmont, CA: Wadsworth.