Journal of Experiential Education • 2006, Volume 29, No. 1 pp. 61–77

Experiential Education in the Classroom and Academic Outcomes: For Those Who Want It All

Bob Ives and Kathryn Obenchain

We conducted a pretest-posttest study using measures of higher order thinking skills (HOTS), and lower order thinking skills (LOTS) in six 12th-grade American Government classrooms taught by three experienced teachers over one semester. One of the three teachers implemented a curriculum in two classes based on experiential education (EE) principles with guidance from the investigators. Students in the EE-emphasized classes demonstrated greater gains in HOTS than the students in the other four classes. There was no difference between the two groups in gains for LOTS. These results suggest that EE instruction in high school classes can promote HOTS more than traditional instruction does with no sacrifice in LOTS.

Keywords: Experiential Education, Academic Achievement, Higher Order Thinking

Bob Ives, Ph.D., is an Assistant Professor of Special Education in the Department of Educational Specialties at the University of Nevada, Reno, USA.
E-mail: rives@unr.edu

Kathryn Obenchain is an Associate Professor of Social Studies in the Department of Educational Specialties at the University of Nevada, Reno, USA.
E-mail: kmo@unr.edu

The concept of accountability and specific accountability movements have affected education for more than 50 years, moving through different eras such as minimum-competency testing and accountability based on standardized tests (see Linn, 2000, for a thorough discussion of the history of accountability movements in American education). The current accountability movement, often identified as the standards-based reform movement, began in the early 1990s and is identified with the Goals 2000: Educate America Act (Abrams & Madaus, 2003; Linn, 2000). While the development of voluntary national standards in the different content areas began a few years earlier (e.g., National Council for Teaching Mathematics first published standards in 1989), this era of standards-based reform is exemplified by individual states' development of academic content standards, as well as the development or adoption of assessments that measure student achievement related to those standards. With the passage in the U.S. of the No Child Left Behind (NCLB) legislation in 2002, the adoption of standards is now required by federal law. The high-stakes consequences of not making "adequate yearly progress" are substantial and may include teacher and administrator reassignment.

Driven by the (NCLB) requirement to measure and report student achievement, at least 25 states have adopted high-stakes tests that serve as graduation gate keepers (Center on Education Policy, 2004). In other words, a single measure of a student's mastery of specified content areas may determine whether or not a student graduates. When graduation is dependent upon achieving a specific test score or scores, those tests literally have high stakes (Amrein & Berliner, 2002). Further, making the decision of whether or not a student graduates based on a single assessment assumes that success on one achievement test should correlate with success on other achievement tests designed to measure the same content and/or skills. Research indicates otherwise (Perkins & Salomon, 1988; Salomon & Perkins, 1989). One study comparing fourth-grade students' results on reading and math sections of the National Assessment of Educational Progress (NAEP) with results on the Texas Assessment of Academic Skills (TAAS), Klein, Hamilton, McCaffrey, and Stecher (2000, p. 7) found that, "the gains on TAAS were much larger than they were on NAEP . . . there was nothing remarkable about the NAEP scores in Texas."

Typically, these high-stakes tests measure lower order knowledge and skills (e.g., recall and comprehension) as opposed to higher order thinking (e.g., analysis, synthesis, evaluation) (Chudowsky & Pellegrino,

2003; National Academy of Education, 1997; Neill, 2003). Believing that these tests prioritize the recall of a breadth of knowledge, teachers adjust their teaching content and instructional practices accordingly. Teachers and administrators want their students to be successful, and if success is determined by a score on a specific exam, teachers will work to help students meet those expectations.

One negative consequence of a high-stakes testing policy is a narrowing of the curriculum as teachers teach to a single high-stakes measure (Abrams & Madaus, 2003; Darling-Hammond, 2004; Marchant, 2004; Pennington, 2004). Narrowing the curriculum includes choosing to focus on the recall of basic information over in-depth understanding as well as focusing primarily on information that teachers believe will be tested. This narrowing may work if there are assurances that only important information is tested and that all important information is tested. This was the goal of the development and assessment of state academic content standards during the most recent era of the accountability movement (Abrams & Madaus, 2003; Linn, 2000). Unfortunately, many state standards are so encompassing of content that it is nearly impossible to develop meaningful assessments that address both the specific details included in the standards, as well as any broad themes and concepts (Chudowsky & Pellegrino, 2003). Rather, teachers use available information such as previous tests and test preparation materials to anticipate what the test will contain, and prepare students accordingly.

Narrowing the curriculum also leads to limited instructional strategies. Teachers choose time-efficient delivery models of instruction (e.g., lecture) over instructional models that promote critical thinking, problem-solving, and inquiry (e.g., experiential education–based models). Studies report that teachers abandon innovative, active, and higher order experiences in favor of rote memorization and drill, believing this is the wise course of action for testing, although not necessarily for student learning (Hillocks, 2002; Marchant, 2004; McNeil, 1990; Pennington, 2004).

The study reported here was initiated by a teacher-participant because of just such a concern. As an experienced and well-educated teacher, she believed that active, experience-based methods would encourage development of the higher order thinking skills that she valued as an educator. However, this teacher was concerned about how to convince her administrator that her students would still master the facts that were being tested. Mathison and Freeman (2003) found similar concerns among New York fourth-grade teachers.

The purpose of this study was to compare the academic achievement (as evidenced by both higher order and lower order thinking) of students taught using experiential education–based instructional practices to students taught using more traditional instructional practices.

Experiential education–based practices in secondary and traditional school settings were of particular interest.

The accountability movement is here for the foreseeable future. However, research can raise questions about the perception that the accountability movement requires certain kinds of classroom teaching in order to ensure success on required tests. Specifically, are there instructional options for those teachers who believe that classroom learning should be more experiential, higher order, thoughtful, connected to the community, and student-driven, but are aware that their students must be successful on measures of lower order thinking?

Academic Achievement

This study defines academic achievement as both higher order and lower order thinking skills achievement. In contrast to the majority of high-stakes assessment instruments used by states and schools that focus on mastery of lower order thinking skills, society in general, and educators, in particular, believe that students should also master higher order thinking skills (Newmann, Bryk, & Nagaoka, 2001). While the acquisition of basic information requires lower level thinking, the ability to evaluate this information for use in new settings, as well as the ability to use it effectively, involves higher order thinking (Kornhaber, 2004).

Higher Order Thinking Skills

Higher order thinking skills (HOTS) is not a well-defined term (Ivie, 1998; Leming, 1998). Attempts to define HOTS have considered abstract thinking, integrating information into systems, and following rules of logic and judgment (Ivie, 1998). HOTS have been equated to problem-solving (Swanson, 2001) and to critical or reflective thinking (Leming, 1998). It is more typical, however, to see HOTS described through a list of subskills. These subskills may include comparison, categorization, inference, prioritizing, analytic perception (Niedelman, 1991), analogical and logical reasoning (Grossen, 1991), question posing, argumentation, system thinking (Dori, Tal, & Tsaushu, 2003), and going beyond the given information into discovery, reasoning, organizing, and argumentation (Torff, 2003). HOTS have also been defined in terms of Bloom's Taxonomy of Educational Objectives (Bloom, 1956) as comprising the analytic, synthetic, and evaluation categories of objectives (Granello & Underfer-Babalis, 2004; Zohar & Dori, 2003).

Lower Order Thinking Skills

Lower order thinking is often defined by what it is not. Lower order thinking does not involve the complex intellectual work of higher order thinking (Newmann, Bryk, & Nagaoka, 2001) or what Ivie (1998) calls

meaningful, as opposed to rote, learning. Rather, lower order thinking includes the memorization and recall of typically arbitrary factual information (Ivie, 1998; Leming, 1998). Specific examples include the recall of the names of American presidents, the recall and application of multiplication tables, and the literal comprehension of literary passages (Swanson, 2001). In Bloom's Taxonomy of Educational Objectives (Bloom, 1956), lower order thinking includes the recall, comprehension, and application categories.

Experiential Education in Academic Achievement

Experiential education (EE) is a philosophy of education, described as "... a process through which a learner constructs knowledge, skill and value from direct experience" (Association for Experiential Education, 1991, p. 1). Based on an examination of the philosophical underpinnings and definitions of experiential education, we identified three essential elements of well-constructed experiential education–based curricula. The identification of these elements came from a synthesis of a variety of literature reflecting the historical foundations of EE, as well as current practices. First, learning should include opportunities for student-direction (Druian, Owens, & Owens, 1980). Second, learning through EE includes curriculum connections to the real world (Rahm, 2002; Shelton, 2000). Critical reflection is the third essential element of EE and permeates every aspect of an EE program. It is described by Dewey (1933/1998) as an internalized inquiry process.

Much of the available research related to experiential education and academic outcomes discusses both the higher order and lower order dimensions of academic achievement. This work, however, often addresses nontraditional settings and has some design limitations. Weinbaum, Gregory, Wilkie, Hirsch, and Fancsali's (1996) report of a three-year project in which students at 10 schools across the United States participated in some form of learning "expedition" found gains in student achievement. Reading and math scores on standardized measures significantly increased in two elementary schools. The results of these studies support the use of EE-based practices. However, the lack of control groups qualifies the results. Wentzel's (1991) study of school-based service-learning practices found that students who were seen by others as more engaged also earned higher grades. It is not clear whether more engaged students earn better grades because of their engagement or because of how they are perceived by others (Goodenow, 1992; Junn, 1991). Set in a traditional school, Reynolds' (1991) study of an experiment-based middle school science program examined student achievement of both science content skills, such as science measurement, graphing (lower order) and science process skills such as hypothesis testing (higher order). Using a pretest-

posttest model, the study found no evidence that the number of experiments would predict higher achievement outcomes on a measure of science content. However, the study did find some evidence that the number of experiments is a predictor for higher achievement outcomes on a science process measure. Also in a traditional school setting, Laney (1989) taught basic economic principles to first-grade students using either a real-life experience or story-telling approach. A six-week delayed posttest showed a statistically significant greater retention of these concepts for the students taught through real-life experiences. Additional research connecting academic achievement with experiential education–based instructional practices in traditional school settings is needed if these practices hope to coexist in an era of accountability.

The purpose of this study was to examine the effects of an experiential education (EE) approach to instruction on academic outcomes in traditionally structured 12th-grade American Government classes. The research literature addressing academic outcomes in experiential education in traditional school settings is scant (Hedin, 1983; Roberts & Yerkes, 2000), and there are continuing calls for more research (Ewert, 1987). Our study makes a contribution to filling that void by addressing the following research hypotheses.

1. Students in classes taught through increased implementation of EE practices will demonstrate a greater improvement in HOTS from pretest to posttest compared to students experiencing more traditional instruction.
2. Students in classes taught through increased implementation of EE practices will demonstrate no difference in changes in LOTS from pretest to posttest compared to students experiencing more traditional instruction.

Methods

Setting and Participants

This study was a collaborative effort among two university faculty members, two graduate students, and three classroom teachers. Each of the three classroom teachers taught two classes of a secondary American Government class. One classroom teacher (experimental teacher) deliberately crafted her curriculum to reflect experiential education practices through close collaboration with one of the faculty members. The other two classroom teachers (control teachers) created and implemented their curricula without any guidance from the faculty members. Data were collected from students and teachers in these six high-school American Government classes over the course of one semester.

The three teachers were interviewed to identify some basic student characteristics. Table 1 reports how the teachers described the students in their American Government classes. Each teacher was teaching two American Government classes, and the table includes all students in both classes for each teacher, not just those who participated in the study. Notably, the experimental teacher's students are overwhelmingly female, unlike the students in the other classes. In addition, while most of the students in all groups were Caucasian, the Control 2 teacher had a larger proportion of Hispanic students and a larger proportion of non-native English speakers.

Table 1
Demographic Data for All Students in the Participating Teachers' Classes

	Experimental	Control 1	Control 2
Total students	34	26	42
Male (%)	18	38	55
Female (%)	82	62	45
White (%)	91	85	60
Hispanic (%)	6	8	36
Asian (%)	0	8	2
African American (%)	0	0	2
Native American (%)	3	0	0
Eligible for Special Education (%)	3	0	4
Non-Native English Speakers (%)	3	0	12

Experiential Education Curriculum Design

The second author of this paper taught a graduate level class in the foundations and applications of experiential education in the K-12 curriculum. The experimental teacher was a student in this class. According to the course syllabus, the course was planned "to explore experiential learning (process of change for the individual) and experiential education (transactive experience between teacher and student) from its historical roots to current practices." In addition, beginning three months prior to the semester of the study and continuing throughout the semester of the study, the experimental teacher and this author discussed the teacher's process in constructing the EE curriculum for her classes. These discussions typically took place more than once each week in person, in telephone calls, and in e-mail. The goal of the graduate course and the

subsequent conversations was to increase this teacher's implementation of EE in her classroom. Procedural fidelity was assessed during the study to confirm that this teacher did indeed implement more EE in her classroom than did the control teachers.

Research Design and Instruments

We conducted a pretest-posttest study with one predictor variable and two outcome variables. The predictor variable was the frequency of experiential education (EE) events in the classrooms. The two outcome variables were (a) basic knowledge of American Government course content (LOTS), and (b) HOTS applied to American Government course content. During the second week of the semester of the study we administered both the HOTS pretest and the LOTS pretest. The two pretests were administered in the same sitting to each class. The same process was followed for the posttests during the last two weeks of the semester.

Experiential Education Events

An Anecdotal Record of Experiential Education Events (AREEE) form to guide classroom observations was developed to document procedural fidelity by measuring the frequency of experiential EE events in all six classrooms. The instrument focused on the three essential elements of experiential education identified earlier: (a) student-directedness, (b) real-world connections, and (c) critical reflection. Following field trials of this observation form and additional discussion, these elements were operationalized in the following definitions.

1. Student directedness was student involvement in decision-making on course content, experiences, assessment, and classroom procedures.
2. Real-world connections were student actions on, or recognitions that they could act on, connections between content and applications outside the classroom.
3. Critical reflection was evidence of student thinking at the evaluation level of Bloom's Taxonomy applied to course content.

Each classroom was observed at least six times throughout the semester by rotating pairs of four investigators trained on using the instrument. For each observation, two observers visited a classroom for roughly an hour and independently recorded observed instances of EE events. The observers recorded four types of information. First, they recorded start and stop times for event intervals in the classrooms, such as the interval for announcements, the interval for working in small groups on an activity, etc. Second, each event was assigned one point for each of the three essential elements of EE (real-life connections, higher order thinking, and

student directedness) observed. Thus a maximum of three points could be assigned to each event for these essential elements.

The third data entry for each EE event was a coding of the event as an "opportunity," or an actual "example" of EE. For example, if students were actively working in small groups on a collaborative project, the decisions they made would be examples of student directedness. However, if students were listening to the teacher describe this assignment in terms of the decisions they will need to make at some other time, this would demonstrate an opportunity for student directedness, but not an actual example. Each observer could assign a maximum of five points for each event interval, one each for the three elements of EE plus one for an example and one for an opportunity. The scores for the two observers were summed for each interval to give a maximum possible score of 10 points. Comparing these scores across classrooms allowed us to assess the relative level of implementation of EE in each classroom. After each observation the two observers would meet briefly to discuss their findings to identify and come to consensus on any disparities.

The fourth type of data recorded with the AREEE were descriptive anecdotal narratives. These qualitative data included a descriptive narrative of what each observer saw and heard during the observation. The specific purpose of these descriptions was to establish procedural fidelity and, therefore, the focus of the narrative was on teacher and student talk in order to more fully describe those opportunities and examples indicating EE elements.

Higher Order Thinking Assessment Instrument

The HOTS assessment was developed by reviewing the Nevada Department of Education's 12th-grade civics standards to create items that addressed that content and also conformed to the analysis, synthesis, or evaluation categories of Bloom's taxonomy (Bloom, 1956). Questions were open-ended and required a written response. The pretest questions were based on standards that all three teachers had covered with these students during the semester before the study. The posttest questions were based on standards covered by all three teachers during the semester of the study. The classroom teachers reviewed our pool of questions and identified those they felt were appropriate for their curricula. From these we selected one question from each Bloom's taxonomy category for the pretest and also one from each category for the posttest.

We developed a scoring rubric based on descriptions of the analysis, synthesis, and evaluation categories of Bloom's taxonomy. We identified two basic components of each of these three categories. Each written response was scored as a 0, 1, or 2 based on how many of the components of the relevant taxonomy category were evidenced in the response. The

goal was to score these responses based on their structure, but not in the accuracy or depth of relevant content knowledge.

We carried out two rounds of interrater reliability checks on sample responses with follow-up discussions. The two investigators had acceptable interrater reliability ($r > .90$) in both rounds. These two investigators independently scored all of the pretest and posttest HOTS responses and then came to consensus on any discrepancies through discussion.

Lower Order Thinking Assessment Instrument

To measure the basic knowledge (LOTS) secondary students have about American Government we developed a multiple-choice test. We used released items from the 1988 and 1998 civics instruments of the National Assessment of Educational Progress (NAEP) that also conformed to the specific Nevada Department of Education's 12th-grade civics standards that all three of the teachers would be addressing in their classes during the semester of the study. The same items were used for the pretest and posttest to give us a measure of the students' gain in basic knowledge across all six classes based on content that they would all be exposed to.

Results

We first conducted analyses of the procedural fidelity at the teacher level. Independent samples *t* tests were conducted to make pair-wise two-tailed comparisons of the means of the AREEE scores for the three teachers. The mean for the experimental teacher was statistically significantly higher than that of each of the control teachers, while there was no statistically significant difference between the means for the two control teachers. Based on these results we combined the data from both control teachers for subsequent analyses. Our expectation was that the experimental data would show more EE events than the control data, so we conducted a one-tailed independent samples *t* test of the equality of the means for these procedural fidelity data (experimental teacher versus combined control teachers). Levene's test for equality of variances was not statistically significant ($F = 2.508$, $p > .11$). Nevertheless, we made the conservative choice not to assume equality of variances. Results of the *t* tests are reported in terms of the *t* statistic, the probability of a Type 1 error, and Cohen's *d* as an effect size measure (Cohen, 1988). For this specific comparison of procedural fidelity data, *t* tests yielded a statistically significant difference in favor of the experimental group with a medium effect size ($t = 2.632$, $p < .02$, $d = .49$).

Student level analyses of the results pertaining to the research hypotheses were conducted on the data from the 36 students for whom we had complete pretest and posttest data. For each hypothesis a repeated measures mixed analysis of variance (ANOVA) was conducted in which

Time (pretest and posttest levels) was the within-groups repeated-measure variable and Group (experimental versus control) was the between-groups variable. We established an alpha level of .05 for the two tests. The assumption of sphericity, which can inflate the *F* statistic for repeated measures tests, was not violated for any of these analyses. In addition, because both hypotheses make predictions comparing changes from pretest to posttest across groups, the Group x Time interaction is the specific result that addresses the hypothesis in each case. We have reported the F statistic (*F*), significance level (*p*), and meta-squared (m^2) effect size measure (Ives, 2003) for each of these tests. We have also compared each of the effect size measures to Cohen's (1988) suggested levels for small (.01), medium (.09), and large (.25) effect sizes. Means and standard deviations for these two tests are reported in Table 2.

Table 2
Means (Standard Deviations) for Tests of the Two Hypotheses

	Experimental		Control	
	Pre	Post	Pre	Post
Higher Order	3.81 (1.72)	3.44 (2.16)	3.50 (1.43)	1.80 (1.51)
Lower Order	18.06 (2.57)	18.50 (2.71)	15.40 (3.19)	16.65 (3.77)

The first research hypothesis was that students in classes taught through increased implementation of EE practices would demonstrate greater improvement in HOTS compared to students experiencing more traditional instruction. Because this hypothesis was directional, we applied a one-tailed test. The Group x Time interaction was statistically significant ($F = 3.276$, $p = .040$, $m^2 = .088$) in favor of the experimental group, which was consistent with the hypothesis. The effect size measure was close to Cohen's suggested level for a medium effect size.

The second research hypothesis of this study was that students in classes taught through increased implementation of EE practices would demonstrate no difference in LOTS compared to students experiencing more traditional instruction. Because this hypothesis was not directional, we applied a two-tailed test. The Group x Time interaction was not statistically significant ($F = .826$, $p = .370$, $m^2 = .024$), which was consistent with the hypothesis. The effect size measure was near Cohen's suggested level for a small effect size.

Discussion

There is some obvious irony in the fact that EE is derived from philosophical ideas about education that are nearly a century old, and yet components of EE are associated with current reform movements across several disciplines. At the same time, it is gratifying that our findings are consistent with those from other disciplines, such as mathematics (Schoenfeld, 2002), in showing that students engaged in a curriculum that emphasizes student-directedness and complex problem-solving over focused practice on lower level fact and skill acquisition show a significant advantage in HOTS with no loss in LOTS. These findings also raise many important questions. Some of these questions are specific to the results of this study while others more broadly apply to issues of teacher education.

The most problematic specific question involves the results for the HOTS instrument. The difference between mean gain scores from the HOTS instrument was statistically significant and favored the students with the EE curriculum. However, both groups actually had lower average scores on the posttest than on the pretest. One obvious and distressing interpretation of these results is that both groups of students were poorer at HOTS at the end of the semester than at the beginning. We reviewed the students' pretest and posttest responses to this instrument and discovered that on the posttest students in both groups were more likely to write something like "I don't know" or draw an unrelated picture or leave the answer space blank instead of responding to the questions as they had for the pretest. Given that there were no significant differences between the two groups in gain scores on the LOTS instrument, and that both groups gained points on the LOTS instrument over the semester, we infer that the most likely explanation for the losses on the HOTS instrument is that both groups found the content of the HOTS posttest questions much more difficult than the content of the questions on the HOTS pretest. We are currently in the process of redesigning the HOTS instrument for follow-up studies to test this hypothesis. Each item in the redesigned instrument will be based on content that should be highly familiar to all students. In this way we expect to have an instrument that is a purer measure of HOTS with less error introduced by question content, and we will be able to counterbalance the administration of the pretest and posttest to further control for error in follow-up studies.

A major concern of the United States Department of Education is the fact that "research principles have not been incorporated into education practice" (Institute of Educational Sciences, 2004, p. 2). One challenge to making this connection is the current accountability movement in public education. The accountability movement, as exemplified by the No Child Left Behind Act (NCLB), mandates goals for student progress and also

requires high-stakes testing to assess progress toward those goals. Unfortunately these high-stakes tests typically assess LOTS such as memorization of facts and procedures, with less focus on HOTS, which tend to be more difficult and time consuming to assess. The possible consequences of poor performance on these tests include possible loss of students, reassignment of teachers, and reorganization of schools. As a result, teachers feel constrained to bypass opportunities to foster students' HOTS in favor of instruction in LOTS.

This same conflict between accountability for LOTS and the broader goal of HOTS exists across content areas. In a study that examined changes in teachers' practices as a result of the New York State Global History and Geography exam, Grant, Gradwell, Lauricella, Derme-Insinna, Pullano, and Tzetzo (2002) found that while a few teachers changed their practice to include more attention to HOTS, "most of the teachers in this study did not tell us that they were doing more engaging work as a result of the test; moreover, several suggested ... pressure to scale back on their current efforts" (p. 511). Similarly, although the National Council for Teachers of Mathematics Standards (2000) promotes flexible problem-solving and conceptual understanding of mathematics, Weiss (1997) found that mathematics teachers support these goals of reform in mathematics education in principle, and yet this support is not reflected in their classroom practices.

Even for teachers who value fostering HOTS in their students and are willing to reduce their instructional focus on LOTS, a second challenge arises. Teachers are not likely to implement new approaches that they have learned about in teacher education programs unless their training in new approaches is continuous, large scale, offers incentives, and can be done without a significantly greater time commitment (Levine, 1994). The difficulty of changing teacher practice surfaced frequently in our study. Although the experimental teacher had a strong personal motivation to implement EE in her classes, she often commented in her notes and meetings with faculty that "it's been a lot of work so that I've many times said, if I just use my lesson plans from last year, my life would be so much easier." Thus teacher dispositions may be a significant hurdle to implementing more effective curricula.

Assessment poses another challenge for implementation of these reform curricula. Teachers are often more comfortable with creating assessment instruments to evaluate basic knowledge as opposed to more abstract understanding and operations. For example, teachers can readily create tasks involving multiple-choice questions, matching, fill in the blanks, and other probes that have objective responses that can be scored as correct or incorrect. In contrast, designing tasks that assess HOTS requires the application of flexible professional judgment on the part of

the teacher in both the creation and the evaluation of the students' responses. This is not to say that teachers don't value HOTS and reform goals in general. They do. However, the challenges of focusing on these goals are superseded by concerns for accountability, lack of adequate time, and lack of confidence in their own training and experience.

The HOTS element of experiential education may also be a subtle puzzle. Scholars regularly emphasize the importance of HOTS (Kosciulek, 2003) and advocate teaching HOTS in a variety of contexts (Baron, 1993; Martin, 1993). As already noted, however, we must recognize that HOTS is not a well-defined term (Ivie, 1998; Leming, 1998). HOTS have been described in terms of lists of subskills such as comparison, categorization, inference, prioritizing, analytic perception (Niedelman, 1991), analogical and logical reasoning (Grossen, 1991), question posing, argumentation, system thinking (Dori, Tal, & Tsaushu, 2003), and going beyond the given information into discovery, reasoning, organizing, and argumentation (Torff, 2003). Alternatively HOTS have been equated to broader terms such as problem-solving (Swanson, 2001) and to critical or reflective thinking (Leming, 1998). In any case, little research evidence explores how HOTS can be taught effectively. In this study students were given opportunities to apply their existing HOTS but there was no explicitly planned HOTS instruction. Do our findings imply that explicit instruction is not needed for students to develop their HOTS? If so, does explicit instruction in HOTS provide any benefits over simply providing opportunities for students to apply their HOTS without instruction? If not, what kind of undocumented HOTS instruction did students in the study receive?

Practical considerations often pose a significant challenge to carrying our controlled studies in classrooms. Follow-up studies should address these challenges. For example, instructional approaches carried out by investigators typically yield larger effect sizes than those carried out by classroom teachers (Swanson, Hoskyn, & Lee, 1999). This difference may be due to greater procedural fidelity, or unintended investigator bias. In addition, random assignment of students and teachers to the experimental and control groups was not practical for this study. Future studies need to control for these intervening variables in this preliminary study.

An emerging research base has begun to look at how reform-based approaches to instruction, including experiential education, affect academic outcomes. Generally these studies have found that reform-based instruction does not detract from students' acquisition of basic knowledge and skills, but it enhances students' HOTS compared to more traditional instructional approaches. The current study expands on this work in that we have found similar results in a new content area, namely civic education. A variety of challenges and questions remain about the implementation of reform-based curricula more broadly. These include the effectiveness

of teacher education and the generalizability of these instructional approaches across different content areas. Nevertheless, the promising results presented here and elsewhere indicate that trying to overcome those challenges and answer those questions is a worthwhile effort.

References

Abrams, L. M., & Madaus, G. F. (2003). The lessons of high stakes tests. *Educational Leadership, 61*(3), 31–35.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18).

Association for Experiential Education. (1991). AEE definition of experiential education. Retrieved November 16, 2004, from http://www.aee.org/customer/pages.php?pageid=47

Baron, J. (1993). Why teach thinking? An essay. *Applied Psychology: An International Review, 42*, 191–214.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain.* New York: Longmans, Green.

Center on Education Policy. (2004). State high school exit exams: A maturing reform. Retrieved February 17, 2005, from http://www.ctredpol.org/highschoolexit/ExitExamAug2004/ExitExam2004.pdf

Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What will it take? *Theory into Practice, 42*, 75–83.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). Hillsdale, NJ: Erlbaum.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 42*, 75–83.

Dewey, J. (1933/1998). *How we think.* Boston: Houghton Mifflin.

Dori, Y., Tal, R. T., & Tsaushu, M. (2003). Teaching biotechnology through case studies—Can we improve higher order thinking skills of nonscience majors? *Science Education, 87*, 767–793.

Druian, G., Owens, T., & Owens, S. (1980). Experiential education: A search for common roots. *Journal of Experiential Education, 3*, 5–10.

Ewert, A. (1987). Research in experiential education: An overview. *Journal of Experiential Education, 10*, 4–7.

Goodenow, C. (1992, April). *School motivation, engagement, and sense of belonging among urban adolescent students.* Paper presented at the American Educational Research Association, San Francisco.

Granello, D. H., & Underfer-Babalis, J. (2004). Supervision of group work: A model to increase supervisee cognitive complexity. *The Journal for Specialists in Group Work, 29*, 159–173.

Grant, S. G., Gradwell, J. M., Lauricella, A. M., Derme-Insinna, A., Pullano, L., & Tzetzo, K. (2002). When increasing stakes need not mean increasing standards: The case of the New York state global history and geography exam. *Theory and Research in Social Education, 30*, 488–515.

Grossen, B. (1991). The fundamental skills of higher order thinking. *Journal of Learning Disabilities, 24*, 343–353.

Hedin, D. (1983). The impact of experience on academic learning: A summary of theories and review of recent research. *IRE Report No. 9 ERIC Document Reproduction No. ED250356.*

Hillocks, J. G. (2002). *The testing trap: How state writing assessments control learning.* New York: Teachers College Press.

Institute of Educational Sciences. (2004). *Cognition and student learning research grants: CFDA 84.305H.* Washington, DC.

Ives, B. (2003). Effect size use in studies of learning disabilities. *Journal of Learning Disabilities, 36*, 490–504.

Ivie, S. D. (1998). Ausebel's learning theory: An approach to teaching higher order thinking skills. *High School Journal, 82*, 35–42.

Junn, J. (1991). Participation and political knowledge. In E. Crotty (Ed.), *Political Participation and American Democracy.* Westport, CT: Greenwood Press.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). What do test scores in Texas tell us? Retrieved November 10, 2004, from http:///www.rand.org/publications/IP/IP202/

Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy, 18*, 45–70.

Kosciulek, J. F. (2003). On critical thinking. *Rehabilitation Education, 17*, 71–79.

Laney, J. D. (1989). Experience- and concept-label-type effects on first-graders' learning, retention of economic concepts. *Journal of Educational Research, 82*, 231–236.

Leming, J. S. (1998). Some critical thoughts about the teaching of critical thinking. *The Social Studies, 89*, 61–66.

Levine, D. U. (1994). Instructional approaches and interventions that can improve the academic performance of African American students. *Journal of Negro Education, 63*, 46–63.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4–16.

Marchant, G. J. (2004). What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science, 104*(2), 2–7.

Martin, D. S. (1993). Reasoning skills: A key to literacy for deaf learners. *American Annals of the Deaf, 138*(2), 82–86.

Mathison, S., & Freeman, M. (2003). Constraining elementary teachers' work: Dilemmas and paradoxes created by state mandated testing. *Education Policy Analysis Archives, 11*, 34.

McNeil, J. D. (1990). *Curriculum: A comprehensive introduction.* Glenview, Illinois: Scott Forsman.

National Academy of Education. (1997). *Assessment in transition: Monitoring the nation's educational progress.* Stanford, CA: National Academy of Education.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Neill, M. (2003). The dangers of testing. *Educational Leadership, 60*(5), 43–46.

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Niedelman, M. (1991). Problem solving and transfer. *Journal of Learning Disabilities, 24*, 322–329.

Pennington, J. L. (2004). *The colonization of literacy education: A story of reading in one elementary school.* New York: Peter Lang.

Perkins, D. N., & Salomon, G. (1988). Teaching to transfer. *Educational Leadership, 46*, 22–32.

Rahm, I. (2002). Emergent learning opportunities in an inner-city youth gardening program. *Journal of Research in Science Teaching, 39*(2), 164–184.

Reynolds, A. J. (1991). Effects of an experiment-based physical science program on cognitive outcomes. *Journal of Experiential Education, 84*, 296–302.

Roberts, N. S., & Yerkes, R. (2000). Experiential education research: Where do we go from here? *The Journal of Experiential Education, 23*, 61–63.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*, 113–142.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31*, 13.

Shelton, J. (2000). A life connected to community: An interview with Jack Shelton. *Active Learner: A Foxfire Journal for Teachers, 5*(1), 164–184.

Swanson, H. L. (2001). Research on interventions for adolescents with learning disabilities: A meta-analysis of outcomes related to higher order processing. *The Elementary School Journal, 101*, 331–348.

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes.* New York: Guilford.

Torff, B. (2003). Developmental changes in teachers' use of higher order thinking and content knowledge. *Journal of Educational Psychology, 95*, 563–569.

Weinbaum, A., Gregory, L., Wilkie, A., Hirsch, L., & Fancsali, C. (1996). *Expeditional Learning Outward Bound: Summary report.* New York: Academy for Educational Development.

Weiss, I. R. (1997). The status of science and mathematics teaching in the United States: Comparing teacher views and classroom practice to national standards. *ERS Spectrum, 15*(3), 34–39.

Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development, 62*, 1066–1078.

Zohar, A., & Dori, Y. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive. *The Journal of the Learning Sciences, 12*, 145–181.