

1-1-2015

Assessing Students' Acquisition of Scientific Reasoning in an Experimental Psychology Class

Gayle S. Morse
Utah State University

Donald F. Graves
The Sage Colleges

Kerry K. Prout
The University at Albany, State University of New York

Jennifer Safford

Follow this and additional works at: <http://scholarworks.merrimack.edu/phs>

Recommended Citation

Morse, G. S., Graves, D. F., Prout, K. K., & Safford, J. (2015). Assessing Students' Acquisition of Scientific Reasoning in an Experimental Psychology Class. *Pedagogy and the Human Sciences*, 5 (1), 41-59. Retrieved from <http://scholarworks.merrimack.edu/phs/vol5/iss1/2>

This Article is brought to you for free and open access by Merrimack ScholarWorks. It has been accepted for inclusion in Pedagogy and the Human Sciences by an authorized administrator of Merrimack ScholarWorks.

*Assessing Students' Acquisition of Scientific Reasoning in an
Experimental Psychology Class*

Gayle S. Morse¹ Donald F. Graves², Kerry K. Prout³, Jennifer Safford

Abstract. This pilot study is an initial exploration of a theoretical rubric proposed to "describe the progress of students' acquisition of scientific inquiry" (Halonen et al., 2003, p. 196), and an application of the utility of the rubric. Twenty-two undergraduates from a woman's college participated in two sections of experimental psychology. Students consisted of sophomores, juniors, and seniors who completed general psychology courses. Consistent with the Halonen et al. (2003) model, results indicated that authentic research experiences in the first phase of the course were positively correlated with changes in scientific thinking in a second phase. In turn, experiences in the second phase were positively correlated with evidence of advanced thinking skills in a third phase. The findings suggest that much of the basic skill knowledge acquired in the beginning lectures, textbook readings, and writing instruction of the course enhanced students' ability to apply that knowledge in later classes and the lab components. Further, the authentic learning experiences were instrumental in fine-tuning the skills learned from the lectures and textbooks readings. As a result, the current authors advocate the use of authentic experiences in teaching research methods, as a way for teachers to transform such classes in a beneficial and systematic way, in order to enhance acquisition of scientific thinking skills and to examine changes in scientific thinking as explicated in the Halonen et al. (2003) model.

Keywords: rubrics, scientific reasoning, pedagogy

¹ Utah State University, ²The Sage Colleges, ³The University at Albany, State University of New York

I.

This is a pilot study designed as an examination of a rubric proposed to "describe the progress of students' acquisition of scientific inquiry" (Halonen et al., 2003, p. 196), and to begin to understand an application that demonstrates the utility of the rubric. Halonen et al. (2003) described a developmental rubric that was formulated by a group of psychology educators working within the American Psychological Association (APA) Psychology Partnership Project (P3). The rubric was an effort to develop a comprehensive system to help educators systematically assess students' acquisition of scientific inquiry skills.

The rubric covered eight domains of scientific inquiry (i.e., Descriptive Skills, Conceptualization Skills, Problem-Solving Skills, Ethical Reasoning, Scientific Attitudes and Values, Communication Skills, Collaboration Skills, Self-Assessment) and specific sub-skill areas within these domains of scientific inquiry. Further, these domains of scientific inquiry and sub-skill areas were broken down by levels of proficiency, ranging from naive before training to expert after training. Halonen et al. (2003) described these domains and skill areas in detail (see Halonen et al., 2003 for further descriptions of these domains and skill areas). The current study examined only a portion of this rubric. Specifically, the study examined the Developing and Integrated Advanced Undergraduate levels of proficiency, across all of Halonen's et al. (2003) eight domains of scientific inquiry. The rubric was applied to an upper-level undergraduate research and design course that implemented an authentic assessment approach. Authentic learning measures are unique because they provide a functional, contextual evaluation of learning development that allows for a more personalized understanding of an individual's capabilities and goals (Bagnato, 2007). Authentic assessment involves examining contextual evidence along with observation and interview techniques to gather data on learning and development (Keilty, LaRocco, & Casell, 2009). In sum, authentic assessment is defined as using "real-world" tasks and meaningful activities that highlight the relevance of material learned and that allows the evaluation of performance on the task, rather than a paper and pencil test only (Halonen et al., 2003). Thus, students are able to experience the connection between the classroom tasks and those future tasks that they will encounter when in graduate school and as professionals in the field. A rubric answers that age-old student question, "Why are they making us to do this?" with "Because it is what all professionals in the field do every day." The rubric allows students to learn how while learning the value of the task and topic. By having students engage in projects that first focused on APA-style writing and basic hypothesis formation, we laid the ground work for acquiring more complex ideas and more complex writing skills. By completing the assigned projects students can learn, practice, and master scientific thinking and design skills in a way that is experiential rather than semantic only.

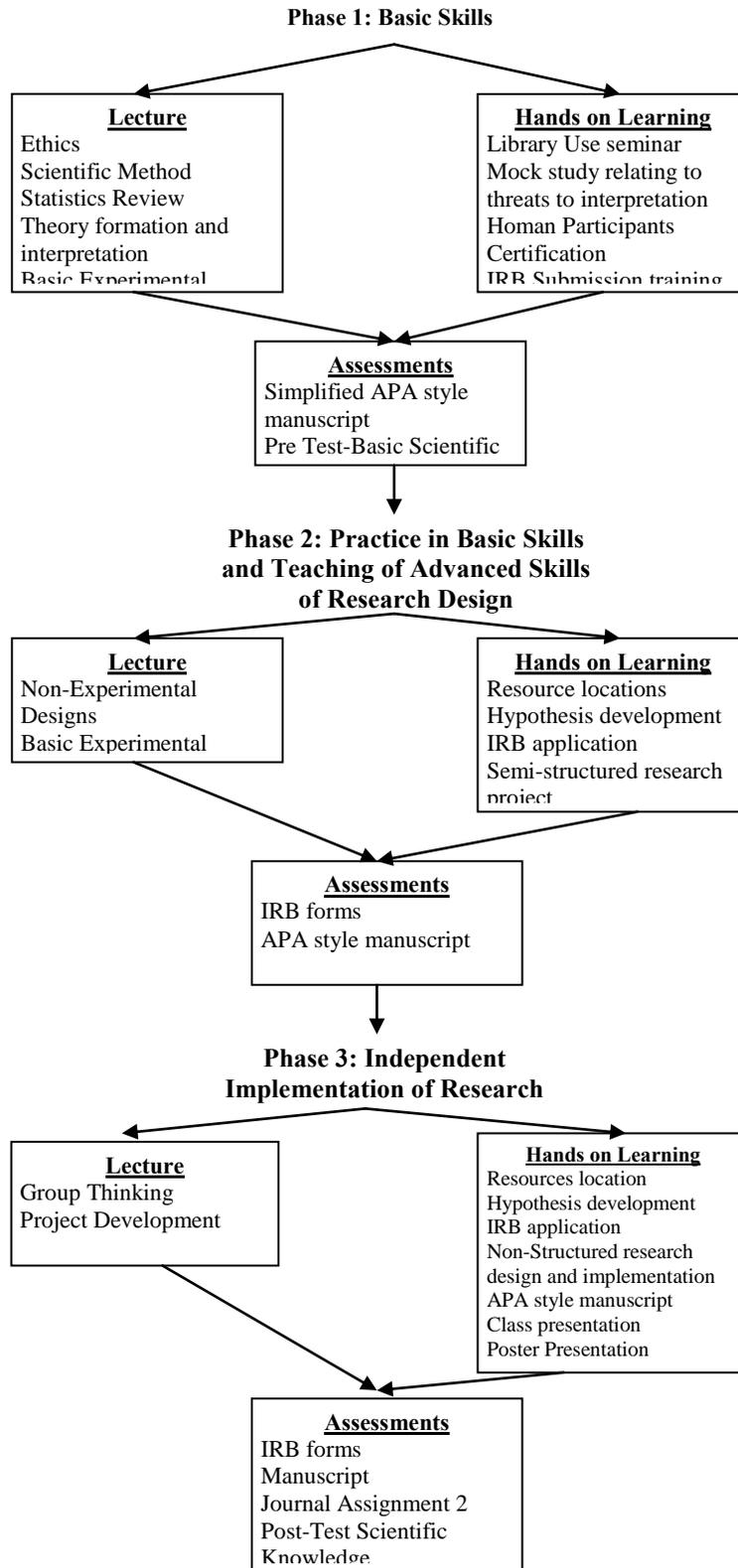
Utilizing the authentic assessment techniques suggested by Grant (1990), we designed classroom exercises that required students to collaborate in groups to design and to implement meaningful research projects, interpret the outcomes, and publicly present their results. We believed that by using the authentic assessment techniques, it would be possible to document a change in the scientific reasoning skills (aka: scientific inquiry skills) in these upper-level undergraduate students from the time that they are introduced to scientific thinking to actual application of those methods of thinking. The change in reasoning ability was measured by using

43 Assessing Students' Acquisition of Scientific Reasoning

a detailed grading procedure for the class experiments and group exercises (Biggs, 2003; Grant, 1990; Meyers & Nulty, 2009).

Changes in scientific inquiry abilities were assessed in accordance with the *Developing* and *Integrated - Advanced* Undergraduate levels of proficiency described in the Halonen et al. (2003) rubric. The Developing Level of proficiency is characterized by emerging application of scientific concepts in methods and conceptualization of projects but with counterintuitive information and frequently seeks supervision. The Integrated - Advanced level of proficiency is characterized by more independence in applying scientific methods and theory and by logically overcoming counterintuitive information. These levels of proficiency were selected because the first two levels of Halonen et al.'s (2003) rubric theoretically would have been achieved by prior student participation in lower-level psychology classes that were prerequisites to the Experimental Psychology class. The two levels of proficiency, which we assumed students achieved prior to the Developing level of proficiency, included: *Before Training* and *Basic Introductory Psychology*. Furthermore, Halonen et al. (2003) indicated that the Developing and then the Integrated - Advanced levels of undergraduate proficiency may provide useful markers of changes that we would expect to see in students as they move from their first exposure to psychology as a scientific discipline to a bachelor level psychology graduate. It was assumed that our sample of undergraduates would reach the Developing level criterion after the basic building blocks of scientific inquiry were presented (e.g., the scientific method, theory formation) and continue on to the Integrated -Advanced Undergraduate level after exposure to and implementation of more advanced methods of reasoning (e.g., between & within groups designs; See Figure 1 for a flow chart of the lecture and experiential events in the class.)

Figure 1. Class Assignment flow chart



45 Assessing Students' Acquisition of Scientific Reasoning

Thus, our course design followed the frequently used approach of presenting basic material followed by expansions on that material using a lecture format. In addition to lectures and textbook readings (Myers & Hanson, 2005a) hands-on learning opportunities that were consistent with Grant (1990) and Biggs (2003) were provided. These learning opportunities allowed the students to use their knowledge as professionals in the field of psychology. This procedure allowed us to assess changes in scientific inquiry abilities with authentic measures such as journal and manuscript writing (Grant, 1990).

The purpose of the pilot project was to examine how we could observe the progress of students' acquisition of scientific inquiry skills using the Developing and Integrating Advanced Undergraduate Levels of Proficiency described by the Halonen et al. (2003). We hypothesized that we could enhance this process using authentic assessment techniques as described by Grant (1990) and suggested by Halonen et al. (2003). This objective supports the rationale of the rubric, which is to guide course development, to design learning experiences that enhance learning, to evaluate course progress, to clearly define precise goals for students, and to reinforce good teaching practices.

II.

Method

Participants

Twenty-two female undergraduates (psychology majors or minors) from two different sections of experimental psychology in a women's college participated in the study. These students were sophomores, juniors, and seniors who had taken the prerequisite psychology courses: Introduction to Psychology and Statistics with Computer Applications. The participants ranged in age between 18 and 37 with a mean age of 20.76 ($SD = 4.05$).

Instructors

Instructors were PhD level psychologists, one a cognitive psychologist (male) and one clinical psychologist (female). They had active research programs, and both taught undergraduate and graduate level research methods classes. The male psychologist also had extensive experience teaching statistical methods at the undergraduate level. They team taught the classes so students had exposure to two different styles of teaching: one a more lecture style and one a more Socratic style.

Class

Previously the class was taught in a standard lecture format with a midterm and a final and no hands on experience. The faculty determined that students would benefit from authentic learning techniques that would enhance more scientific thinking. Class proceeded as a mixture of readings, lecture, and classroom activities. The classroom activities benefited from the familiarity

of class members and faculty due to the makeup of the school and psychology department. Thus, a strong group work ethic was fostered within and around the class environment.

Classroom Materials

The textbook was Myers and Hansen's (2005a) *Experimental psychology* (6th ed.), which is a traditional experimental text in that it begins with chapters on basic concepts such as the scientific method, ethics, and hypothesis formation and then goes into non-experimental and experimental designs. The text served as the primary reference for the course and lectures expanded on the topics in the text. Lectures used overhead screen slide projections and a conversational style. Students were given off-slide examples and encouraged to ask questions. The slides were not made available to the students in an electronic format, thus students had to take notes. The instructors used an additional source as well, Langston's (2005) *Research Methods Laboratory Manual for Psychology* (2nd), which contained summaries of research that related to the common sections of an experimental course. For example, one chapter discussed a one-way randomized multiple group design that described Alloy and Abramson's (1981) judgment of control task. The manual presented a summary of the research, potential readings, potential variations of the study, and computer software to conduct studies on the topic within a classroom environment. This software served as the basis for the class projects and as a mode of data collection both for students and for instructors. Throughout the course peer-reviewed articles supplemented these two sources of materials. (See references in procedure section).

Measures

The measures utilized included multiple methods of assessment such as exam grades (traditional multiple-choice exams), research lab reports, journal entries, and an objective pre- and post-test. This course was presented in three phases: Phase 1-Basic skills, Phase 2- Practice and learning, and Phase 3- Independent implementation of skills (See Figure 1).

Phase 1: Basic skills measures.

1. The grade for a simplified APA-style manuscript based on an in-class research project- Project 1.
2. The score for pre-test of Basic Scientific Knowledge at the end of Phase 1.
3. The grade for the first Journal Assignment.

Phase 2: Practice and learning measures.

1. The grades for completed IRB forms for the second research project- Project 2
2. The grade for an APA-style manuscript based on Project 2 and closely supervised by faculty.

Phase 3: Independent implementation of research measures.

1. The grade for completed IRB forms for Projects 3 and 4.
2. The grade for an APA-style Manuscript based on Project 3.

47 Assessing Students' Acquisition of Scientific Reasoning

3. The grade for the second Journal Assignment.
4. The grade for an APA Poster Presentation to Faculty based on Project 4.
5. The score for a Post-test of Scientific Knowledge at the end of Phase 3.

Estimate of acquisition of scientific inquiry skills.

1. Journal 1 and 2 entries were reviewed for statements that reflected the Developing and Integrating Advanced Level of Proficiency of the rubric (Halonen et al., 2003). The number of statements in each category served to measure the difference from Journal 1 after the first half of the semester to Journal 2 after the final half of the semester.

Procedures**Phase 1 - Basic Skills.**

Phase 1, conducted during the first four weeks of class, was designed to review basic skills obtained in previous courses such as Ethics, Statistics, and Introduction to Psychology. Phase 1 was delivered primarily in a lecture-style format with some hands-on projects designed for illustrative purposes. The study requirements for this phase included chapters (Myers & Hanson, 2005a) discussing APA ethical guidelines; APA-style reading; components involved in research with human participants such as informed consent, debriefing, the Institutional Review Board (IRB), experimental validity, and hypotheses. During this time, we also provided meaningful activities to provide context for the skills discussed in the lectures. For example, one project utilized horoscope interpretation to illustrate internal and external validity. The participants also attended a session at the library in which they received instructions about reading and retrieving empirical and literature articles. Course instructors then introduced a form (Appendix A) method for reading and taking notes on each section of a professional research paper. The participants used these forms throughout the semester as a tool in understanding and explaining assigned articles that contributed to their hypotheses for group projects.

Students were also required to complete an IRB submission and obtain certification in Human Subjects Research for Project 1. Project 1, a simple survey study, was done step by step with the instructors who explained each step with clear examples. The study was modeled after the correlational research example (chapter 3) provided in Langston's (2005) research methods laboratory manual. The course instructors provided articles for the literature review (i.e., Baun, Bergstrom, Langston, & Thoma, 1984; Cohen, & Williamson, 1991; Friedmann, Katcher, Lynch, & Thomas, 1980) and the logical relationships between the articles and the survey's hypothesis were discussed during lectures. The students then wrote their literature review sections and developed their hypotheses together in class. Students were required to create an informed consent and debriefing script using existing sample templates provided by the Institutional Review Board and upgraded to reflect their new hypotheses and the current topic. Next, students practiced administering consent forms, the survey, and debriefing with students from their Experimental Psychology class. The students then were guided through the data analyses (a series of correlational analyses), writing the methods, and results section of their paper in APA-style and format.

Finally, during the last class of the basic skills component, students completed an objective measure, a pre-test designed to assess scientific reasoning knowledge. The pre-test was developed from the test bank of their class text (Myers & Hanson, 2005b). Multiple-choice questions were selected that required students to apply the knowledge they had reviewed thus far. Students were informed these tests were not part of their official class grade, but were asked to think carefully about the questions and take their time to respond as thoughtfully as possible.

Journal Assessment.

A set of journal entries was collected at the end of this phase. Participants were required to have one journal entry per class period. They were instructed to complete the first section of the journal entry by describing either material from the lecture, a reading assignment, or a project that piqued their interest. In the second section, they were instructed to write a thoughtful response that tied the area of interest to what they had learned thus far (See Table 1 for examples of Developing and Integrated Proficiency Domains statements).

49 Assessing Students' Acquisition of Scientific Reasoning

Table 1

Illustrative Journal Statements for the Developing and Integrated Proficiency Domains

Domains	Developing	Integrating
	<i>Undergraduate (basic classes completed - Introductory Psychology and Statistics)</i>	<i>Advanced Undergraduate</i>
Descriptive Skills	<p>“In my opinion, media violence alone is not enough to elicit aggressive behavior... takes into account home life, economic status, peers, school life, etc.”</p>	<p>“One of the major benefits of carrying out a participant observer study is that the people being studied will likely be more comfortable with someone who appears to be from their group, compared to someone that they know is observing them and whom is not a part of their group or society.”</p>
Conceptualization Skills	<p>“At the heart, Type I error is that we don’t want to make an unwarranted hypothesis, so we exercise a lot of care by minimizing the chance of its occurrence. For instance, it’s like saying a woman is pregnant, when in all reality she is not, not exactly an error you want to make.”</p>	<p>“The manner in which the sample accurately represents the population is representativeness. The greater the representativeness, the greater the generalizability of the research will be. Researchers try to increase the generalizability of their research to increase external validity.”</p>
Problem Solving Skills	<p>“I was also interested in how people tend to have over confidence in their judgments and how people look for examples to confirm their own biases and tend to disregard information to the contrary.”</p>	<p>“The Rosenthal Effect is a form of experimenter bias... describes ways in which an experimenter’s behavior towards his/her participant changes according to the expectations they have of the volunteers.”</p>

Ethical Reasoning

“Because investigators of child abuse cannot actually abuse a randomly chosen group of children, they must instead compare children who already have a history of abuse with children who do not (ethics).”

“It was evident that the researcher didn’t use an informed consent at the end for permission to release the data collected. Moreover, he had developed close friendships with these people. I would imagine that the participants felt betrayed and were emotionally injured.”

Scientific Attitudes and Values

“The wording of survey questions is very important. The wording must be very clear and you must be very careful not to have more than one idea per question.”

“Although we tried our best to pick neutral words for the experiment in order to prevent any emotional factors from playing an unwanted part, we believe that there are still several possible factors that could potentially confound the experiment.”

Collaboration Skills

“After gaining a better understanding of the variables, we thought about the correlations and tried to find a question.”

“When designing our experiment... We decided to create four conditions, two consisted of the highest stress words and the other two consisted of the least stress words... We were trying to create a design that would measure the effect of emotional words on a scale while trying to keep the stimuli grouped.”

51 Assessing Students' Acquisition of Scientific Reasoning

Communication Skills	“The three articles discussed the relationship between pet ownership and health, and also the relationship between stress and health.”	“However, what is more likely the case is that mood states (especially depression) are in fact cyclic. For example, I may take a difficult chemistry test and not pass, which would understandably put me in a depressed mood.”
Self-Assessment	“I had been horrified of statistics before I took the course. I soon realized that they are not that scary. I find that some studies are clear with analyzing the results statistically.”	“For example, in doing our experiment, my group noticed multiple confounds that to account for would have taken redesign and revision that time would not allow for.”

Phase 2: Practice and Learning.

Phase 2 was designed to offer students practice in research methodology, ethics, hypotheses development, and statistical interpretation. The chapter study requirements for this phase included chapters (Myers & Hanson, 2005a) discussing survey research and sampling, statistics, research report writing, analyzing results, inferences and evaluating the study, and basics of experimentation and non-experimental designs.

Project 2 required students to complete the same aspects of a research paper as in Project 1. In this project, the students expanded a predesigned research methods lab that required the manipulation of the independent variable. Additionally, this project allowed for random assignments for three different conditional groups. The study was modeled after the one-way randomized multiple groups design research example (chapter 11 and accompanying software) provided in Langston's (2005) research methods laboratory manual. The project was based on the induction of mood using Velten's (1968) mood statements to create a three-level independent variable, which was combined with Alloy and Abramson's (1981) judgment of control task. The judgment of control task allowed a mixture of one-factor hypothesis to be created as well as several correlational hypotheses. Thus, students could develop varied hypothesis. They were required to have at least one one-factor hypothesis and one correlational hypothesis. As in Project 1, instructors provided three initial papers for the literature review (i.e., Alloy, & Abramson, 1981; Seligman, & Maier, 1967; Velten, 1968) and students needed to obtain two additional papers to guide the development of their hypotheses. Hence, students were required to apply the skills presented in the library usage seminar and the hypothesis development sections of the lecture and the text. By creating and supporting their own hypothesis, they were exposed to the complexities of hypothesis development and had to consider the theoretical and practice issues of creating workable scientific hypothesis. They completed their literature reviews, developed

hypotheses, and wrote their introduction and methods sections, independently. This time, the data analysis, results, and discussion sections were worked on in class with the instructor, and papers were completed in APA style. Students also practiced completing IRB protocols, consent forms, debriefing statements, and standardized administration of experimental procedures. This project allowed students to apply methods of controlling extraneous variance that were discussed in this phase and the previous phase. This phase was accomplished after completion of this study, which was a structured pre-planned experiment designed to allow the participants hands-on practice interacting with human participants in an ethical and professional way and to experience the relationship between hypotheses formation, research design, and interpretation.

Journal Assessment.

Students did not turn in any journal assignments during Phase 2.

Phase 3: Independent Implementation of Research.

Phase 3 offered students an opportunity to design two studies independently using instructor-provided experimental tasks (based on software that accompanies chapters 4 and 7 in Langston, 2005) that allowed students to manipulate the variables to better answer the relationships they chose to examine. The projects in Phase 3 required students to conduct an independent literature search, adding to the two articles that faculty provided before they formulated their hypotheses, thus allowing them to demonstrate the scientific reasoning skills learned previously in the course, to choose their design, and deal with data collection concerns. The study requirements for this phase included chapters (Myers & Hanson, 2005a) discussing correlation and quasi-experimental designs, between- and within-subjects designs, factorial designs, and small group designs. During Phase 3, the third project, a Lexical Decision Task (based on Langston, 2005, chapter 7 and accompanying software) was assigned, and two articles were provided (i.e., Halberstadt, & Niedenthal, 1997; Niedenthal, & Setterlund, 1994) and each student group was required to obtain two more articles that supported their independently developed hypothesis. For this project, students were required to conduct their work independently with limited supervision. The students completed a literature review, formed their hypotheses, completed IRB forms, created informed consent and debriefing forms, and created their computer stimuli within the computer program. After the IRB approved their study, they completed it with the other class serving as participants, as well as students recruited from the general college population. After data collection was completed, students independently completed their data analysis, results, and discussion sections. The projects were presented orally in class with audio or visual support, and finally a manuscript was submitted in APA-style format.

During Phase 3, students also completed their 4th project called *The Stroop Project* (based on Langston, 2005, chapter 4 and accompanying software). The students were only provided one paper to read, the original Stroop (1935) paper. They then had to obtain three or four more papers to develop their literature review, hypotheses, and methods section. They were required to use the Stroop Effect to understand the relationship between their chosen independent and dependent variables. Students developed this project independently similar to Project 3 with the exception of the manner in which they presented their experiments. For this project, the participants were required to create an APA-formatted poster and participate in a poster session for the Psychology

53 Assessing Students' Acquisition of Scientific Reasoning

Department faculty and students. The students needed to be prepared to discuss the experiment material and answer questions. This project was also submitted in an APA-style manuscript format.

At the end of Phase 3, journal assignments were collected again, and a post-test, similar to the pre-test, was administered. The questions and answers on the post-test were presented in a different randomly selected order than the pre-test.

Journal Assessment.

At the end of Phase 3, journal assignments were collected for the second time.

Coding Activities.

Two sets of five journal entries, or informal papers, were used to evaluate the participants' integrated understanding of the material. The first set of five journal entries was collected after the first phase of the class, and the second set of five journal entries was collected after the last phase of the class. Independent readers, who were graduate student research assistants in the community counseling program, reviewed the journals to evaluate the progress of students' acquisition of scientific inquiry skills. Each reader was randomly assigned journals with no indication of whether they had journals from the first or final phase of the class. The readers were instructed to identify statements from the journals that matched the definition of the Developing or Integrated-Advanced Undergraduate levels of proficiency (Halonen et al., 2003). Readers were given prepared scoring sheets that defined the specific skills for each domain. They then selected all instances of writing that represented evidence of the specific domain. The final statements included comments such as those found in Table 1. An example from the Problem Solving Skills domain of a Developing level statement is: "This taught me that when conducting an experiment I need to consider all the dynamics of the situation." In contrast, a corresponding Integrated-Advanced undergraduate level statement for this same domain is: "I feel there are too many issues involved with using internet surveys, that the benefits are just not desirable enough to take such a gamble." This item shows the complexity in the level of student response.

III.

Results

Data were first examined to evaluate the assumptions underlying a normal distribution and found to be within tolerances. Then, the data were analyzed using bivariate correlations focusing on grades obtained for each project or journal (thus, higher values indicated greater competence). Finally, the total number of scientific statements at each level of proficiency was evaluated across journal 1 and 2 with a two-way repeated measures ANOVA to determine if journals could be used as a measure of change in scientific reasoning. Each analysis in the project was held to a .05 level of significance.

Table 2 Correlation Matrix of Measure Scores

	Projects						Pre-test
	Project 1	Journal 1	Journal 2	Project 2	Project 3	Project 4	
Project 1	---						
Journal 1	.567**	----					
Journal 2	.311	.670**	----				
Project 2	.088	.170	.679**	----			
Project 3	-.095	-.036	.615**	.689**	----		
Project 4	.192	.248	.442*	.513*	.209	----	
Pre-test	.105	.259	.220	.169	.043	.364	---
Post-test	.375	.162	-.117	-.348	-.217	.016	.133

Note: * is significant at the .05 alpha and ** is significant at the .001 alpha level.

The correlation matrix was examined to understand the relationships among the measured outcomes (Table 2). Results indicated that the grades of the pre-test and post-test had no statistically significant relationship with any other outcome ($p > .05$). Project 1 was only correlated with Journal 1 ($r(22) = .567, p = .006$), indicating that the lectures and readings served to help students hone their ability to apply APA writing skills and basic hypotheses development in a manuscript (Developing Phase skills). It was not surprising that Project 1 did not correlate with Project 2, 3, or 4, given the drastic change in focus between Project 1 (title page, introduction, methods, and reference sections, which focus on basic APA style and use of logic) and Projects 2, 3 and 4 (full manuscripts, which focused more on advanced APA-style and statistical logic). Journal 1, which occurred at the end of Phase 1, was related to the Journal 2 ($r(21) = .67, p < .001$) which occurred at the end of Phase 3). This correlation may be interpreted as evidence that scientific knowledge gained early in the course was maintained and enhanced by further lectures, reading, and experiences.

Phase 2, Project 2 was correlated with Journal 2 ($r(22) = .679, p < .001$), Project 3 ($r(22) = .689, p < .001$), and Project 4 ($r(22) = .513, p = .015$) all of which were in Phase 3. In Phase 3, Journal 2 was correlated with Project 3 ($r(22) = .615, p = .002$) and Project 4 ($r(22) = .442, p = .039$) while Project 3 was not statistically related to Project 4. These findings indicated that Project 2 (a full manuscript) was correlated with scientific thinking in Journal 2, which related to increases in Project 3 and Project 4, while the shift in focus between Project 3 (a pre-prepared

55 Assessing Students' Acquisition of Scientific Reasoning

study in which the students picked and supported their own hypothesis) to Project 4 (a completely student prepared study based on the general Stroop (1935) paradigm) rendered the projects unrelated.

Table 3: *Mean number of statements output in journal entries one and two as defined by Halonen et al.'s (2003)*

Domains and Skills Areas	Level of Development			
	Developing		Integrated	
	1st	2nd	1st	2nd
Descriptive Skills	2.00	.91*	.27	.91*
Conceptualization skills	1.27	.18*	.27	1.00
Problem-solving skills	1.45	.64*	.73	1.82*
Ethical-reasoning	1.45	.18*	.36	.82
Scientific values and attitudes	.91	.46	1.27	1.73
Communication skills	.80	.50	.40	1.00
Collaboration skills	.73	.09*	.00	1.27*
Self-assessment	.82	.46	.00	.55*
Total Scientific Statements	9.55	3.00*	3.27	9.18*

Note: * inter-level (developing or integrated) comparison is significant at $p = .05$

Analysis of Scientific Statements found within the Journals.

The mean number of statements from journal 1 and journal 2 were examined in an effort to understand the students' acquisition of scientific inquiry skills (see Table 3). For each domain a two-way repeated measures ANOVA with journal (1st or 2nd hand in time) by skill level (developing vs. integrated) was conducted. Based upon the analysis of the frequency of scientific statements from these journals, there appears to be a pattern of movement from a Developing level of scientific inquiry to a more advanced-integrated and sophisticated level of scientific inquiry. For the "total scientific statements" analysis there was a significant interaction between Journal and Skill Level ($F(1, 10) = 20.64$, $MSE = 20.67$, $p = .001$), such that journal 1 included more Developing level statements than journal 2, while journal 2 included more Advanced-Integrated level statements than Developing level statements. These results suggest an overall shift in the level of scientific thinking across the journals, where students were using more Integrated level scientific-minded thinking by the end of the course. Of the eight "domains and skills areas" defined by Halonen et al. (2003) six showed a shift from Developing level statements dominating the response to Integrated level statements dominating the response (see Table 3). The other two domains "Scientific values and attitudes" and "Communication skills" failed to show the pattern observed in the overall analysis. Inter-rater reliability for the total number of statements as measured by correlation was broader line but significant, $r(9) = .629$, $p = .038$.

IV.

Discussion

This study was designed as a pilot study to examine a rubric proposed to "describe the progress of students' acquisition of scientific inquiry" (Halonen et al., 2003, p. 196). Our results indicate that students likely do progress from the Developing level to the Advanced-Integrated level while taking an Experimental Psychology course. In particular, progress is seen when students are provided a course design that follows the frequently used approach of presenting basic material followed by expansions on that material using lecture, textbook readings as well as the less frequently used authentic experiences and measures. The addition of authentic experiences and measures to standard approaches may be transformative in that it helps teachers create courses that enhance students' acquisition of scientific inquiry skills. This finding was supported by positive correlations between early projects and later projects, as well as positive correlations between early projects and journals assignments that measured the number of scientific statements. Specifically, Project 1 (a partial manuscript) was related to scientific thinking in Journal 1 that carried over to Journal 2. This finding suggest that the skills learned while writing the introduction and methods section of a paper for Project 1 may have facilitated students thinking about the logical process of science, which was later reflected in their increased use of Advanced-Integrated scientific statements and decreased reliance on Developing statements from Journal 1 to Journal 2. So, it appears that the rubric may have captured some element of students' progress from the developing level to the integrated level of scientific inquiry in a course like that administrated here.

The authentic learning measures used within this study (e.g., hands on exercises, the APA style manuscripts) provided a functional, contextual evaluation of learning development as detailed by Halonen et al.'s (2003) rubric. By utilizing class materials that are real-world and

57 Assessing Students' Acquisition of Scientific Reasoning

relevant, constructive and interlinked, and that require engagement with progressively higher order cognition (Biggs, 2003), we were able to measure skills changes in six of Halonen et al.'s domain and skills areas, as well as an overall shift in thinking from a Developing level to an Advanced-Integrated level. As no control condition was included within this study we were unable to determine whether these authentic measures were the causative component of learning, but we believe they contributed significantly to the learning experience over and above lecture and textbook readings found in traditional lecture format Experimental Psychology courses. Confidence in this belief is increased by Meyers and Nulty's (2009) finding that students' that engaged in authentic measures produced work that indicated significantly greater proportions of responses displaying multi-structural, relational, and abstract levels of thinking compared to students from the previous year who had not been taught with authentic measures. Thus, it appears that a change in reasoning ability was measured by using a detailed grading rubric of the class experiments and group exercises.

The strength of the current study lies in the use of a carefully described theory as the basis of this investigation. Further, the design and longitudinal nature of the present study help show change over time. Hence, we have a high degree of confidence that the correlations found between early projects and later projects are indicative of skill acquisition and mastery. Finally, the organization, extensive set of measures, and format of the procedures enhanced the collection of the data.

The primary limitation of this pilot study is the small sample, which carries some threats to validity. However, the small sample size allowed for the use of a more comprehensive set of measures and procedures than might otherwise occur. In addition, the longitudinal nature of this design provides internal control as the same participants are sampled at different points in time. The information gleaned from the current study will allow us to formulate a more concise, more efficient set of measures and procedures to expand upon for a future study.

Our use of a convenience sample consisting of advanced undergraduate students could impair our ability to generalize our data. Nonetheless, the sample did allow us to begin studying a relatively complex rubric for learning, teaching, and assessing scientific inquiry in psychology, in a systematic way. As a practical matter these data also suggest that three rather than four research projects may be adequate authentic assessments to help students understand the material in an experimental psychology class.

References

- Alloy, L. B., & Abramson, L. Y. (1981). Induced mood and the illusion of control. *Journal of Personality and Social Psychology*, 41, 1129-1140.
- Bagnato, S. J., (2007). *Authentic Assessment for Early Childhood Intervention: Best Practices*. London, UK: Guilford Press.
- Baun, M. M., Bergstrom, N., Langston, N. F., & Thoma, L. (1984). Physiological effects of human/companion animal bonding. *Nursing Research*, 33, 126-129.
- Biggs, J. (2003). *Teaching for Quality Learning at University* (2nd Ed.). Berkshire, UK: Buckingham Society for Research into Higher Education and the Open University Press.
- Cohen, S., & Williamson, G. J. (1991). Stress and infectious disease in humans. *Psychological Bulletin*, 109, 5-24.
- Friedmann, E., Katcher, A. H., Lynch, J. J., & Thomas, S. A. (1980). Animal comparisons and one-year survival of patients after discharge from a coronary care unit. *Public Health Reports*, 95, 307-312.
- Grant, W. (1990). *The case for authentic assessment*. Eric Digest, ERIC Clearinghouse on tests measurements and evaluation Washington DC., American Institutes for Research Washington DC.
- Halberstadt, J. B., & Niedenthal, P. M. (1997). Emotional state and the use of stimulus dimensions in judgment. *Journal of Personality and Social Psychology*, 72, 1017-1033.
- Halonen, J. S., Bosack, T., Clay, S., McCathy, M., Dunn, D. S., Hill, G. W., et al. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30, 196-208.
- Keilty, B., LaRocco, D. J., & Casell, F. (2009). Early interventionists' reports of authentic assessment methods through focus group research. *Topics in Early Childhood Special Education*, 28(4), 244-256.
- Langston, W. (2005). *Research Methods Laboratory Manual for Psychology* (2nd). Toronto, Canada: Thomson-Wadsworth Publishers.
- Meyers, N. M. & Nulty, D. D. (2009). How to use (five) curriculum design principles to align authentic learning, environment, assessment, and students approaches to thinking and learning outcomes. *Assessment and Evaluation in Higher Education*, 34, 565-577.
- Myers, A., & Hansen, C. H. (2005a). *Experimental psychology* (6th ed.). Belmont, CA: Wadsworth Publishers.
- Myers, A., & Hansen, C. H. (2005b). *Experimental psychology* (6th ed.). Exam View Test Bank. Belmont: CA, Wadsworth Publishers.
- Niedenthal, P. M., & Setterlund, M. B. (1994). Emotion congruence in perception. *Personality and Social Psychology Bulletin*, 20, 4001-411.
- Seligman, M. E. P., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology*, 74, 1-9.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Velten, E. C. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy*, 6, 473-482.

Appendix A

Article Review (Langston, 2005)

Author(s): _____

_____ Year of Publication: _____

Title of Article: _____

_____.

Journal Name: _____

Vol. # _____, Page #s. _____ - _____.

Topic and Purpose of the Study: _____

Hypothesis: : _____

Method: : _____

Results: : _____

Discussion: _____

Criticisms/Alternative Explanations/Future Research _____
