CHUN-TING CHEN and HSIAO-CHING SHE

THE EFFECTIVENESS OF SCIENTIFIC INQUIRY WITH/WITHOUT INTEGRATION OF SCIENTIFIC REASONING

Received: 11 February 2013; Accepted: 20 September 2013

ABSTRACT. This study examines the difference in effectiveness between two scientific inquiry programs—one with an emphasis on scientific reasoning and one without a scientific reasoning component—on students' scientific concepts, scientific concept-dependent reasoning, and scientific inquiry. A mixed-method approach was used in which 115 grade 5 students were administered the scientific concept test, scientific concept-dependent reasoning test, and scientific inquiry test before, 1 week after, and 8 weeks after instruction. In addition, students' scientific inquiry worksheets in the classroom were collected and evaluated. Results indicated that the experimental group outperformed the control group, regardless of scientific concept test, scientific concept-dependent reasoning test, and scientific inquiry test. Moreover, the classroom inquiry worksheets results demonstrated that the experimental group generated a significantly greater number of testable hypotheses, correct hypotheses, and correct evidence-based scientific explanations and a higher level of scientific reasoning than did the control group.

KEY WORDS: evidence-based scientific explanations, formulating hypotheses, integration of scientific reasoning, scientific inquiry

INTRODUCTION

Scientific inquiry has been considered one of the major goals in science education since the 1960s (National Research Council, 2001). The Program for International Student Assessment (PISA) 2006 defined scientific inquiry as the central process of science, which includes various components of scientific processes and scientific explanations (Organization for Economic Co-operation and Development, 2006). Inquiry has been considered a very important and efficient strategy for promoting students' science learning. In addition, many studies have viewed fostering students' scientific reasoning as one of the central goals of science education (American Association for the Advancement of Science, 1993; National Research Council, 1996). Despite considering scientific reasoning as critical for scientific inquiry (National Research Council, 1996; Lawson, 2003), none of these studies has demonstrated that scientific reasoning could actually facilitate students' competency in scientific inquiry. Therefore, this study specifically examined

International Journal of Science and Mathematics Education (2015) 20: 1–20

learning with and without an emphasis on scientific reasoning to explore the impact on students' learning of science and their competency of inquiry.

INQUIRY

The National Science Foundation (2000) defined inquiry as an approach to learning that involves a process of exploring the natural or material world and that leads to asking questions, making discoveries, and then rigorously testing those discoveries in the search for new understanding. These essential features of inquiry would help students to develop a clearer and deeper knowledge of science concepts and processes (National Research Council, 2000). Some studies have suggested that generating hypotheses, theories, or interpretations for the phenomena being investigated is critical for developing conceptual understanding of science (Muukkonen, Lakkala & Hakkarainen, 2005). Scientific inquiry teaching originated with efforts to engage students in the same thinking processes and activities used by practicing scientists (Rutherford, 1960). It did so by placing students in investigative settings in which they might develop scientific concepts and ideas from their own experience, that is, engaging students in a scientific version of empiricism (Edwards & Mercer, 1987). For scientists, important practices include being able to formulate researchable questions, design and conduct informative investigations, and formulate persuasive arguments.

A previous study reported that an inquiry approach facilitates students asking more and better questions than the traditional laboratory approach (Hofstein, Navon, Kipnis & Mamlok-Naaman, 2005). It has noted that students have substantial problems with the inquiry processes, such as difficulty in stating a testable hypotheses, choosing the right variables, specifying the relationship among these variables, designing conclusive experiments to test their hypotheses, drawing the correct conclusions from experiments, linking experimental data with hypotheses, and interpreting the results with appropriate theories (De Jong & Van Jollingen, 1998). It is important to provide students with coaching, situation-specific guidance, and expert participation during the inquiry process (Lakkala, Muukkonen & Hakkarainen, 2005) although the obstacles are obvious. Based on this previous research, the current study focuses on providing students with frameworks for their scientific inquiry processes.

Reasoning and Scientific Inquiry

Lawson (2003) suggested that scientific inquiry is followed by cycles of hypothetico-deductive reasoning. King, Whelan, Jones, Reiser, Bryant, Muggleton, Kell et al. (2004) noted: "A widely accepted view of science is that it follows a hypothetico-deductive process. Scientific expertise and imagination are first used to form possible hypotheses, and then the deductive consequences of these hypotheses are tested by experiment. We employ the logical inference mechanism of abduction to form new hypotheses, and that of deduction to test which hypotheses are consistent (p. 248)."

Lawson (2004) proposed that scientific inquiry seems to operate in a hypothetico-deductive way that involves the elements of making an initial puzzling observation; using abduction to spontaneously generate one or more hypotheses; assuming that the hypotheses under consideration is correct; conducting a test; comparing predicted and observed results; and repeating the procedure until a hypotheses is generated, tested, and supported on one or more occasions and its competing alternatives have been tested and rejected. However, Kuhn, Amsel & O'louohlin (1988) suggested a different point of view about reasoning: that inductive reasoning is clearly central to what scientists do, no matter whether or not deductive logic plays a role in scientific thinking. We agree with Kuhn et al.'s claims that both inductive and deductive reasoning are important for science inquiry.

The National Research Council (1999) specified science reasoning skills as the ability to define a scientific question, plan a way to answer the question, analyze data, and interpret results. Koslowski (1996) addressed that reasoning is central to scientific inquiry and should be understood as subordinate to the procedures of controlling variables. These studies point out that reasoning involves the whole process of inquiry, regardless of defining a scientific question, controlling variables, planning a way to answer a question, designing an experiment to test hypotheses, analyzing data, and interpreting results.

Russ, Scherr, Hammer & Mikeska (2008) identified that causal reasoning is involved in constructing explanations for scientific understanding. Other research presents evidence that knowledge of mechanism is, in fact, what helps us identify and understand the causal structure of the world (Ahn & Kalish, 2000). Russ et al. (2008) further suggested that a mechanism is important to scientific explanation and that reasoning about mechanisms is an important aspect of science inquiry. Koslowski (1996) suggested that scientific reasoning consists largely of giving a mechanism that "explains the process by which a cause brings about an effect" (p. 13). Abrams, Southerland & Cummins (2001) claimed that students should give mechanistic explanations that identify physical causes and the process of a phenomenon. Russ et al. (2008) advocated the use of mechanism to explain phenomena generally and described them as identifying the process between causes and effects. We agree with these claims that scientific reasoning is pivotal for constructing evidence-based scientific explanations and further generate a mechanism to explain its causality for these inquiry results and phenomena. These studies serve as our basis to design our inquiry learning with the emphasis of scientific reasoning.

Although promoting scientific reasoning is important for science learning, many scientific inquiry tasks given to students do not reflect the core attributes of authentic scientific reasoning (Chinn & Malhotra, 2002). Scientific reasoning skills can be developed through training and can be transferred (Adey & Shayer, 1994; Chen & Klahr, 1999). Training in scientific reasoning may also have a long-term impact on students' academic achievement (Adey & Shayer, 1994). Therefore, we consider that it might be meaningful to engage students in authentic scientific inquiry with emphasis on engaging students in the use of scientific reasoning practice.

PURPOSE

This study presents a conceptual framework to guide the design of a framework for inquiry. The design includes two scientific inquiry programs-one with an emphasis on scientific reasoning and one without a scientific reasoning component. This study had four specific aims. First, to compare the difference in effectiveness between two scientific inquiry programs-one with an emphasis on scientific reasoning and one without a scientific reasoning component- on students' conceptual construction, scientific concept-dependent reasoning, and scientific inquiry ability. Second, how inquiry learning with/without the integration of scientific reasoning affects the quality of hypotheses formulation and evidence-based scientific explanations generation by students. Third, to examine whether experimental group's students' scientific reasoning ability made progress across time. Finally, to explore what are the relationships among students' scientific conceptions, scientific concept-dependent reasoning, and scientific concept-dependent inquiry.

Subjects and Procedures

A total of 115 grade 5 students were recruited from four average-achievement classes of an elementary school located in Northern Taiwan to participate in this study. Students ranged between 10 and 11 years old. Four classes of students were randomly assigned to experimental and control groups. Two classes (*n*=64; 34 males and 30 females) were presented with a framework for inquiry with scientific reasoning (experimental group), and the other two classes (*n*=57; 30 males and 27 females) received a structure for inquiry without scientific reasoning (control group) for three units of science across 3 months. Pretests, posttests, and retention tests, including the scientific concepts test (SCT), scientific concept-dependent reasoning test (SCDRT), and scientific concept-dependent inquiry test (SCDIT) were administered to all students before instruction, at 1 week, and at 2 months after the instruction. In addition, both groups' inquiry learning worksheets were further analyzed and yielded information on formulating hypotheses, generating evidence-based explanations, and scientific reasoning. All the data collected were entered into SPSS 13.0 for further analysis.

Design of Framework for Inquiry Content: Units on Sound, Heat Transfer, and Stars

A five-person panel composed of three science teachers and two science educators was involved in the development of the units on sound, heat transfer, and stars. These three units were covered in developing a series of frameworks for inquiry learning with/without scientific reasoning events. Each unit covered its major concepts: sound, including frequency and amplitude of sound, how to make sounds of different frequencies, and how sound propagates in different media such as air or water; heat transfer, including heat conduction of metal, heat convection of water, heat radiation across substances of different color; and stars, including constellations, constellation pictures, constellation disks, brightness of stars, color and temperature of stars, and rotation of stars.

The design of each inquiry unit was intended to provide specific frameworks for students to formulate hypotheses, identify variables, and generate evidence-based scientific explanations. The difference between the experimental group and control group was to integrate inquiry learning with or without scientific reasoning, respectively. First, we provided an inquiry-learning framework for asking students to formulate hypotheses, identify variables, make conclusions based upon experiment data and results, and generate evidence-based scientific explanations. Second, a set of driving questions were provided for students to make predictions and generate their arguments before formulating a hypothesis that would employ both inductive and deductive reasoning abilities to provide a possible causal mechanism. This would activate and facilitate students to formulate a workable hypothesis. Students were then required to identify the controlling variables after formulating hypothesis, which they will need to employ their deductive reasoning. It also provides an opportunity for students to exchange their individual perspectives while they are designing the experiment and performing it. Third, to provide evidence-based scientific explanations for a series of driving question after collecting data and making conclusions, which requires students to employ causal reasoning to generate plausible evidence-based scientific explanations. Both groups received the same frameworks for formulating hypotheses and identifying variables. The whole process allows students to examine what they already know, activate their prior knowledge, and reason what the possible answers are. Students designed the experiment in the laboratory and performed it according to their designs developed in groups. For the control group, the same driving questions were presented to them as part of the instruction, but they were not required to answer the question or to provide their justification.

Instruments

Scientific Concepts Test

The SCT is a two-tier, multiple-choice, diagnostic instrument that was developed to measure the students' level of scientific concept understanding related to stars, heat transfer, and sound before, directly after, and nine class periods after receiving the units for the inquiry learning program. Each unit covers its major concepts (specifics provided earlier). Content validity was established by the previously mentioned five-person panel, who ensured that the items were properly constructed and relevant to the five units on learning inquiry materials we developed. Each unit consisted of ten items, resulting in a total of 30 items; the highest possible score is 30. The internal consistency, as measured by Cronbach's α, of the SCT was 0.60 for the pretest, 0.74 for the posttest, and 0.81 for the retention test.

Scientific Concept-Dependent Reasoning Test

The SCDRT is a two-tier, multiple-choice, diagnostic instrument developed for this study to measure students' scientific content-dependent reasoning before, directly after, and 10 weeks after instruction. It requires students to employ both inductive and deductive reasoning in learning the related scientific content in this inquiry program involving stars, heat transfer, and sound units. Content validity was established by the previously mentioned five-person panel who ensured that the items were properly constructed and relevant to the three units on learning inquiry materials we developed. Each unit consisted of eight items, resulting in a total of 24 items. Each item contained two tiers; the first tier is to choose the answer and the second tier is to use the thinking ability mentioned above. Students needed to answer both tiers correctly to receive 1 point; the highest possible score is 24. The Cronbach's α of the SCDRT was s 0.74 for the pretest, 0.83 for the posttest, and 0.96 for the retention test.

Scientific Concept-Dependent Inquiry Test

The SCDIT is a diagnostic instrument developed for this study to measure the students' inquiry competency-dependent on scientific concepts. It requires students to employ their scientific concepts and reasoning to formulate hypotheses, identify independent variables, identify dependent variables, make conclusions, and provide evidence-based explanations. These were evaluated before, directly after, and 2 months after instruction. Content validity was established by the previously mentioned five-person panel who ensured that the items were properly constructed and relevant to the six units on learning inquiry materials we developed. There were six scenarios, covering the three stars, heat transfer, and sound units, with a total of 32 items. The highest possible score was 32. Cronbach's α of the SCDIT was 0.92 for the pretest, 0.90 for the posttest, and 0.96 for the retention test.

Quality of Hypotheses Formulation, Evidence-Based Scientific Explanation, and Scientific Reasoning

The quality of hypotheses, evidence-based scientific explanation, and scientific reasoning were measured based upon their inquiry activity at the classroom across the three units.

Formulate Hypotheses

Two aspects were used to measure the nature and extent of formulating hypotheses: correctness of the hypotheses and the testability of the hypotheses. Correctness of hypotheses was defined as the hypotheses students generated having correct relationships between dependent and independent variables. Three categories were used to determine the correctness of hypotheses: correct, partially correct, and incorrect. For instance, “the longer the metal string, the deeper the sound produced” was coded as correct; “when the length of the metal strings differ, the sounds of the instrument become deeper or sharper” was coded as partially correct; “the shorter the metal string, the deeper the sound produced” was coded as incorrect. The definition for testable hypotheses is that the hypotheses students generated can be tested scientifically. The testability of hypotheses was coded as 2, 1, and 0 according to their testability. For instance, “the more water in the test tube, the lower the pitch of sound made by the tube” was coded as 2 points because the hypothesis clearly described relationships between dependent and independent variables and can be tested scientifically; “the amount of water in the tube influences the pitch of the tube” was coded as 1 point for partially describing relationships between dependent and independent variables and the testability being less clear; and “would water influence the pitch of the tube?” was coded as 0 point as it both lacked dependent/independent variables and cannot be tested scientifically.

Evidence-Based Scientific Explanations

Three categories were used to determine the correctness of evidence-based scientific explanations that students generated for a series of driving questions across three units. They were given after students’ made their conclusions according to their experiment data and results. The categories were: correct, partially correct, and incorrect in terms of evidence-based scientific explanations. A correct evidence-based scientific explanation is using correct scientific principles/theories to develop a scientific explanation regarding evidence the students found from their experimental results; for instance, the pitch of the instrument was determined by the length of the metal string, so the longer the string, the deeper the sound. A partially correct evidence-based scientific explanation is making a partially correct scientific explanation regarding evidence the students found from their experimental results; for instance, the length of the string would influence the pitch of the sound. An incorrect evidence-based scientific explanation is either making an incorrect scientific explanation regarding evidence the students found from their experimental results or explaining the results incorrectly; for instance, the metal string would make a sharper sound and a plastic string would make a louder sound.

Scientific Reasoning

The nature and extent of the scientific reasoning students used was measured and compared with their reasoning ability for a similar set of driving questions raised before formulating hypothesis and after making conclusions across the three units. Three categories of scientific reasoning, modified from Hogan, Nastasi & Pressley's study (2000), were used in this study: generativity, elaboration, and justification to measure the nature of the scientific reasoning. Generativity (G) is the number of a student's observations, ideas, or conjectures to explain an observation. Elaboration (E) is the number of scientific concepts used to describe and explain the concepts. Justification (J) is the number of aspects based on evidence, inference, or experiments used to support the ideas or assertions.

RESULTS

Scientific Concepts Test

The experimental group students made significant progress from pre- to post-SCT ($T_{(\text{post–pre})}=7.98$, $p<0.001$) and retention-SCT ($T_{(\text{post–pre})}=6.52$, $p<0.001$). The control group made significant progress only from pre- to post-SCT ($T_{(\text{post–pre})}=5.38$, $p<0.001$). One-factor MANCOVA was conducted to examine the effects of instructional approaches and of using post- and retention SCT scores as the dependent measures, with pre-SCT scores as the covariate. Table 1 summarizes the results of the one-factor MANCOVA, indicating that instructional approaches (Wilk's Λ=87, $p<0.001$) had a statistically significant effect on the performance of post- and retention-SCT.

The univariate F (one-factor ANCOVA) was performed to examine the effect of the instructional approaches on post- and retention SCT indepen-

TABLE 1

One-factor MANCOA for scientific concept, scientific concept-dependent reasoning, and scientific concept-dependent inquiry tests

	Wilk's Λ	*df*	*F*	*Effect size (η^2)*
Scientific concept test				
Pretest	0.60	2	37.80*** (0.000)	0.403 (L)
Instructional approach	0.87	2	8.40*** (0.000)	0.130 (M)
Scientific concept-dependent reasoning				
Pretest	0.62	2	33.95*** (0.000)	0.377 (L)
Instructional approach	0.83	2	11.30*** (0.000)	0.168 (L)
Scientific concept-dependent inquiry test				
Pretest	0.42	2	78.05*** (0.000)	0.582 (L)
Instructional approach	0.76	2	17.32*** (0.000)	0.236 (L)

*$p<0.1$; **$p<0.01$; ***$p<0.001$

dently. It indicated that the effects for instructional approaches on both post-SCT scores ($F = 4.56$, $p = 0.035$) and retention-SCDIT scores ($F = 16.21$, $p < 0.001$) were significant. Thus, the students' post- and retention SCDIT were significantly affected by the instructional approach. Post-hoc analysis for the main effect suggests that the experimental group performed significantly better than the control group on the posttest and retention test (experimental group > control group, $p_{(\text{post})} = 0.035$, and $p_{(\text{retention})} = 0$) (Table 2).

Scientific Concept-Dependent Reasoning Test

The experimental group students made significant progress from pre- to post-SCDRT ($T_{(\text{post–pre})} = 8.94$, $p < 0.001$) and retention SCDRT ($T_{(\text{post–pre})} = 7.93$, $p < 0.001$). The control group made significant progress only from pre- to post-SCDRT ($T_{(\text{post-pre})} = 8.87$, $p < 0.001$). One-factor MANCOVA was conducted to examine the effects of instructional approaches and of using post- and retention-SCDRT scores as the dependent measures, with pre-SCDRT scores as the covariate. Table 1

TABLE 2

One-factor ANOVA for scientific concept, scientific concept-dependent reasoning, and scientific concept-dependent inquiry tests

	Number	*Mean*	*SD*	*F*	*Effect size (η^2)*	*Post hoc*
Scientific concept test						
Posttest						
Control group	54	15.57	5.16	4.56* (0.035)	0.039 (S)	*E* > *C* (0.035)
Experimental group	62	17.65	4.56			
Retention test						
Control group	54	13.15	6.09	16.21*** (0.000)	0.125 (M)	*E* > *C* (0.000)
Experimental group	62	17.37	5.45			
Scientific concept-dependent reasoning						
Posttest						
Control group	54	25.91	8.67	1.59 (0.210)	0.014 (S)	*E* > *C* (0.210)
Experimental group	62	28.39	8.32			
Retention test						
Control group	54	20.59	9.34	19.09*** (0.000)	0.145 (L)	*E* > *C* (0.000)
Experiment group	62	27.87	8.89			
Scientific concept-dependent inquiry test						
Posttest						
Control group	54	21.61	11.80	6.15* (0.015)	0.052 (S)	*E* > *C* (0.015)
Experimental group	62	25.26	10.38			
Retention test						
Control group	54	20.94	10.03	23.26*** (0.000)	0.171 (L)	*E* > *C* (0.000)
Experimental group	62	26.26	9.80			

E experimental group, *C* control group
*$p < 0.1$; **$p < 0.01$; ***$p < 0.001$

summarizes the results of the one-factor MANCOVA, indicating that instructional approaches (Wilk's $\Lambda = 0.83$, $p < 0.001$) reached a statistically significant effect on the performance of post- and retention SCT.

The univariate F (one-factor ANCOVA) was performed to examine the effect of the instructional approaches on post- and retention SCT independently. It indicated that the effects for instructional approaches retention-SCDIT scores ($F = 19.09$, $p < 0.001$) were significant. Thus, the students' retention SCDIT was significantly affected by the instructional approach. Post-hoc analysis for main effect suggests that the experimental group performed significantly better than the control group on the retention test (experimental group > control group and $p_{(\mathrm{retention})} = 0$) (Table 2).

Scientific Concept-Dependent Inquiry Test

The experimental group students made significant progress from pre- to post-SCDIT ($T_{(\mathrm{post-pre})} = 6.61$, $p < 0.001$) and retention-SCDIT ($T_{(\mathrm{post-pre})} = 9.20$, $p < 0.001$). A similar result was found for the control group who made significant progress from pre- to post-SCDIT ($T_{(\mathrm{post-pre})} = 3.53$, $p = 0.001$) and retention SCDIT ($T_{(\mathrm{post-pre})} = 4.95$, $p < 0.001$). One-factor MANCOVA was conducted to examine the effects of instructional approaches and of using post- and retention-SCDIT scores as the dependent measures, with pre-SCDIT scores as the covariate. Table 1 summarizes the results of the one-factor MANCOVA, indicating that instructional approaches (Wilk's $\Lambda = 0.76$, $p < 0.001$) have a statistically significant effect on the performance of post- and retention-SCDIT. Therefore, the following main effect for the instructional approach was performed.

The univariate F (one-factor ANCOVA) was performed to independently examine the effect of instructional approaches on post- and retention-SCDIT. It indicated that the effects for instructional approaches on both post-SCDIT scores ($F = 6.15$, $p = 0.015$) and retention-SCDIT scores ($F = 23.26$, $p < 0.001$) were significant. Thus, the students' post- and retention-SCDIT were significantly affected by the instructional approach. The post-hoc analysis for main effect suggests that the experimental group performed significantly better than the control group on both the posttest and retention test (experimental group > control group, $p_{(\mathrm{post})} = 0.015$, and $p_{(\mathrm{retention})} < 0.001$) (Table 2).

Stepwise Regression Analysis

The stepwise regression method was used to explore whether the post-SCT, SCDRT would be most important factor for predicting the post-SCDIT scores and whether retention SCT, SCDRT would be the most important factor for predicting the retention-SCDIT scores. Results indicated that the best predictor for post-SCDIT scores was the post-SCDRT scores. The standardized regression coefficient for post-SCDRT was 0.62, which accounts for 38 % of the variance in post-SCDIT scores. Results also showed that the best predictors for retention-SCDIT scores were retention SCT and retention SCDRT. The standardized regression coefficient for retention-SCT and SCDRT were 0.36 and 0.36 and accounted for 87.8 % of the variance in retention-SCDIT scores (Table 3).

Quality of Classroom Inquiry Learning

The nature and extent of students' classroom inquiry learning were analyzed according to the three aspects: hypotheses formulation, evidence-based scientific explanations, and level of scientific reasoning.

Formulate Hypotheses

Table 4 and Fig. 1 summarizes the one-factor repeated measures of ANOVA that were conducted to examine the effects of instructional approaches and any increase on their correctness of hypotheses generated from unit 1 to unit 3. The significant effect on the correctness of hypotheses was found for units ($F=174.85$, $p<0.001$) and interaction between instructional approach and units ($F=7.63$, $p=0.001$). The mean scores of the correct hypotheses in unit 3 was significantly higher than unit 2 ($p<0.001$) and unit 2 was greater than unit 1 ($p<0.001$). For the partially correct hypotheses, the results indicate only which units reached a significantly different level ($F=90.90$, $p<0.001$). It also shows that the mean score of partially correct hypotheses in unit 3 was significantly higher than in unit 2 and unit 1 ($p<0.001$, $p<0.001$). For the incorrect hypotheses, the results show that instructional approach ($F=8.51$, $p=0.004$), units ($F=47.11$, $p<0.001$), and interaction between instructional approach and units ($F=5.19$, $p=0.008$) all reached a statistically significant difference level. The control group outperformed the experimental group in the mean score of incorrect hypotheses at units 1 and 3.

The testability of hypotheses was analyzed according to their testable properties (Table 5). The one-factor repeated measures of ANOVA was conducted to examine the effects of instructional approaches and any increase from unit 1 to unit 3 for formulating testable hypotheses. The results indicated that both instructional approaches ($F=9.88$, $p=0.002$) and units ($F=139.49$, $p<0.001$) had a statistically significant effect on their performance of formulating testable hypotheses. No interaction was observed for units and instructional approach. The post hoc test suggests

TABLE 3

Stepwise regression for relationships among scientific concept, scientific reasoning, scientific concept-dependent reasoning, and scientific concept-dependent inquiry tests

	Standardized regression coefficients	*Multiple R*	R^2	*Cumulative % of variance explained*	*T*	*Confidence intervals*	
						Lower bound	*Upper bound*
Post-scientific concept-dependent inquiry test							
Post-scientific concept-dependent reasoning	0.62	0.62	0.39	38.0 %	8.46*** (0.000)	0.622	1.002
Retention-scientific concept-dependent inquiry test							
Retention-scientific concept	0.36	0.65	0.42	41.8 %	3.14* (0.002)	0.223	0.985
Retention-scientific concept-dependent reasoning	0.36	0.69	0.47	87.8 %	3.14* (0.002)	0.139	0.616

$^*p<0.1$; $^{**}p<0.01$; $^{***}p<0.001$

TABLE 4

One-factor repeated measure ANOVA of the correctness of formulating hypothesis and providing evidence-based scientific explanations

			Post hoc	
	F	*Effect size* (η^2)	*Unit*	*Instructional approach*
Formulate hypothesis				
Correct				
Unit	174.85*** (0.000)	0.605 (L)	2 > 1*** (0.000)	
Instructional approach	1.39 (0.240)	0.012 (S)	3 > 1*** (0.000)	
Unit × instructional approach	7.63*** (0.001)	0.063 (M)	3 > 2*** (0.000)	
Partial correct				
Unit	90.90*** (0.000)	0.439 (L)	3 > 1*** (0.000)	
Instructional approach	0.69 (0.407)	0.153 (L)	3 > 2*** (0.000)	
Unit × instructional approach	1.90 (0.153)	0.016 (S)		
Incorrect				
Unit	47.11*** (0.000)	0.290 (L)	1 > 2*** (0.000)	*C* > *E* (0.004)
Instructional approach	8.51** (0.004)	0.068 (M)	1 > 3*** (0.000)	
Unit × instructional approach	5.19** (0.008)	0.043 (S)		
Evidence-based scientific explanations				
Correct				
Unit	179.43*** (0.000)	0.616 (L)	2 > 1*** (0.000)	*E* > *C* (0.000)
Instructional approach	15.95*** (0.000)	0.139 (L)	3 > 1*** (0.000)	
Unit × instructional approach	2.00 (0.138)	0.018 (S)	3 > 2*** (0.000)	
Partial correct				
Unit	42.99*** (0.000)	0.272 (L)	2 > 1*** (0.000)	
Instructional approach	5.32* (0.023)	0.044 (S)	3>1*** (0.000)	
Unit × instructional approach	2.33 (0.100)	0.020 (S)		
Incorrect				
Unit	27.93*** (0.000)	0.194 (L)	1>2*** (0.001)	
Instructional approach	1.87 (0.174)	0.016 (S)	1>3*** (0.000)	
Unit × instructional approach	4.53* (0.015)	0.038 (S)	2>3*** (0.000)	

E experimental group, *C* control group, η^2 *effect size*
*$p < 0.1$; **$p < 0.01$; ***$p < 0.001$

that the experimental group outperformed the control group ($p = 0.002$), and units 3 and 2 were significantly higher than unit 1 ($p < 0.001$, $p < 0.001$).

Evidence-Based Scientific Explanations

Table 4 summarizes one-factor repeated measures of ANOVA results for evidence-based scientific explanations. It indicated that instructional approaches ($F = 15.95$, $p < 0.001$; $F = 5.32$, $p < 0.05$) and units ($F = 179.43$, $p < 0.001$; $F = 42.99$, $p < 0.001$) all reached statistically significant effects on the performance of generating correct and partial correct evidence-based

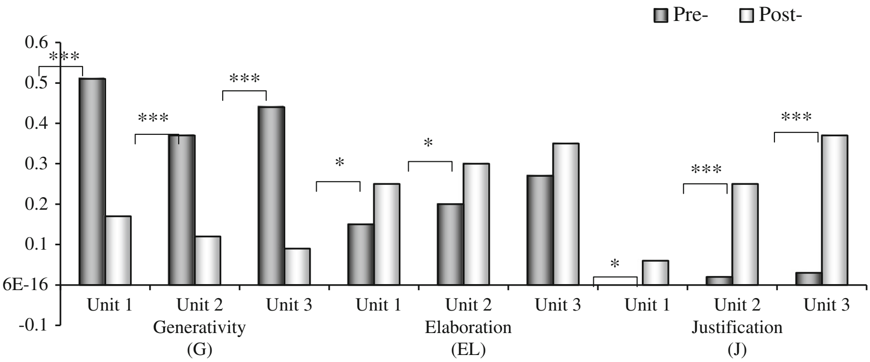

Figure. 1. The descriptive statistics and *t* test of level of scientific reasoning in pre- and postdriving questions before and after learning across three units

scientific explanations. The post hoc test suggested that the experimental group outperformed the control group on their performance of correct evidence-based scientific explanation mean scores ($p < 0.001$). Both correct and partially correct results also revealed that unit 3 was significantly higher than unit 1 ($p < 0.001$), and unit 2 was greater than unit 1 ($p < 0.001$). For the pattern of incorrect evidence-based scientific explanation, only the units reached a statistically significant difference level ($F = 42.99$, $p < 0.001$; $F = 27.93$, $p < 0.001$). The post hoc indicated that the scores of unit 1 are greater than unit 2 and 3 ($p < 0.001$), and the unit 2 is greater than unit 3 ($p < 0.001$) for the incorrect evidence-based scientific reasoning.

Level of Scientific Reasoning

Figure 1 indicated that the experimental group students used more G level (lowest) of scientific reasoning before the inquiry than after it and they all

TABLE 5

One-factor repeated measure of ANOVA for formulating testable hypothesis

			Post hoc	
	F	*Effect size* (η^2)	*Units*	*Instructional approach*
Formulating testable hypothesis				
Units	139.49*** (0.000)	0.548 (L)	2 > 1*** (0.000)	*E* > *C* (0.002)
Instructional approach	9.88** (0.002)	0.079 (M)	3 > 1*** (0.000)	
Units × instructional approach	0.49 (0.589)	0.004		

*$p < 0.1$; **$p < 0.01$; ***$p < 0.001$

reached a statistically significant difference level across the three units ($t_{(\text{unit }1)} = -7.52$, $p < 0.001$; $t_{(\text{unit }2)} = -6.81$, $p < 0.001$; $t_{(\text{unit }3)} = -9.95$, $p < 0.001$). Interestingly, students used more E level (middle) of scientific reasoning after the inquiry than before it across units, and they also reached a statistically significant difference level at units 1 and 2 ($t_{(\text{unit }1)} = 2.35$, $p = 0.022$; $t_{(\text{unit }2)} = 2.93$, $p = 0.005$; $t_{(\text{unit }3)} = 1.82$, $p = 0.073$). Moreover, students used more J level (highest) of scientific reasoning after the inquiry than before it, and they also reached a statistically significant difference level across all units ($t_{(\text{unit }1)} = 2.79$, $p = 0.007$; $t_{(\text{unit }2)} = 7.64$, $p < 0.001$; $t_{(\text{unit }3)} = 10.82$, $p < 0.001$).

Discussion and Conclusions

Our quantitative data demonstrated that both control and experimental groups made significant progress from pre- to posttest for scientific concepts, which supports Muukkonen et al.'s (2005) study that inquiry is critical for developing conceptual understanding of science. In addition, the experimental group significantly outperformed the control group in their posttest and retention test of scientific concept construction. The possible explanation is to foster students' scientific reasoning indeed reduce the obstacles raised by previous studies, such as students' scientific reasoning ability is still underdeveloped and limits their ability to efficiently work on systemic inquiry and construct more sophisticated scientific knowledge (Hanauer, Jacobs-Sera, Pedulla, Cresawn, Hendrix & Hatfull, 2006).

On the other hand, the quantitative results of scientific concept-dependent reasoning indicated that both the control and experimental groups made significant progress from pre- to posttest, which somehow supports Koslowski (1996) that reasoning is central to scientific inquiry and Lawson's (2003) suggestion that scientific inquiry operates as hypothetic-deductive reasoning. Moreover, the experimental group significantly outperformed the control group in their posttest and retention test of scientific concept-dependent reasoning. This indicates that integrating scientific reasoning with inquiry does improve students' performance in scientific concept-dependent reasoning, which supports that scientific reasoning skills can be developed through training and can be transferred (Adey & Shayer, 1994; Chen & Klahr, 1999).

The quantitative results of scientific concept-dependent inquiry indicated that both the experimental and control groups made significant improvement from pre- to posttest and from posttest to retention test, which supports that inquiry as an approach to learning that leads to asking questions, making discoveries, and then rigorously testing those discoveries in the search for new understanding (National Science Foundation, 2000). Moreover, the experimental group performed significantly better than the control group on the posttest and retention test of scientific concept-dependent inquiry. This clearly demonstrates that both groups received a framework for identifying testable scientific hypotheses, identifying variables, and generating evidence-based scientific explanations, which leads to the improvement of their inquiry ability. This result adds evidence to the suggestions from previous studies that scientific reasoning is critical for promoting scientific inquiry (Chinn & Malhortra, 2002; Lawson, 2003).

The regression results indicated that the best predictor for post-scientific concept-dependent inquiry test was the post-scientific concept-dependent reasoning test only. This indicated that scientific concept-dependent reasoning is the only critical factor to predict students' scientific concept-dependent inquiry ability. This supports the idea that underdevelopment of students' scientific reasoning ability may limit their ability to work efficiently on systemic inquiry (Hanauer et al., 2006; Schauble, Klopfer, & Raghavan, 1991). The best predictor for retention scientific concept-dependent inquiry test was the retention scientific concept and followed by retention scientific-dependent reasoning. Our study tends to support the idea that students' scientific concept-dependent reasoning is more critical for their inquiry ability than scientific domain knowledge immediately after instruction.

The results also demonstrated that students who received inquiry instruction that emphasized scientific reasoning indeed made greater progression for the correctness and testability of hypotheses and evidence-based scientific explanations. A possible interpretation for the correctness and testability of hypothesis is that a set of driving questions was provided for students to make predictions and generate their arguments before formulating hypothesis in which they would employ both inductive and deductive reasoning ability to connect their prior knowledge and provide a possible causal mechanism and, therefore, activate and facilitate them to formulate a more correct and workable hypothesis. Similar interpretation can be applied for evidence-based scientific explanations is that students were required to employ causal reasoning to generate plausible evidence-based scientific explanations for a series of driving questions after collecting data and making conclusions. The whole process allows students to examine what they already know, activate their prior knowledge, and reason what are the possible answers.

Our qualitative and quantitative results consistently reveal that integrating scientific reasoning with inquiry is critical improve students' scientific inquiry, science concepts understanding, and scientific reasoning. Several implications was provided in the following. First, it is important to note that a well-theoretically designed study can improve students' ability within a period of time, regardless of their gender and level of academic achievement. These notable results open a new avenue for science educators to foster students' inquiry ability by providing the framework for emphasizing scientific reasoning with inquiry. Second, it would be efficient to provide students with a clear scaffolding framework and step-by-step inquiry process for students to engage in the inquiry activity, which would reduce all the possible obstacles. Third, the results shed light on combining inquiry with an emphasis on reasoning would not just enhance students' ability to formulate hypotheses but also facilitate their evidence-based scientific explanations ability, which is the core of the PISA test. Finally, the limitations of this study is only examining its effect on elementary students' science learning, which needs further study to confirm whether it also can apply for middle school or high school science learning.

ACKNOWLEDGMENTS

This research is based in work supported and funded by National Science Council (NSC 98-2511-S-009-006-MY3).

REFERENCES

Abrams, E., Southerland, S. & Cummins, C. (2001). The how's and why's of biological change: How learners neglect physical mechanisms in their search for meaning. *International Journal of Science Education, 23*(12), 1271–1281.

Adey, P. & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.

Ahn, W. K. & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge, MA: MIT Press.

American Association for the Advancement of Science (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Chen, Z. & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Chinn, C. & Malhotra, B. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175–218.

De Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179–201.

Edwards, D. & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Methuen.

Hanauer, D., Jacobs-Sera, D., Pedulla, M., Cresawn, S., Hendrix, R. & Hatfull, G. (2006). Teaching scientific inquiry. *Science, 314*(5807), 1880–1881.

Hofstein, A., Navon, O., Kipnis, M. & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching, 42*(7), 791–806.

Hogan, K., Nastasi, B. & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*(4), 379–432.

King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G. K., Bryant, C. H., Muggleton, S. H., Kell, D. B. & Oliver, S. G. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature, 427*, 247–252.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

Kuhn, D., Amsel, E. & O'louohlin, M. (1988). *The development of scientific thinking skills*. New York, NY: Academic.

Lakkala, M., Muukkonen, H. & Hakkarainen, K. (2005). Patterns of scaffolding in computer-mediated collaborative inquiry. *Journal of Mentoring and Tutoring, 13*(2), 281–300.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25*(11), 1387–1408.

Lawson, A. (2004). A reply to Allchin's Pseudo history and Pseu doscience. *Science & Education, 13*(6), 599–605.

Muukkonen, H., Lakkala, M. & Hakkarainen, K. (2005). Technology-mediation and tutoring: How do they shape progressive inquiry discourse? *Journal of the Learning Sciences, 14*(4), 527–565.

National Research Council (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council (1999). *New strategies for America's watersheds*. Washington, DC: National Academy Press.

National Research Council (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academies Press.

National Research Council (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.

National Science Foundation (2000). *An introduction to inquiry*. Washington, DC: Author.

Organization for Economic Co-operation and Development (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA*. Paris: Author.

Russ, R. S., Scherr, R. E., Hammer, D. & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499–525.

Rutherford, F. J. (1960). The role of inquiry in science teaching. *Journal of Research in Science Teaching, 2*(2), 80–84.

Schauble, L., Klopfer, L. E. & Raghavan, K. (1991). Students’ transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28*(9), 859–882.

Institute of Education
National Chiao Tung University
1001 Ta-Hsueh Rd, Hsinchu, 300 Taiwan, Republic of China
E-mail: hcshe@mail.nctu.edu.tw