Supplemental Material

CBE—Life Sciences Education

Stevens e*t al.*

SD 1

(Hurley's adaptation (2008) of Weiss Observation Protocol)

OBSERVATION PROTOCOL FOR THE CREATE MODEL[1]

Purpose: To gather accurate information about how a program actually operates, particularly about its processes.

Background Information

Observer _____ Date of Observation _____

Duration of Observation:

___ 1 hour ___ 2 hours ___ 3 hours ___ half day ___ whole day

Location: _____ Travel time: _____

Title of Course: _____

Name of Implementer: _____

Total Number of Attendees: _____ Female Attendees: _____ Male Attendees _____

Part I: Context Background & Activities

This section provides a brief overview of the session being observed.

[1] Protocol is adapted from the Local Systemic Change Observation Protocol developed by Iris Weiss (1997) for the National Science Foundation.

I. Session Context

In a few sentences, describe the session you observed. Include whether or not the observation covered a partial or complete session.

II. Session Focus

Indicate the major intended purpose(s) of this session based on the information provided by the Implementer.

III. Classroom Environment

Make a drawing of the classroom setup, indication major furniture, equipment, doorways, and locations of both students and implementer.

IV. Instructional Activities ***(Check all activities observed and describe, as relevant)***

A. Indicate the major instructional resource(s) used in this session.

____ print materials

____ hands-on materials

____ technology/audio-visual resources

____ other instructional resources (please specify)

B. Indicate the major way(s) in which participant activities were structured.

____ as a whole group

____ as small groups

____ as pairs

____ as individuals

C. Indicate the major activities of implementer in this session.

____ formal presentations by implementer *(describe focus):*

____ formal presentations by participants *(describe focus):*

____ hands-on/investigative/research activities *(describe):*

____ problem-solving activities *(describe):*

____ proof & evidence *(describe):*

____ reading/reflection/written communication *(describe):*

____ technology-based activities *(describe):*

____ assessment activities *(describe):*

____ other activities *(please specify):*

D. Comments

Please provide additional information you consider necessary to capture the activities or context of this session. Include comments on any feature of the session that is so salient that you need to get it "on the table" right away to help explain your ratings.

Part II: Ratings

Using the information collected and observed in Part I, rate each of the key indicators from 1 (not at all) to 5 (to a great extent) by circling the correct response. Use 6 (don't know) when not enough evidence exists to make a judgment and 7 (N/A) when you consider the indicator inappropriate for the purpose and context of the session. Similarly, there may be entire rating categories not applicable to a particular session. You may list additional indicators that you consider appropriate for the session and rate them.

Use your "Ratings of Key Indicators" from A to inform your "Synthesis Ratings" from B and indicate in "Supporting Evidence for Synthesis Ratings" from C what factors were most influential in determining your synthesis ratings.

I. Design of Instruction

A. Ratings of Key Indicators

	Not at all				To a great extent	Don't know	N/A
1. The strategies in this session were appropriate for accomplishing the implementer's purpose	1	2	3	4	5	6	7
2. The session effectively built on students understanding of the CREATE process	1	2	3	4	5	6	7
3. The instructional strategies & activities used in this session reflected attention to participants experience, preparedness, & learning styles	1	2	3	4	5	6	7
4. The design of the session reflected careful planning & organization	1	2	3	4	5	6	7
5. The design of the session encouraged a collaborative approach to learning	1	2	3	4	5	6	7
6. The design of the session provided opportunities for students to consider applications of learning to science content	1	2	3	4	5	6	7
7. The design of the session incorporated tasks, roles, & interactions consistent with a spirit of investigation	1	2	3	4	5	6	7
8. The design of the session effectively built on student understandings of the nature of	1	2	3	4	5	6	7

science

9. The design of the session effectively built on student understandings of the science content	1	2	3	4	5	6	7
10. The design of the session appropriately balanced attention to multiple goals within the CREATE structure	1	2	3	4	5	6	7
11. Adequate time & structure were provided for reflection	1	2	3	4	5	6	7
12. Adequate time & structure were provided for participants to share experiences & insights	1	2	3	4	5	6	7
13. _____	1	2	3	4	5		

B. Synthesis Rating

1	2	3	4	5
Design of the session was not at all reflective of best practices for CREATE				Design of the session extremely reflective of best practices for CREATE

C. Supporting Evidence for Synthesis Rating

II. Instruction/Implementation

A. Ratings of Key Indicators

	Not at all				To a great extent	Don't know	N/A
1. Initiation of session effectively referred back to key concepts from prior session & tied them to current session	1	2	3	4	5	6	7
2. The session effectively incorporated instructional strategies appropriate for the purposes of the CREATE session & the needs of the learners	1	2	3	4	5	6	7
3. The implementer effectively used questioning strategies that are likely to enhance the development of conceptual understanding (e.g., emphasis on higher-order questions, appropriate use of "wait-time," identifying perceptions & misconceptions)	1	2	3	4	5	6	7
4. The pace of the session was appropriate for the purposes of CREATE and the needs of the learners	1	2	3	4	5	6	7
5. Routines and transitions were orderly and efficient and resulted in minimal time off task	1	2	3	4	5	6	7
6. Science concepts were explicitly mentioned in the lesson to promote a deeper understanding of content	1	2	3	4	5	6	7
7. NOS skills were explicitly mentioned in the lesson to promote a deeper understanding of	1	2	3	4	5	6	7

the process							
8. Opportunities for critical thinking were ample and provided time to reflect on newly integrated knowledge for application to a novel situation (moving higher on Bloom's Taxonomy)	1	2	3	4	5	6	7
9. The implementer's background, experience, and/or expertise enhanced the quality of the session	1	2	3	4	5	6	7
10. The implementer's management style/ strategies enhanced the quality of the session	1	2	3	4	5	6	7
11. Active participation of all was encouraged and valued	1	2	3	4	5	6	7
12. There was a climate of respect for students' experiences, ideas, & contributions	1	2	3	4	5	6	7
13. Interactions reflected collaborative working relationships among students & between Implementer & students	1	2	3	4	5	6	7
14. The implementer's language & behavior demonstrated sensitivity to variations in students:							
a. experience &/or preparedness	1	2	3	4	5	6	7
b. gender, race/ethnicity, culture	1	2	3	4	5	6	7
15. The implementer monitored students formally &/or informally and consistently focused on student understanding of content and skills	1	2	3	4	5	6	7
16. Implementer adjusted lesson through varied strategies & activities to provide deeper understanding of content & skills	1	2	3	4	5	6	7

17. Degree of closure or resolution of conceptual understanding was appropriate for the purposes of the session & the needs of learners	1	2	3	4	5	6	7
18. Implementer's ability to teach science using CREATE	1	2	3	4	5	6	7
19. _____	1	2	3	4	5		

B. Synthesis Rating

1	2	3	4	5
Instruction/ Implementation of the session not at all reflective of best practices for CREATE				Instruction/ Implementation of the session extremely reflective of best practices for CREATE

C. Supporting Evidence for Synthesis Rating

III. Science Content

A. Ratings of Key Indicators

	Not at all				To a great extent	Don't know	N/A
1. Disciplinary content was appropriate for the purposes of the CREATE session and the backgrounds of the students	1	2	3	4	5	6	7
2. The science content was appropriately presented/explored within the CREATE process	1	2	3	4	5	6	7
3. Content area was portrayed as a dynamic body of knowledge continually enriched by conjecture, investigation, analysis, & proof/ justification	1	2	3	4	5	6	7
4.Depth of attention to science content was appropriate for the purposes of the session & the needs of learners	1	2	3	4	5	6	7
5. Breadth of attention to science content was appropriate for the purposes of the session & the needs of learners	1	2	3	4	5	6	7
6. Appropriate connections were made to real world contexts through CREATE process	1	2	3	4	5	6	7
7. Students' ability to identify & understand important ideas of science	1	2	3	4	5	6	7
8. _____	1	2	3	4	5		

B. Synthesis Ratings

1	2	3	4	5
Science content during session was not at all reflective of best practices for CREATE				Science content during session extremely reflective of best practices for CREATE

C. Supporting Evidence for Synthesis Ratings

IV. Nature of Science

A. Ratings of Key Indicators

1. The nature of science was appropriately presented/explored within the CREATE process	1	2	3	4	5	6	7
2. The nature of science was portrayed as presuming that things & events in the universe occur in consistent patterns that are comprehensible through careful, systematic study[2]	1	2	3	4	5	6	7

[2] Rutherford, F. J. & Ahlgren, A. (1990). *Science for all Americans.* New York: Oxford University Press.

3. Students' understanding of the nature of science	1	2	3	4	5	6	7

B. Synthesis Ratings

1	2	3	4	5
NOS content during session was not at all reflective of best practices for CREATE				NOS content during session extremely reflective of best practices for CREATE

C. Supporting Evidence for Synthesis Ratings

SD 1. Outside Evaluator's modification of Weiss Observation Protocol for evaluation of CREATE implementers' teaching. See text for discussion.

SD2

Student population	N	Question A								
		Logical Justifications					Illogical justifications			
		Pre mean (SD)	Post mean (SD)	Signif	ES		Pre mean (SD)	Post mean (SD)	Signif	ES
1	17	2.5 (1.3)	3.1 (1.5)	ns	--		0.3 (0.8)	0.2 (0.5)	ns	--
2	11	2.5 (0.9)	2.4 (0.9)	ns	---		0.3 (0.6)	0 (0)	ns	--
3	14	1.9 (1.3)	3.1 (1.6)	0.040	0.9		0.6 (0.5)	0.4 (0.6)	ns	--
4	29	1.5 (1.0)	2.4 (1.4)	0.005	0.8		0.4 (0.4)	0.4 (0.7)	ns	--
5	12	2.2 (1.6)	2.7 (1.9)	ns	--		0.6 (0.8)	0.4 (0.7)	ns	--
6	11	1.9 (1.5)	2.4 (1.5)	ns	--		0.5 (0.7)	0.3 (0.5)	ns	--
7	14	3.1 (1.4)	3.7 (0.9)	0.050	0.5		0.1 (0.3)	0 (0)	ns	--

Student population	N	Question B								
		Logical Justifications					Illogical justifications			
		Pre mean (SD)	Post mean (SD)	Percent change	ES		Pre mean (SD)	Post mean (SD)	Signif	ES
1	17	1.0 (1.1)	2.5 (1.2)	150%	1.3		1.3 (1.2)	0.2 (0.5)	0.001	1.2
2	11	1.4 (1.2)	3.1 (1.4)	63%	1.3		1.0 (0.9)	0.2 (0.4)	0.020	1.1
3	14	1.3 (1.3)	3.0 (2.1)	130%	1.0		0.9 (0.7)	0.6 (0.7)	ns	--
4	29	0.3 (0.5)	1.8 (1.3)	500%	1.5		1.7 (0.6)	0.8 (0.7)	0.000	1.3
5	12	0.7 (0.8)	3.1 (1.8)	343%	1.7		0.7 (0.4)	0.7 (0.6)	ns	--
6	11	1.5 (1.4)	2.4 (1.6)	60%	--		1.4 (1.1)	0.5 (0.7)	0.010	1.0
7	14	1.6 (1.1)	3.6 (1.3)	125%	1.6		0.9 (0.7)	0.1 (0.4)	0.003	1.4

Student population	N	Question C								
		Logical Justifications					Illogical justifications			
		Pre mean (SD)	Post mean (SD)	Signif	ES		Pre mean (SD)	Post mean (SD)	Signif	ES
1	17	2.2 (1.4)	1.9 (1.2)	ns	--		0.9 (0.9)	0.3 (0.6)	0.037	0.9
2	11	2.3 (1.1)	2.7 (0.7)	ns	--		0.3 (0.5)	0 (0)	ns	--
3	14	2.2 (1.1)	3.1 (1.4)	0.001	0.7		0.2 (0.4)	0.1 (0.4)	ns	--
4	29	1.4 (1.0)	1.9 (1.1)	0.025	0.5		0.6 (0.6)	0.5 (0.8)	ns	--
5	12	1.4 (1.6)	1.4 (1.4)	ns	--		0.7 (1.0)	0.6 (0.5)	ns	--
6	11	1.8 (1.0)	1.1 (1.3)	ns	--		0.7 (0.9)	0.8 (0.9)	ns	--
7	14	1.5(1.3)	3.1 (1.4)	0.004	1.3		0.6 (0.6)	0.1 (0.3)	0.027	1.1

Student population	N	Question D								
		Logical Justifications (mean, SD)					Illogical justifications (mean, SD)			
		Pre mean (SD)	Post mean (SD)	Signif	ES		Pre mean (SD)	Post mean (SD)	Signif	ES
1	17	0.9 (0.4)	1.2 (0.9)	ns	--		1.5 (1.1)	1.3 (1.0)	ns	--
2	11	1.4 (0.5)	1.7 (1.0)	ns	--		1.2 (0.9)	0.7 (0.9)	ns	--
3	14	1.7 (0.8)	1.9 (1.0)	ns	--		1.2 (1.0)	1.0 (0.8)	ns	--
4	29	0.4 (0.7)	1.3 (0.7)	0.000	1.2		1.0 (0.8)	1.2 (0.6)	ns	--
5	12	1.2 (0.7)	1.4 (1.0)	ns	--		1.1 (0.9)	0.8 (0.7)	ns	--
6	11	1.3 (1.0)	1.0 (0.9)	ns	--		1.4 (0.9)	1.2 (1.0)	ns	--
7	14	1.4 (0.9)	2.2 (0.4)	0.009	1.3		1.1 (1.0)	0.6 (0.6)	ns	--

SD 2a-d. Critical Thinking Test (CTT) scores for individual implementer campuses. CTT questions A, C, D were from the Field Tested Learning Assessment guide; question B was written by the PIs for the CTT. ES, effect size. Each question was used previously at CCNY (Hoskins et al., 2007). CI 1, 3 and 7 were full-semester implementations; the others ranged from 4-10 weeks. CI 1, 2 and 7 are private institutions; the others are public. Significance determined using paired t test (Excel); p values < 0.05 were considered significant.

SAS survey—individual CIs compared with a CCNY CREATE course on attitude/ability categories

Decoding primary literature	Pre mean (SD)	Post mean (SD)	Significance (Wilcoxon Signed Rank Test)	ES
CCNY spring 2009	11.7 (3.4)	16.1 (2.3)	0.000	1.5
1	13.6 (2.5)	15.2 (2.2)	ns	0.5
2	15.1 (2.5)	17.5 (1.8)	0.004	1.3
3	12.4 (2.8)	15.6 (2.1)	0.003	1.2
4	12.8 (3.1)	13.6 (3.3)	ns	--
5	14.4 (2.7)	16.0 (1.7)	ns	--
6	11.7 (2.6)	15.1 (3.1)	0.01	2.4
7	14.1 (2.1)	16.2 (1.3)	0.08	1
Interpreting data	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	9.9 (1.8)	12.7 (1.4)	0.001	1.6
1	10.4 (2.1)	11.2 1.6)	ns	-
2	11.8 (1.0)	12.3 (1.5)	ns	-
3	10.2 (1.9)	12.0 (1.7)	0.003	1.2
4	10.3 (2.1)	10.2 (2.4)	ns	-
5	11.7 (1.8)	12.3 (1.4)	ns	-
6	8.6 (2.1)	11.4 (2.2)	0.025	0.9
7	10.9 (1.4)	12.4 (1.3)	0.017	0.9
Active Reading	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	12.9 (2.4)	16.3 (2.7)	0.001	1.3
1	14.6 (1.4)	15.9 (1.4)	0.002	0.8
2	15.8 (1.8)	17.3 (1.7)	0.041	0.7
3	13.5 (1.8)	15.3 (1.9)	0.009	0.9
4	13.9 (2.1)	15.3 (1.7)	0.003	0.8
5	14.7 (1.7)	16.1 (1.4)	0.020	0.8
6	12.0 (2.3)	15.3 (2.9)	0.005	1.5
7	13.9 (1.7)	16.2 (2.0)	0.002	1.4
Visualization	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	9.7 (2.1)	12.5 (1.4)	0.000	1.6
1	9.8 (2.2)	11.7 (0.7)	0.008	0.8
2	10.3 (2.1)	11.9 (1.5)	0.02	0.8
3	9.6 (2.0)	11.3 (2.0)	0.008	1.0
4	10.2 (1.9)	10.7 (2.3)	ns	--
5	10.9 (2.2)	12.6 (0.9)	0.049	--
6	8.4 (2.7)	11.2 (2.1)	0.006	1.4
7	10.3 (2.1)	12.5 (1.0)	0.005	1.0
Thinking like a scientist	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	9.6 (2.2)	12.7 (1.8)	0.000	1.6

1	11.4 1.8)	12.6 (1.3)	ns	0.6
2	12.5 (1.6)	12.6 (1.2)	ns	--
3	10.3 (1.7)	12.5 (1.7)	0.001	2.0
4	10.6 (2.1)	10.5 (2.0)	ns	--
5	10.7 (1.6)	12.8 (0.9)	0.005	1.0
6	10.6 (2.1)	12.0 (2.5)	0.01	0.9
7	10.4 (1.1)	12.3 (1.4)	0.003	1.4

SD 3a: Outcomes for individual campuses on Student Attitude Survey; outcomes from the same survey administered to a single CREATE course taught to upper-level (third and fourth-year) students at CCNY, a MSI, in spring 2009 by one of the authors (SGH, an experienced CREATE teacher) are included for comparison. Campuses 1-7 were taught by first-time CIs. On five of the six attitude/ability factors defined in our previous study, students on multiple campuses showed significant precourse/postcourse gains, with moderate to large ES. The extent of change varied on different campuses, but was overall in the same range and with same ES magnitudes as seen at CCNY). A sixth factor, "research in context" showed significant change at CCNY with moderate ES, but no change on the other campuses. These findings argue that student self-rated attitudes and abilities change significantly over a CREATE semester, both at a MSI (CCNY) and at private liberal arts colleges (implementation 1, 2), public universities (implementations 3-6) and in a cohort from a private R1 institution (implementation 7). N for CCNY 2009=26; for implementation 1 = 16; implementation 2 = 13; implementation 3 = 15; implementation 4 = 31; implementation 5 = 12; implementation 6 = 11; implementation 7 = 14. Statistical analysis performed using VassarStats calculator for Wilcoxon Signed-Rank test (http://vassarstats.net/wilcoxon.html).

SAS survey—individual CIs compared with a CCNY CREATE course on epistemological belief categories				
Certain knowledge	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	7.9 (1.4)	8.6 (1.5)	0.015	0.5
1	8.8 (0.9)	7.7 (1.3)	0.01	0.9
2	9.0 (1.1)	9.2 (0.8)	ns	--
3	8.5 (1.7)	7.4 (1.0)	0.01	0.7
4	8.1 (1.1)	7.8 (1.3)	ns	--
5	7.8 (2.0)	7.5 (1.4)	ns	--
6	8.1 (1.5)	7.8 (1.6)	ns	--
7	8.9 (2.9)	8.7 (0.9)	ns	--
Creativity	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	4.2 (0.7)	4.6 (0.6)	0.036	0.6
1	4.1 (0.9)	4.7 (0.4)	0.049	0.7
2	4.3 (0.8)	4.6 (0.6)	ns	--
3	4.1 (.6)	4.5 (0.5)	0.049	0.4
4	4.2 (0.6)	4.2 (0.7)	ns	--
5	4.1 (0.8)	4.4 (0.5)	ns	--
6	4.5 (0.5)	4.5 (0.9)	ns	--
7	4.2 (0.5)	4.8 (0.4)	ns	--
Sense of scientists	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	3.0 (0.8)	3.7 (0.7)	0.012	0.9
1	3.6 (0.9)	3.8 (0.7)	ns	--
2	3.5 (1.1)	3.4 (0.8)	ns	--
3	2.7 (0.8)	3.5 (1.0)	0.035	0.8
4	2.7 (0.9)	2.7 (0.8)	ns	--
5	3.7 (0.6)	3.8 (0.7)	ns	--
6	2.5 (1.0)	3.7 (0.4)	0.001	1.4
7	3.8 (0.8)	3.9 (0.5)	ns	--
Known outcomes	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	4.0 (0.9)	4.4 (0.9)	ns	--
1	4.1 (0.8)	3.9 (0.8)	ns	--
2	3.6 (1.0)	3.7 (1.0)	ns	--
3	3.7 (1.0)	3.5 (1.1)	ns	--

4	4.1 (0.5)	4.2 (0.7)	ns	--
5	3.2 (1.3)	4.0 (0.8)	0.049	0.7
6	3.2 (0.8)	3.9 (1.0)	0.049	0.7
7	3.6 (0.9)	3.9 (0.6)	ns	--
Collaboration	**Pre mean (SD)**	**Post mean (SD)**	**Significance (Wilcoxon Signed Rank Test)**	**ES**
CCNY 2009	4.3 (0.6)	4.6 (0.6)	ns	--
1	4.6 (0.5)	4.2 (1.3)	ns	--
2	4.8 (0.4)	4.8 (0.4)	ns	--
3	4.2 (0.5)	4.7 (0.6)	0.005	1.1
4	4.2 (0.6)	4.3 (0.8)	ns	--
5	4.5 (0.5)	4.5 (0.5)	ns	--
6	4.3 (0.6)	4.4 (0.7)	ns	--
7	4.8 (0.4)	4.7 (0.5)	ns	--

SD 3b . Individual implementation outcomes on SAS statements that address students' epistemological beliefs about science. Outcomes for individual campuses are compared with outcomes on the same survey among CCNY students in a CCNY course taught by one of the authors (SGH). As in our previous study at CCNY, fewer changes were seen in epistemological beliefs than in student self-rated attitudes and abilities (Hoskins, Lopatto, Stevens 2011). Notably however, for five of the seven epistemological categories, changes in the direction of more mature beliefs were seen on at least one campus, most with moderate to large ES. Overall students in four of the seven CREATE implementation cohorts, including private and public, full and part-semester implementations, showed significant change in at least one epistemological category.

Results were more variable for the statements related to students' beliefs. In four implementations, some significant change was seen in such beliefs. Three implementations produced changes in students' sense of scientists, and in two implementations, students shifted to more mature views of whether science is creative and whether knowledge is certain. Students' sense of the extent to which knowledge is certain, whether science is collaborative showed significant change in one implementation. Most epistemological changes occurred in full-semester implementations. The CCNY 2009 cohort showed significant change in four of the epistemological measures. As in the implementation cohorts, overall fewer changes and lower ES were seen for epistemological categories than were seen in self-rated abilities of the same student cohorts. Taken together, these data support the conclusion that as at CCNY, students in the majority of CIs undergo some epistemological maturation in their CREATE course.

The changes seen in the implementations taught by first-time CREATE faculty are comparable in magnitude to those seen in the CCNY 2009 CREATE course taught by an experienced CREATE faculty member (SGH). Taken together with Table 2a, these data support the idea that first-time teachers can teach CREATE courses that have strong effects both on students' self-rated abilities and attitudes, and on their core beliefs about the nature of scientific knowledge. In previous work we saw no pre/postcourse changes in any comparable factors in a comparison group of non-CREATE students taking a semester long sophomore-level physiology course taught by a CCNY colleague (Gottesman and Hoskins, 2013) ; we thus attribute the shifts seen to the effects of the CREATE pedagogical style rather than to changes that would happen "naturally" in a student cohort during any semester-long college science course.