DOCUMENT RESUME

ED 373 702 IR 016 712

AUTHOR	Black, John B.; And Others
TITLE	Assessing Student Understanding and Learning in Constructivist Study Environments.
PUB DATE	94
NOTE	10p.; In: Proceedings of Selected Research and Development Presentations at the 1994 National Convention of the Association for Educational Communications and Technology Sponsored by the Research and Theory Division (16th, Nashville, TN, February 16-20, 1994); see IR 016 784.
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Archaeology; Comparative Analysis; *Computer Assisted Instruction; Computer Graphics; *Computer Simulation; *Constructivism (Learning); Courseware; Educational Assessment; *Educational Technology; Elementary School Students; Evaluation Methods; Grade 6; Intermediate Grades; Problem Solving
IDENTIFIERS	Columbia University NY Teachers College; Explanations; Learning Environments; New York (New York)

ABSTRACT

Teachers College of Columbia University (New York) and the Dalton School, an independent school in New York City, have collaborated on the Dalton Technology Project and its "Archaeotype" program which presents students with a graphic simulation of an archeological site. Students simulate digging up the artifacts, use reference sources to learn about the history, and apply their knowledge in the simulation. Comparing the ability of "Archaeotype" students to investigate and make conclusions with that of students who did not use the "Archaeotype" program serves as a test of learning and understanding from the simulation. Subjects were 20 sixth graders, who were compared with 20 from another independent school. Students used a simulation unfamiliar to both groups. Results show an impressive ability on the part of "Archaeotype" students to create explanations of observations and argue for the validity of those explanations using a mixture of their own ideas and terms and the technical terminology and concepts in the simulation. "Archaeotype" students did not excel in data representation, an area in which the simulation might be strengthened. One table presents analysis results. (Contains 8 references.) (SLD)

Title:

Assessing Student Understanding and Learning in Constructivist Study Environments

Authors:

John B. Black
Robert McClintock
Clifford Hill

Teachers College
Columbia University

116TH ST. AND BROADWAY
NEW YORK, NY 10027

For the last couple of years, Teachers College, Columbia University and the Dalton School (an independent school in New York City) have collaborated on the Dalton Technology Project. This project aims to use networked multimedia workstations to produce an environment that supports student studying in groups using authentic materials and contexts. This approach to education constrasts sharply with the usual approach which has students working individually to passively receive knowledge from teachers and textbooks using artificial problems. The project shares many features with the developing constructivist approaches to instructional design (e.g., Jonassen, 1991; Bednar, Cunningham, Duffy and Perry, 1991; Collins, Brown and Newman, 1990; Cognition and Technology Group at Vanderbilt, 1990; Spiro, Feltovich, Jacobson and Coulson, 1991), but it differs from them in emphasizing design for ***study*** as opposed to design for instruction. Thus, we strive to create "a place for study in a world of instruction" (McClintock, 1971).

Seven Principles of Study Design

In addition to developing the particular study systems for different subject areas in the Dalton Technology Project, we have been trying to specify what the underlying design principles are for such an approach. In doing this we draw inspiration both from Cognitive Science (e.g., Brown, Collins and Duguid, 1989) and from hermeneutic interpretation theory (e.g., Palmer, 1969). From this effort, we have come up with the following seven study system design principles:

1. **Text:** Present students with particular cultural objects (events, writings, images, artifacts, scores, observations, experiments, etc.), the origin and meaning of which will confront them as obscure, a challenge to the understanding.

2. **Context:** Provide students with open-ended access to contextual materials that may help to clarify and interpret the cultural objects presented to them and provide pathways leading from the particular object to the comprehensive assemblage of pertinent materials. On the one hand, the context must be immediate, and on the other hand it should include everything.

3. **Engagement:** Situate the presentation of the text and context -- both the challenging cultural objects and their contextualizing resources -- in such a way that students will grasp strong ownership of the on-going effort to interpret the material.

4. **Cooperation:** Have students collaborate in their quest for interpretative understanding, learning to empathize with the interpretative actions of their peers.

5. **Inclusivity:** Use cognitive apprenticeship to show students how to enlarge the scope and power of the contextual materials they bring to bear on interpreting the text, moving the interpretation toward that ideal condition in which all significant contextualizing materials have been taken into account.

6. **Abstraction:** Encourage students to bring significant contexts to bear upon multiple, different cultural objects to prepare them to transfer their interpretative skills to novel problems.

7. **Diversity:** Encourage students to situate complex cultural objects in many different significant contexts to prepare them to develop the cognitive flexibility of understanding things from many points of view.

An example program will serve to illustrate these principles, then we will discuss how to assess student understanding and learning in these kinds of study environments. In the *Archaeotype* program, students study ancient Greek and Roman history by using observations of simulated archaeological digs to construct interpretations of the history of these sites, while drawing upon a wide variety of background information. The *Archaeotype* program (implemented in Supercard on Macintosh computers), which is the earliest and most fully-developed of the Dalton Technology Project programs, presents the students with a graphic simulation of an archaeological site, then the students study the history of the site through simulated digging up of artifacts (the **text**), making various measurements of the artifacts in a simulated laboratory, and relating the objects to what is already known using a wide variety of reference materials (the **context**). The students work **cooperatively** in groups, while the teacher models how to deal with such a site then fades their involvement while coaching and supporting the students in their own study efforts (**inclusivity**). The students develop ownership of their work by developing their own interpretations of the history of the site and mustering various kinds of evidence for their conclusions (**engagement**). By arguing with the other students and studying related interpretations in the historical literature, they get a sense of other perspectives (**diversity**). By going through the process a number of times bringing each contextual background to bear on a number of differnt artifacts, the students learn and understand the general principles behind what they are doing (**abstraction**).

Assessing Student Understanding and Learning

So, what might students get from an educational experience like *Archaeotype* that they wouldn't get from a regular class, and what might they get from a regular class that they wouldn't get in *Archaeotype*? In a regular class on Greek and Roman history, the students would probably learn more facts about history (because they are devoting all their time to learning such facts) than the *Archaeotype* students would learn, but the *Archaeotype* students would probably remember the facts they do learn longer and have a greater understanding of them and historical reasoning. Thus if given an objective test of memory for Greek and Roman history facts at the end of the course, a standard class would probably do better than an *Archaeotype* class, but a year or two later the *Archaeotype* class would probably do better. More importantly, if we examined essays arguing for some historical conclusion, then we would expect the *Archaeotype* students to be much more sophisticated than the regular students (in fact, the reports from current *Archaeotype* students seem quite sophisticated in terms of language, argument structure, citations, etc.) -- and thus demonstrate a much deeper understanding of historical facts and reasoning. We are in the midst of conducting such an investgation of content learning, but do not have the results to report yet.

However, more than these particulars of the topic area for a class, an *Archaeotype*-type educational experience should teach students to examine any situation, make relevant observations and measurements, organize these materials, search out related bodies of knowledge, organize all this information and use it to draw compelling conclusions and make useful recommendations. Thus, the strongest test of student learning and understanding from *Archaeotype* would be to compare their ability to investigate and make conclusions and recommendations in an entirely different and unrelated situation to the ability of students who have not had an *Archaeotype* experience to do the same. That is what we did in the study reported here.

In the study we conducted, the students were given a booklet describing four psychology experiments examining how people remember lists of words. The students had to examine the basic obeservations, report on the results of the studies, find the patterns, devise explanations and argue for those explanations. They were also given some background readings in the psychology of memory. The Dalton students who had been through the *Archaeotype* program were compared to students from the Grace Church School (who also had some data-analysis experience from going through *The Voyage of the Mimi* program from Scholastic Publishers).

Method

Participants

The experimental group was 20 sixth-grade students who had participated in the *Archaeotype* program at the Dalton School, an independent school located on the east side of Manhattan. The control group was 20 sixth-grade students who attended the Grace Church School, an independent school also located on the east side of Manhattan.

Materials

Students in the two groups were given a ten-page document (the assignment booklet) divided into two parts. The first part described the results of four memory studies as follows:

(1) in study 1 subjects listened to 20 words spoken at the rate of one word per second and then immediately recalled them
(2) in study 2 subjects listened to the same words spoken at the rate of one word every three seconds and then immediately recalled them
(3) in study 3 subjects listened to the same words spoken at the rate of one word per second but recalled them only after performing an unrelated 30-second task
(4) in study 4 subjects listened to a different 20 words (many of which were semantically related) spoken at the rate of one word per second and then immediately recalled them.

The second part of the document provided background readings on technical concepts such as short-term memory and long-term memory. Students were asked to use these readings to interpret the results of the four studies and to present their interpretations, along with practical recommendations for improving memory, in a written report.

Procedure

Administering the Materials and Collecting Student Reports

The study was conducted in two 2-hour sessions (for a total of 4 hours) spread over two adjacent days. On the first day, the experimenter passed out the assignment booklets, the students paired up, the experimenter read the instructions on the first page of the assignment booklet, then the experimenter ran a demonstration of the kinds of memory studies described in assignment booklets. In the demonstration the experimenter read a list of 20 words then the students wrote down their recall of them and the experimenter conducted a short discussion of what the results were. This demonstration was done so that the students could see what the studies described in the assignment booklets were like. After the demonstration, the students proceeded to work on the assignment in groups of two. While doing the assignment the students were free to use any of the resources in the Dalton and Grace Church School buildings (computers, libraries, etc.) including asking the experimenter for clarification and information questions (the same experimenter conducted all sessions). At the end of the 2-hour period on the second day, the students handed in their reports and all the work they had done in folders. The experimenter then lead a half-hour discussion of the study.

Analysis of Student Reports

We devised a rubric for evaluating three dimensions of the student reports pattern recognition, argumentation, and data representation. Given the emphasis on data interpretation in the *Archaeotype* program, we accorded the most weight to the dimension of argumentation, as indicated by the following distribution of points:

(1) pattern recognition (20 points)
(2) argumentation (30 points)
(3) data representation (10 points)

In principle, students could receive a total of 60 points, though we should point out that the rubric was designed to reflect what might be described as expert responses to the task. This emphasis on high standards is in keeping with the larger movement in educational reform that is often referred to as *authentic assessment*.

<u>Pattern Recognition.</u> Students received 1-2 points for describing each of the following intra-study patterns:

(1) in study 1 the pattern of *last words/first words/middle words* (with *middle words* highly attentuated)
(2) in study 2 the pattern of *last words/first words/middle words* (with *middle words* more developed)
(3) in study 3 the pattern of *first words/middle words/last words*(with *last words* highly attenuated)
(4) in study 4 the pattern of *last words/words grouped in semantic categories* (with *last words* relatively attenuated)

In addition, students received 1-2 points for describing each of the following cross-study patterns that relate to number of words recalled:

(5) more words were recalled in study 2 than in study 1
(6) fewer words were recalled in study 3 than in studies 1 and 2
(7) more words were recalled in study 4 than in studies 1, 2, and 3

In effect, the number of words recalled in the studies can be ranked in the following order:
study 4 > study 2 > study 1 > study 3

Apart from these major patterns, students received 1-6 points for noticing other significant patterns (i.e., 1-2 points up to three patterns): for example, in studies 1 and 2 when middle words were recalled, they often formed associative pairs (e.g., *cup/water*); or in study 4 the most salient semantic categories were those involving fruit and animals as opposed to those involving furniture and transportation (i.e., words in these categories were recalled not only more frequently but earlier in the sequence); and within the various categories, certain words which function as prototypes, tended to be recalled first: for example, *coat* for the category of clothing and *chair* for the category of furniture.

Explanation and Argumentation. Students were expected to draw on the background readings to develop arguments supporting hypotheses about the patterns they observed in the four studies. As a consequence, arguments that drew appropriately on the background readings were awarded 1-4 points each, whereas arguments, which did not draw on the background readings, were awarded 1-2 points each. Here are local arguments that could be used in interpreting major patterns in the four studies:

(1) in study 1 short-term memory explains the fact that the last words are the first recalled
(2) in study 2 increase in time - and thus deeper processing in long-term memory - explains the fact that more words can be recalled (especially, the middle words that can be meaningfully associated)
(3) in study 3 the intervening 30-second task is used to explain not only the fact the last words are no longer recalled first (i.e., short-term memory is no longer operating) but fewer total words are recalled (i.e., long-term memory is diminished as well)
(4) in study 4 the presence of semantically related words is used to explain the fact that not only are more words recalled but the sequence in which they are recalled (i.e., semantically related v 's tended to be grouped).

In addition to local argumentation, students were given credit for global argumentation (e.g., these four studies suggest that meaningful associations among individual words is the most powerful factor in word recall). They were given 1-2 points if such argumentation was presented without the background readings, 1-4 points if it was presented with the background readings.

As to the final recommendations in the report, students were given 1-4 points for grounding them in the data (e.g., ample time should be provided so that meaningful associations can be formed between the items to be remembered) and 1-4 points for grounding them in the background readings (e.g., meaningful associations should be developed so that material can be transferred from short-term memory to long-term memory).

Students were also given 1-2 points whenever they displayed legitimate forms of alternative explanation for the same phenomena (for example, in study four the fact that *cat* tended to occur early among the recalled words could have been explained by the fact that it was among the last words presented (i.e., short-term memory) and/or the fact it serves as a prototype of the 'animal' category (i.e., members of such a category, as mentioned, tend to occur before members of 'furniture' or 'transportation' categories).

Data Representation. Students were given credit if they used numerical and/or graphic methods to represent major patterns in the four studies. With respect to numerical methods, they received 1-2 points if they calculated the means for significant patterns such as

(1) the total number of words recalled in each study
(2) the number of *first words, middle words,* and *last words* recalled in studies 1-3
(3) the number of words recalled in the semantic categories as well as the number of *last words* recalled in study 4.

Students received an additional 1-2 points if they used these means to establish significant proportions such as

(1) the relative weighting of *first words, middle words,* and *last words* that were recalled in studies 1-3
(2) the relative weighting of *last words* and associated words (i.e,. those in the semantic categories) that were recalled in study 4.

As to graphic methods of representation, students were given 1-6 points for appropriate use of such methods. These methods include bar graphs that represent the proportions of different kinds of words recalled in the four studies. With respect to studies 1-3, the line graph of proportion recalled ploted against serial position (usually called "the serial position curve") could have been used to represent the major patterns constituted by *first words/middle words/last words.* Alternatively, they could have used a flow chart to represent the input/output relations for short-term and long-term memory in these studies. With respect to study 4, they could have used tree-structures to represent membership in the major semantic categories.

Results

We present the results in Table 1. The numbers in this table are the means for the *Archaeotype* group and the Control group. The total possible score overall was 60 points, although this represents all that could conceivably be found, not what any pre-college student could attain -- only a specialist in the psychology of memory would have a chance of getting all these points. Thus, the important aspect of these numbers is not their absolute value, but how the *Archaeotype* and Control groups compare. This comparison is striking: in total (the first column in Table 1), the *Archaeotype* group scored 31% higher than the Control group (25.2 vs 19.2 -- out of a possible 60), and this difference was very statistically significant, $t(38)=2.22$, $p<.02$. To do this statistical analysis and the others reported later, we assigned each student the score of the report created by the group (here, each group is a pair) that they were in, then calculated a t test to see how big the difference between the means of the *Archaeotype* student scores and the Control student scores were compared to the variance of these scores within the *Archaeotype* group and within the Control group.

Table 1
Quantitative Analysis of Reports Written by Students in the *Archaeotype* Group and the Control Group

Representation	Total	Pattern Recognition	Explanation and Argumentation	Data
Archaeotype Group	25.2	10.6	13.8	0.8
Control Group	19.2	9.6	8.0	1.6

As described earlier, this overall total score breaks down into subscores for recognizing the patterns in the observations (Pattern Recognition), explaining the patterns and arguing for those explanations (Explanation and Argumentation), and converting the observations into forms that could provide insight (Data Representation). This breakdown shows that the overall *Archaeotype* superiority was almost totally caused by a 73% higher performance for the *Archaeotype* students in the important Explanation and Argumentation area (13.8 vs 8.0 -- out of a possible 30 points). Statistically also, this is a highly significant difference, $t(38)=3.34$, $p<.001$. There was also a slight difference in favor of the *Archaeotype* students in the Pattern Recognition scores (10.6 vs 9.6 -- out of a possible 20), but that difference was not even close to being statistically significant so we have to discount it, $t(38)=0.76$, $p>.2$.

The Data Representation scores held two surprises for us. The first surprise is that they were so low (16% and 8% of the possible, compared to 27%-53% of the possible in the other areas): neither the *Archaeotype* students nor the Control students used means, proportions, graphics nor diagrams in their discussions -- they merely talked about one condition described in the experimental materials being greater than another. The second surprise is that the Control students scored better than the *Archaeotype* students (1.6 vs 0.8 -- out of a possible 10) to a significant degree, $t(38)=1.95$, $p<.05$. However, the Control advantage was totally due to these students putting the observations into a database program on the computers (part of *Microsoft Works*, which they were accustomed to using) and calculating means. For example, one pair of students in the control group displayed the database shown in Appendix C, Figure 5. This use of databases was a potentially valuable move, but the control students did not exploit this analysis for Pattern Recognition and Explanation-Argumentation. The *Archaeotype* students did not show comparable use of database or spreadsheet programs and thus scored lower on Data Representation. Taken together these results show that the students both need to have experience using computer programs for manipulating data, but they also need practice using them meaningfully as part of their work in analyzing authentic tasks.

Discussion

The results showed an impressive ability on the part of the *Archaeotype* students to create explanations of observations and argue for the validity of those explanations using a mixture of their own terms and ideas, and the technical terminology and concepts provided by background readings in a research literature. They also did well in recognizing patterns in the observations, but not significantly better than the control group we compared them to. In fact, the similar performance of the Dalton School *Archaeotype* students and the Grace Church School Control students on the Pattern Recognition portion of the assignment provides assurance that the two groups were comparable, which makes the much higher performance of the *Archaeotype* students on Explanation-Argumentation all the more impressive. However, we need to also recognize that the basic patterns in the observations the students were analyzing were fairly easy to see -- particularly, after the demonstration and discussion conducted by the experimenter in the beginning of the sessions. It may be that if the patterns being searched for had been less apparent then there would have been more of a difference in Pattern Recognition between the *Archaeotype* students and the Control students. In fact, a study we have done comparing performance on another program with a similar design (Galileo which teaches science to high school students through astronomy) found pattern-recognition differences when the patterns were much harder to see.

The *Archaeotype* students actually did worse than the Control students in Data Representation, although both groups scored rather low in this area. It is disappointing that the *Archaeotype* students did not use even such rudimentary ways of representing data as counts, means and proportions. At least some students in the Control group managed to do some counting and means through entering the observations into a computer database program they were accustomed to using. Ideally, the students would even have used visualization techniques like graphs and diagrams to reveal patterns in the observations and to argue for their explanations. *Archaeotype* would seem a natural context within which to introduce the powerful idea of representing information in different forms to gain insight.

References

Bednar, A.K., Cunningham, D., Duffy, T.M., and Perry, J.D. (1991). Theory into practice: How do we link? In G.J. Anglin (Ed.) *Instructional technology: Past, present and future*. Englewood, CO: Libraries Unliminted, Inc.

Brown, J.S., Collins, A. and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, **19**, 32-42.

Cognition and Technology Group at Vanderbilt (1990). Anchored instruction and its relation to situated cognition. *Eductional Researcher*, **20**, 2-10.

Collins, A., Brown, J.S., and Newman, S.E. (1990). Cognitive apprenticeship. In L.B. Resnick (Ed.). Knowing, learning and instruction. Hillsdale,NJ: Erlbaum.

Jonassen, D.H. (1991) Objectivism versus Constructivism: Do we need a new philosophical paradigm? *Educational Technology Research and Development*, 39, 5-14.

McClintock, R. (1971) . Toward a place for study in a world of instruction. *Teachers College Record*, **72**, 405-416.

Palmer, R.E. (1969) *Hermeneutics*. Evanston,IL: Northwestern University Press.

Spiro, R.J., Feltovich, P.J., Jacobson, M.J. and Coulson, R.L. (1991). Cognitive flexibiltiy, constructivism, and hypertext. *Educational Technology*, **21**, 24-33.