Selecting Evaluation Items for Judging Concept Attainment in Instructional Design

David Richard Moore
Ohio University

Abstract

Instructional strategies for successfully teaching concepts are found throughout the instructional design literature. These strategies primarily consist of presenting learners with definitions, examples, and non-examples. While examples are important presentation instruments, theorist suggests that examples should not be re-used in the assessment phase of instruction. The rationale being that encountered examples could be memorized thus activating different cognitive processes than those required for concept attainment. Consequently, test items referring to encountered examples may have less value in assisting evaluators in discerning whether or not a learner has attained a target concept. In this study test items consisting of encountered examples and un-encountered examples were examined. There appears to be evidence supporting the notion that examples are not sufficient discriminators for judging a learner's level of concept attainment.

Introduction

Instructional theorists often advise practitioners to use different examples when testing for the attainment of a target concept than were used during the instructional phase (Merrill, 1983; Merrill, Tennyson, & Posey, 1992). The rationale for this procedure is based upon the belief that previously used examples could be learned by rote memory instead of being meaningfully integrated into one's semantic structure, leading to an arbitrary and unstable retention (Ausobel, 2000). The use of un-encountered examples in testing has been found to be more robust than those of encountered examples (Penn, Nedeff, & Gozdzik (2000).

This study has attempted to find performance differences between encountered examples and un-encountered examples by evaluating the percent of correctly categorized instances after concept instruction. Performance was judged by percent of correct responses, as well as the response times. It was expected that the results would provide guidance to instructional designers in selecting appropriate test items for evaluating concept attainment. It is proposed that re-using examples of concepts in both instructional and testing situations will not produce valid results for assessors; if this were not the case one should see no difference in an assessment between examples that have been encountered and those that have not.

Conceptual knowledge domain

Generally speaking concepts are groupings of ideas or entities based upon shared features or attributes. Rand (1990) suggests that concepts can be defined as, "the mental integration of two or more units which are isolated according to a specific characteristic(s) and united by a specific definition" (p.10). The ability to regard separate entities as members of a single class has been thought to be critical in cognitive efficiency and may be a precursor to higher-order thinking (Hunt, 1962; Rosch, 1999). Concepts, as a knowledge domain, have been isolated in instructional design models as

being unique enough to demand specific strategies for their instruction (Merrill, 1983; Smith & Ragan, 2005).

The instructional design literature discusses two primary categories of concepts. Concrete Concepts are ostensively defined, meaning that can be identified by pointing to specific entities having a tangible presence (i.e., physical objects). However, Defined Concepts are abstract categorizations, often defined in terms of their relationship to other concepts (i.e. democracy, liberty, friend). Defined concepts may be more challenging to learn because of their abstract and less tangible nature (Smith & Ragan, 2005).

Concepts are efficiently organized into hierarchical structures that signify their relationships to other concepts. (Merrienboer, 1997). These structures form taxonomies that illustrate how super-ordinate, successive/coordinate, and subordinate concepts relate to one another (Merrill et al., 1992). For example, the concept bear is subordinate to the concept animal and super-ordinate to the concepts grizzly bear and polar bear. Other animals such as tigers and jackals might be considered to be successive concepts of the target concept "bear". Understanding these relationships is critical in the attainment of a target concept (Merrienboer, 1997).

Concept instructional methods

Concept attainment is often taught using methods such as supplying definitions and providing participants with a series of examples and non-examples (Klausmeier & Feldman, 1973). Likewise, evaluating a learner's attainment of a target concept is often based upon a learner's ability to supply definitions, list attributes, and classify instances of the target concept. (Klausmeier, 1974; Merrill, Tennyson, & Posey, 1992; Merrill & Tennyson, 1971). Additionally, comprehension may be displayed by asking the learner to identify a target concept's relationship to its hierarchical structure (Merrienboer, 1997).

Producing a definition or identifying an instance of a target concept that one has previously observed often isn't considered a sufficiently valid form of assessment (Feldman & Klausmeier, 1974; Merrill, Tennyson, & Posey, 1992; Merrill, Olsen, & Coldeway, 1976). The nature of language encourages multiple meanings and thus classifying concepts requires more than a definition, it requires examples to provide clarity and specificity. However, inferring a conceptualization from examples alone is often difficult or incorrect because their attributes may encourage a learner to over-generalize or under-generalize the concept without the guidance of a definition (Feldman & Klausmeier, 1974; Merrill et al., 1992). Consequently, definitions and examples are often codependent tools in concept instruction.

A series of divergent examples are useful in concept instruction, particularly when unique attributes are isolated and highlighted (Feldman, 1972). These examples can help learners generalize a concept to the widest of circumstances. Feldman and Klausmeier (1974) found that the presentation of examples was superior to definitions alone while the presentation of both definitions and examples in tandem was superior to either in isolation; a result confirmed by Litchfield (1987).

Response time and its role in cognitive function

Response time is a variable that has been used extensively in cognitive research and has been described as a significant variable in learning about mental processes (Rosch, 1978; Luce, 1986). A number of cognitive processes have been studied by tracking participant response times. Shepard and Metzler (1971) determined that it took participants longer to mentally rotate three-dimensional objects when the angle of initial

rotation was greater than another similar object. Rosch (1978) found that response time in a category membership verification task was a function of how close an example was to a prototype (an example with typical features). Bornstein and Monroe (1980) confirmed this finding; however, their study, as well as many studies on using response time, didn't involve potentially meaningful learning material (Ashby, Boynton, & Lee, 1994; Ausobel, 2000). Hornke (2005) confirmed a pattern of participants taking longer to respond to wrong answers than for correct ones, indicating that confusion and uncertainty has an impact on response times.

Beer and Diehl (2001) have suggested that response time consists of three pieces: individual differences, treatment effects, and error. In other words, some people naturally respond quicker than others, or the task encourages different response times, or differences can be attributed to random error. Computer tracked data is particularly appropriate for isolating these elements (Hornke, 2005) because of the relative ease in which participants can be randomly assigned to different treatments and the capability of collecting accurate and detailed measurements of response time.

One would expect that test items that provide an opportunity for immediate rote recall would require a smaller response time than ones that require an extensive semantic search. To that end, this study constructed an experiment to record the times associated with responding to examples that have been encountered and examples that are new to the learner.

Method

Purpose

In this experiment, two different types of test items (encountered examples of the target concept, un-encountered examples of the target concept) were examined and compared. The research hypothesis was that participants would be able to correctly answer questions that incorporated encountered examples more so than un-encountered examples. Additionally, it is proposed that participants will answer encountered examples faster than un-encountered examples. The difference in both percentage correct and response times should indicate that the two assessment types of examples are not instructionally equivalent. The target concept used in this study (Georgian Style of Architecture), was concrete in nature and chosen for its relative obscurity in common knowledge. Data was collected on the percent of questions participants could correctly answer as well as how long it took for them to answer these questions. It is hypothesized that participants will correctly identify examples of the target concept that have been encountered more so than examples of the target concept that have not been encountered and that response times will be faster for encountered examples than for un-encountered examples.

The two categories used for testing are displayed in Table 1.

Table 1
Expected results

Test category	Expected results
Encountered examples of the target concept	Participants should be more successful in this test category because of the identical relationship between the presented item and previously seen items, regardless of the degree of concept attainment. It is also expected that participants would respond to questions in this category quickly because of their recent exposure.
Un-encountered examples of the target concept.	Participants should be successful in this test category to the degree in which they have mastered the concept domain. Because of the unfamiliarity of the examples, participants should do less well than when answering encountered examples. It is also expected that participants will take substantially longer to respond to this category because they must apply a conceptualization scheme instead of short-term memorization effects.

Participants

The study consisted of 30 graduate students from the College of Education; on a background survey, all rated their prior knowledge of the presented topic as being minimal.

Materials

The study was conducted by means of a computer-based instruction module designed and developed in the Macromedia Authorware™ programming environment. The instrument included a method of delivering instruction on the target concept, followed by an assessment. The instruction included descriptions, definitions, and examples of Georgian architecture (Figure 1 and Figure 2).

Figure 1
Screenshot of instruction

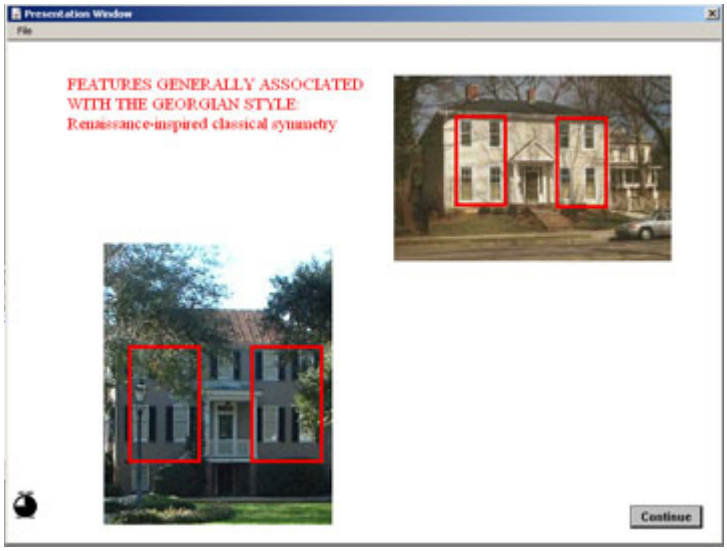


Figure 2
Screenshot of instruction

An assessment instrument was immediately delivered that evaluated participant's ability to correctly classify encountered, as well as un-encountered examples of the Georgian style Figure 3 and Figure 4). The instrument was delivered over the Internet. Data was collected and recorded to a database through a secure Internet connection.

Figure 3
Screenshot of assessment

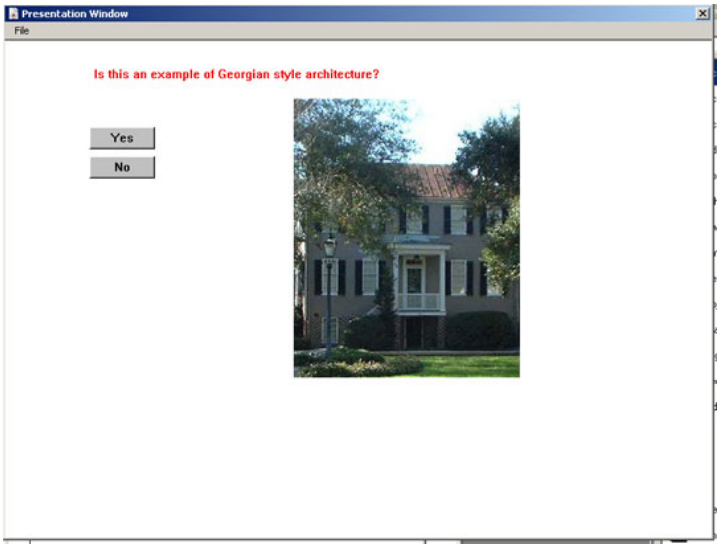
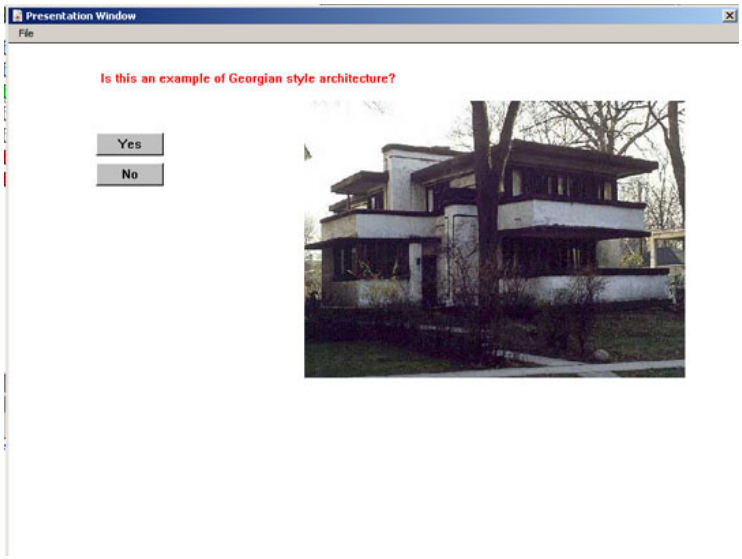


Figure 4
Screenshot of assessment



Procedure

The assessment involved a straight-forward presentation of encountered and non-encountered examples. Participants were asked to respond “yes” or “no” to the following question, “is this an example of the target concept in which you have just received instruction.” This study used an Internet delivered experimental treatment; participants were introduced to the experiment and asked to log onto the web-delivered module. The study used a, within-participants, after-only, experimental design (Christensen, 2001). All participants received the same instruction and same assessment items. The instruction consisted of five screens describing the primary characteristics of the Georgian Style. The screens where available for the participants for a total of three minutes, followed by an

assessment instrument. The assessment presented a picture of a building and asked participants whether it was an example of the Georgian style or not. The assessment was presented as a series of randomized examples from the test item categories as well as non-examples consisting of coordinate concepts.

Two hypotheses tested were 1) participants would correctly classify encountered examples better than un-encountered examples and 2) that participants would have a faster response time for examples that they had encountered compared to those that they had not encountered. It was expected that encountered examples would create a sequencing effect on the test scores and times (Christensen, 2001).

Results

Participant's data on the two categories of assessment were collected. The Mean Percent Correct data is displayed in Table 2 and the Mean Response Time data is displayed in Table 3.

Table 2
Mean Percent Correct of Test Items Evaluating Target Concept Attainment

Groups	*N*	*Mean*	*S.D.*
Encountered examples of the target concept	30	0.84	0.192
Un-encountered examples of the target concept	30	0.38	0.081

Table 3
Mean Response Time (Seconds) for Each Category

Groups	N	*Mean*	*S.D.*
Encountered examples of the target concept	30	5.427	2.932
Un-encountered examples of the target concept	30	5.503	3.734

Statistical tests used

The data was then submitted to a series of standard one-way ANOVA statistical tests to provide a basis for determining if the differences between the assessment item scores and times were significant. Uni-variate analyses were performed for each pairings.

The data from the ANOVA computations comparing correct answers of encountered examples of the target concept vs. un-encountered examples of the target concept resulted in significant effects $F(1,59) = 146.11$, $p < .05$; while ANOVA computations comparing response times did not demonstrate significant differences. As was expected participants answered correctly a high percentage of the time when presented with examples of the target concept encountered in instruction (M=.84). Results indicate that participants were better able to positively identify previously encountered examples than they were at positively identifying new examples of the target concept that had not been part of the instruction (M=.38). This result may indicate that participants gained only a tenuous grasp of the target concept. The results of response times are less conclusive; significant differences were not found between response times of encountered and un-encountered examples.

Discussion

The hypothesis that learners would perform better on encountered examples than with un-encountered examples due to the difference in the nature of the learner task is supported by the percent of correct scores reported. Participants, as expected, displayed a sequencing effect when responding to the assessment. However, participants clearly were unable to consistently correctly identify un-encountered examples of the target concept. The results confirm the theory that encountered examples represent a different cognitive task than un-encountered examples particularly when there is a minimal delay between instruction and assessment (Feldman & Klausmeier, 1974; Merrill, Tennyson, & Posey, 1992; Merrill, Olsen, & Coldeway, 1976). Participants scored significantly better on encountered items. Using encountered examples could provide assessors with a false sense of a learner's level of concept attainment.

However, contrary to other studies, response time doesn't appear to be significant indicator of concept attainment in this case. It was expected that participants would respond to encountered examples faster than they would un-encountered examples. The lack of significance may be attributable to the unfamiliarity of the assessment instrument or perhaps the artificialness of the content. However, instructional designers should be wary of making an inference between response times and concept attainment. Studies with a longer time between instruction and assessment may show more divergence in the respective response times.

The results of this study imply that instructional designers must be careful and vigilant when attempting to create congruency (Smith & Ragan, 2005) between objectives, instructional materials, and assessments. By reusing examples designers may be testing for unintended domains of learning i.e. verbal knowledge vs. conceptual knowledge (Gagne, 1973). Similar questions may be asked of test items in other learning domains such as principles and procedures. Additionally, research studies that extend the time between the instructional event and assessment may provide further data on the relationship between the uses of encountered vs. un-encountered examples.

References

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response time with multidimensional stimuli. Perception & Psychophysics, (55) 11-27.

Ausobel, D.P. (2000). The acquisition and retention of knowledge: A cognitive view. Kluwer Academic Publishers: Boston

Beer, A.L., & Diehl, V.A. (2001) The Role of Short-Term Memory in Semantic Priming. Journal of General Psychology, 128 (3), 329-350.

Bornstein, M. H., & Monroe, M.D. (1980). Chromatic information processing: Rate depends on stimulus location in the category a psychological complexity. Psychological Research, (42) 213-225

Christensen, L.B. (2001). Experimental Methodology, 8th Edition, Allyn and Bacon: Boston

Feldman, K.V. (1972). The effects of number of positive and negative instances, concept definition, and emphasis of relevant attributes on the attainment of mathematical concepts. Technical Report No. 243, Report from the Project on Conditions of School Learning and Instructional Strategies, Madison, WI

Feldman, K.V., & Klausmeier, H.J. (1974). The effects of two kinds of definition on the concept attainment of fourth and eighth grade students. Journal of Educational Research, (67) 219-223

Gagne, R. M. (1973). The conditions of learning. New York: Holt, Rinehart and Winston, Inc.

Hornke, L.F. (2005). Response time in computer-aided testing. Psychology Science, (47) 2, 280 - 293

Hunt, E. B. (1962). Concept learning: An information processing problem. John Wiley Sons: New York

Klausmeier, H.J., & Feldman, K.V. (1973). The effects of a definition and a varying number of examples and nonexamples on concept attainment. Technical Report No. 280, Report from the Project on Conditions of School Learning and Instructional Strategies, Madison, WI,

Litchfield, B.C. (1987). The effect of presentation sequence and generalization formulae on retention of coordinate and successive concepts and rules in computer-based instruction. (Doctoral dissertation, Florida State University, 1987). Dissertation Abstracts International 49, 486A

Luce, R.D. (1986). Response Times: Their role in inferring elementary mental organization. Oxford University Press: New York

Merrienboer, J. (1997). Training complex cognitive skills: A four-component instructional design model for technical training. Englewood Cliffs, New Jersey: Educational Technology Publications.

Merrill, M.D., & Tennyson, R.D. (1971). The effect of type of positive and negative examples on learning concepts in the classroom. Final Report, Project No. O-H-014, Contract No. OEG-8-70-0006 (509), U.S. Department of Health, Education, and Welfare

Merrill, M.D. (1983). Component Display Theory, In Reigeluth, C.M., Instructional-design theories and models: An overview of their current status, Lawrence Erlbaum Associates, Publishers, Hillsdale NJ.

Merrill, M.D., Olsen, J.B., & Coldeway, N.A. (1976). Research Support for the Instructional Strategy Diagnostic Profile, Technical Report No. 3, Courseware, Inc.: Orem UT.

Merrill, M.D., Tennyson, R.D., & Posey, L.A. (1992). Teaching concepts: An instructional guide, Englewood Cliffs, New Jersey: Educational Technology Publications

Penn, J.H., Nedeff, V.M., & Gozdzik, G. (2000). Organic Chemistry and the Internet: A New Approach to Homework and Testing Using the WE_LEARN System, Journal of Chemical Education, 77 (2), 227-31.

Rand, A. (1990). Introduction to objectivist philosophy, In H. Binswanger & L. Peikoff (Eds.), 2nd Edition, New York: Meridian

Rosch, E. (1998).Principles of categorization readings in cognitive science, a perspective from psychology and artificial intelligence. Allan Collins & Edward E. Smith, Morgan Kaufmann Publishers, San Mateo, California, 1988, pp 312-322.

Rosch, E. (1999). Reclaiming concepts. Journal of Consciousness Studies. (6)11-12, 61-77

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three dimensional objects. Science, 171, 701-703.

Smith, P.L., & Ragan, T.J. (2005). Instructional Design. John Wiley & Sons: Hoboken, NJ