Theory-Based Tool for Direct Evaluation of Educational Materials

Kelsey A. Lemmon

Ball State University

Department of Psychological Science
Abstract

The current study introduces a new direct evaluation method for discriminating the strengths and weaknesses of educational materials. Six areas or “standards” of good materials were identified in the literature and used to create a 33-item survey. The measure scores a variety of educational materials from textbooks to multimedia web-based tools on a scale of 1-25 in each of the six standards. Two experiments were conducted to determine the reliability and validity of the new measure. In experiment 1, a group of expert judges completed the evaluation for three different materials: a textbook, an e-book and a viziswap module. The scores were not reliable between judges but they were consistent with previous research indicating increased student exam scores and student attitude surveys. In experiment 2, a group of undergraduates completed the survey twice after a two-week delay. The results of experiment 2 indicated that the survey does not have test-retest reliability. The current measure needs further research to produce a reliable measure but is useful in identifying the strengths and weaknesses of educational materials.
Theory-Based Tool for Direct Evaluation of Educational Materials

Importance of Educational Materials

For many years, educational materials have been used to increase students’ acquisition and retention of information in academic settings. Research has demonstrated the advantages of using external representations in a variety of different learning situations. Teachers assigning a textbook, employees viewing a training video, or a child putting together a toy with directions are all examples of using knowledge representations to learn. Tversky & Lee (1999) discussed the utility of external representations as learning materials in terms of the possible information processing advantages. “Whereas human information processing is limited, both in number of items (memory) and in number of operations (processing), external representations are virtually unlimited” (Tversky & Lee, 1999). External representations allow people to have access to a great amount of information at once.

With the emergence of the WWW and the ever-expanding variety of innovative technologies available, new kinds of learning materials are constantly being created. The wide range of available materials often make it difficult for educators to discriminate good from bad. There is a need for new tools to evaluate emerging interactive, multimedia-learning materials for comparison to each other and to more traditional methods.

Evaluation of Educational Materials

Educators have used several different methods of evaluating educational materials. All methods have notable strengths and weaknesses that impact their
ability to measure and compare the value of materials. This review will summarize the current methods and discuss some of the different aspects that influence reliability and validity. It will also introduce a new tool that can be used to directly evaluate materials.

Chumley-Jones, Dobbie, and Alford (2002) conducted an empirical review of the different methods of evaluation used to compare computer-assisted instruction to other materials. The two most common methods of evaluating materials were studies evaluating knowledge gains (i.e. testing) and student attitudes (pg. S86). Other evaluations that reflect on the quality of educational materials include real-world outcomes and direct analysis.

**Testing.** One of the most logical ways to evaluate an educational material is by measuring the approximate amount of information that was learned with use. The rationale for this measure is that if one material leads to a higher amount of knowledge acquisition and retention then it has a higher benefit/cost ratio and can be considered more “effective”. Retention can be measured in many ways. Some of the methods often employed by educators include multiple choice tests, short answer questions, and essays.

The weakness of estimating the amount of knowledge gained as a measure for educational materials is that students often compensate when their needs are not met. Most students are motivated to achieve a certain minimum score in order to be satisfied with their performance. Regardless of the difficulty of the course or the effectiveness of the educational materials, they will often strive for the same minimum course grade. Consequently if an educational tool falls short, students
seek out their own better tools or seek out help from other people. Researchers often find it difficult to determine if the knowledge gain is the result of course materials or not.

**Student attitudes.** Student attitudes are another form of evaluation that helps to demonstrate the effectiveness of educational materials. Although students may not be experts in the area of study, they are usually aware of how difficult it was to understand the information. Students cannot always correctly evaluate the amount of information they are learning, but they can shed light on ease of use. Students usually have a clear memory of the times when searching a tool was tedious or frustrating. Frustration is an indication that the cost of using a tool is exceeding the benefit.

The biggest weakness of a student attitudes evaluation is that students can only tell us what they like or dislike. It is difficult to determine what student preference is actually indicating. Some research (Mayer, 2009; Boling & Robinson, 1999) suggests that educational materials students enjoy may not always be the best tools for the learning situation at hand. For example, students tend to enjoy videos in the classroom even if the video does a poor job teaching the topic. The enjoyment of using multimedia may be disproportionate to the amount of learning benefits experienced or may even distract the student from the lesson objectives. Evaluators can sometimes phrase questions so that they encourage responses that indicate the quality of educational materials. For example, asking a student to answer the question “Do you think the multimedia helped you learn compared to the typical text?” should provide more useful information than simply asking “Did
you like using the multimedia?"

Student evaluators may also be significantly influenced by personal bias against a course. A student could be angry due to a disagreement with the instructor or a simple lack of interest in the subject causing he or she to have a negative attitude towards every aspect of the course. The opinion of students is influenced by many factors that may or may not be related to the quality of the educational material used.

Real-world outcomes. The development of new skills from using educational materials is a real-world outcome that is predictive of effectiveness. Skills such as problem solving or typing can be tested after completion of a course for a more immediate understanding of what a student gained from a course. However, the development of the skill cannot necessarily be attributed to the educational material in question unless students are tested immediately before and after use.

Comparing the relative life success of individuals that used different educational materials can indicate the long-term effectiveness of materials over time. In general, students who use educational materials that induce a high level of understanding can more easily apply what they learn to real-world situations and should therefore be able to function more easily on a daily basis, get better jobs, and have more success in their careers. This type of evaluation is especially helpful in fields where students are expected to learn certain skills rather than accumulate a wide range of knowledge (e.g. artists, business, mechanics).

Using relative life success for evaluation has several weaknesses. Universities sometimes find it difficult to keep track of alumni extensively over time. After
graduation, universities may not consider tracking student success to be a beneficial investment. Even if educators are successful in tracking students after graduation, it is difficult to attribute an outcome with so many independent variables (SES, student personality, student experience, other courses, etc.) to the educational material in question. This type of evaluation is best at determining the strengths and weaknesses when the two groups of students have had significantly different forms of education overall. For example, a study, comparing Harvard graduates to community college graduates, would most likely indicate that those graduating from Harvard earned significantly greater salaries than the community college graduates. However, even in this case it is debatable whether the affect is due to personal background of individuals or education.

**Direct evaluation.** Direct evaluation involves comparing the body of information and method of presentation between different educational materials to determine how learning outcomes are affected. The most important advantage of direct evaluation is that it eliminates the middleman by getting information directly from the material rather than through the student. Evaluating through learning outcomes, attitudes, or real-world outcomes introduces uncontrollable variables relative to student differences. Student’s motivation, level of prior knowledge, and other external references can dramatically alter the outcome of an evaluation. No one learner is ever directly equivalent to another. By directly evaluating an educational material the number of variables that are introduced is minimized.

A limitation of using direct evaluation is that educators often disagree on what they believe constitutes good educational materials for a several reasons: they
disagree on what information is important or they disagree on which methods are most effective. While student differences become irrelevant, differences between expert opinions can cause problems. Bias on the part of the evaluator is a significant possibility. Each individual has a personal bias concerning what information is the most important to learn and how to best communicate that information to students. Research indicates that it is difficult to grade from expert opinion because such a substantial among of variability exists between individuals (Acton, Johnson, & Goldsmith, 1994; Ruiz-Primo & Shavelson, 1996, Williams, 1998). Comparing expert opinions on what information is important is unreliable. Comparing the opinions of educators unfamiliar with the theories and research on effective methods of teaching is problematic. Educators are often experts in their own field of study, but lack the information necessary to make judgments as to what methods positively influence learning.

The ability of direct analysis to assess the comparative value of educational materials is also limited. Many direct analysis evaluations simply label a material “good” or “bad”. This kind of label has limited value in a world where thousands of different kinds of materials are available. To make direct analysis of educational materials useful, it is necessary to have a comparative basis for relating the tool to other such materials. A tool with the ability to score materials on an ordinal or ratio scale is needed to allow for easy comparison to others.

Considering many different components separately in order to make a conclusion about the how the material as a whole contributes to acquisition and retention of knowledge allows for a more complete evaluation. What constitutes
good educational materials falls into several different overlapping categories that can be difficult to distinguish. Mayer (2009) and Clark & Mayer (2008) used a check sheet based on laboratory research of multimedia. The direct evaluation provides a good example of evaluating multiple aspects, but it is limited to use with multimedia. The method also does not take into account several important categories such as feedback.

**Direct evaluation in the current study.** The current study introduces a new method of direct evaluation that addresses these problems in the following way. First the evaluators are not asked to make judgments about what information should or should not be included; they will only indicate what is present and evaluate quality. Second, Evaluators will be using a consistent “set of standards” that is supported with numerous methodologically sound theories and empirical studies to evaluate each material. Each evaluator will independently indicate to what degree the material met the criteria specified and will not be asked if he or she believes the method was appropriate.

In order to allow for a comparative analysis, the evaluation will score each material on a scale ranging from 1-25. The ordinal scale rating can be used to compare the effectiveness of educational materials overall or within one of the standard areas. The range of the scale creates a relatively wide amount of possible variability compared to most direct evaluations.

The method of direct evaluation considers six areas or standards that research indicates are of importance in education. The research was used to make some basic assumptions about what constitutes the best possible educational materials.
The standards will cover three issues of content and three issues of usability. Content issues assess the quality of the information presented. Usability issues assess the basic “user-friendliness” of the tool, which allows for better access to the information a student is trying to learn. The “set of standards” will contain the following six sections.

**Appropriate Mode of Communication**

For an educational material to be effective, it is necessary to determine the best possible way to communicate information clearly and efficiently to the student attempting to learn that information. Verbal, picture, video and non-verbal audio are modes of communication that have different strengths and weaknesses in academic settings. Goodhue and Thompson’s task technology fit model states that, “Task technology fit is the degree to which a technology assists an individual in performing his or her portfolio of tasks. More specifically, TTF is the correspondence between task requirements, individual abilities, and the functionality of the technology” (Goodhue & Thompson, 1995). Determining which form of communication (technology) is best for the task at hand should be decided based on which technology/task combination produces the best learning outcome.

Finding the appropriate mode to communicate the desired information to a student depends on several different factors: the type of information being communicated, the level of experience the student has had previously, and the kind of knowledge or skill the student is expected to come away with. Differences in these areas make different modes of communication more suitable for use in any given situation. For example a student who is taking an introductory college course
with the goal of getting a broad understanding of the topic might benefit from a
different form of communication than a graduate student in an advanced course
trying to understand complex and specific theories. Successful communication of
ideas is dependent on identifying the type of knowledge, level of prior experience,
and goals of the mode of communication most suited to the context of the learning.

**Verbal.** Verbal information is defined as any linguistic representation of
knowledge including text and oral narration. Verbal information has traditionally
been the most common form of communication used in education. Once a student
has mastered the basics in an area and has a strong mental model of concepts, tools
such as video and picture become less useful. Verbal instruction is best suited for
students that are analyzing the complex details or debating theoretical topics that
cannot be explained through real-world examples. Freeman (1924) comments “The
contrast which is drawn here is between concrete experience on one hand, and
comparing, analyzing, and generalizing operations on the other hand. The
contention is that these latter ways of working over experience and of converting
raw experience into thought are very much facilitated by language” (pg. 69). Visual
information is limited in its ability to elaborate on complex topics.

For example in learning models of cognitive theory, a diagram outlying each
cOMPONENT OF THE THEORY MAY BE HELPFUL TO A LEARNER UNFAMILIAR WITH COGNITION. IF AN ADVANCED STUDENT WITH A CLEAR MENTAL MODEL OF COGNITIVE THEORY IS ASKED TO THEORETICALLY EXAMINE THE DIFFERENT STRENGTHS AND WEAKNESSES OF A MODEL, A LINGUISTIC REPRESENTATION IS PREFERABLE. ALTHOUGH EMPIRICAL RESEARCH IN THIS AREA HAS BEEN SOMEWHAT LACKING IN DEMONSTRATING THE RELATIONSHIP, “SOME OF THE MOST
distinguished contemporary psychologists... Lev Vygotsky, A. R. Luria, and Jerome Burner, for example, have all pointed out that higher cognitive functions such as analysis and synthesis seem to develop most fully only with the support system of verbal language" (Emig, 1977). Communicating information through language is most effective when instructing students with a high level of prior knowledge in learning theoretically advanced ideas.

Basically every educational material used in academia includes verbal information of some kind. Consequently, most make fairly good use of verbal information and generally use it appropriately. The biggest problem encountered is overuse of verbal communication when another form would be more relevant and appropriate. Research has indicated that in several areas picture, video, or non-verbal audio information is more helpful to learners than verbal. Thus a general principle is “what doesn’t fit into the categories of being best taught through picture, video, and non-verbal audio, is best expressed through language.” Such judgment requires an understanding of the appropriate use of pictures, video, and other non-verbal media.

**Picture.** Pictures, graphics, or diagrams are defined as any single visual representation of knowledge not containing motion that involves more than simple text. The strength of pictures is that they are particularly helpful to learners when the subject being learned requires specific spatial information for understanding. For example, a surgical student learning the specific parts of the body would be able to learn the location of organs within the chest cavity much faster from a diagram.
that illustrates the location then from a list of organs describing where the location is verbally.

Pictures also help students learn hands-on applied tasks rather than memorizing certain facts or concepts. Hands on applied tasks often require a learner to have specific spatial information in order to complete the task. Brunye, Taylor, & Rapp (2008) conducted a series of experiments examining the affect of a dual coding on students learning to assemble a toy. Fifty-two college students received partial course credit for completed order verification, recall, and object assembly tasks. The assembly task was taught through text only, picture only, or multimedia (text and picture). Results of this experiment suggest that multimedia has an advantage over single format in order verification and recall tasks or tasks that required memory. However, the results of the object assembly task consistently demonstrated the utility of picture learning, with or without accompanying text (Brunye et. al., 2008). Results of this study suggest that dual coding in multimedia demonstrations may lead to better memory. However, in actually completing the task (less reliance on memory), a picture was sufficient to express the specific, spatial information needed to complete the task.

Najjar (1996) conducted a meta-analysis concerning the use of multimedia to further educational outcomes. One study done by Bartram (1980) was an investigation into teaching students spatial information through text and picture. College students were asked to get from a starting point to an end destination using a minimum number of buses. Researchers presented the participants with either a description of different bus routes or a map of different bus routes. Students who
received the maps learned the bus routes and got to their destination more quickly than students given descriptions. The data suggests that spatial representations are the preferred representation when the information requires spatial knowledge (Najjar, 1996).

**Video.** Video or animated graphics are defined as any non-static collection of pictures including slide shows that imply motion and time. Video communication can provide students with multiple spatial representations of concepts and how they change over time. It has the ability to easily manipulate time and space in order to teach concepts where real life examples cannot be observed. For example, a cell is too small to see with the naked eye and cellular respiration is a process that is too fast for a human being to observe. However, through video it is possible to communicate this concept using a model that is large enough to be seen and a process slow enough to be understood. This ability makes video useful to learners in several different situations.

Video is the most helpful mode of communication to create a context for new knowledge. Human beings often rely on our keen sense of vision to help us understand context. Frederick Freeman (1924) comments that "(When) pupil's... lack some of the concrete experience which was essential to full comprehension of the instruction, this experience was in a measure supplied by the various types of visual material which were employed" (pg. 27). Karppinen (2005) compiled a list of theoretical prospective on meaningful learning using digital and online videos. Many educators have stressed that meaningful learning needs to be situated or anchored in authentic, relevant, and/or realistic contexts. Visual education helps when
information is representing and simulating meaningful real world situations, problems, or contexts and when representing the beliefs, perspectives and stories of others (Karppinen, 2005).

Picture information can sometimes provide a context in which learners can construct their own knowledge. However, the concepts being learned must usually be visually simple or stagnant over time in order for students to gain context from a picture. Earlier I gave the example of a surgeon studying a diagram of the chest cavity, which represents something that will remain fairly consistent over time. If the surgeon were learning the process of performing a surgery, this learning would be best supported by a video or slide show that can demonstrate spatial movement over time.

Video has a positive effect on student motivation. The Cognition and Technology Group at Vanderbilt (1990) conducted an experiment investigating their anchored instruction model and the situated cognition model (Brown, Collins, & Duguid, 1989) when comparing students with video instruction and students without video instruction. Fifth grade students were placed into two groups. The experimental group learned language and social studies skills through instruction anchored in explorations of the movie The Young Sherlock Holmes or Oliver Twist. The control group received daily instruction of the same concepts using several different micro contexts (e.g. story A, story B). The participants who received instruction rooted in the rich context of one of the videos were more motivated to learn and seemed to have more ambitious learning goals for themselves than those who did not view the videos. The data also indicated that students instructed
through video spontaneously used more of the lessons target vocabulary words when it became relevant (The Cognition and Technology Group at Vanderbilt, 1990). In other words, students had a better understanding of the vocabulary the instruction was designed to teach, demonstrated by their ability to apply the words in real-world situations.

Video is also important when students are required to make a judgment about a concept rather than simply recalling facts. Concepts that involve students forming a personal opinion are important in all subjects especially in areas like art or literature. When forming an opinion, it is important that the student is provided with unbiased information that does not lead the reader to form a particular conclusion. Verbal information often provides a biased account of events. Students feel more comfortable expressing an opinion about a situation they can see for themselves rather than someone else’s account of the situation.

Lim & Benbasat (2000) studied the amount of confidence students had in the answer they gave to a certain question when they learned the information through text or video. Eighty students with an average age of 25 were asked to complete 14 tasks. Seven of the tasks were classified as analyzable tasks that required the student to recall facts and seven were classified as less-analyzable tasks that required the student to make judgment about answer. The participants were divided equally into two groups, learning through video and learning through text. Researchers measured the perceived equivocality, (confusion or lack of understanding) for each task on a 5 point Likert scale. Results showed that the equivocality of analyzable tasks was not different between the two different
learning materials. However, the perceived equivocality of less-analyzable tasks learned through video was significantly different than the perceived equivocality of these tasks learned through text. Researchers concluded that learning through video reduces confusion when the student is asked to give his or her opinion (Lim & Benbasat 2000).

**Non-verbal audio.** Non-verbal audio is defined as any audio representation of knowledge that does not include linguistic information. Examples of non-verbal audio would include music and sound effects. This type of communication is most important in certain fields where recognizing or understanding certain non-verbal audio is essential. For example, a student studying music, bird song, or learning to recognize auditory cues of some kind can learn more efficiently when these stimuli are available.

**Dual Coding**

When a student is trying to learn information that requires him or her to recall the information at a later time, the process is greatly helped by identifying multiple connections between the novel information and preexisting conceptual representations. The more connections a person makes between the new idea and prior knowledge, the more routes exist to retrieve that information when it is needed later. Dual Coding Theory states that, “Two classes of phenomena are handled cognitively by separate subsystems, one specialized for the representation and processing of information concerning nonverbal objects and events, the other specialized for dealing with language... They are independent in the sense that either system can be active without the other or both can be active in parallel”
Human beings have the ability to simultaneously process visual and verbal information. When visual and verbal information is encoded in the brain the possible connections a student can make between old material and novel material is greatly increased.

Gellevij, Meij, Jong & Pieters (2002) investigated the difference in knowledge retention of students using either multimodal (picture and text) instruction or unimodal (text only) instruction to learn complex software application. The researchers hypothesized that due to Dual Coding Theory the students who received the multimodal instruction would have a stronger mental model of the program than those participants only receiving unimodal instruction. The participants consisted of 44 teacher education students chosen for their level of knowledge in physics, basic computer skills, and of the program being used. Each participant completed three chapters of a manual (either multimodal or unimodal) in two different sessions. After each chapter participants completed a test. Researchers concluded that the multimodal instruction led to participants creating a significantly better mental model than unimodal instruction. Participants in the multimodal group also had significantly better accuracy and speed when identifying window elements and objects. A weakness in this study is that no latent measure of retention was administered. It is hard to conclude anything about the long-term effects of multimodal instruction based purely on these results (Gellevij et. al., 2002).

Another subject of interest in dual-coding research is whether or not all dual coding methods have an equal effect on learning. Mayer (2009) discusses the modality principle of multimedia learning that states, “People learn more deeply
from pictures and spoken word (narration) than from pictures and printed words.”

In 17 out of 17 laboratory tests performed by Mayer and his colleges, students
performed better on problem-solving transfer tasks when animation (video) or
graphics (picture) was presented with narration rather than on screen text (Mayer,
2009). When information can be divided between auditory and visual systems, the
cognitive load associated with learning the information is significantly reduced.

Brunken, Steinbacher, Plass, & Leutner (2002) assessed the level of cognitive
load that is associated with learning information from either simultaneously
presented text and related pictures and an audiovisual presentation. Ten female
psychology students with an average age of 22.8 years participated in the study.
Each completed an acquisition phase followed by two tasks. In the acquisition
phase, the participants were instructed on how the human cardiovascular system
works through a program that presented the information in two formats,
audiovisual (verbal narrative) and visual only (verbal information provided in text
on the screen). The amount of time the participants were given to view the program
was standardized for both formats. The first task given to the participants was an
assessment of the knowledge acquisition about the information learned in the
program. The secondary task was done simultaneously with the first task and
involved pressing the space bar any time that the participant noticed a letter
changing colors on the screen. The second task is considered to be a measure of the
cognitive load answering the first questionnaire requires. The amount of time it
took each participant to notice the color change was considered to be representative
of how heavy the cognitive load of the first task was. The results showed that
reaction time for the second task was significantly lower when the participant had received the audiovisual format of instruction. Participants that received the necessary information through both auditory and visual channels were able to retain more information (Brunken et. al., 2002).

Feedback

Whether or not a tool is capable of providing helpful and appropriate feedback is an important factor in the effectiveness of the educational material. Not all forms of feedback work equally well in different situations. Much like the modes of communication, it is important to determine what type of feedback is best depending on the information being learned and the ultimate goal of the lesson. Mason & Burning (2001) comment that, “Consistent with earlier research on feedback in conventional learning settings, the findings on feedback in CBI indicate that there is no clear-cut ‘best’ type of feedback in computer based instruction for all learners and learning outcomes. The challenge therefore is to identify the type of feedback, which will assist in correcting initial errors in understanding and help prevent inaccurate information from being encoded.” One of the greatest strengths of interactive multimedia educational tools is their ability to personalize feedback to each student’s individual learning experience. Although appropriate feedback depends on the individual student and the individual learning situation, there is evidence that some types of feedback are generally more helpful to the learner than others.

Immediate vs. delayed feedback. There are several different aspects of feedback that can be manipulated. One of those aspects is the amount of time
between the submittal of information from the student and when they receive the feedback. In most situations, experts agree that immediate feedback increases the acquisition and retention of information more than delayed feedback. It is easier for a student to apply the feedback to the information being learned when the time between submission and feedback is minimized. Mason & Burning (2001) comment, “Immediate feedback... will assist in correcting initial errors in understanding and help prevent inaccurate information from being encoded” (Mason & Burning, 2001).

Dihoff, Brosvic, & Epstein (2004) investigated the relationship of feedback timing to determine the affect it has on learning. Eighty undergraduate students were randomly assigned to four conditions: (1) control condition received no feedback, (2) end of test condition received correct solutions after exam, (3) delayed feedback condition received correct solutions the next day, (4) immediate feedback condition received correct solutions immediately upon reply. A latin square design was used to counterbalance response format order. Results suggested that immediate rather than delayed feedback results in the greatest increases in retention, confidence, and the ability to identify-incorrect and correct responses (Dihoff, et. al., 2004).

Berry (2009) conducted an experiment investigating the effect of immediate feedback in senior nursing students enrolled in ‘Nursing: Children and Families with Health Deviations’. Historically, the course had a lot of trouble engaging students in course material. Sixty-five students participated. The clicker group was given five multiple-choice questions to answer with the clickers during class and then feedback consisted of the professor discussing the answer and incorrect options
immediately. In the control group, the students were given a take-home essay quiz to be completed previously to each class meeting that they did not receive immediate feedback on. The results showed that clickers significantly increased the amount of student involvement in the course material. Students also scored significantly higher with the use of clickers on exam 2 and the overall course grade (Berry, 2009). The specificities for enrolling in this nursing course suggest that immediate feedback is affective in a specific area where students have a wide background of knowledge in the area. Other studies have indicated that this is also true in introductory courses (Morling, McAuliffe, Cohen, & DiLorenzo, 2008).

**Elaborated vs. simple feedback.** Another aspect that has caused some debate among experts is the actual content of the feedback provided. Deciding the correct amount of information to include in feedback to a student is somewhat of a balancing act. Elaborated feedback should include a clear and concise summary of how the work can be improved. Shute (2008) reviewed the corpus of research that has been conducted examining the relationship between feedback and different learning outcomes. Shute asserts the view that effective feedback provides two types of information: verification and elaboration.

Verification is defined as a simple statement that reflects the “correctness” of an answer. Epstein, Lazarus, Calvano, Matthews, Hendel, Epstein, and Brosvic (2002) conducted a study using a test-retest method with 70 Introduction to Psychology undergraduate students. Participants were randomly assigned to complete a multiple choice trivia test with either a Scantron form or an immediate feedback assessment technique (IFAT) indicating if the answer picked was correct or
incorrect. The participants were than randomly assigned to complete a retest after a delay of either one day or one week. The mean scores of participants that used the IFAT format were significantly higher than those participants who used the Scantron. The results suggest that giving students immediate feedback as to the “correctness” of the answer provided helps students to increase retention of important information (Epstein, et. al., 2002). Immediate verification prevent inaccurate encoding from taking place causing increased the probability of the student remembering the correct answer at a later date.

The elaboration of feedback is more complicated because it is always a balancing act to decide what is appropriate to include and what is not. Elaboration may address the topic, address the response, discuss particular errors, provide examples, or give gentle guidance. In general, “response-specific feedback appears to enhance student achievement, especially learning efficiency, more than other types of feedback, such as simple verification or ‘answer until correct’” (Shute, 2008, pg. 159). However, feedback that provides too much detail or throws too much information at a student simultaneously may discourage a student and make them feel overwhelmed. It is important to provide the appropriate amount of information to increase understanding without burdening the student with extraneous details.

Meyer, Wijekumar, Middlemiss, Higle, & Lei (2010) investigated the use of elaborated feedback to increase reading comprehension. One hundred eleven fifth and seventh graders were instructed on how to structure different writing assignments using intelligent tutoring of the structural strategy (ITSS). Participants were randomly assigned to the ITSS condition, received tutoring in the form of
instructional emails including multiple aspects such as feedback on last lesson, encouragement, daily assignments, and additional instruction, or simply completed the regular school reading program. ITSS was further divided into elaborated, advanced tutor responses or simple responses about the accuracy of student work. An example of an elaborated feedback response might be “Your structure, main idea, and details are correct. Great job! But your signaling words were incorrect. Using the chart as your guide, rewrite the signaling words” (Meyer et al., 2010, Pg. 69). In the simple feedback condition, the tutor responded “good job” for recalling 60% of a text and “try again” if less than 60% was correct (Meyer et al., 2010).

Meyer et al. (2010) gave students three tests: a pretest, a posttest, and a four-month delayed posttest of the material. Results indicated that students who received ITSS with elaborated feedback substantially increased the relative level of reading comprehension in both the immediate posttest and the four month delayed posttest. Elaborated feedback appears to benefit both below-grade-level readers and readers with stronger reading skills although the effect is less substantial in readers who already scored high on the pretest. The results of this study indicate the importance of providing feedback that includes verification of the “correctness” of an answer as well as providing elaboration that guides the learner in improving his or her future responses (Meyer et al., 2010).

**Structure**

As discussed earlier, in order for meaningful learning to take place a student needs to have the ability to place information within a context. The educational materials that are used in classrooms play an important role in this process. It is the
responsibility of the creators of educational materials to provide tools that structure information and concepts in a logical way to promote learning. Students learning novel information need an experienced individual to provide guidance for understanding and organizing the concepts that are most important. Interactive multimedia that is structured appropriately has the ability to take the place of that “experienced individual” to some extent. Several ways that educational materials can provide this guidance is by providing objectives/directions and organizing information within materials by related concepts.

Guidelines. Most educators agree that it is important to provide learners with guidelines that structure novel information in a meaningful way. The objectives or directions for use of an educational tool provide students with an endpoint to work towards. Objectives need to specifically address two issues; the specific concept that the educational material is attempting to teach and the way in which the concept being learned will aid in the eventual completion of the course objectives. Nilson (1996) comments that objectives should be detailed and clear. Stating the specific topics that the student is meant to learn from the material not only gives them a goal to work towards, but guides them from concept to concept when learning new information. McKeachie, Chism, Menges, Syinicki, & Weinstein (1994) state that objectives should facilitate student learning that is appropriate to the course setting. In order to motivate students to complete individual assignments, it is important that they understand the role the single assignment plays in the eventual completion of the course.
**Context.** Providing a context for novel information greatly increases the probability that a student will be able to retrieve the information from memory when needed. Students need to make connections between novel information and previous knowledge in order to comprehend a topic. Structuring educational materials to provide context for information will help guide students to make these connections.

Bradford & Johnson (1972) completed a series of experiments that investigated the effect of providing the context of information on comprehension and recall. Fifty participants were placed into five different groups; no context (heard passage once), no context (heard passage twice), context after (heard passage, than saw picture), partial context (saw a mixed up picture and heard passage), and context before (saw appropriate picture and heard passage). Participants completed an acquisition phase where they learned the information in the way specified through their group and then performed a comprehension task and a recall task. Results confirmed the hypothesis of the researchers in that the group given the context before the passage performed significantly better than both no context groups in both the comprehension and recall tasks. These results indicate that there is a significant benefit to learning information that is in context. Educational materials have the potential to provide the context of information depending on how they are designed (Bradford & Johnson, 1972).

Maguire, Frith, and Morris (1999) conducted an experiment that used PET scans to image the brains of people asked to remember and comprehend stories with varying levels of context. Thirteen males ranging in age from 25-43 years with
a mean age of 31.9 years participated in the study. Two sets of independent variables were used to create six conditions. The first independent variable was the type of story: standard or unusual. The second independent variable was the picture shown before the story: no picture, irrelevant picture, or relevant picture. The six conditions consisted of two different picture/story combinations: (C1) no picture, unusual story, irrelevant picture, unusual story, (C2) irrelevant picture, unusual story, irrelevant picture, unusual story, (C3) irrelevant picture, unusual story, relevant picture, unusual story, (C4) relevant picture, unusual story, relevant picture, unusual story, (C5) No picture, standard story, no picture standard story, and (C6) relevant picture standard story, relevant picture standard story. During each story the participants were given a PET scan. Immediately following the PET scan participants were asked to rate their level of comprehension on a scale from one to seven. Memory was scored by counting the total number of “idea units” recalled for each story when participants were asked to recall out loud as much of the story as possible (Maguire et. al., 1999).

The results of the Maguire et. al.’s (1999) comprehension and memory tests consistently indicated that those participants who got the standard story outperformed those who got the unusual study and those who got a relevant picture outperformed those who got no picture or an irrelevant picture. C3 reported the biggest difference between the comprehension scores after story one and the comprehension score after story two. Results indicate that unusual stories can still be understood as long as a context is provided to aide the reader. The PET scans indicated that different areas of the brain were activated when the student was
provided with a mental framework (context) and when stories were repeated (Maguire et. al., 1999). Processing of information when an educational material provides context is a fundamentally different mental operation than processing information without context. Context also improves recall and overall understanding of content.

**Organization.** The best way for an educational material to put information into the appropriate context is by organizing that information in a logical or meaningful way. McKeachie et. al. states that, “People can learn and remember much more when their learning fits into an organization... The important thing is that students find some way of structuring the material. Students with more background and ability can do this even in relatively unstructured situations, but in courses where the material is new to the students it is probably important that the teacher provide ways of organizing material” (pg. 280-281) Organizing materials by grouping related concepts together helps students to put the information they are attempting to learn into the appropriate context.

Bower, Clark, Lesgold, & Winzenz (1969) investigated the affect of organization on recall. Participants were given a list of words to memorize that were either in random order, alphabetical order, or hierarchically organized. Each participant completed 4 rounds of an acquisition phase where they were asked to memorize a list of words and a recall phase where they were asked to remember as many words as possible. Results concluded that participants who received the lists that were organized hierarchically were able to recall 2-3 times the number or words as those participants who received the random and alphabetically organized
lists (Bower, et. al., 1969). Grouping concepts and information with others that are related helps students to put that information into context and consequently improves learning. Organization of material within educational tools can have a profound effect on the amount of information a student can remember and helps create a mental model to build future information on.

Richard Mayer (2009) discusses several important considerations of organization that are important to learning. The two principles deal with the contiguity or flow of the organization. The spatial contiguity principles states “Students learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen” and the temporal contiguity principle states “Students learn better when corresponding words and pictures are presented simultaneously rather than successively” (Mayer, 2009). Both principles suggest that information should be organized near by and at the same time as other information concerning the same or related topics.

Mayer (2009) also discusses another consideration when organizing material is the use of headings or key concept titles within a material to signal to the reader what information is most important. Mayer’s signaling principle states “people learn better when cues that highlight the organization of essential materials are added”. In other words labeling sections in which different key topics are of most importance helps students to associate and organize the information he or she just learned to the main topic it is related to. In five out of six laboratory tests participants performed better on transfer tests when they received multimedia that used signaling. The experiments also seemed to suggest that when used sparingly, when
readers have low reading skills, and when information is disorganized (Mayer, 2009).

**Addresses Individual Differences**

The amount of individual differences within a sample of random people makes generalizing any experiment to the entire population difficult. A certain amount of variation exists within every classroom where all students use the same educational materials. It is nearly impossible for any teacher or professor to adapt a lesson around each individual student’s ability and background due to constraints of time and resources. Most experts agree, “Learning and studying are always influenced by students’ prior knowledge, conceptions, and interests” (Karppinen, 2005). Interactive multimedia materials have the potential to be much more flexible and adaptive than traditional materials.

Multimedia interfaces have the ability to record and interpret certain aspects of a student’s performance instantly, such as response time and percent of correct answers. Always having up to date information in these areas creates the opportunity for multimedia materials to direct students to information they may be struggling with while still giving them the option to override suggestions. If interactive multimedia can accurately assess a student’s prior knowledge, than it would be able to accommodate to his or her needs. Though the machine will probably never be as beneficial as a human tutor, it can provide much-needed individual guidance to students in a large classroom setting when professors do not have time to modify the learning experience for each student.
Adaptive navigational support is a technology that accommodates the linearity of the information within a material to the individual user. Brusilovsky (2000) provided a brief overview of adaptive navigational support in multimedia educational tools. The goal of the adaptive navigational support is to adapt link presentation and functionality to the goals, knowledge, and other characteristics of an individual user. Adaptive navigational support can be implemented in one of the following ways; direct guidance, link sorting, link annotation, and link hiding, disabling, and removal. Direct guidance refers to the system visually outlining a link on the page as the “next best” page. Link sorting refers to a system that sorts the links of a page according to the user model and displays the links in order from best to worst. Link annotation refers to when a link is created from a word, phrase, or picture that provide further description related to that topic. Link hiding, disabling and removal refers when a system visually removes links that are not helpful to the user at that time (Brusilovsky, 2000).

Empirical research in this area has yielded mixed results on the use of adaptive navigational support in education. Adaptive navigational support is still in the early stages of development and many experiments not supported the utility of addressing individual differences. Brusilovsky & Eklund (1998) conducted a study that investigated the use of link annotation in educational hypermedia. Twenty-five teacher education students were recruited from the University of Technology in Sydney Australia to participate in the study. Participants learned two chapters on databases and spreadsheets using a multimedia program with or without adaptive annotated links. The study yielded mixed results. Originally the participants that
used the program containing the adaptive annotated links performed worse on the subject tests. However, results also suggested that the program was not used correctly by most of the participants. Students were not using the suggested links provided to them and it was suggested that the added complexity of the annotated links increased the cognitive load and distracted from the content when users did not effectively utilize the links suggested (Brusilovsky & Eklund, 1998).

Experts continue to emphasize the need for further research in this area. Many of the problems are thought to be the result of poor design and novice participants. Research that investigates the components of successful adaptive navigational support is needed to create better systems. Plowman (1996) argued that the way in which interactive multimedia transforms traditional narrative structures affects student comprehension and cognition. The researcher suggests that disturbing the linearity of a narrative line by allowing it to be suspended or altered at different points is likely to increase the initial cognitive demand for the user. However, it is possible to diminish the negative effects of highly fractured narration with some simple guidelines. Plowman concluded that machine interaction should be simple, any tasks should be short discrete units which arise logically from the narrative, the balance of different media components should be considered (avoid overreliance on text), and visual and auditory continuity should be promoted with links between elements (Plowman, 1996).

Another possibility for multimedia materials to address individual differences between students is the ability to allow students to allot different amounts of time to different topics. The segmenting principle introduced by Mayer
(2009) states that “People learn better when a multimedia message is presented in user-paced segments rather than as a continuous unit”. Students who are able to navigate their own path through multimedia software can spend more time with the topics they find difficult to understand and skim over topics where they already have a good understanding. Allowing a student as much time as he or she needs to consider a slide, picture, video, or page gives a student who has not quite been able to form a cognitive model from the information more time to do so.

Carincross & Mannion (2001) analyzed different approaches that could be used to create a more effective interactive multimedia environment. One aspect of multimedia they discussed was delivery control (allowing the user to direct the pace of his or her own learning). “Non-linearity... allows a user greater navigational control and freedom. Using the flexibility offered... allow(es) the learner to tailor both the presentation of information and access to this, depending on their needs” (pg. 163). As long as some source of organization or structure is provided, allowing students to control the pace of progression through the learning environment can provide greater flexibility to focus on important materials (Carincross & Mannion, 2001).

One of the reasons textbooks are so popular in academia is that they give the user delivery control. The learner has complete control over time spent on certain topics and repetition of information not yet understood. Students who do not understand a topic can reread a section as many times as necessary until they achieve insight. The problem that can arise is when students don’t know what topics they need to allot extra time to because they are unaware of the critical pieces of
information they are missing. The strength of interactive multimedia is that it may be able to guide the user in identifying the areas where further attention is needed.

Another possible benefit of interactive multimedia is the option to add special features that can adapt to ESL users and users with disabilities. Multimedia can easily include audio or text options in multiple languages. It can also include subtitles for hearing impaired students or voice activation for students with impaired fine motor skills. Traditional methods such as textbooks and lectures can adapt in some ways by providing note takers or printing versions of a textbook in multiple languages. However, the adaptive nature of interactive multimedia makes providing these options easier.

From a strictly empirical standpoint, the benefit of designing interactive multimedia materials to adapt the presentation of information to the individual user is currently inconclusive. More research is needed to determine the potential of materials that include these capabilities, as well as a carefully laid out organization and clearly defined objectives, so that the student has some control but information is still presented in a structured learning environment. From a cognitive prospective, designing materials to address individual differences such as focusing on problem areas and allowing repetition of information not yet understood should encourage greater acquisition and retention of course materials. Past research has indicated that having a human tutor implement these strategies does help students learn. The focus of future research in this area should be to design a material that can address individual differences in a way that benefits the user.

**Interactivity**
One problem that can arise when using traditional forms of education such as lecture and textbooks is that they encourage students to take a passive role in the learning experience. Because we cannot see what is going on inside a student’s mind, it is difficult to determine whether or not that student is actively engaging in the learning process, that is, really analyzing, synthesizing, or engaging in other higher level cognitive processes. McKeachie (1994) comments on the utility of active learning stating “Active learning works not only because it helps motivation and feedback but also because active learners are more likely to be attentive and to be thinking about the topic, relating new knowledge to previous learning, and elaborating the implications of what they have learned”. Interactivity refers to the ability of an educational material to encourage students to take an active role in learning. Many multimedia materials can be specifically designed to require interaction that is conducive to active learning.

One mistake that is often made in designing and creating interactive materials is the assumption that turning a page or clicking to a new slide creates an interactive environment for learning. Unfortunately creating interactive learning materials is not that simple. Interactivity only increases retention and acquisition of knowledge when the task or input requires the student to perform some kind of cognitive function for completion. Mayer (2009) states “Meaningful learning outcomes depend on cognitive activity of the learner during learning rather than the learner’s behavioral activity during learning” (pg. 3). It does not apply to functions that can potentially become an automatic process because these actions easily turn into a passive task and do not require any higher level cognitive processing.
Interactive features include things such as answer till correct feedback, links to helpful information, completing activities, and directing pace of information. Some of the previous principles in this review, such as feedback and addressing individual differences, contribute to the level of interactivity. If interactivity within a material can fall into one of the other standards than what is the point of giving the material an interactivity score? This question is a valid one in that this review attempts to separate the different components of educational materials and measure them in order to make a judgment about the whole. The current study separates interactivity from other categories for several reasons: first because the goal of interactivity is different than other standards and second because interactivity often involves other features of the material that do not fall under any of the other categories.

The goal of feedback and addressing individual differences are in general terms to prevent inaccurate encoding of information and to providing individualized support for students with different needs. The goal of interactivity is to encourage as much higher-order thinking as possible so that students take an active role in the learning process. Karppinen (2000) summarizes some of the potential channels for creating these educational materials, “Interactive multimedia resources can allow pupils to direct their own pace of learning, input their own views and ideas interact with other learners across the globe, revisit learning points easily and create their own multimedia notebook for future use” (pg. 235). Multimedia tools can be designed to compel a student to take an active role in his or her own learning producing a better learning outcome.
Schaffer & Hannafin (1986) completed a study investigating the affect of progressively interactive multimedia learning tools on recall. Ninety-eight high school students were asked to learn material through a video only, video-plus-questionnaire, video-plus-question-plus feedback, and full interactivity (feedback included branches to video sections covering topics where student responses indicated inaccurate learning). Participants than immediately completed a recall test. Results of this study indicated a significant relationship between the level of interactivity and the amount of information a student retained. A predictable recall pattern was found in that a higher level of interactivity predicted a higher level of recall. Researchers indicated that it was possible the increased level of recall was actually due to the increased time spent with the items rather than the interactivity itself. However, time constraints aside, a significant relationship exists between the level of interactivity and performance on the recall task (Schaffer & Hannafin).

Increasing interaction with learning materials to help students pay attention to information has become a serious concern as more and more students are getting degrees through the WWW or are enrolled in large lecture courses with little human interaction. Students in these settings are even more likely to take a passive role in their learning. Computer activities such as games (which learning materials can often resemble) and web surfing are often passive activities. Increasing student motivation in these settings is important because it is easy for a student to take a passive role in their education when they are not in a socially engaging classroom. Boling and Robinson (1999) investigated motivation and test scores of distance education students instructed through individual study, cooperative learning, and
interactive multimedia. One hundred fifteen undergraduate students participated. Results indicated a significant relationship between the motivation of students and use of interactive multimedia. On the posttest scores, students who were taught through cooperative learning got the highest scores on the posttest. Researchers hypothesis that the failure of interactive multimedia to increase the posttest scores of students is due to the fact that the learners were allowed to work their way through material anyway they wanted without a purpose or objective (Boling & Robinson, 1999). The ability of interactive multimedia to increase motivation of students is an important benefit in distance education and large courses. Although interactive multimedia does not appear to be as effective as interaction with other students or professors, it still encourages active learning more than listening to a lengthy lecture or clicking through endless slides of information.

In addition to feedback and addressing individual differences, materials can use other means of prompting students to take an active role in learning. For example, the multimedia evaluated in experiment 1 was used in a design class where students were given the task of creating a design using different shapes and colors. Students were given time to experiment with options in an interactive workspace to form their own insights about affective design. This activity involves using higher order thinking but does not provide feedback or adapt for individual differences. Judging how well a material utilizes interactive aspects to encourage higher order thinking involves considering feedback, addressing individual differences, and interactive tasks such as the design workspace.
Allen (1998) presented several techniques that can be used to exploit the educational potential of the web. Some of the techniques suggested include virtual environments and interactive simulations. Virtual environments imitate aspects of real life situations to help students understand how the theories and information learned fit into real environments. The design workspace discussed previously is an example of a virtual environment that imitates the process of creating a design in real life. The student uses the raw materials of multiple shapes and colors to create a design in the same way an actual designer does. Interactive simulations focus on more specific and focused subjects than virtual environments. Interactive simulations are most often activities such as experiments that are designed to help the learner understand one concept rather than offer a realistic experience. Allen concluded that the use of interactive multimedia has the potential to offer new perspectives in otherwise static representations of information (Allen, 1998).

Hypotheses

The current study aimed to establish the reliability and validity of the new direct evaluation method by testing a group of four hypotheses: two concerning reliability and two concerning validity. The first hypothesis (h1) is that the method would demonstrate test-retest reliability. Participants judging a material at one point in time will be able to reliably evaluate that material at a significantly later point in time. The second hypothesis (h2) is that the scores given to materials would be reliable between individuals evaluating the same materials. Experts in the area will give similar scores to the same educational material. The third hypothesis (h3) is that higher ratings in the six standards would be related to higher student exam
scores. The fourth hypothesis (h4) is that higher ratings on the six standards would be related to positive student attitudes.

**General Method**

**Overview**

The *Set of Standards Evaluation Sheet* was developed over a period of approximately 1 year prior to the experiments. The goal was to create a survey instrument that would provide a reliable and valid measure for discriminating quality of pedagogical materials. An initial version was developed from the literature review. The group of expert judges reviewed the survey and suggested revisions. Four individuals who created Viziswap modules or wrote textbooks reviewed the survey and it was revised after each meeting. Experiment 1 and 2 were completed to provide an initial assessment of the reliability and validity of the newly developed instrument.

**Measure**

The *Set of Standards Evaluation Sheet* was a 33-item survey. The survey was divided into the six sections: a *Modes of Communication* section (MC) containing 10 questions, a *Dual Coding* section (DC) containing three questions, a *Feedback* section (FB) containing six questions, a *Structure* section (S) containing eight questions, an *Addressing Individuals Differences* section (ID) containing four questions, and an *Interactivity* section (I) containing two questions. Each section was divided into two subsections rating the material: subsection one (QN) yielded a quantitative score and subsection two (QL) yielded a qualitative score. QN rated each standard on a 5-point Likert scale from 1-5 (1=Never or 0-19%, 5=Always or +80%) that scored how
often each standard was used throughout the material. QL rated each standard on a 5-point Likert scale from 1-5 (1=Never, 5=Always) that scored how well each standard was used throughout the material. QL ratings also had a Not Applicable option if the standard is not included within the material at all. In that case the item was not taken into account in the final score of the material. QN consisted of one question for each standard. QL consisted of a range of one to nine questions. Upon completion of the survey, each standard yielded a score out of 25 ranging from 1 (not included at all or used incorrectly in all categories) to 25 (included throughout and used correctly in all categories). A draft of the instrument is shown in the Appendix.

**Experiment 1**

Experiment 1 was conducted to determine if how a material scored on the Set of Standards Evaluation Sheet was consistent with previous research indicating differences in departmental exam scores and student attitudes. Inter-rater reliability among expert judges was correlated.

**Method**

**Participants.** Five expert judges completed an evaluation for three different materials. Expert judges consisted of one professor, three graduate students, and the principal investigator who were all familiar with the cognitive principles and literature used to support the six standards in the evaluation. All were affiliated with Ball State University. Judges did not receive an incentive for completing the evaluations.
**Materials.** Three different educational materials were evaluated in this experiment. The materials were used in a Journalism 103 and Design 130 course at Ball State University. The chapter or lesson evaluated was on the topic of Principles of Design. Material one (M1) was a textbook called *Graphic Communications Today* 3rd edition used prior to the fall 2004 semester. The 9th chapter was evaluated. Material two (M2) was an e-book used between fall 2004 and spring 2010. Material three (M3) was a Viziswap module, a type of multimedia educational learning material developed at Ball State University, used the fall 2010 and spring 2011 semester. A list of tasks and the *Set of Standards Evaluation Sheet* were used for evaluation of each material.

**Tasks.** The list of tasks was a 14-page document containing a list of questions and short evaluations covering all six standards. The tasks were designed to allow each judge to get aquatinted with the material well enough to make an accurate rating. The list contains two sections. Section one was completed while working through the material and the information pertained to objectives, pictures, videos, and feedback. Section two was completed after using the material and the information pertained to all of the six standards.

**Set of standards evaluation sheet.** The *Set of Standards Evaluation Sheet* used in this experiment contained the original 33-item survey and one question that indicated the material being evaluated.

**Procedure.** The expert judges completed a three-phase evaluation process for the materials. During phase one (P1) the list of tasks was completed while working with the material. During phase two (P2) the judges rated the material
using the set of standards through the online testing service *Inqsit*. The judges repeated P1 and P2 for each of the three educational materials that were evaluated. Judges completed the three evaluations over a 21 day period and were able to contact the principal investigator at any time for questions concerning procedural issues and clarity of the standards. During phase three (P3) participants meet as a group and discussed the reasoning behind some of the differences in final ratings of the material to provide further incite on how to improve the tool.

**Results**

**Reliability.** Judges rated M1 in each of the six standards: MC(8.48), DC(12.88), FB(1), ID(4.59), and I(2.52). M2 was rated in the six standards: MC(10.89), DC(11.96), FB(13), S(19.89), ID(9.6), and I(3.98). M3 was rated in the six standards MC(13.51), DC(20.5), FB(19.58), S(19.89), ID(17.87), and I(10.2). All scores had 25 possible points. As shown in figure 1, the ratings generally increased from M1 to M2 and from M2 to M3 with the exception of the DC score between M1 and M2. See Table 1 for the QN and QL ratings for each material. One-way *ANOVA*s on each of the six standards were not found to be significant. The results suggest that the *Set of Standards Evaluation Sheet* is not reliable between judges.

During P3 of the evaluation the judges met and discussed some of the different interpretations of the materials and tool that caused inconsistent ratings. Most of the inconsistencies in the data seemed to be the result of one of the following reasons: format, a misunderstanding of the directions, a misunderstood, vague, or subjectively worded item. The results suggest that the *Set of Standards*
Evaluation Sheet still needs a considerable amount of refinement and clarification of the language within the directions and items in order to produce a reliable measure.

**Validity.** Previous research indicated that students using Viziswap modules in the Design 130 class performed better on the departmental exams than students who had not used it. Previous research on student attitudes surveys suggests that students preferred the Viziswap modules to the typical textbook. The results of the current study are consistent with the previous research although not statistically significant. Because the results here were not significant, no statistical measures of validity were calculated.

**Discussion**

In experiment 1, five expert judges determined the relative scores of M1, 2 and 3 that was used to cover Principles of Design in journalism and design courses at Ball State University. The results indicated that scoring across judges was not statistically reliable.

Previous research indicated that the students’ scored higher on departmental exams when they had used a Viziswap module in class rather than an e-book or textbook. Previous research indicated that students enrolled in the design and journalism classes preferred using Viziswap to the typical textbook. The results of the current study were consistent with these findings as the scores in each of six standards are higher for M3 than for M1 and 2. The results of the new direct evaluation method are consistent with other more conventional forms of evaluation that indicate student acquisition and retention of information can be positively influenced by the growing capabilities of new multimedia learning materials.
Reliability. The lack of reliability among expert judges was the result of several factors. A small number of participants completed the survey. Only five expert judges were able to complete the evaluation for M1, 2, and 3. The group of individuals had varying amounts of background in cognitive psychology, knowledge of the six standards, and knowledge of journalism and design principles. Judges also varied in age and level of education.

The relatively new nature of the measure suggests that alterations are necessary to produce a reliable tool. Further investigation into the inconsistencies in the data in P3 suggested that clarity and subjectivity in items and directions might have led to problems in rating materials. Several areas of the Set of Standards Evaluation Sheet were identified as unclear or misleading to the evaluators. A post-experiments draft of the survey was revised to address these problems and made changes in the following areas.

Format. The format of the Set of Standards Evaluation Sheet caused some inconsistencies between judges. The option of rating items Not Applicable on any QL rating caused confusion for some judges. The option of Not Applicable was provided for a situation when the QN rating is 0-19% or Never and there is nothing to rate QL items over. Some participants over used the option. In the post-experiments draft of the survey if the evaluator answers 0% or Never the program automatically skips the QL items to avoid possible confusion. Consequently, no items will have a Not Applicable option and unanswered questions will not be counted in the final score.
With the addition of these format changes another problem that is caused is that when viewing the QL items the evaluator would lose the definition of the mode. The problem was solved by providing the option to click on a standard in order to receive the definition if desired. In the post-experiments version of the survey, the standard name in each QL section can be clicked on to provide the definition.

**Misunderstood directions.** Misunderstanding the directions and definitions given on the survey was one problem that caused inconsistencies in the judges’ ratings. The QN1 item for MC listed the four modes of communication evaluated and then asked the judge to fate the number of modes used in the material. The separation of directions from the item led participants to answer inconsistently. For example when rating the Vizibook, one participant did not remember that Non-Verbal Audio was not considered a different mode unless it was non-verbal and marked narration and text as separate forms of communication. The post-experiments draft of the measure asks about each mode individually and defines the mode within the each item to avoid this mistake.

**Misunderstood or vague items.** Another problem that led to inconsistencies between judges was items that were vague. For example, in the MC section the first QL item was worded “(MODE) is/are relevant to the concepts being communicated.” The word relevant was too vague to discriminate between information that added to the quality of the lesson and information that did not. The post-experiments draft was worded “(MODE) is/are important to understanding the concepts being communicated.”
The final video question caused variability between the judges because it asked about the quality of the video if the purpose of the video is to help the student form an opinion that is not relevant to all videos used throughout a lesson. In the post-experiments version the question was changed to an if-then statement worded, “If the purpose of the video is to help the student form an opinion or argue a viewpoint, than it presents both sides of the argument in a non-biased way.”

**Subjectively worded items.** The QL interactivity item was worded in a subjective manner. The original question asked how often the evaluator used critical thinking throughout the lesson. This item led judges to answer in a subjective way. The post-experiments version worded the item “Does the material require the student to use higher order thinking to work through or complete the lesson.” This wording is intended to prompt a less bias response.

**Experiment 2**

Experiment 2 conducted two rounds of testing on Psyc 100 participants to determine the test-retest reliability of the *Set of Standards Evaluation Sheet.*

**Method**

**Participants.** One hundred seventy-six participants were recruited from the Psyc 100 subject pool at Ball State University and completed round one (r1) of experiment 2. Thirteen participants were removed initially for completing the survey in less than three minutes. Participants in r1 reported being in the age range of 18-19 (55.5%), 20-21 (36%), 22-23 (6.1%), and 24+ (2.4%). Seventy-four participants were female. Participants reported being in their 1st (50%), 2nd (31.7%), 3rd (11%), 4th (4.9%), and 5th+ (2.4%) years of college. Twenty-three
participants completed the survey again for round two (r2) of the evaluation. Four participants were removed because they completed the evaluation on the wrong chapter. One more participant was removed during evaluation for leaving a large number on items blank. The remaining 18 participants in r2 reported being in an age range of 18-19(33.3%), 20-21(55.6%), 22-23(5.6%), and 24+(5.6%). Seven participants were female. Participants reported being in 1st (33.3%), 2nd (50%), 3rd (11.1%), and 5th+ (5.6%) years of college. Participants received an incentive of one hour of research credit in the Psyc 100 course for both r1 and r2.

**Materials.** Each participant was asked to evaluate a chapter in his or her course textbook that had already been tested over in class. Participants who returned for r2 were asked to rate the same chapter as they did in r1. The *Set of Standard Evaluation Sheet* was used to evaluate the textbook in r1 and r2.

**Set of standards evaluation sheet.** The *Set of Standards Evaluation Sheet* used in this experiment was a 39-item survey in r1. It contained the 33 original items, three demographic questions, two questions that indicated the chapter and title of the textbook being evaluated, and one guiding question. In r2, the survey contained the same items with the exception of the demographic questions already provided in the r1 evaluation.

**Procedure.** In r1 the participants were recruited to complete the study online through the Ball State University experiment sign-up website. The participants completed the survey through the online testing service *Inqsit*. After completing the survey, the participants who were interested in completing the survey again for more credit were asked to enter their e-mail addresses to be
further contacted for participation in r2. Participants were recruited via e-mail to participate in r2. A two-week delay separated the closing of r1 and the opening of r2. The survey completed in r2 randomized the order of the six standards sections. Items within the sections remained in order to avoid confusion.

Results

Reliability. The responses to r1 of the survey were correlated to the responses in r2 to determine the test-retest reliability of the measure. Results indicated a significant, positive relationship between r1 and r2 in the following items and scores: picture 1, \( r(16) = 0.484, p < 0.05 \), picture 2, \( r(16) = 0.476, p < 0.05 \), video 1, \( r(16) = 0.535, p < 0.05 \), video 3, \( r(16) = 0.486, p < 0.05 \), DC 2, \( r(16) = 0.666, p < 0.01 \), speed of FB, \( r(16) = 0.608, p < 0.01 \), FB 1, \( r(16) = 0.76, p < 0.01 \), FB 3, \( r(16) = 0.686, p < 0.01 \), QL FB, \( r(16) = 0.663, p < 0.01 \), objective 2, \( r(16) = 0.471, p < 0.05 \), and ID 1, \( r(16) = 0.648, p < 0.01 \). The # MC, verbal 1, verbal 2, verbal 3, video 2, QL MC, TOT- MC, # DC, DC 1, QL DC, TOT-DC, # FB, FB2, TOT-FB, # S, objective 1, objective 3, organization 1, organization 2, organization 3, organization 4, QL S, TOT-S, # ID, ID 2, ID 3, # I, QL I, and TOT-I items and scores did not indicate any positive correlation from r1 to r2. Only 11 out of the 45 possible items and scores were positively correlated.

A significant, positive relationship was indicated between some items within the same section in both r1 and r2. The r1 verbal 1 item indicated a significant positive correlation to verbal 2, \( r(16) = 0.83, p < 0.01 \) and to verbal 3, \( r(16) = 0.704, p < 0.01 \). Verbal 2 was positively correlated to verbal 3, \( r(16) = 0.594, p < 0.01 \). In r2, verbal 1 item was positively correlated to video 2, \( r(16) = 0.833, p < 0.01 \) and video...
$r(16) = 0.7, p < 0.01$. Verbal 2 was positively correlated to verbal 3, $r(16) = 0.735, p < 0.01$. In addition to the verbal section, the video section, feedback section, objective section, individual differences section, and interactivity section all indicated a significant positive relationship to the other items within the section in at least one of the rounds.

**Discussion**

**Reliability.** The results of experiment 2 indicated that the current *Set of Standards Evaluation Sheet* does not have high test-retest reliability with this group of participants. The lack of reliability could be the result of several different factors. First, the survey contained ambiguous wording that led to participants to interpret the items differently at different times. The subjective, misunderstood, and vaguely worded items that were identified by the expert judges in experiment 2 could also be responsible for the large amount of the variability between $r_1$ and $r_2$. Considering the fact that the expert judges had problems understanding items it seems likely that student participants would have been confused by some items.

The few items that were significantly correlated contain subject matter that students are more likely to be familiar with. Students hear the terms picture, video, feedback, and objectives on a daily basis. Some of the other standards that discuss topics like dual coding, individual differences, and interactivity require students to make judgments about concepts they may never have heard of before. Giving a reliable response to something that is addressing a novel concept is a difficult task. Students trying to complete the survey may not have understood what they were rating or may have become so frustrated with the survey that they gave up.
Another important consideration in interpreting the results of this experiment is that students received an incentive of one hour of research credit for completing the survey regardless of the actual time spent. As mentioned in the participant section, 13 participants were deleted from r1 for completing the survey in less than 3 minutes. Realistically speaking, it should take a minimum of at least 10 minutes to complete the Set of Standards Evaluation Sheet reliably with the possible exception of an evaluator who is a highly trained expert of both the pedagogy and subject matter. Participants in the current study also completed the evaluation online. The lack of a controlled environment and supervision makes it likely that the participants would have not take the survey seriously or would have been distracted by other things.

These two weaknesses of the assessment process create a strong motivation for students to complete the survey as quickly as possible. The large, positive correlation among items within the same section suggests that participants were answering according to a response style. Results suggest that completing the evaluation requires both a general understanding of pedagogy and subject matter as well as motivation. Because of the cognitively demanding items that the tool assess, the best results could be expected when participants have some sort of personal motivation (e.g. developer of new educational materials or professor choosing a material for a course) driving completion of the survey.
Implications

Hypotheses. The purpose of the current study was to develop a new method of direct evaluation and determine its reliability and validity through a series of experiments. H1 and h2 were not supported by the current study suggesting that the new measure cannot always produce a reliable score for materials. H3 and h4 were supported by the current study, but no concrete predictions as to the validity of the measure can be made due to the lack of reliability.

Organization of standards. The completion of the current study led investigators to form new conclusions on the organization of the six standards. The adjustments made to the survey after the completion of experiment 1 and 2 reorganized the six standards into two categories: passive learning standards and active learning standards. In the post-experiments version of the survey, the standards were organized in these sections.

Passive standards. Passive standards consist of MC, DC, and S. These areas represent important features of educational materials that provide students with necessary information but do not engage the student in active learning. These standards are all important in providing students with a quality learning material.

Active standards. Active standards consist of FB, ID, and I. These areas represent features that help prompt students to take an active role in learning. These standards are the areas in which new interactive, multimedia educational materials can become far superior to traditional methods such as text and lecture style presentations. In past years the active learning was provided by class
discussion and individual’s questions with instructor answer. In modern academic institutions the number of students per classroom is steadily increasing and one-on-one student instructor interaction is minimal.

**Strengths and Limitations**

**Strengths.** A strength of the new measure is the extensive amount of background research done on each of the six standards. A significant amount of time was spent collecting and organizing literature before any conclusions were formed as to what constitutes good educational materials. The extensive literature review would give the new measure a high degree of construct validity if the reliability can be increased.

Another strength was the process of multiple experts of pedagogy reviewing the measure and the subsequent revisions working out inconsistencies and adding clarification to items. The small, personal nature of the participant group in experiment 1 allowed for the P3 investigation into what caused the inconsistencies in the data. The opinion and experiences of all these individuals gave valuable feedback on how to improve the tool to make it both more user friendly and reliable.

Previous research made available some more traditional methods of evaluation that measured the affect M1, 2, and 3 had on student exams and student attitude surveys for comparison to the new measure’s results. Lastly, the flexibility in the time (large window, no limit) and location (access from any computer with internet) for the survey made it easy for participants to complete.

**Limitations.** The current study was limited by the time and resources available to the principle investigator. This study was completed as part of an
undergraduate thesis project with a one year time limit. More time was needed to develop a reliable and valid form of this complex measure. The participants available to complete the measure and form conclusions on the reliability were not ideal. In experiment 1, the results were limited by the lack of a homogenous group of experts with background knowledge of design. The participants in experiment 2 completed the survey as part of a research requirement and had no background knowledge or instruction on pedagogy or psychology.

**Future Research**

The current study suggests that more research is needed in order to develop a reliable and valid method of direct evaluation. Due to limitations of time and resources the development of the current measure is incomplete. Further refinement of format, directions, and items is needed. Future research should investigate how participant differences impact evaluations using the Set of Standards Evaluation Sheet. Research seems to indicate that the evaluation may take an individual with both knowledge of the subject matter covered by the material and effective use of pedagogy in order to produce a reliable score. A homogeneous sample of participants with knowledge of pedagogy would be ideal in future research. Future research that implements some of the changes suggested by P3 of experiment 1 should be completed.

Future research should consider the possibility of more or less standards being necessary to form a complete evaluation of educational materials. The current study identified six areas in which research has indicated strong evidence about what constitutes good educational materials. However, it is possible that more or
less areas may be the ideal number to measure in order to gain as complete a picture as possible of the quality of a material. For example, researchers debated whether or not to include interactivity as a separate standard of educational materials because it is in essence a measure of how many active learning strategies were utilized (including FB and ID). The post-experiments version of the Set of Standards Evaluation Sheet kept the I section of the evaluation because investigators identified active learning strategies that were not included under FB or ID. It is also possible that more standards exist that are important to learning that were not included in this survey.

Conclusions

Reliability. The current study suggests that further work is needed to increase the reliability of the new measure. Both experiments failed to demonstrate reliability either between judges rating a common material or test-retest measures completed by the same individual. Results of the research indicate that misinterpretation of direction and items on the Set of Standards Evaluation Sheet play a role in the lack of reliability. Considering both the expert judge and undergraduate participant groups used in experiment 1 and 2, results suggest that an ideal group of judges should have both knowledge of the subject the material is intended to teach and of the empirical evidence supporting the six standards. Difficulties in understanding directions and items suggest that completion of the survey either in a proctored environment where participants can ask questions or using a trained administrator could increase the reliability of the measure.
Validity. The validity of the new measure cannot be concluded based on these experiments because the survey was not reliable. The results of the current study being consistent with departmental exam scores and student attitudes surveys and the completion of a thorough literature review suggest that if reliability could be increased than the validity of the measure would be high.

Use of current measure. Although it is clear that the current measure still needs further work to produce a reliable measure, the *Set of Standards Evaluation Sheet* still has practical use in discriminating between and improving different educational materials. Scores in the six standards provide valuable information to educators about which material will help students meet course objectives and complement the teaching style of individual professors.

The six principles covered by the *Set of Standards Evaluation Sheet* could also be useful to the developers and creators of new educational materials. The tool can identify possible ways to improve materials. The tool outlines the advantage of active learning that is offered by new multimedia learning materials. The current study suggests that capacity to encourage active learning is the reason interactive, multimedia materials can outperform traditional methods in some situations. The new measure can help identify both passive and active learning standards that have room for improvement.
References


Table 1

*Mean quantity and quality ratings of the six standards in the textbook, ebook, and viziswap*

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>DC</th>
<th>FB</th>
<th>S</th>
<th>ID</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>ViziQN</td>
<td>3.2</td>
<td>5</td>
<td>4.8</td>
<td>4.8</td>
<td>4</td>
<td>3.4</td>
</tr>
<tr>
<td>ViziQL</td>
<td>4.2</td>
<td>4.1</td>
<td>4.1</td>
<td>4.1</td>
<td>4.5</td>
<td>3</td>
</tr>
<tr>
<td>EbookQN</td>
<td>2.8</td>
<td>4.6</td>
<td>5</td>
<td>4.8</td>
<td>2.4</td>
<td>1.8</td>
</tr>
<tr>
<td>EbookQL</td>
<td>3.9</td>
<td>2.6</td>
<td>2.6</td>
<td>4.1</td>
<td>4</td>
<td>2.2</td>
</tr>
<tr>
<td>TextQN</td>
<td>2</td>
<td>4.6</td>
<td>1</td>
<td>4.2</td>
<td>1.6</td>
<td>1.4</td>
</tr>
<tr>
<td>TextQL</td>
<td>4.2</td>
<td>2.8</td>
<td>-na-</td>
<td>3.4</td>
<td>2.9</td>
<td>1.8</td>
</tr>
</tbody>
</table>
Figure 1. The relationship of six standard’s mean ratings among the Textbook, Ebook and Vizswap module on Principles of Design.
Appendix

Set of Standards Evaluation Sheet

Content Issues- Refers to how much information is provided and how the information is communicated within the lesson.

Lesson- Refers to a group of all of the educational materials used to teach a certain topic, often a chapter in a textbook perhaps supplemented with handouts, a DVD, or a website.

Modes of Communication

- Definitions of Modes of Communication:
  - Verbal- Any linguistic representation of knowledge; includes text and oral narration.
  - Picture- Any single visual representation of knowledge not containing motion and involves more than simple text; includes photos, graphics, and diagrams.
  - Video- Any non-static collection of pictures including slide shows that imply motion.
  - Non-verbal audio- Any audio representation of knowledge that does not include linguistic information; including music and sound effects.

- Indicate the number of modes of communication used throughout the lesson:

<table>
<thead>
<tr>
<th>Total modes of communication</th>
<th>1 mode</th>
<th>2 modes</th>
<th>3 modes</th>
<th>4 modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

- Indicate the quality of the VERBAL information provided:

<table>
<thead>
<tr>
<th>Verbal information is relevant to the concepts being communicated</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verbal information is used to communicate in-depth analysis of abstract</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Verbal information is used to communicate concepts where the majority of students are expected to have an appropriate amount or prior experience to relate novel information to

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictures are relevant to the concepts being communicated</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Pictures are used to communicate concepts that require specific spatial information or detailed step-by-step information for understanding</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

- Indicate the quality of the VIDEOS provided:

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Videos are relevant to the concepts being communicated</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Videos are used to communicate concepts where the majority of students in the course are expected to have little prior experience</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Videos present both sides of an argument used by the student to express an opinion or argue a viewpoint about the information</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
- Indicate the quality of NON-VERBAL AUDIO provided:

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-verbal audio is relevant</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>to the concepts being</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>communicated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Dual Coding**

*Definition:* Visual information (picture/video) that is presented simultaneously with verbal information (text/narration) on the same subject

*Example:*

- Indicate the percent of visual information that includes verbal information simultaneously

<table>
<thead>
<tr>
<th></th>
<th>0-19%</th>
<th>20-39%</th>
<th>40-59%</th>
<th>60-79%</th>
<th>+80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual Coding</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
• Indicate the quality of the dual coded information

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>The verbal information and the visual info complement each other (one increases the understanding of the other)</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>The verbal information that is dual coded uses oral narration rather than text</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Feedback**

*Definition:* A response to work the user submits that provides either verification or elaboration provided by the professor/program

*Types of Feedback:*

- *Verification:* A statement that expresses the “correctness” of an interaction.

- *Elaboration:* A statement that makes some sort of response specific comment on submitted work other than a simple statement of “correctness.”

• Indicate the percent of questions in the lesson that provide feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>0-19%</th>
<th>20-39%</th>
<th>40-59%</th>
<th>60-79%</th>
<th>+ 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

• Indicate the speed of feedback

<table>
<thead>
<tr>
<th>Feedback is provided or can be retrieved...</th>
<th>Immediately</th>
<th>Slight Delay (no more than 30 sec.)</th>
<th>Delay (no more than 2 days)</th>
<th>Extended Delay (no more than 1 week)</th>
<th>Significant Delay (longer than 1 week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicate the</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
### Speed of Feedback

- Indicate the quality of feedback

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback provides verification of the “correctness” of answer</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Feedback provides “response specific feedback” that identifies why a wrong answer was incorrect or why a correct answer was correct</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Feedback provides information that gives guidance on how to answer correctly in the future</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Feedback provides an example of a correct response if the wrong answer is given</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Usage Issues** - Refers to how easy the tool is to use or if working with the material is frustrating to the user.

### Structure

**Definition** - The general organization, order, and flow of the information within the lesson

- Indicate how well the information within the material was organized

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>The information within the material flows easily between the key concepts</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
- Indicate the quality of the objectives provided for the lesson (if any)

  - *Objectives*- An isolated section educational material that contains a list of “things to know” by the end of a lesson

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives are provided before the student begins working with the material</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Objectives address how the information being learned will aid in the eventual completion of the course objectives</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Objectives provide a clear endpoint with enough detail for user to know what they should understand by the end of the lesson</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

- Indicate the quality of the organization

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information is organized in a logical meaningful way</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>All of the information is integrated with or nearby other information covering the same concepts</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>All of the information is presented simultaneously with information covering the same concepts</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>The material provides headings and bolded key words throughout the lesson to assist the users organization of information over time</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Addresses Individual Differences

*Definition* - Features of the educational material that support the uniqueness of the user

- *Examples of Addressing Individual Differences*: include but are not limited to...
  - Providing suggesting links or suggesting additional materials based on input of user
  - Providing information broken down into manageable chunks
  - Skipping a section already understood based on input of user
  - Ability to repeat a section not yet understood
  - Ability to stop or pause on section and choose when to continue
  - Providing optional definitions of words within the material for users with different reading skills
  - Providing foreign language options for ESL users
  - Providing options for users with disabilities

- Indicate the adaptability of the material to the individual learner

<table>
<thead>
<tr>
<th>Material is appropriately adapts to the uniqueness of the user</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

- Indicate the quality of the adaptability

<table>
<thead>
<tr>
<th>Tool utilizes adaptive support in that information is changed or “adapted” by the input of the user</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Content is broken down into manageable chunks of information

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

The user of the material controls the delivery pace of the chunks of information

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Interactivity**

*Definition:* Features that require the user to take an active role in the lesson in order to continue through the remainder of the information that go beyond tasks that can become automatic processes (e.g. beyond simple page turning or clicking “next”)

- **Examples of Interactivity:** include but are not limited to...
  - Submitting a response for feedback
  - Examples of addressing individual differences that require user to take an active role in education
  - An interactive area that facilitate the development of skills with or without the submission of work for feedback

- Indicate how often the tool uses methods to encourage students to take an active role in the learning process

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material requires the user to consistently complete an activity, complete a question, make a decision about when or how to continue through the lesson</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
- Indicate how often the tool provided information with a high cognitive demand in that your work prompted a deep sense of understanding.

<table>
<thead>
<tr>
<th>Did the material required you to use critical thinking skills in order to work through or complete the lesson</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>