ED 340 000   CS 010 720

AUTHOR   Burnham, Catherine A.; Anderson, Thomas H.
TITLE   Learning To Sew on a Button by Reading a Procedural
              Text. Technical Report No. 543.
INSTITUTION   Illinois Univ., Urbana. Center for the Study of
              Reading.
SPONS AGENCY Office of Educational Research and Improvement (ED),
              Washington, DC.
PUB DATE   Dec 91
CONTRACT   G0087-C1001-90
NOTE   81p.
PUB TYPE   Reports - Research/Technical (143)

EDRS PRICE   MF01/PC04 Plus Postage.
DESCRIPTORS   Adults; Grade 7; Grade 10; High Schools; Information
              Sources; Reader Text Relationship; *Reading
              Comprehension; Reading Research; Regression
              (Statistics); *Technical Writing; *Text Structure
IDENTIFIERS   Text Factors

ABSTRACT
        Two studies examined the validity of a model of
procedural document processing, and the relationships among document
features, reader characteristics, and successful completion of a
button-sewing task. The first study tested three information sources.
Subjects were 12 seventh-grade, 12 tenth-grade, and 12 adult students
who used 2 commercially published texts and 1 experimental text. The
experimental text helped older subjects perform better; adult
students performed best overall. The second study tested instructions
written to encourage more self-testing and self-correcting. Subjects
were 105 high school students. Amounts of self-correcting were
increased, but overall button-sewing quality did not improve
significantly. Effects of the experimental text features were not
significant, but analyses of regression performed on the
without-instruction button-sewing scores, sewing improvement scores,
sewing time, and number of corrections revealed significant effects
from gender, grade level, reading ability, coordination, prior
experience, and sewing time. (Five figures of data are included; the
button-sewing texts, rating scales, and 112 references are attached.)
(Author/RS)

CENTER FOR THE STUDY OF READING

Technical Report No. 543

LEARNING TO SEW ON A BUTTON BY READING A PROCEDURAL TEXT

Catherine A. Burnham
Eastern Illinois University - Charleston
Thomas H. Anderson
University of Illinois at Urbana-Champaign

December 1991

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

1991-92 Editorial Advisory Board

James Armstrong

Diana Beck

Stacy Birch

Diane Bottomley

Clark A. Chinn

Judith Davidson

Irene-Anna N. Diakidoy

Colleen P. Gilrane

Barbara J. Hancin-Bhatt

Richard Henne

Carole Janisch

Bonnie M. K rr

Raymond Langley

Jane Montes

Marty Waggoner

Janelle Weinzierl

Hwajin Yi

Abstract

Two studies were conducted to test the viability of a model of procedural document processing offered by Guthrie, Bennett, and Weber (1990), and the relationships among document features, reader characteristics, and successful completion of a button-sewing task. The first study tested three information sources. Subjects were 7th-grade, 10th-grade, and adult students who used two commercially published texts and one experimental text. The experimental text helped older subjects perform better; adult students performed best overall. The second study tested instructions written to encourage more self-testing and self-correcting. Subjects were 105 high school students. Amounts of self-correcting were increased, but overall button-sewing quality did not improve significantly. Effects of the experimental text features were not significant, but analyses of regression performed on the without-instructions button-sewing scores, sewing improvement scores, sewing time, and number of corrections revealed significant effects from gender, grade level, reading ability, coordination, prior experience, and sewing time.

LEARNING TO SEW ON A BUTTON
BY READING A PROCEDURAL TEXT

Analysis of data from the National Assessment of Educational Progress (NAEP) survey of young adult literacy revealed that workers spend more time on the job in reading manuals and guides, or "reading-to-do," than in any other kind of reading (Guthrie, Schafer, & Hutchinson, 1989). In fact researchers (Diehl & Mikulecky, 1980; Sticht, 1975; Sticht, Fox, Hauke, & Welty-Zapf, 1977) report that workers spend an average of nearly 2 hours a day engaged in reading job-related procedural documents in order to perform a task. These reading-to-do tasks require an understanding of technical language, diagrams, and illustrations (Chang, 1983; Cranney, Rowley, & Stonehocker, 1984). However, recent NAEP surveys have shown that only one fifth of young adults sampled demonstrated the advanced reading skills necessary for understanding complex texts of the sort that might be found in a technical working environment (Applebee, Langer, & Mullis, 1987).

Why do people have difficulty reading-to-do? Possibly because many of them have never received direct instruction in how to read instructions (Henk & Helfeldt, 1987). Hayes and Henk (1986) found, for example, that none of approximately 100 high school readers remembered receiving any formal training in following written instructions. Students also may not have sufficient exposure to procedural documents to become adept at using them, with or without direct instruction. The sample of 17-year-olds in the 1986 NAEP survey reported reading procedural materials (books on how to do, make, or repair something) approximately 1.1 times per week (Guthrie, Schafer, & Wang, 1990). These findings may help explain why good and average readers follow instructions well only about 85% of the time, while poor readers are lucky to succeed 50% of the time (Fox & Siedow, 1980).

It appears, then, that while reading-to-do tasks abound in the workplace, students in general are not well prepared to perform such tasks. Unfortunately, little is known about the cognitive processes that underpin the comprehension of procedural documents, although the cognition underlying reading comprehension has been a major research focus for many years. This research focus, however, has been on the processes that students employed when they read narrative and/or expository texts (see Kintsch, 1985; Mandler & Johnson, 1977; Meyer, 1980; Trabasso, Secco, & van den Broek, 1984, for models of reading processes based on these genre). Formulating and testing models of procedural text/document processing have apparently not attracted much research attention.

There are signs, however, that the situation is changing. Diehl, Mills, Birkmire, and Mou (1989) tested models of narrative text on eight procedural documents to determine if a good transfer was possible. Specifically, they investigated whether text models useful in predicting importance of idea units in narrative texts could successfully predict importance of idea units in procedural documents. If models useful in explaining reader comprehension of narrative text also explain comprehension of procedural text, research on a new model specific to procedural text may not be necessary.

The two models Diehl et al. (1989) chose to test were the "structure" model of Trabasso et al. (1984), which emphasizes the structural aspect of the relationships among text units, and the "process" model of Kintsch (1985), which emphasizes the longer presence of important ideas in short-term memory. While the structure model was somewhat more accurate than the process model in predicting importance of idea units in the procedural documents, Diehl et al. were not satisfied that either model worked very well in the new text genre. Furthermore, the structure model was not as effective as it had been in previous research with narrative text. Diehl et al. concluded that the major difficulty in transferring models of narrative text to procedural text might stem from the fact that a reader must construct a representation of both the task (a mental portrayal of the activities to be performed) and the text in a procedural document, and that models for narrative text do not include the representation of the task.

Another difficulty of trying to use on procedural texts models that were developed and tested on narrative texts is the dependent measures problem. Recall is a typical dependent measure used in studies of expository or narrative texts. In a study investigating characteristics of procedural documents, Mark and Bracewell (1989) reported findings that called into question the use of typical recall measures in determining the comprehension of instructions. They obtained low recall results across text conditions and phases of the task, but high performance results. They also found a strong ceiling effect. They suggested that "functional comprehension"--a measure that combined performance and comprehension--may be a more relevant measure when the text is procedural, due to the requirement for the reader to perform, or use the text not simply to process the information in it.

In the Diehl et al. (1989) and Mark and Bracewell (1989) studies, possible problems were found in trying simply to transfer what we know about the structures of expository or narrative texts to procedural documents. To address these problems, Guthrie, Bennett, and Weber (1990) have proposed a model of procedural text comprehension very different from the models examined by Diehl et al. and by Mark and Bracewell.

The so-called Guthrie model requires two elements that seem to be missing in models built on expository text--transformation of declarative knowledge into action, and complex interactions among the reader, the document, and the outcome. In explaining the transformation process, the Guthrie model draws heavily on Anderson's (1982) model of the acquisition of procedural knowledge. In the Guthrie model, movement from declarative knowledge (knowing "that"), to document-dependent action (action guided by references to the information source) to automatic action (document-independent action resulting from practice, or knowing "how") is suggested to explain the cognition underlying procedural learning. Guthrie et al. refer to this entire process as "transforming procedural text" but accept it as only a partial explanation of the processing of procedural information.

The complex interactions proposed in the Guthrie model as necessary to process procedural documents arise from the interaction of information available in the document with certain cognitive activities. These cognitive processes are believed to be different from the cognitive processes required by nonprocedural documents (see also Diehl & Mikulecky, 1980). The elements of the model follow.

Sources of Information:

1. Exposition of the outcome (an overview of the process in prose, or a statement of purpose containing information about the nature of the outcome),

2. Procedural steps (a list of separate, executable, required actions),

3. Representation of the outcome (a description of the anticipated outcome whether in the form of pictures/illustrations, or text), and

4. The workspace (which includes partially assembled equipment or half-drawn figures).

Cognitive processes:

1. Forming a conceptual model (creating a mental idea of how to do the task that can help the reader use the written materials),

2. Encoding procedures (identifying separate steps and entering them into memory),

3. Self-testing (asking oneself "Do I understand what I've read and have I done it correctly?"), and

4. Self-correcting (repairing mistakes identified by self-testing).

Guthrie, Bennett, and Weber (1990) suggest that optimal combinations of information and cognition include (a) using the exposition to help form the conceptual model of the task; (b) using the written steps to help encode the procedures (identify and execute the steps); and (c) using the graphic representation of the outcome (pictures or drawings) to foster self-testing. They do not suggest an optimal combination that might foster self-correction activities because of a lack of research into this type of cognitive activity.

The Guthrie model is appealing for several reasons. It applies only to procedural documents, for example. And it makes a specific point of referring to documents instead of text, thus recognizing the necessary presence of illustrations or other graphic material. It also recognizes that a reader of procedural material has access to external resources, and is not limited to internal cognitive processing alone (e.g., when Tab A does not fit into Slot B the reader can see it, touch it, take it apart, etc.). It acknowledges that the reader is involved in complex activities to combine text, illustrations, and actions that must lead to a very specific outcome. In essence, the model recognizes the unique characteristics of procedural documents and the unique demands they make on the reader, as well as certain testing and correcting behaviors that may be necessary when a reader/performer is trying to reproduce a specific outcome.

The outstanding omission in this model is any reference to equipment with which the reader/performer may be dealing, the manipulanda required by the procedure. However, this equipment may be included in what the model terms the workspace, which includes "partially assembled equipment" (p. 22), but Guthrie et al. do not make the definition of the workspace clear. The workspace has been referred to by others as a source of "extralinguistic cues" (Diehl & Mikulecky, 1980, p. 225), a definition that may correspond to that used in the Guthrie model.

Although development of models of procedural document processing is limited, there are studies that have investigated some of the elements of procedural documents. Our review of these studies is organized around three components of the Guthrie model--representation of the outcome (an information source), forming a conceptual model, and encoding of procedures (two cognitive processes).

Representation of the Outcome

Guthrie et al. suggest that the "representation of the outcome" (p. 19)--a description of the anticipated outcome whether in the form of pictures/ illustrations, or text--is one of the information sources available to the reader. The reader performing an assembly task may require rather detailed drawings, while a reader involved in locating a shopping mall in a strange city may be instructed by a series of written steps, that is, sentences. Research related to this representation can be classified under two headings--graphics (pictures, drawings, or diagrams) and text.

Graphics. The term "graphics," as used here, reflects the Random House Dictionary definition. That is, it includes pictures (visual representations of objects or people), or more specifically, drawings (visual representations of objects or people characterized by being linear; not photographs) and diagrams (visual representations of information that "outline and explain the parts, operation, etc., of something") (Stein, 1980, p. 242).

Memory for pictures (photographs) is superior to memory for words, and memory for pictures may be almost perfect (Haber, 1970; Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970). The use of drawings instead of text may allow faster location and identification of information in memory by the reader (Seymour, 1974), and drawings can contribute to understanding by providing nonredundant context information (Bransford & Johnson, 1972; Rigney & Lutz, 1976).

These characteristics of pictures may make readers prefer to use them when following instructions. LeFevre and Dixon (1986) reported that subjects overwhelmingly preferred to use a graphic representation (a drawing) instead of a textual one to help them solve shape-classification or series-completion problems. Other researchers (Curran & Mecherikoff, 1979; Dwyer, 1971, 1972; Loftus & Bell, 1975; Readence & Moore, 1981; Schoff & Robinson, 1984) report conflicting findings about the efficacy of line drawings and/or photographs to convey information. The choice between these types of graphics seemingly depends on the nature of the task. Diagrams, another type of graphic, also appear to be powerful conveyors of information (Larkin & Simon, 1987). When information is spatial in nature, a diagram may be especially well suited to represent it, thus minimizing what Guthrie et al. refer to as the transforming of text into action.

Not only are pictures or drawings memorable, popular, and efficient at conveying information; they also have been found to be versatile, that is, able to convey different types of information. Several studies (Bieger & Glock, 1984, 1986; Schorr & Glock, 1983; Stone, 1977a, 1977b) have investigated and expanded Mandler and Parker's (1976; see also Mandler & Johnson, 1976) taxonomy of the types of information that can be conveyed by pictures (inventory, spatial location, descriptive, and spatial composition), enlightening our knowledge of how pictures convey meaning. For example, Schorr and Glock (1983) concluded that instructions containing explicit, or detailed, operational information (e.g., "Slide a short rod . . . .") led to greater accuracy of assembly than instructions including only general operational information (e.g., "Move a short rod . . . ."). Although graphics seem to be very powerful informative devices, when used alone they appear to be insufficient to encourage optimal task performance(see Hayes & Henk, 1983, 1984; Stone, 1977a, 1977b; Stone & Glock, 1981). It may be that the nature of the task seriously impacts the effects of combining text with pictures.

Stone (1977a) and Bieger and Glock (1984, 1986) have also specifically studied spatial and contextual information, reporting that spatial information presented in text can lead to fewer errors in assembly tasks, while spatial information presented in pictures can lead to shortened assembly times. Bieger and Glock (1986) suggest, too, that the "trade-off benefit" (p. 185), that is, increased speed versus reduced errors, may have important educational implications for instructional design (see also Booher, 1973, cited in Reynolds & Booher, 1980).

In summary, it seems that using pictures can contribute to improved processing of procedural documents, but because pictures can convey different types of information, combining text with pictures must be carefully done. It may be, too, that the nature of the task must be taken into account when designing instructional texts.

Text. When studying the text component of procedural material, a frequently used dependent measure is response time (see Gagne, 1985). Response times have been found to vary according to sentence structure (Seymour, 1974). For example, sentences of the type "The square is inside the circle" increased subjects' speed in visualizing and in making a simple drawing over the time required for sentences of the type "Inside the circle is a square" or "The circle has a square inside." Action-first sentences, such as "Draw a circle above a square" have been found to be read more quickly than condition-first sentences, such as "Above a square draw a circle." (Wright & Wilcox, 1978). Imperatives also seem to shorten response time. For instance, "Please make the circle blue" contributed to shorter response times than "Can you make the circle blue?" (Clark & Lucy, 1975).

The reader's success with instructional text may hinge somewhat on the structure of the sentences when sentences are used to represent the outcome performance.

Forming a Conceptual Model

Guthrie et al. refer to forming a conceptual model (a mental idea of how to do the task that can help the reader use the written materials) as a necessary cognitive process. For example, if the task is to open a window, the performance may be influenced by the conceptual model of the performer. If one's model included the fact that windows slide, one would look for a handle in a different location than if one's model indicated that windows push outward (see Norman, 1989). Guthrie et al. believe that forming the conceptual model can be enhanced by activating certain "schemata" (see Anderson & Pearson, 1984). The Guthrie model includes two necessary schemata: (a) an outcome schema that allows the reader to know where she or he is headed, and (b) a procedural schema that allows the reader to know how to get there.

Outcome schema. Outcomes may be performances or objects. Dixon (1982) presented readers with instructions to make adjustments among buttons or knobs on a control panel for an electronic device, for example, "Turn the left knob when the alpha meter reads 20," and "When the alpha meter reads 20, turn the left knob," and concluded that plans are organized around actions and that condition information is only remembered in relation to particular actions. When the outcome is a performance, presenting the outcome first seemed to increase true effectiveness of processing the document.

The outcome of procedural instructions may be an object as well. Dixon (1987a) reported that when organizational information (information about the outcome) was presented before component step information (information about the action), directions were processed more efficiently. For example, the sentence "You can make a wagon by drawing a long rectangle with two circles underneath" took less time to read than "By drawing a long rectangle with two circles underneath you can make a wagon" (Dixon, 1987c).

Procedural schema. Instructions generally consist of information that must be processed step by step. The manner in which those steps are presented may affect the ease of processing the instructions. For example, Smith and Goodman (1984) found differences in processing time and efficiency of processing when instructions for assembling a simple electric circuit were presented in three different formats--one linear and two hierarchical. Subjects who received the hierarchical versions, which contained both explanatory material and steps (and which were, therefore, much longer), read the steps faster, recalled and executed them more accurately, and showed better transfer than subjects who received the linear version. In both of the hierarchical versions, the authors provided a schema for the relationships among ideas, actions, and component pieces (see also Dixon, 1987a; Graesser, 1978).

Encoding the Procedure

Encoding the procedure, or identifying "separate, executable steps" (Guthrie, Bennett, & Weber, 1990, p. 9) and "entering them into memory" (p. 5), is one of the cognitive processes hypothesized in the Guthrie model to be important in processing procedural documents. Procedures, by definition, contain some representation of separate actions, whether done verbally or graphically (e.g., maps, diagrams, blueprints, drawings). The reader must transform the verbal/graphic information to visual-spatial information upon which she or he can then act. Helping the reader encode efficiently can reduce the number and complexity of these transformations.

Following written directions has been found to impose a heavy cognitive load (Glover, Harvey, & Corkill, 1988; Glover, Timme, Deyloff, Rogers, & Dinell, 1987), so most people appear to iterate. That is, a step is encoded, executed, checked in various ways, and then the reader moves on to the next step, having consulted both text and graphics (Hayes & Henk, 1984). Researchers have identified a typical pattern of looking at picture-text amalgams (Berlyne, 1958; Loftus & Bell, 1975; Mackworth & Morandi, 1967; Stone & Glock, 1981; Zinchenko, Chzin-Tsin, & Tarakanov, 1963, cited in Mackworth & Bruner,

1970), describing the process in terms of locating 'gist' information first and then alternating between reading and comparing pictorial information to textual information. What the reader may be doing, instead, is the iterative process referred to above, which is required, in part, by the limited size of short-term memory.

Tools found to be helpful to readers in encoding instructions include listing (Carliner, 1987; Frase & Schwartz, 1979; Hartley, 1981; Schoff & Robinson, 1984), and the many formulas and guidelines used in military procedure-training aids and technical manuals (Braby, Hamel, & Smode, 1982; Curran & Mecherikoff, 1979; Johnson, 1976; Kern, Sticht, Welty, & Hauke, 1975; Siegel, Lambert, & Burkett, 1974; Terrell, Ewell, Scott, & Braby, 1983).

Summary

It appears, then, that we know something about the inadequacy of models of narrative text when applied to procedural documents (Diehl et al., 1989), and that perhaps measurement methods useful with nonprocedural documents transfer poorly to procedural materials (Mark & Bracewell, 1989).

Investigations into the effects of drawings (LeFevre & Dixon, 1986), and diagrams (Larkin & Simon, 1987) have helped elucidate effective ways to support and encourage the processing of procedural documents. The effects of various combinations of text and graphics on comprehension of procedural documents have been investigated, yielding some support for various treatments of written instructions.

Dixon's (1982, 1987a, 1987b, 1987c) investigations of the influence of action and condition information, and the order of presenting such information in written instructions, and Smith and Goodman's (1984) conclusions about the benefits of using explanatory material in hierarchically structured written directions have contributed to our understanding of the effects of outcome and procedure schemata. These researchers' conclusions seem to support the Guthrie model's hypothesis that providing both outcome and procedural schemata helps the reader of a procedural document.

Presenting procedural information in list form has been found to be an effective method of encouraging the reader's encoding of the procedure (Carliner, 1987; Frase & Schwartz, 1979; Hartley, 1981). Much of the military research into optimal presentation techniques for technical material has also focussed on helping the reader encode the procedure, with varying degrees of success (Duffy, Post, & Smith, 1987).

Our first study examined the usefulness of three of the Guthrie model's hypothesized information sources and two of the cognitive processes included in the model--forming a conceptual model and encoding the procedure. Our second study examined the model's remaining two cognitive processes--self-testing and self-correcting.

STUDY 1

This study tests certain components of Guthrie, Bennett, and Weber's (1990) transformational model of procedural document processing. Specifically it investigates the effects of altering the exposition of the outcome, the procedural steps, and the representation of the outcome.

Method

The Task

The task chosen for study was a button/shank sewing task. It was chosen for several important reasons: It was relatively easy to do without rich prior knowledge in a technical area, it was common enough to be included in most of the curriculum materials in a content area, it was challenging enough to produce a diversity of quality in the finished product, it was short enough to complete within an hour or so, and

it required small, inexpensive equipment. Another reason it was chosen is that one of the authors was both an experienced sewer and an experienced sewing teacher.

The button/shank task is a more complicated one than first impressions would suggest. The student must not only sew the button to the fabric, but must, in the process, create a shank--a reinforced column of thread that lifts the button off the surface of the fabric, allowing space for the layer of the garment contaiṅing the buttonhole.

Subjects

Subjects for this study were typical 7th-grade, 10th-grade, and postsecondary readers. Typical means a student who reads at about grade level. The 7th- and 10th-grade students' reading levels were determined from their scores on the reading component of Scientific Research Associates tests used in their school district. The adult subjects' reading levels were determined from their scores on the Stanford Diagnostic Reading Comprehension test (Karlsen, Madden, & Gardner, 1976). Adult students who participated in the experiment read at Grade 12+ level, or "grad level," the term used when scoring the Stanford Diagnostic Reading Comprehension test to indicate a person who reads at approximately Grade Level 13. Twelve students, 6 male and 6 female, made up the sample from each grade level.

The 7th- and 10th-grade students were selected from a rural high school and a rural junior high school by personnel in the attendance center (7th grade) or counseling center (10th grade). The school district from which both groups were chosen required one quarter of home economics for 7th-grade students, so some of the subjects had taken home economics prior to testing, or were taking it at the time of the testing. Seven 7th-grade students had taken home economics prior to testing. Five 10th-grade students claimed prior experience, but none mentioned the 7th-grade class as providing that experience.

Seventh-grade students were required to return signed parental consent forms to the researcher to participate in the experiment; 10th-grade students provided their own verbal consent. Seventh-grade students were removed from classes during morning hours over a 3-day period for the researcher's convenience. Tenth-grade students were taken from study halls throughout the school day over a four-day period. Adult students were selected on a volunteer basis from an adult basic education center. Each adult was paid $5 to participate in the experiment.

Materials

Three versions of the button/shank instructions text were used in this study. Two were chosen from commercially published textbooks--one the authors identified as low quality (Text A, Appendix A) and another identified as higher quality (Text B, Appendix B), from experience using both of them in undergraduate clothing courses. An experimental text for the task was developed by the authors (Appendix C). Several versions were written and edited before pilot testing the final draft on several people.

The experimental instructions began with a short expository paragraph to orient the reader, and a line drawing of the finished product, as Guthrie et al. suggest to help the reader form a conceptual model of the task. Line drawings were used to accentuate the details (see Dwyer 1971, 1972; Readence & Moore, 1981), and to provide what Guthrie et al. call the representation of the outcome. Heeding Dixon's (1982) conclusions, actions were mentioned first where possible, and in accord with the suggestion to provide organizational information prior to component step information (Dixon, 1987a, 1987b, 1987c) several steps were worded to indicate the goal state toward which the reader was working prior to the required actions. Steps were separated and numbered (Frase, 1981; Carliner, 1987; Hartley, 1981) to help the reader encode the process.

11

Advice was removed from the steps, and "pointers" were presented separately but on the same page to encourage problem prevention and problem solving. For example, rather than including the instruction to "not pull the thread so tight that the toothpick will not slide back and forth" in the action/step information, it was placed in the "Pointer" column, the far right third of the page, as a warning to the student, and to help eliminate problems later in the procedure.

Each subject was provided with the following items: a 5" x 5" piece of white cotton/polyester broadcloth; a 5/8" black sew-through button with four holes; a hand sewing needle ("sharp" or "between," size 7-9) threaded with a knotted, double strand of #9342 red polyester Mölnlycke thread; a round toothpick; a blue chalk pencil; a small pair of embroidery scissors; and a sample buttonhole worked in a fabric swatch. Some of the equipment was unknown to a few of the subjects (e.g., the chalk pencil).

Dependent Measures

The dependent variables in this study were without-instructions sewing time, with-instructions sewing time and without- and with-instructions button ratings from four evaluation scales. Both without- and with-instructions sewing times were determined by timing the students' performance on videotapes. The time started when the subject first purposively picked up a piece of equipment and ended when the subject announced she or he had finished, or picked up the buttonhole sample to attach the button to it. Thus, with-instructions sewing time includes both reading and sewing time, while the without instructions sewing time did not include reading time because there was nothing to read during that phase of the experiment (see later discussion).

A scoring procedure was developed and evaluated that would allow quantification of the 72 button-sewing products (without- and with-instructions buttons from 36 subjects). First, the authors developed a 7-point rating scale for each of four important button/shank sewing criteria--shank, security, appearance front, and appearance underside (see Appendix C)--which reflected possible variations associated with each criterion.

Second, these verbal descriptions of ratings scales (from a low score of 1 to a high score of 7) were used by two judges to rate a representative sample of 18 of the 72 buttons. Then based on the results of a correlational analysis and a comparison of means between the two judges' ratings, the scales were found to need revision. Thus, "prototype" buttons were added, so that a judge could not only read the verbal descriptions of each scale rating, but also see an example illustrating that scale value. For example, there was a button illustrating what a "Shank value of 2" looked like. When the scaling procedures were tested with three other judges, more consistent ratings among the judges were found. A sample of the Pearson Correlation Coefficients follows: Shank scale: Judges 3 and 4, $r = .837$; Judges 3 and 5, $r = .942$; Judges 4 and 5, $r = .871$; Security scale: Judges 3 and 4, $r = .798$; Judges 3 and 5, $r = .889$; Judges 4 and 5, $r = .772$; Appearance Front scale: Judges 3 and 4, $r = .904$; Judges 3 and 5, $r = .819$; Judges 4 and 5, $r = .809$; Appearance Underside scale: Judges 3 and 4, $r = .829$; Judges 3 and 5, $r = .829$; Judges 4 and 5, $r = .886$. Also, the difference among the mean ratings of the three judges on the four scales was not significantly different.

Procedures

Prior to beginning a session with a student, the researcher explained to the student that she or he would be videotaped while performing a simple sewing task, and that she or he would be asked to think-aloud while sewing. A short discussion about thinking-aloud occurred at this point. The think-aloud strategy and the videotaping were used so that more could be learned about the cognitive activities of the students, as well as their opinions about the various text versions. Each student's face and hands, as well as the instruction set and equipment were visible in the camera's eye.

After the pre-sewing briefing, each student was given the materials described earlier and two pieces of fabric--one plain and one containing a buttonhole. The student was then requested to sew the button onto the plain piece of fabric in such a way that it could be attached to the second piece of fabric by using the buttonhole. This activity was thought to provide a measure of how much each student knew about sewing on buttons prior to reading the text.

After the without-instructions button-sewing task was completed, each student was given one of the three versions of the instructional text and replacement materials and was requested to use the written instructions while sewing another button. She or he was also reminded to continue to think-aloud. Following the with-instructions sewing task each student was debriefed with a short explanation of the purposes of the experiment.

Each subject sewed both buttons during the same observation. Seventh-grade students were videotaped in an ?' x 10' windowless room in their school that contained a large desk, student chair, small end table, videocamera on tripod, and the researcher. Tenth-grade students were videotaped in an 8' x 8' room with one window in the counseling center of their school. The room contained only a desk and student chair, the videocamera, and the researcher. The adult students were videotaped in a 18' x 20' windowless room that contained several large tables pushed aside to leave one for the student to use, and several chairs, as well as the videocamera and the researcher. The experiment took place over a 2-week period.

Scoring

Each of the 72 buttons was rated on the four evaluation scales by one of the authors. These scale values ranged from 1 to 7 on each scale.

Research Design

A 3 x 3 x 2 factorial design (3 text levels, 3 age levels, and 2 gender levels) was used to analyze the effects of the instructions on each of the four without-instructions scale scores. All of the independent factors in this design were between-groups comparisons.

A similar 3 x 3 x 2 factorial design was used to analyze the with-instructions scores, with the exception that the without-instruction scores and sewing time were used as covariates.

Independent variables in the statistical analyses were: (a) three levels of instructional text (Text A--lower quality, Text B--higher quality, and experimental text), (b) two levels of gender, and (c) three age levels (7th grade, 10th grade, and adults).

Covariate measures included without-instructions time (time measured from when the subject first purposively picked up a piece of equipment and ending when the subject announced she or he had finished), and without-instructions ratings of the buttons on four evaluation scales.

Dependent variables included the without-instructions sewing times, with-instructions sewing times, the without-instructions button ratings on the four scales, and the with-instructions button ratings on the four scales.

All statistical analyses were done using the SYSTAT program (Wilkinson, 1989). The Bonferroni procedure (see Wilkinson, 1989) for making post hoc comparisons was used to identify significant differences among the three text levels and three age levels.

The videotapes were transcribed into written documents to help determine how the texts were processed and how the equipment was used.

Results

This section presents the results of analyses of sewing times of the without-instructions buttons, followed by the results of analyses of the without-instructions button ratings. The results of analyses of sewing times for the with-instructions buttons and the with-instructions button ratings follow. The results of the study of the videotapes are presented last.

Without-Instructions Measures

Sewing Time

When without-instructions sewing times were used as the dependent variable in an analysis of variance, no significant main effects or interactions were found among any of the independent variables. The various experimental groups started the experiment largely equivalent to each other on the time taken to sew on a button.

Ratings

When the button shank scores were used as dependent measures in an ANCOVA, there was a significant effect due to age level, $F(2, 18) = 5.486, p = .014$. Post hoc analyses showed that the 7th-grade students' mean score ($M = 2.417, SD = 1.505$) was significantly higher than the 10th-grade students' mean score ($M = 1.083, SD = .289$), $F(1, 18) = 10.971, p = .004$. There were no other significant differences in shank quality among the three age levels, text types or genders.

However, gender had a significant effect on each of the other scales, unlike text type and age level. On the security scale the males ($M = 2.389, SD = 2.118$) scored significantly lower than the females ($M = 4.611, SD = 1.720$), $F(1, 18) = 16.327, p = .001$. On the front appearance scale, the females ($M = 4.333, SD = 1.715$) again outscored the males ($M = 2.944, SD = 1.955$), $F(1, 18) = 7.022, p = .016$. On the underside appearance scale, the same result was obtained (males: $M = 1.889, SD = 1.231$; females: $M = 3.500, SD = 1.339$). $F(1, 18) = 16.490, p = .001$, with females creating better underside appearances.

Significant interactions. There was a significant interaction between text and age level in the shank scale, $F(4, 18) = 4.200, p = .014$, but not in the other scales. This was due to the fact that some of the 7th-grade students had some instruction on button sewing that affected their without-instruction product.

Summary of analyses of without-instructions measures. There were no significant time differences among the groups on the without-instructions button ratings. Overall button/shank quality was higher among the female subjects, and an interaction between text and age level was noted. Only the shank without-instructions measure showed significant differences due to age level. So, it appeared that on the without-instructions measures, only gender had a systematic, significant impact on the overall quality of the finished product.

With-Instructions Measures

With-instructions measures included sewing time and the scores on the four rating scales. Analyses of covariance were done on the with-instructions scores of the buttons, using the without-instructions rating score as covariate for its corresponding with-instructions rating score. These covariance analyses were performed even though we were concerned about the effect of the systematic relationship between

14

gender and the without-instructions measure on all of the scales, and the nuisance interaction between grade level and text type on the shank without-instructions score.

Sewing Time

When with-instructions sewing times were used as the dependent variable in an analysis of variance, the following results were obtained. The text effect was significant, $F(2, 18) = 11.829, p = .001$. No significant effects were found due to age level or gender and there were no significant interactions.

Bonferroni test results revealed that students using the experimental text ($M = 15.264, SD = 8.310$) took significantly more time than those using either Text A ($M = 6.199, SD = 3.122$), $F(1, 18) = 15.330, p = .001$, or Text B ($M = 4.944, SD = 1.781$), $F(1, 18) = 19.865, p = .000$. Using the experimental text required students to spend almost three times as much sewing time as using either of the other two texts.

Ratings

Shank scale. When an analysis of covariance was used on the shank with-instructions rating scale, using the without-instructions shank score as covariate, main effects were revealed for both text, $F(2, 17) = 11.266, p = .001$, and age level, $F(2, 17) = 6.543, p = .008$, but not for gender. No significant interactions occurred.

Students who used the experimental text made better shanks than students who used either Text A or Text B. Post hoc comparisons using the Bonferroni procedure revealed that the text effect occurred between Text A ($M = 2.667, SD = 2.015$) and the experimental text ($M = 5.667, SD = .778$), $F(1, 17) = 18.229, p = .001$, and between Text B ($M = 3.083, SD = 2.193$) and the experimental text, $F(1, 17) = 15.459, p = .001$, with shanks made by students using the experimental text being of higher quality. Scores of shanks made by students using Text A and Text B did not differ significantly from each other.

The age level effect was examined in post hoc comparisons using the Bonferroni procedure. Only the adults ($M = 5.083, SD = 1.929$) made better shanks than the 10th-grade students ($M = 2.750, SD = 2.179$), $F(1, 17) = 12.504, p = .003$. The buttons of the adults and 7th graders did not differ from each other, nor did those of the 7th and 10th graders.

Security scale. When an analysis of covariance was used on the security with-instructions ratings, using the without-instructions security score as covariate, main effects were revealed for text, $F(2, 17) = 16.730, p < .0001$, and age level, $F(2, 17) = 12.638, p < .0001$. A significant interaction occurred between age level and gender, $F(2, 17) = 4.341, p = .030$.

The security of buttons sewn by students using the experimental text was better than that of buttons sewn by students using either of the other texts. Bonferroni test results revealed significant differences only between security of buttons made by students using Text A ($M = 3.333, SD = 1.670$) and those using the experimental text ($M = 6.0, SD = 1.206$), $F(1, 17) = 29.269, p < .0001$, and between buttons made by students using Text B ($M = 3.667, SD = 2.387$) and those using the experimental text, $F(1, 17) = 21.27, p < .0001$.

Using the Bonferroni procedure revealed that the adult students ($M = 5.750, SD = 1.545$) scored significantly higher than both the 7th-grade students ($M = 3.500, SD = 2.195$), $F(1, 17) = 24.563, p < .0001$, and the 10th-grade students ($M = 3.750, SD = 2.006$), $F(1, 17) = 10.699, p = .005$ on the security evaluation.

The interaction between age level and gender, $F(2, 17) = 4.341, p = .030$, indicated that as the age of the male subjects increased, so did the quality of their with-instructions button/shanks. (These with-instructions security rating means and standard deviations are simple gain scores that have been adjusted based on the without-instructions security ratings. Males: 7th grade: $M = 2.861, SD = .459$; 10th grade: $M = 4.401, SD = 1.160$; Adults: $M = 6.041, SD = 1.791$.) The female subjects' gain scores did not display the same magnitude of increase (7th grade: $M = 4.049, SD = 1.223$; 10th grade: $M = 3.730, SD = 1.916$; Adults: $M = 4.918, SD = 1.252$).

Appearance front scale. When an analysis of covariance was used on the appearance front with-instructions ratings, using the without-instructions appearance front score as a covariate, main effects were revealed for age level, $F(2, 17) = 8.812, p = .002$. Neither text nor gender contributed to significant effects, nor were there any significant interactions. The covariate (without-instructions appearance front rating) contributed a significant amount of the variance, $F(1, 17) = 12.485, p = .003$.

Bonferroni test results revealed that students using the experimental text ($M = 4.667, SD = 1.969$) performed significantly better than students using Text A ($M = 3.667, SD = 2.015$), $F(1, 17) = 5.643$, $p = .030$, on the front appearance of the buttons, but not better than students who used Text B. There were no differences between the appearance front scores of buttons sewn by students using Texts A and B.

The adult students made better looking buttons than either the 7th-grade or the 10th-grade students. Bonferroni test results revealed significant differences only between the 7th-grade students ($M = 3.5$, $SD = 2.236$) and the adult students ($M = 5.5, SD = 1.314$), $F(1, 17) = 12.265, p = .003$, and between the 10th-grade students $M = 3.167, SD = 1.749$) and the adult students, $F(1, 17) = 14.148, p = .002$).

Appearance underside scale. When an analysis of covariance was used on the appearance underside with-instructions ratings, using the without-instructions appearance underside score as a covariate, main effects were revealed for text, $F(2, 17) = 6.785, p = .007$. Neither the effects due to age level nor those due to gender differences reached significance. There were no significant interactions.

Bonferroni test results revealed that students using the experimental text ($M = 4.167, SD = 1.749$) significantly outscored students using either Text A ($M = 2.167, SD = 1.115$), $F(1, 17) = 13.372, p = .002$, or Text B ($M = 3.0, SD = 1.477$), $F(1, 17) = 4.878, p = .041$. There was not a significant difference between the buttons sewn by students using Text A and Text B.

Summary of analyses of with-instructions button ratings. Using the experimental text required the students to spend about three times as long on the task as either Text A or Text B. As was hypothesized, those students also made better button shanks.

Both text and age level created significant main effects on the with-instructions shank and security scales. Using the experimental text enabled students to make better shanks and more secure buttons; using Text A or Text B did not help students do this. Significant differences in shanks occurred only between the adult students and the 10th-grade students. The adult students were able to make more secure buttons than both the 7th- and the 10th-grade students.

Only age level created a significant main effect on the front appearance of the buttons. The adult students made better looking buttons than either the 7th- or 10th-grade students. On the appearance front scale the overall text effect was not significant, but students using the experimental text made better looking buttons than students using Text A. Text B appeared to be adequate.

Only the text factor showed a main effect on the underside appearance of the buttons. The students who used the experimental text made better looking undersides.

The only significant interaction found in the with-instructions measures analyses was between age level and gender on the security scale. This was a partial result of the adult male subjects' showing a large improvement score when given instructions for the task.

Summary of Comparisons Across Independent Variables

Comparisons of with-instructions scores revealed that Text A was very poor in helping students make good button shanks, that Text B was only slightly better, and that the experimental text was very effective for the 10th-grade and the adult students. The 7th-grade students only showed improvement on the shank scale when using Text A and the experimental instructions; otherwise their with-instructions scores differed little from their without-instructions scores. The 10th-grade students also benefitted only from using the experimental text. The adult students were able to use all the texts effectively.

Both genders benefited from using instructional text for the button-sewing task, with the males showing much greater improvement scores. A group of 10 students did not benefit from using any instructions. Of these, 5 were male and 5 were female; 4 were 7th-grade students, 5 were 10th-grade students, and 1 was an adult; 6 had used Text A, 3 used Text B, and 1 used the experimental text.

Qualitative Data

Examining the qualitative data from the videotapes revealed individual differences among some of the subjects in coordination and in the various strategies used to follow the instructions. About two thirds of the adult students exhibited better coordination than did the remaining one third. The women generally exhibited better needle handling and used the equipment more effectively than the men. Among the younger subjects only about half could be identified as having good coordination for this task. For example, some of the 7th-grade male subjects had great difficulty handling the equipment; they tended to leave the fabric and button on the surface of the table instead of holding it in their hands. They also dropped or lost the toothpick more frequently than did most other subjects.

Several text processing strategies used by subjects were observed on the videotapes. For example, many of the 7th-grade students read by tracing the lines of text with their fingers; only a few 10th-grade and adult students did this. The majority of 7th- and 10th-grade students also did a lot of rereading of the text. Look-backs (flipping through the preceding pages) could also be seen on many of the videotapes at places where subjects identified a mistake, or found the text difficult to understand. Five adults said that they were relying more on the pictures than on the text. One adult male subject read ahead in the text--an unusual strategy in this sample of subjects.

Guessing was also used as a strategy as evidenced by the language of several subjects, "I guess . . .," "I don't know what that means, I guess it means . . .," "I don't know . . . . I'll just do . . .". Among the subjects whose with-instructions scores were lower than their without-instructions scores, 8 frequently said that they were guessing. Of these, 5 used Text A, 2 used Text B, and 1 used the experimental text.

Another strategy used by some subjects was reading the entire text prior to starting the task. Of the 12 adults, 11 were observed doing this. Only 5 7th-grade students and 5 10th-grade students read the entire text first.

Some subjects did not read the entire instructional text, even while executing the steps. Four subjects (2 7th-grade students, 1 using Text A and 1 using Text B, and 2 10th-grade students, also 1 each using Texts A and B) did not read the introductory paragraph, but started immediately with the sewing instructions. Four subjects skipped reading some of the steps (3 7th-grade students and 1 10th-grade

student; 2 used Text A and 2 used Text B). Two adult women and 1 10th-grade male subject read the instructional text aloud, as requested, but paid very little attention to it.

Discussion

From the analyses of the quantitative data it was concluded that the experimental text seemed to be more effective in teaching students how to sew on buttons with shanks, and that the adult students benefited more than either the 7th-grade or the 10th-grade students. Text A and Text B were found to be more alike than different in their effects on students. The performances of the 7th- and 10th-grade students were very similar. While a significant text effect was expected in this study, the significant effect of age level was unexpected.

Comparisons Among the Three Texts

When the three texts were compared, possible reasons for the results were found. These comparisons will be discussed here as they relate to elements of the Guthrie model.

The Guthrie model suggests that procedural documents must include exposition of the outcome, or introductory material to enable the reader to form a conceptual model of the task--a mental idea of how to do it. Neither Text A nor Text B had adequate introductory exposition. Text A had two sentences of irrelevant material about sewing buttons so they "stay put" and no definition of the term "shank." Text B had four sentences somewhat more related to the task, but which used the term "shank" without a definition. Neither Text A nor Text B showed a finished shank labeled as such. Among the subjects who used these two texts were 4 subjects who did not read the introductory material, but instead, began with the instructions. Their with-instructions scores were lower than their without-instructions scores, which seemed to indicate that reading even inadequate exposition might contribute to improved performance.

In contrast, the experimental instructions included an expository paragraph, in which both the definition and purpose of a shank were given, and a drawing of a finished button/shank with the shank clearly labeled. The statistical analyses supported the conclusion that the experimental instructions were more helpful to students than either of the other two texts; this might be a result, in part, of differences in the expository material.

The Guthrie model also suggests that the use of a procedural schema can help the reader encode the steps. No evidence of any procedural schema could be found in either Text A or Text B--that is, neither had either numbered sequences or information presented as distinct steps. In both texts the instructions were presented in paragraph form, with no indentation, underlining, or other typographic tool used to identify separate steps. The lack of a procedural schema seemed to create problems for some subjects using Texts A and B; a few revealed in their thinking-aloud that they had lost their place and were trying to relocate in the text. Transcripts also revealed that many subjects using Texts A and B reread previously used information, or information out of order.

A procedural schema was used in the experimental instructions. The experimental instructions were divided into four major segments (positioning button, forming stitches, making the shank, and securing thread), and individual steps in each segment were numbered and clearly separated from each other. Subjects who used the experimental instructions did not verbalize as much difficulty in maintaining their place in the text or in understanding and following the order of the steps as subjects who used either of the other two texts.

The representation of the outcome, whether conveyed by pictures or in sentences, is an information source hypothesized in the Guthrie model as essential to successful procedural document processing.

The three texts used in this experiment differed greatly in how the outcome is represented, both pictorially and verbally. Both Texts A and B had only two illustrations each--apparently too few to convey to the reader the various needle, button, and toothpick positions and movements. In Text A there were letter labels on the photographs, but one was hidden in the shadow of the toothpick, and in Text B neither drawing had a number or letter label to help the reader decide where to look. Text A had photographs, which reproduced poorly, even in the original, and Text B had line drawings. In all four illustrations the underside of the fabric was not shown and little perspective was given to help the student calculate hand positions relative to button and fabric.

In contrast to Texts A and B, the experimental instructions included 15 line drawings. The two sides of the fabric were differentiated from each other, the button location marking was visible, and many needle and button positions required by the task were illustrated. Some of the drawings also showed that the needle pierced the fabric and emerged on the underside, which neither Text A nor Text B included. The with-instructions button scores seemed to support Guthrie et al. in their call for adequate graphic representation of the outcome, in that the experimental instructions enabled students to make better button/shanks.

The sentences used in Texts A and B to represent the outcome also seemed to be inadequate. In these two texts many vague terms were used, and few definitions were given. Instructions are equally vague (e.g., "Sew through holes of button sewing over pick" in Text A, and "Wind thread firmly around stitches to make shank" in Text B). Among the 10 subjects whose with-instructions scores were lower than their without-instructions scores, 5 used Text A, and 2 used Text B, and all 7 of these stated that they were guessing what to do while following the instructions. The guessing was probably related to the vagueness of the instructions and the lack of definitions.

Conversely, students who used the experimental text were able to find a definition of "shank" in the first paragraph. They also could read specifically stated sentences in which there was little chance to misinterpret the information (e.g., "Repeat Steps 5, 6, and 7 using the same two holes until you have 3 or 4 stitches around the toothpick.") The fact that students who used this text made better with-instructions buttons than the other students seemed to support the Guthrie model's demand for adequate verbal representation of the outcome.

In developing the experimental text, we made an effort to avoid the problems associated with inadequate exposition, the lack of procedural schema, and inadequate representation of the outcome. Although the experimental instructions were lengthy, and as previously noted, required almost three times as long for processing as either Text A or Text B, it appeared that there were characteristics of the experimental text that were helpful to students. More adequate exposition (including a definition of "shank"), the picture of a finished shank labeled as such at the start of the instructions, the use of many line drawings and numbered steps all appear to have contributed to improved task performance.

The improved scores of students using the experimental text must be examined in light of the increased time to perform the task, however. If a student takes more time, it is likely that a better product will result. The question to be considered then, is whether it matters that students take much more time to sew on buttons, if they do it well. Considering that most students will probably improve with practice, perhaps sewing the first few buttons slowly becomes less important. The importance of the time factor probably also varies from task to task.

Comparisons Among Age Levels

As previously reported, the male adult students seemed to benefit more from instruction in general, and all the adult students benefitted from the experimental text in particular. The adults seemed more aware of helpful reading strategies, such as reading the entire text prior to starting. Five of the 12 adult

subjects also said that they had used the pictures more than the text. The younger subjects used the text about as much as they used the pictures, although the pictures were heavily used in problem solving.

There were differences among age groups in thoroughness, too. The only subjects who did not read the entire text were 6 younger students (4 7th- and 2 10th-grade students. All 6 were part of the group whose with-instructions scores were lower than their without-instructions scores.) The differences in thoroughness of reading might indicate age level differences in using procedural documents.

The statistical analyses and the videotape data supported the conclusion that adults appeared much better able to use procedural text. Perhaps 7th- and 10th-grade students are not taught enough about how to use procedural text. Perhaps they encounter it too infrequently and thus are not afforded the opportunities to practice. The differences in physical coordination between adults and children might also account for the observed differences in performance. For example, the younger students seemed to have more trouble with tangling thread and knots than did the adult students. Adults, too, might have more experience with buttons, and greater understanding of the entire concept of "attaching." Many of the 7th- and 10th-grade students seemed unaware of the idea of security, or how to achieve it--their buttons were attached with one or two stitches, instead of five or six, as were many of the buttons sewn by the adult students. Some of the adult students also mentioned a concern with the appearance of the button, which only a few of the younger students did.

A possible problem in this study might arise from the language in the experimental text. The experimental instructions were pilot tested with adults. In so doing, the language in the instructions might have been biased against younger, less able readers.

The Effects of Prior Experience

In observing the videotapes we identified other reasons which might explain the results. Many of the adult women and some of the adult men displayed behaviors that indicated they came to the task with fairly rich prior experience. Their needle handling and button/fabric holding techniques showed their ability and their coordination. The 7th- and 10th-grade students exhibited these abilities far less frequently. Many of the 7th-grade students were completely unable to hold everything (button, fabric, needle, and toothpick) simultaneously. Younger students spent more time retrieving and replacing the toothpick than the adults did, and more time just looking at the materials.

Summary

In summary, this study supported part of the Guthrie model. The three information sources tested (exposition of the outcome, the procedural steps, and the representation of the outcome) appeared to be important influences on two cognitive processes hypothesized to be necessary for successful task performance--forming a conceptual model of the task, and encoding the procedure. Good introductory exposition and a picture of a labeled shank, accompanied by precise graphic and verbal representation of the outcome in the experimental text seemingly helped the reader, as indicated by improved with-instructions button/shank scores. Using the listing strategy to indicate and number steps in the procedure in the experimental text apparently helped the readers encode the process, too.

A second study was undertaken to continue the examination of the Guthrie model. Specifically, it examined the hypothesized cognitive processes used by readers to monitor comprehension of written instructions.

STUDY 2

After studying the relationships between the three information sources and two of the cognitive processes hypothesized to be necessary for successful processing of procedural documents, and finding support for Guthrie et al.'s suggestions, further study seemed warranted. The readers' use of the two cognitive processes, self-testing and self-correcting, was studied.

Metacomprehension

The Guthrie model suggests that individuals reading a procedural document need encouragement to use two types of cognitive processes: self-testing (recognizing when understanding has failed) and self-correcting (knowing what to do at those times and being able to do it) (see also Schorr, 1982). These activities have elsewhere been discussed as part of metacognition (Flaveil, 1979), that is, knowledge and awareness of, and the ability to control and adjust one's own cognitive processes (Mayer, 1987; Reeve & Brown, 1984). Others refer to these cognitive processes as necessary reading comprehension processes (Gagne, 1985; Guthrie, 1983). Procedural documents, the Guthrie model suggests, must foster such activities, perhaps more than expository texts (see also Glock et al., 1984).

Winograd and Johnston (1980) included the ability to self-check and self-correct in their definition of a fluent reader. Self-interrogation and self-regulatory activities of metacognition (Brown & DeLoache, ·1977) include goal setting, strategy selection, checking outcomes, planning the next move, monitoring effectiveness (testing, revising, and evaluating one's strategies), and remediation (Brown, 1978; Gagne, 1985) or the use of "fix-up" strategies (Wagoner, 1983).

We have evidence that young and/or poor readers do not spontaneously self-regulate their reading comprehension (Brown, 1975, 1980; Markman,1979; Reeve & Brown, 1984; Trepanier, 1982), and that adults and college students, and better readers in general, are aware of their cognition and able to spontaneously self-regulate, albeit inefficiently at times (Baker & Anderson, 1982; LeFebvre-Pinard, Bouffard-Bouchard, & Decary, 1983; Phifer & Glover, 1982; Smith, 1985). Baker and Anderson (1982) and Smith (1985) reported that comprehension monitoring may also reflect great individual differences.

We do not, however, have much evidence of self-regulation during the use of procedural documents. Most of the comprehension monitoring research has tried to identify strategies used by readers and to document their use. Typical tasks are listening, recall of lists or categorization of items on a list; typical texts used are expository or narrative (see Brown & DeLoache, 1977; Garner & Reis, 1981; Palincsar & Brown, 1984; Trepanier, 1982; Wagoner, 1983). Determining if earlier findings transfer to a new text genre seems both appropriate and necessary.

Although good readers of prose actively monitor their comprehension (Baker & Anderson, 1982; Garner & Reis, 1981; LeFebvre-Pinard et al., 1983; Smith, 1985; Wagoner, 1983), it may be that readers of procedural documents require more frequent, more precise, and better timed self-monitoring (Guthrie, Bennett, & Weber, 1990) due to the specific nature of the outcome of procedural document processing (Mark & Bracewell, 1989). It is difficult to create a correct outcome without seeing that each step in the process is done correctly (Hayes & Henk, 1986).

Comprehension monitoring research using procedural text. Only one study of comprehension monitoring was located in which the text was procedural in nature. In this experiment (Schorr, 1982), the application of comprehension monitoring strategies was studied using written directions to assemble a model. Three text versions were used--text only, pictures only, and text combined with pictures. Differences in monitoring activities used by college students were documented through videotaped protocols, interviews, and prompted recalls. Schorr developed a taxonomy of metacognitive strategies applicable to procedural tasks; the taxonomy included planfulness, ways of following the instructions,

evaluation of errors, and reactions to mistakes or problems. However, the relationship between the strategies and effectiveness of performance was not clear; those who used both a greater variety and a greater number of strategies did not perform most efficiently. Subjects who used the text-only and pictures-only versions of instructions were the ones who used a greater number and a greater variety of comprehension monitoring strategies.

While Schorr's (1982) results helped identify comprehension monitoring strategies that might be unique to procedural documents, they also yielded confusing conclusions about the efficacy of using those strategies. Perhaps the comprehension monitoring processes used while reading procedural documents are more complex than Schorr's taxonomy captured, or perhaps they relate to document features such as the presence of text or pictures. It appeared that the increase in the use of comprehension monitoring strategies in this experiment might be confounded with the differences in the three text versions.

Comprehension monitoring is a name for a category of processes. Specific strategies used by readers include self-testing and self-correction. Studies investigating readers' use of these strategies are reviewed below.

Self-testing processes. Readers must be able to recognize barriers to understanding before they can make corrections, that is, they must become aware of the level of their understanding (Pitts, 1983). Activities that lead to this awareness have been called "goal checking" (Gagne, 1985) and "evaluation" (Baker, 1979a), both meaning assessment of understanding. There is some evidence that self-questioning techniques can be taught and used effectively to this end (Andre & Anderson, 1978-79; Palincsar & Brown, 1984).

Not surprisingly, a developmental trend in learning to goal-check has been identified (Harris, Kruithof, Terwogt, & Visser, 1981). A reader must know something about her or his own meta-memory processes before being able to choose and apply an appropriate strategy (Brown, 1975).

Some adult readers seem to know at least one strategy specifically concerned with regulating progress toward the reading goal or assessing the level of their understanding and recall during studying (e.g., self-testing, evaluating understanding, or evaluating recall) (LeFebvre-Pinard et al., 1983). Law students have also been shown to use a form of self-testing while reading legal case write-ups. Frequently they admitted "I'm confused," apparently having used some type of self-testing to arrive at that conclusion (Lundeberg, 1987).

In a study by Brown (1979b), college students confronted with textual inconsistencies spent greater amounts of time rereading, a correcting behavior, apparently made necessary by their having noted that they did not understand. This seemed to indicate self-testing during reading. In another study (Schorr, 1982), videotaped protocols, interviews, and prompted recalls of college undergraduates were analyzed, and results documented readers' use of self-checking activities while reading instructional materials (also see Glock et al., 1984).

Self-correcting behaviors. Collins and Smith (1980) suggested six remedial activities useful when comprehension fails: ignore the obstacle and read on, suspend judgment, form a tentative hypothesis (guess), reread the current sentence, reread previous context (look back), and ask an expert. Mature readers have been found to spontaneously use some of these--guessing, rereading, reading on, and ignoring trivial confusions. (Baker, 1979a; Baker & Anderson, 1982; Hare & Pulliam, 1979). Older, better comprehenders tended to use more look-backs than other good and poor comprehenders (Garner & Reis, 1981; Smith, 1967), and good readers also looked ahead for help (DiVesta, Haywood, & Orlando, 1979). The use of fix-up strategies such as ignoring and reading on, suspending judgment, and guessing (Pitts, 1983) may cause serious problems in the performance of a procedure, or, in the case

22

of asking an expert, completely remove the need for the instructional materials. Look-backs, reading ahead, and rereading may prove to be more useful in procedural documents.

Summary

Study of comprehension monitoring, specifically, self-testing and self-correcting behaviors, while of great interest in various research settings, seems to be somewhat limited in the types of tasks and texts studied. However, Schorr (1982) has suggested a taxonomy of comprehension monitoring strategies used by readers while processing procedural documents, which presents possible new metacognitive processes not previously identified as such.

Researchers studying procedural document processing may be able to replicate the findings reviewed above from studies of comprehension monitoring of readers using nonprocedural documents. However, the differences between procedural and nonprocedural documents and the required processing of each may invalidate those findings. The requirements of reading-to-do tasks may create vastly different cognitive demands compared to tasks that require only reading. Seeking to understand comprehension monitoring during the use of procedural documents can contribute to both the understanding of metacognition itself and the creation of more useful procedural documents. It is to this end that Study 2 is directed.

Method

This study focused on the self-testing and self-correcting cognitive activities of readers. The button-sewing task used in the first experiment was also used in this study.

Subjects

Subjects were 105 high school students who volunteered to participate and whose most recent reading scores were at or above the 50th percentile. One exception was made for a 10th-grade female subject whose score was at the 48th percentile. The subjects resided in four small rural towns in Central Illinois.

Subjects' reading scores, both percentiles and grade equivalents, were obtained from test results made available by each school. The Iowa Test of Achievement Proficiency was used by two of the participating schools; the Iowa Tests of Basic Skills by two schools; and the Stanford Achievement Test by one of the participating schools.

Materials

Three versions of the experimental button/shank instructions used in Study 1 were used in this study. The same experimental text was used with only minor changes made as a result of analyzing the think-aloud protocols. Specifically, a few sentences were reworded to be more clear, and five of the pictures were refined or slightly changed.

The two new versions of the text included features that would enable the researcher to learn about the subject's self-testing and self-correcting cognitive activities. One text version was written with *3 interruptions*, and another with *5 interruptions*--places in the text where the student was instructed to stop and evaluate her or his progress with the task.

Another version of the text was written so that once the subject was interrupted, she or he could *simply compare* her or his button to a picture, or the subject was required to *rate her/his button* against the drawing by circling a score on a furnished 5-point scale which ranged from a low score of 1 ("hardly

matches at all") to 5 ("perfect match"). Subjects were also provided with the same equipment as in Study 1.

Other written materials used in the study included a 4-item questionnaire given to each subject at the beginning of the button sewing, and a 4-item post-task questionnaire. The first document asked the students to indicate their prior experience with button sewing, that is, to record on a scale how many buttons they had ever sewn, and whether they had ever been taught how to sew on a button or to make a shank. The last question asked the students to mark on a 4-point scale the importance they attached to knowing how to sew on a button. The second document was made up of four questions about making shanks, for example, "Why must a button have a shank?" The first questionnaire was included in the study to provide information about the subjects that might influence their performance. The post-task questionnaire was included to provide information about fact learning which might occur during the performance of the procedure.

Procedure

The experiment took place over a 4-week period in April and May 1990. Each subject who agreed to take part in the experiment was given the sewing equipment listed above and sewed the first button, with no instructions, in groups ranging from as small as 1 student to as large as 40 students spread out in a large study hall room. They were asked simply to sew the button on the best way they knew how, or to use previously learned procedures, and told that the button would serve as an indication of their prior knowledge about the task. Many students were curious about the presence of the toothpick with the sewing equipment. Their questions were deferred to the second button-sewing session.

Subjects pretested in groups were instructed by the researcher and, in one school, by the study hall teacher, not to talk to or watch each other during the without-instructions button-sewing session. Desks and/or students were placed approximately 2'-3' apart. When students had finished sewing the first button they were given a 4-item questionnaire that included questions about the their prior experience with button sewing and the value they attached to the task.

Following the without-instructions button sewing, each subject was assigned to one of the five experimental groups (based on the text version that they used) in the following way. The buttons sewn by the students were categorized into those sewn by males and females, and rank ordered by the researcher. Beginning with the best button, each button in a consecutive group of five was assigned randomly to one of the five treatment groups by using a table of random numbers.

The session began by reminding the subject that she or he would be videotaped while sewing the button to the fabric using the provided instructions. Then the subject read a cover sheet that provided an explanation of the instructions. The cover sheet was used to make the subjects aware of the interruptions in the text and the rating scales, where appropriate. The subject also read that she or he could start over at any time during the procedure.

Following the reading of the cover sheet, the researcher took a few minutes to clarify what had been read, if necessary. Each subject was then asked to read the instructions aloud while sewing the button, and also to do any rereading out loud. Videotaping was done with very little interchange between the subject and the researcher. If a subject asked the researcher a question about the sewing task, the response was that she could not answer questions about the sewing; if a problem of another sort arose, the researcher found a solution (e.g., re-knotting a thread, furnishing a new toothpick, saying "do the best you can," etc.). After the subject finished sewing the button, she or he was given the post-task questionnaire made up of four questions about shanks.

Research Design and Statistical Analysis

The model used to analyze most of the data was a 3 x 2 x 2 x 4 x 2 x 4 incomplete factorial design. Four of the factors were between-groups factors: three levels of self-testing (0, 3, and 5), two levels of opportunity to self-correct (required, not required), two levels of gender, and four grade levels. Two of the factors were within subjects factors: two levels of instructions availability (without- and with-instructions) and four levels of button-rating scales (shank, security, appearance front, and appearance underside). The design was incomplete (all cells were not filled) because it was logically impossible for a group of students to get the text materials that did not have the rating scales and yet require them to make the ratings! Also included in this analysis was a covariate, discussed below.

Other dependent variables (besides the button ratings) tested in this design were: number and type of adjustments, answers to four post-task questions, a composite self-assigned rating score for each subject who used a rating-required text, sewing time for the with-instructions button, and a coordination measure for each subject.

Variables serving as covariates were: coordination, grade-equivalent reading scores, and sewing time for the with-instructions buttons.

All possible main effects and interactions were tested. Because the interaction between the two text variables was not logically possible in the design discussed above, another strategy was used. A so-called reduced design was developed. In this design, the students who received the 0-interruptions treatment were not included in the analysis. Thus, it was a complete factorial design (2 x 2 x 2 x 4 x 2 x 4) and enabled a test for the interaction between the two text variables. Since there was not the slightest indication of a significant interaction between the two text variables (e.g., $p = .568$), only the full design was used in all ensuing analyses with the omission of the impossible interaction term in the model equation.

While all possible main effects and interactions were tested, the one of most interest was the interaction between the without- and with-instructions factor and the factors that manipulated text variables. The without-/with-instructions factor can be thought of as pre-/post-treatment conditions, and if its interaction were significant, then it would be evidence that the text variable factor was influencing the amount of change between the without- and with-instructions levels. Other statistics of interest were the one indicating a main effects significant difference between the without- and with-instructions measure, and the ones indicating possible interactions with the gender factor. The latter would indicate a differential effect on the button-sewing task depending on whether the student was a male or a female.

In those cases where a multiple comparison statistic was needed, for example, when determining which among the four grade levels was significantly different from another, a Bonferroni statistic (discussed in Wilkinson, 1989) was used.

At times a variable, such as the measure of coordination, was used as a dependent measure in one analysis, and as a covariate in another. This strategy was a function of the research questions that were being asked. For example, when coordination was used as a dependent measure, it was a test to determine whether males or females are better coordinated, or whether the older students were better coordinated than the younger ones. At other times, coordination was used as a covariate to reduce error variance in the analysis since the measures of coordination and of button sewing were rather closely related.

Another type of analysis was implemented to determine the influence of the several student characteristic variables (the predictors) on their sewing performance (the criteria). Namely, what was the relationship between predictors (students' grade level, reading ability, coordination, prior sewing

experience, value attached to the task, gender, and sewing time on the with-instructions button), and criteria (their ability to sew a button without instructions, and also the difference between the scores when they did not use instructions and when they did). This was done by using a stepwise analysis of canonical regression (since there were m: .tiple criteria and multiple predictors). In this analysis, only one variable (and its interactions with other previously added variables) was added to the regression equation at a time, and the amount of error variance reduction in predicting the criteria was noted. Then another variable (and its interactions) was added and the additional reduction in error variance was noted.

The order of entering the variables was determined by first computing the simple correlation of all predictor variables with the criteria variables, and then rank ordering the predictors from the weakest relationship to the strongest relationship with the criteria. The variables were entered into the stepwise regression analysis from the weakest to the strongest.

Dependent Measures

Dependent variables in this study are: without- and with-instructions button ratings on four rating scales, the number and type of adjustments (a type of correction) made by each subject, answers to four post-task questions, a composite score developed from a subject's rating's of her/his button while using the rating-required text versions, sewing time for the with-instructions button, and a measure of each subject's coordination at the task.

Dependent variable values were obtained in the following ways. The button-scoring proced· e developed in Study 1 was used in Study 2 to obtain the without- and with-instruction button ratings. Both the written scales and the button examples used in Study 1 were used to evaluate the buttons in Study 2.

A typology of adjustments was created based on the examination of the videotapes. Activities used by subjects to repair errors or solve problems were noted and written on each subject's instruction set. After observing approximately one fifth of the videotapes, categories were collapsed to retain only major adjusting activities. Adjustments were divided into three main categories--those related to the text of the instructions, those related to the equipment, and questions asked of the researcher.

Adjustments related to the text included rereading immediately following the first reading, rereading previously read material following some intervening activity (either sewing or reading), looking back at previous text or illustration with no reading, looking ahead at either text or illustration with no reading, and reading ahead. Subjects were instructed to reread audibly to enable the researcher to identify these correcting behaviors; occasionally the researcher probed the subject to determine if inaudible rereading had occurred.

Adjustments related to the equipment included late execution (doing something later than originally instructed to, and usually to enable execution of current instruction, or to solve a problem), undo/redo (undoing some previously made needle, toothpick, button, or fabric movement, followed by redoing it, including turning the fabric over, and also including redoing by itself with no undoing), undo/untangle (undoing some previously made motion, including untangling thread and trimming threads), new start (beginning again with a new button), and checking the toothpick for freedom to slide under the stitches (included because so many subjects checked the toothpick in this way, and not only at places where advised in the "Pointer" column to do so).

The question-asking adjusting behavior constituted the third major class of adjustments, and was included because it seemed to represent a unique solution to any problem encountered. Some subjects apparently wanted verification of their interpretation of what had been read. The question-asking

adjusting behavior also seemed to provide insight into the apparent decision of the subject to defer to perceived expert authority, the perception that the subject had not understood the instruction and might not even on rereading, and as a comment on the clarity of the instructions.

A 4-question post-task questionnaire was included in the study to provide a measure of fact learning that might occur during the execution of the procedure. Four simple, basic questions were written to determine if students had learned the most fundamental facts about making shanks from an instructional text. Questions included the following:

1. What is a shank?
2. Why must a button have a shank?
3. Why did you use the toothpick? and
4. Why was it important to make sure the needle went through the X marking on each stitch?

After the subjects finished sewing the with-instructions button, they were given the questionnaire and wrote their responses. The answers to the questions were scored by first identifying the best answer to each question and counting the number of unique ideas in those four answers. A maximum possible score for each question was thus found. All the answers were then scored with maximum scores as follow: Question #1=4; Question #2=2; Question #3=3; Question #4=3.

A self-assigned rating composite score for each subject using a rating-required version of the text was created by finding the mean of the ratings each subject assigned her or his with-instructions button at the rating interruptions in the text. Only three ratings were used so that scores of subjects using both the 3- and the 5-interruptions versions could be combined.

Sewing time, to the nearest second, for the with-instructions button was measured by watching and timing the videotaped sewing performances with a stop watch, eliminating time spent reading.

A coordination measure was obtained for each subject by watching the videotaped performance of each subject at Step 5. Step 5 is the first step in the procedure that required the subject to hold all of the equipment simultaneously--fabric, button, and toothpick--while positioning the needle and pushing it through the button and fabric. A simple scale of 1 = poor coordination at this task to 5 = excellent coordination at this task was used. Certain patterns emerged that seemed to indicate varying levels of coordination. For example, students whose coordination score was 1 either dropped the toothpick and/or button several times or could not hold the button and fabric in their hands'simultaneously, but tried to leave it on the surface of the table. Students whose score was 2 devised a way to slide the fabric and button to the edge of the table and then managed to hold the toothpick on the button while trying several times to pierce the fabric with the needle. Students whose coordination scores were either 4 or 5 demonstrated greater manual dexterity, and also showed some of the needle handling techniques of experienced hand sewers.

Independent Variables

The independent variables included the manipulated text features (the number of interruptions and the requirement/option to rate the button/shank), the gender of the subject, and the grade level of the subject. There were five different texts that varied in these systematic ways. Text 1 had no interruptions, and thus no option to rate the buttons, Text 2 had 3 interruptions and required the students to rate their buttons, Text 3 had 3 interruptions and gave the students the option to rate their button, Text 4 had 5 interruptions and required the students to rate their buttons, and Text 5 had 5 interruptions and gave the students the option to rate their button.

Covariate Measures

Covariate measures included a grade-equivalent reading score for each subject, the coordination measure described above, the value attached by the subject to the task obtained from the pre-sewing questionnaire, and the subject's prior experience as reported on the pre-sewing questionnaire.

The grade-equivalent reading score was obtained in the following way. Percentile scores were available for all students and grade-equivalent reading scores were available for most students. Those that were not available were estimated from the percentile scores using regression techniques. All subjects' grade-equivalent reading scores were adjusted to 1990 levels.

The value assigned to the task by the subject was indicated on a 4-point scale on the pre-sewing questionnaire. Students marked whether they thought knowing how to sew buttons was very unimportant, unimportant, important, or very important. Responses were coded from 1 (very unimportant) to 4 (very important) for entry into the data file.

Prior experience with button sewing was also indicated by each subject on the pre-sewing questionnaire. Subjects indicated whether they had sewn no buttons prior to testing, from 1 to 5 buttons, 5 to 10 buttons, or more than 10 buttons. Responses were coded into four quartiles--1 button, 3 buttons, 7 1/2 buttons, and 15 buttons.

Results

Subject Characteristics

The measures described here include gender and grade level, the grade-equivalent reading scores, the measure of coordination, prior experience, and the value attached to the task.

Gender and Grade Level

Fifty-two males and 53 females participated in the study--24 9th-grade students; 39 10th-grade students; 26 11th-grade students; and 16 12th-grade students.

Grade-Equivalent Reading Scores

Percentile scores were available for all students and grade-equivalent reading scores were available for most students. Those that were not available were estimated from the percentile scores. All subjects' grade-equivalent reading scores were adjusted to 1990 levels by adding the months between the administration of the reading test and the experiment. The mean grade-equivalent reading score for the sample was 13.1 (SD = .464). For the male subjects, the mean grade-equivalent reading score was 12.8 (SD = 1.71); for the females the mean score was 13.3 (SD = 1.98).

Coordination

A measure of coordination was taken based on the student's performance on Step 5 of the text instructions, as discussed previously. Differences in coordination due to gender were significant, $F(1,75)$ = 14.64, $p < .001$. The female subjects (M = 3.20, SD = .91) displayed significantly greater coordination at this task than the male subjects (M = 2.40, SD = .82). The only subjects who received a coordination score of 5 (excellent) were females, while the only subjects who received a coordination score of 1 (poor) were males.

No main effects of the number of interruptions in the text, the requirement to rate the product, or grade level were found. Apparently, coordination differences were fairly evenly distributed among the experimental treatment groups.

Prior Experience

The subjects were asked in the pre-sewing questionnaire about their prior experience with the button-sewing task. The sample was composed of approximately 40% novice button sewers (those having no prior experience), 33% with very little experience (1-5 buttons), 9% with some experience (5-10 buttons), and only 18% experienced button sewers (more than 10 buttons prior to testing).

When prior experience was used as the dependent variable in an analysis of variance, no main effects from the number of interruptions in the texts, the requirement to rate the product, or grade level were significant. Significant gender differences were revealed, however, $F(1, 97) = 7.979, p = .006$, with the female subjects ($M = 2.53, SD = 1.07$) having more prior experience than the males ($M = 1.67, SD = 1.04$). Grade-level differences in prior experience were not significant, and the interaction effect between gender and grade level was not significant. The interaction between gender and requirement to rate the product was significant, $F(1, 75) = 4.07, p = .047$, however. Male subjects using the no-rating-required texts ($M = 1.23, SD = .43$) had a lower mean score for prior experience than male subjects using the rating-required texts ($M = 1.95, SD = 1.3$). Females using the no-rating-required texts ($M = 2.68, SD = 1.16$) had a higher mean score for prior experience than female subjects using the rating-required texts ($M = 2.4, SD = 1.04$).

Value Attached to the Task

Subjects were asked to answer a question on the pre-sewing questionnaire about the importance of knowing how to sew on buttons. Few subjects regarded the task as having high value (3%), while almost half (48%) termed it "unimportant" or "very unimportant." More females (75%) than males (23%) called it "important," but those who saw it as "very important" were males. Eleven percent of the male subjects called the task "very unimportant"; none of the females rated the task "very unimportant."

When value attached to the task was used as the dependent variable in an analysis of variance, gender differences were found to be significant, $F(1, 75) = 7.333, p = .008$. The female subjects ($M = 2.76, SD = .43$) valued the task more highly than the males ($M = 2.19, SD = .67$). No significant main effects were identified for number of text interruptions, requirement to rate the product, or grade level. The interaction between gender and the requirement to rate was also significant, $F(1, 75) = 5.355, p = .024$. Male subjects using the no-rating-required texts attached lower value to the task ($M = 1.955, SD = .72$) than male subjects using the rating-required texts ($M = 2.45, SD = .76$), while female subjects using the no-rating-required texts ($M = 2.842, SD = .38$) attached higher value to the task than female subjects using the rating-required texts ($M = 2.74, SD = .45$).

Without-Instructions Button Ratings

Each subject's without-instructions button was evaluated on the four rating scales developed and tested in the first study--shank, security, appearance front, and appearance underside. An analysis of covariance was used to determine the effects of the two text variables (number of interruptions and requirement to rate the product), grade level and gender on the four ratings from each button. The grade-equivalent reading score was used as a covariate measure.

Shank Scale

The without-instructions button scores on the shank scale clustered closely around the low end of the scale. The without-instructions shank scores did not show significant differences due to the number of interruptions in the text, the requirement to rate the product, or the grade level of the subject.

Security Scale, Appearance Front Scale, and Appearance Underside Scale

Without-instructions scores on these three scales did not show significant differences due to the number of interruptions in the text, gender, grade level, grade-equivalent reading score or to the requirement to rate the product.

Summary

Rating scores from the without-instructions buttons revealed that the subjects knew more about security and the appearance of buttons than they knew about shanks. There was a consistent gender difference on all four scales, that is, the female subjects performed at higher levels than the males, but it was not a statistically significant difference on any scale. The groups formed around the "requirement to rate" text feature seemed to be largely equivalent. Grade-level differences were not significantly different on any of the four rating scales, either.

In summary, then, random assignment to the treatment groups was effective in that it generally controlled for the effects of the manipulated variables--both text features (number of interruptions and requirement to rate the product)--, and the controlled variables--(gender and grade level).

Relationship Between Without-Instructions Button Ratings and Subject Characteristics

After evaluating the without-instructions buttons and observing the influence of some of the subject characteristics variables, a stepwise analysis of regression was used to determine which of the subject characteristics had a significant impact on the without-instructions button-sewing performances. It had been expected, after analyzing the results from Study 1, that differences might be revealed that were attributable to differences in gender or grade level, but no hypotheses had been advanced that included characteristics such as coordination, prior experience, grade level, or reading ability.

The initial step in the analysis was to compute the simple correlation coefficient between each listed variable and the button rating scores. The four coefficients (one of each button scale) were averaged to determine an estimate of the linear relationship between the measure of each student characteristic and the criteria. Then in the step-wise analysis of canonical regression, each variable, and its associated possible interactions, was introduced into the analysis in ascending order--that is, variables with the weakest correlation to the without-instructions button ratings were entered first, and variables with the strongest correlations were entered last. After all variables and interactions had been entered, a final analysis was run in which only the significant variables were included.

Significant results due to the effects of grade level, coordination, gender, and prior experience (and their respective interactions) were identified. Subjects' without-instructions button ratings appeared sensitive to several subject characteristics other than their reading ability. Coordination accounted for a significant effect on the appearance front ratings of the without-instructions buttons, $F(1, 88) = 4.255$, $p = .042$, with appearance improving as coordination improved. Significant interactions are described next.

The interaction between gender and grade-equivalent reading score had a significant effect on the security ratings of the without-instructions buttons, $F(1, 88) = 7.173$, $p = .009$. This was due to the males outscoring the females in the second quartile, but scoring lower than the females in all three other reading ability quartiles (see Figure 1).

[Insert Figure 1 about here.]

The interaction between gender and grade-equivalent reading score was significant on the appearance underside scores, too, $F(1, 88) = 5.638$, $p = .020$. The male subjects in the lowest quartile outscored the females in that quartile, the males in the second quartile scored at the same level as the females in the second quartile, but the females outscored the males in both of the third and fourth quartiles (see Figure 2).

[Insert Figure 2 about here.]

The interaction between gender and sewing time showed a significant effect on the security ratings of the without-instructions buttons, $F(1, 8) = 6.461$, $p = .013$. This was due to the fact that the females outscored the males in the three fastest quartiles, but the males outscored the females in the slowest quartile (see Figure 3).

[Insert Figure 3 about here.]

Sewing Time for the With-Instructions Buttons

The videotaped sewing performances of the subjects were timed with a stop watch, and time spent reading the instructions was excluded to obtain sewing time for the with-instructions button. For use in some comparisons, sewing times were divided into quartiles.

An analysis of covariance was made on the sewing times data, using the two text features (number of interruptions and requirement to rate), gender, and grade level as independent variables, and coordination as a covariate measure. No main effects were identified from any of the independent variables. A significant interaction between grade level and gender was identified, $F(3, 74) = 3.187$, $p = .029$).

Adjustments

The adjustments made by each subject were identified, counted, and classified. The total number of adjustments was classified into three main types--those related to the text, those related to the equipment and questions asked by the subject of the researcher.

Among the adjustments related to the text, the most frequently used type of adjustment was rereading of previous text (205 instances), followed by rereading immediately following the first reading (100), looking back (60), reading ahead (55), and looking ahead (16). The most frequently made adjustment related to the equipment was late execution (63 instances), followed by undo/untangle (49), checking the toothpick for freedom to slide under the stitches (48), undo/redo (45), and new start (21). There were 87 instances of asking a perceived expert (the researcher) for help.

Because the question under investigation dealt with the effects of the various texts on the adjusting activities of the students, an analysis of variance was made by text, instead of by number of interruptions and/or opportunity to rate. In this analysis, the total number of adjustments was the dependent variable, with the five experimental text versions represented in the independent variable "text." This analysis was different than earlier ones in that the two text variables (two levels of requirement to rate

and three levels of number of interruptions) were collapsed into one text factor with five levels (one of each distinct text condition). This enabled the use of various contrasts to detect differences in experimental treatments. Each of the probability statements below is based on the results of a Bonferroni post hoc contrast. Grade level and gender were also entered into the analysis as independent variables. The "text" variable was not significant overall (p = .193), but post hoc tests, reported below, revealed several significant differences between the five text versions. Grade level and gender effects did not reach significance, and no interactions were significant. When post hoc tests were made to investigate possible differences between text versions in their effects on student adjusting activities, the following was found.

Interrupting the students, whether or not the rating was required (M = 7.915, SD = 5.67), was found to have encouraged the subjects to make significantly more adjustments than not interrupting them at all (M = 5.1, SD = 3.726), $F(1, 74)$ = 5.556, p = .021. The difference between the number of adjustments made by students using the text with no interruptions and those using the 3-interruptions texts (M = 8.146, SD = 6.106) was also significant, $F(1, 74)$ = 4.449, p = .038). The 5-interruptions texts (M = 7.683, SD = 5.265) also encouraged significantly more adjustments per subject, $F(1, 74)$ = 4.399, p = .039, than did the control text. When the number of adjustments made by students using the texts that required ratings (M = 7.884, SD = 5.811) was compared with the control text, the difference was significant, $F(1, 74)$ = 5.709, p = .019). Merely providing the opportunity to examine the button did not encourage significantly more adjustments than providing no opportunity. It appeared that students could be encouraged by the text features to make more adjustments while executing the procedure, and that requiring the rating was more effective in encouraging adjustment than simply providing the opportunity to adjust.

Overall, Step 16 (securing the thread) was the step where the greatest number of adjustments was made. Step 8 (making stitches and ending with the needle on the underside of the fabric) and Step 2 (taking the first stitch in the fabric) also were difficult steps. In the text with 0 interruptions, Step 2 was the location of the greatest number of adjustments. In both of the 3-interruptions texts, Step 10 (making the last stitches and preparing to remove the toothpick) was the step at which subjects made the most adjustments. In the 5-interruptions text with no rating required, the greatest number of adjustments at a particular step occurred on Step 8. In the 5-interruptions text, Step 16 was the place where the greatest number of adjustments occurred.

Step 17 ("Cut the thread near the needle. Separate the two strands and tie a knot.") posed problems for some subjects, who appeared to find the instruction ambiguous. Of the 105 subjects, 81 tied some kind of knot on the surface of the fabric under the button as instructed. Of these, 51 tied a double knot, 16 tied a single knot, 10 tied a triple knot, 3 tied a quadruple knot, and 1 tied five knots. Eleven subjects tied no knot of any sort, 10 knotted the two loose ends of the thread together and then cut the thread off at the button, and 3 fastened off in other ways. Two subjects remarked, as they cut off the knot in the loose end, "I know this is wrong, but that's what it said to do."

Composite Self-Assigned Rating Scores for Subjects Using the Rating-Required Texts

To learn as much as possible about the relationship between student performance and the text variables, the self-assigned ratings obtained from the students' instructions sets were examined. The three ratings from the instruction set of each subject who used a 3-interruptions rating-required text and the three ratings from the instruction sets of subjects using the 5-interruptions rating-required text which fell at the same locations in the procedure were compared in an analysis of variance to determine if there were significant differences among the three self-assigned ratings. None were found so a composite self-assigned rating score representing the three self-assigned ratings was created (the mean of the three ratings) for each subject who used a rating-required version of the text. This composite self-assigned rating score was used as the dependent variable in two analyses described next.

An analysis that investigated the relationship between the two text variables (number of interruptions and the requirement to rate the product) and the students' ability to rate their own work was done to determine if the text variables affected the self-assigned ratings differentially. The composite self-assigned rating score was the dependent variable; the independent variable was a restricted text variable that represented only the two texts used in the rating-required format (one with 3 interruptions and one with 5); gender and grade level were also used as independent variables. No main effects of the text variables, gender, or grade level were revealed, nor did any interactions reach significance.

To obtain a clearer picture of the subjects' performance, the self-assigned composite score was also used as a covariate measure in an analysis of the with-instructions button ratings. When the with-instructions button ratings were used as the dependent variable with gender as the independent variable and the composite self-assigned rating score as a covariate in analyses of covariance, gender differences failed to reach significance, as did the possible interaction between gender and the composite rating score. The covariate accounted for a significant amount of the variance on three scales--shank, $F(1, 40) = 4.200, p = .047$, security, $F(1, 40) = 10.315, p = .003$, and appearance front, $F(1, 40) = 14.179, p = .001$. It appeared that the subjects had a rather clear idea of how well they were doing on their buttons, since their self-assigned rating composite scores were so closely related to their with-instructions button ratings.

Factors Related to the Differences Between Without-Instructions Button Ratings and With-Instructions Button Ratings

One of the main questions being investigated in this study was the following: "Did the text variables differentially influence how much the subjects' performances improved, or the differences between the ratings of the without-instructions buttons and the with-instructions buttons?" That is, were the students aided in improving their product by the text features of any of the experimental texts? As discussed earlier, the research design included a pre-/post-treatment conditions factor (the without-/with-instructions factor), which was included to help answer the question about subject improvement.

An analysis of covariance was done to learn about the improvement scores. The eight button ratings (without- and with-instructions on four scales) were the dependent variables, with the two text variables (number of interruptions and requirement to rate the product) as the independent variables of greatest interest, and gender and grade level as other independent variables. Grade-equivalent reading score was used as a covariate.

The eight repeated measures button variables were partitioned into two "trials factors"--a term used by the SYSTAT analysis software. The first trials factor had four levels, one corresponding to each of the four rating scales. The second trials factor had two levels, one corresponding to the without-instructions condition and one corresponding to the with-instructions condition. The second trials factor was of particular interest because it was a test of improved performance when the instructions were used. In addition, significant interactions between the second trials factor and the text variables can serve as evidence that the text treatment affected the improvement of performance.

Initially, analysis of covariance indicated that the subjects' with-instructions mean scores on the four rating scales differed significantly from their corresponding without-instructions ratings, $F(3, 222) = 3.331, p = .020$. This reflected the significant difference among the means of the four scales (shank $M = 3.1$, security $M = 4.74$, appearance front $M = 4.49$, appearance underside $M = 3.80$). The significance test on the second trials factor (the improvement scores) failed to show an overall improvement on all four scales taken simultaneously ($p = .104$). However, there was a significant ($p = .05$) trials factor interaction, which indicated that there was greater improvement on some scales than on others. The greatest improvement among the four was on the shank scale (without-instructions M

= 1.133, with-instructions M = 5.048). The least improvement was on the appearance underside scale (without-instructions: M = 3.514, with-instructions: M = 4.086).

Gender was a significant determinant of subject performance (i.e., on the improvement scores), $F(1, 74)$ = 4.821, p = .031, with females outscoring males on every scale, both without- and with-instructions, but the male subjects *improved* more than the females. The gender differences also significantly influenced the interaction between the improvement scores and the rating scales, $F(3, 222)$ = 3.593, p = .014, that is, the improvement on a particular scale was related to the gender of the subject.

Grade level also significantly influenced improvement scores, $F(3, 74)$ = 3.867, p = .013). Twelfth-grade students improved their button scores when using instructions an average of 2 points on the rating scale, the largest increase; 10th-grade students showed the least improvement between without- and with-instructions buttons (1.3 points). Eleventh-grade students averaged an increase of 1.6 points, and 9th-grade students showed a mean increase of 1.8 points.

The number of interruptions in the text and grade level interacted with the improvement scores and the scales, $F(18, 222)$ = 1.667, p = .047. The 9th-grade students performed the best when they used the text with 0 interruptions while all three upper grade levels did the best with the 5-interruptions text. The 11th- and 12th-grade students, in fact, both did their worst when using the text with 0 interruptions, and the 9th-grade students did their worst when they used the 5-interruptions versions of the text. The largest improvements continued to be made on the shank scale. Other contributors to this interaction include the following.

In the three lower grades on the appearance-front scale, some of the without-instructions scores were higher than some of the with-instructions scores. For example, the without-instructions appearance-front scores of the 9th-grade students who were later assigned to the control text were higher than those of the 9th-grade students who used the 5-interruptions text. Similar scores were obtained by the 10th-grade students who were assigned to use the 5-interruptions texts. Their without-instructions front appearance scores were higher than both the without- and with-instructions front appearance scores of those students who used the 0- and the 3-interruptions texts. The 10th-grade students' with-instructions appearance underside scores and the 12th-grade students' underside appearance scores showed similar patterns (see Figure 4).

[insert Figure 4 about here.]

Grade-equivalent reading score was a significant covariate in this analysis, $F(1, 74)$ = 10.747, p = .002). The upper two quartiles in reading scores showed improvement scores of 2.0 and 1.9 points on the rating scale respectively, while the lower two quartiles only showed improvements of 1.1 and 1.25 points on the rating scale, respectively.

Analysis of Regression

A stepwise analysis of canonical regression was made on the without-/with-instructions button-rating differences--the improvement scores. The object was to more clearly identify the relative importance of the student characteristics on the improvement scores. (Because of the lack of any significant effect of the two text variables on the various comparisons in the previous statistical analyses, they were dropped from the list of variables under study.) Variables were entered into this analysis in the same manner as was used in the previously reported analyses of canonical regression, that is, those with the weakest overall correlation first, followed by those with progressively stronger correlations with the rating scores.

An analysis of canonical regression was made on the improvement scores. In this analysis, reading ability contributed significantly to the variability among the differences between the without- and with-instructions button ratings scores. Two other subject characteristics also significantly accounted for the improvement scores--coordination and time spent to sew the with-instructions button.

The only interaction identified in the analysis of regression that reached significance on the with-instructions shank scores was one between grade level and prior experience, $F(3, 52) = 2.870$, $p = .045$. The 9th-grade students' with-instructions shank scores were the highest among all four grade levels at the level of greatest prior experience, the reverse for the 9th-grade students' with-instructions shank scores at the level of least prior experience. The 12th-grade students consistently scored higher than all of the other grade levels at all prior experience levels except the level of greatest prior experience. The 11th-grade students outscored the 10th-grade students at the extremes of the prior experience scale, but the 10th-grade students outscored the 11th-grade students at the middle two levels of prior experience (see Figure 5).

[Insert Figure 5 about here.]

Sewing time spent on the with-instructions button was further investigated when the analysis of regression revealed its significant effects on the improvement scores. When sewing time was used as a covariate measure in an analysis of covariance of the button ratings, a significant interaction between the improvement scores and sewing time was revealed, $F(1, 72) = 7.639$, $p = .007$. Without- and with-instructions button ratings were compared by sewing time quartiles.

Post-Task Questions

After each subject finished sewing the with-instructions button, she or he answered the four basic questions about button shanks listed earlier.

In the statistical analysis, the four questions were treated as dependent variables, with the two text variables (number of interruptions and requirement to rate the product) as independent variables. Gender and grade level were also included as independent variables, and the grade-equivalent reading score was used as a covariate. The covariate was used because success at the question-answering task was likely to be related to reading ability.

The mean scores on the four questions represent the proportion of correct ideas (as compared to the best answer) contained in the answers given by the subjects. The mean scores were as follows: Question #1-M=26.7; Question #2-M=38.6; Question #3-M=31.1; Question #4-M=16.2. In other words, on Question 2 the average answer contained about 40% of the information in the best answer to Question 2. Question 3 (Why did you use the toothpick?) elicited the greatest proportion of correct information, and Question 4 (Why was it important to go through the center of the X?) elicited the least.

In the analysis of covariance, the text variables did not show significant main effects on the correct response rates to any of the post-task questions. Grade-level effects were not significant overall. None of the differences in correct response rates due to a gender-grade level interaction were significant.

Evidently, aside from learning why to use the toothpick (Question 3), learning facts about button sewing was not accomplished by many of the students while performing this task. Only the 12th-grade students demonstrated moderate success on one of the four questions (Question #2-M=50.0).

Discussion

A Model of Procedural Text Processing

The two studies described in this report were undertaken to test Guthrie, Bennett, and Weber's (1990) proposed transformational model of procedural text processing in which successful processing of procedural documents, or successful execution of a procedure, is claimed to be related to the reader's use of several information sources in the document combined with certain hypothesized cognitive processes. The Guthrie model was chosen for testing because it appears to be the best explanation of a reader's cognitive activities while reading procedural documents, and it includes references to essential external sources of information and to the necessary specificity of the outcome. Study 1 dealt with the external sources of information, and Study 2 dealt with two suggested cognitive activities--self-testing and self-correcting. The sources of information and hypothesized cognitive processes were listed earlier.

Guthrie et al. hypothesized that optimal combinations of information sources with cognitive processes included using the exposition to form the conceptual model of the task, using the written steps to help encode the procedures, and using the representation of the outcome to foster self-testing. They did not advance any hypothesis about how to foster self-correcting. The Guthrie model also lacks any reference to objects the reader/performer may be using (the equipment required by the procedure) and any reference to sources of information that may help the reader eliminate confusions (such as other expository text or a teacher/colleague).

The three information sources we tested in Study 1 (exposition of the outcome, the procedural steps, and the representation of the outcome) appear to be important contributors to successful task performance. Subjects who used the experimental text that contained relevant and accurate exposition, clearly labeled procedural steps, many clear and informative illustrations of the outcomes, and clear, precise prose, did better at the task than did subjects who used either of the two commercially prepared texts.

The Experimental Text Features in Study 2

To test the hypotheses that procedural document processing could be enhanced if students engaged in increased amounts of self-testing and self-correcting, Study 2 introduced two new text features--a varying number of self-testing interruptions (opportunities) and a requirement/option to rate the button product.

It was anticipated that using the experimental texts would affect the quality of the button product in the following ways. If the model were correct, then the students who engaged in greater amounts of self-testing (by using the 3- and 5-interruptions texts) would have a more successful execution of the procedure--that is, their buttons would receive higher scores on the four rating scales--than the buttons of students who engaged in lesser amounts of self-testing. The greater the increase in self-testing, the greater the increase in the button quality was the expected relationship. Because Guthrie and his colleagues did not specifically offer a hypothesis about the relationship between the amount of self-correcting done by the reader and successful task execution, but hinted at a direct relationship between the two, the study tested the hypothesis that greater amounts of self-correcting would lead to better button products, too.

It was anticipated that using the experimental texts would affect student behavior in the following ways. The students using any text other than the control text were expected to engage in more self-correcting activities than the students using the control text. The students using the 5-interruptions texts were expected to engage in more self-correcting than the students using the 3-interruptions texts, too. Those students who were required to rate their buttons were expected to engage in more self-correcting than the students who were not required to rate their buttons.

Subjects who used the 5-interruptions texts were also expected to spend more time sewing the button than subjects using the other text versions, because it was anticipated that they would make more adjustments. Subjects who were required to rate the button were also expected to take longer than subjects not required to rate the button because they, too, were expected to make more adjustments.

Effects of the Instructions on the Button Ratings

Did the Instructions Help?

When the effects of the text variables on the button rating improvement scores were examined, the results indicated that the instructions apparently gave the students little help. There was little overall improvement, with the exception of the shank scale.

Although the students in Study 2 did not improve overall when using instructions, they did improve significantly on the shank scale. Only 7 subjects indicated on the pre-sewing questionnaire that they had ever been taught to make a shank, and 92 subjects scored a 1 on their without-instructions shank, while, on the with-instructions shank, only 7 subjects scored a 3 or lower. This seems to indicate that the instructions were effective in helping the students learn a discrete new skill--making the shank itself. Although the shank improvement scores were significant, because they represented only one aspect of the button-sewing performance, any effects on the improvement scores due to improvements on shanks alone were lost in the lesser improvements on the other three rating scales.

The low improvement scores may be a result of the characteristics of the task. This task did not include easily separable parts, and some of the most useful execution strategies, apparently, were undoing or redoing and/or starting over, also some of the most difficult and time consuming strategies. The task also included equipment that was difficult for some subjects to handle well. There might also have been a perception of the task itself that affected performance. For example, a few of the male subjects admitted to feeling very ill at ease with the button.

In trying to explain the apparent ineffectiveness of the instructions in Study 2, a comparison of the without-instructions ratings between the two studies was made, using the scores of the 10th-grade students, the only grade level tested in both studies.

The 10th-grade students in Study 2 had higher without-instructions ratings than the 10th-grade students in Study 1 on all scales except the shank scale, where the two sets of scores were quite similar. Because the students in Study 2 started the experiment with higher scores, their improvement potential was limited, unlike the students in Study 1, whose without-instructions scores were so low that they had the potential of great improvement. This seems to be the major contributor to the low improvement scores in Study 2.

Possible explanations of the difference between the without-instructions button ratings of the 10th-grade subjects in the two studies include the following. A range of reading scores above the 50th percentile was represented in the 10th-grade subjects in Study 2, while in Study 1 all subjects read only at approximately grade level. Reading scores also represent variations among the subjects other than actual reading ability. These variations might well be responsible for the differences between the without-instructions ratings in the two studies.

The difference in without-instructions ratings might also relate to the group testing of subjects without instructions in Study 2, which was not a feature of Study 1. However, only 9 of the 10th-grade subjects in Study 2 were tested in groups too large to be monitored individually, as in Study 1.

Effects of the Text Variables on the Improvement Scores

The with-instructions button ratings in Study 2 were higher than the without-instructions ratings, but the two text variables (number of interruptions and requirement to rate the product) did not produce significant effects overall. There was one significant interaction related to the number of interruptions in the text, though.

Grade level influenced ability to use the interruptions in the text to advantage. The 9th-grade students did best when not interrupted at all, but all the other grade-level groups did best when interrupted frequently (five times). This may relate to a reduced ability of younger students to maintain concentration, or to keep mental track of their location in the procedure. Perhaps it relates to their overall experience, or lack of experience, with written instructions. It also may reflect the fact that the 9th-grade students were poorer readers than readers in the upper grades. Whatever the cause, the results seem to indicate that younger students may not benefit from interruptions in the text designed to allow progress to be checked in the same ways as older students.

Effects of the Text Variables on Metacognition

The main metacognitive behaviors of interest in this study were the number and type of adjustments made by a subject while engaged in performing the task, and their relationship to the text variables. The results indicated that the subjects had indeed made more adjustments when they used both the 3- and the 5-interruptions versions of the text, and that, when required to rate the product, they had also made more adjustments than when not required to rate the button.

The results revealed that while the subjects using the two 5-interruptions texts made more adjustments than the subjects using the control text, they did not make more adjustments than the students using the 3-interruptions texts. Originally, the hypothesis was that more interruptions would lead to more adjusting, and that more adjusting would lead to higher improvement scores. What appeared to happen was that being frequently interrupted (i.e., using the 5-interruptions texts) allowed the student fewer opportunities to make a mistake that required adjusting. In other words, short periods of unmonitored sewing, instead of leading to more adjusting, reduced the need for adjusting, possibly by providing frequent detailed progress checkpoints.

For example, the 5-interruptions texts provided a picture of the button position following Step 3, which allowed some subjects to remove it and turn it over if they had incorrectly positioned the button. The 3-interruptions texts did not include this checkpoint and did not stop the student until after Step 10, at which point the necessary adjustment was considerably more difficult to do. The adjustments counts at Step 10 show that the students using the 3-interruptions texts engaged in significantly more ($c^2 = 11.53$, $a = .05$) adjusting at Step 10 than did the students using the 5-interruptions texts. The requirement to rate the product was effective in encouraging the students to make significantly more adjustments than simply providing the opportunity to examine the button. Apparently, constraining the subjects to evaluate the button product contributed to more effort on the part of the subject, and may be a valuable addition to written instructions for certain types of tasks. It should be noted here, though, that the value the subject attached to the task interacted with the requirement to rate the button in the formation of the experimental treatment groups, and this imperfect effect of random assignment clouds this conclusion.

If the increase in the number of adjustments encouraged by the use of certain of the experimental texts is seen as helpful to the execution of the task, then it appears that the subjects benefitted by using the experimental texts. However, if the increase in the number of adjustments and the low improvement scores are considered together, a conclusion about improving written instructions may be more difficult to draw. It appears that the nature of the task itself may be related to both the adjusting activities and

the low improvement scores. The characteristics of the subjects may also be involved as well, and will be discussed later.

The button-sewing task may have characteristics that impeded student improvement. For instance, undoing a misplaced stitch was a difficult adjustment for many students to make, because it involved carefully replacing the wide end of the needle through narrow fibers of the fabric in exactly the same hole. Some students persevered but others gave up after a few tries. Another difficult adjustment, which some subjects did not attempt, was untangling a knot in the thread. There were a few subjects who spent over 10 minutes simply working on a major thread tangle, but they were the exception. Another procedural task, such as an assembly task with plastic or metal parts that could be easily rearranged (see Schorr, 1982; Stone, 1977a, 1977b), might lead to very different adjusting results, while the procedure followed in using a recipe, in which separate elements cannot be separated after combination, might lead to results similar to those of the present study. The button-sewing task also involves equipment, and may be very different from a task such as map reading, which requires no physical coordination skills but only reading ability, around which the model being tested was developed.

Effects of the Text Features on the Sewing Times

When the sewing times for the with-instructions button were examined, no main effects from either the number of interruptions in the text or the requirement to rate the product were identified. Apparently sewing times on the with-instructions buttons was not related to either of the manipulated text variables. Requiring the subject to stop and examine or rate the button seemingly did not cause the student to spend less time sewing the button, nor did it seem to necessitate that the subject take more time to sew the button. Sewing times are somewhat difficult to interpret. Obviously, if a subject makes more adjustments than another, the time required will be longer. Apparently, though, the sewing times in Study 2 were independent of the text features.

Effects of the Text Features on Learning

The learning of facts while following written instructions was not the goal of this experiment. However, to try to learn more about the effects of the manipulated text features, the researcher asked each subject to answer four post-task questions. The results revealed that the text variables did not create differences in the answers that students gave to any of the post-task questions. One interaction involving the requirement to rate the product and gender was significant, again manifesting the imperfect effects of random assignment. Only the 12th-grade students demonstrated moderate success in answering two of the four questions.

Apparently, learning facts about buttons and shanks was not accomplished by many subjects in this study. This is not surprising, because, in the explanation provided to the subjects about the experiment, fact learning was not stressed. Subjects were told that the instructions were being tested, not their ability to learn while using instructional text. While the subjects were told that there would be four post-task questions, the questions were treated as a minor part of the overall task.

The performance on the post-task questions seemed to indicate that performing the task occupied the attention of most of the students. It may be that following written instructions required that more attention be given to executing actions than to remembering facts, or that having to execute procedures may occupy such a large portion of the student's attention that memory for facts suffers. It may also be true that the students simply took the researcher at her word and tried to follow the instructions instead of trying to remember why buttons must have shanks. Also, the fact that the test only had four items probably meant that it was not very valid or reliable. Thus, the test itself may not have been a very sensitive index of knowledge increments.

At the very least, the students' performance on the post-task questions indicates that an author of instructions cannot assume that the subject learns facts about the procedure. If it is desired that students learn facts related to a procedure, they apparently need to be alerted ahead of time to include that as a part of the overall task.

The Influence of the Subject Characteristics

Little is known about the factors that influence the reading-to-do task. Although this study investigated the impact of two text variables on task performance, a range of student characteristics was also measured. The influences of these characteristics on the without-instructions button ratings, the button rating improvement scores, the sewing times for the with-instructions button, and on the adjusting behaviors of the students will be discussed here.

The subject characteristics that significantly influenced the without-instructions button ratings were grade level, coordination, gender, and prior experience. These results do not seem unusual, given the absence of written instructions, and the already identified gender effect. Apparently, this task required physical ability, and subject performance was also influenced by what the subject already knew about button sewing. It is somewhat surprising that reading ability was not found to be a significant influence on performance, if reading ability is thought of as an index of overall scholastic ability.

Those subject characteristics found to have exerted significant influences on the button rating improvement scores were reading ability, coordination, and sewing time for the with-instructions button. Grade level, prior experience, and value attached to the task did not contribute significantly to the improvement scores. These results support the assumption that reading ability is an important subject characteristic and that it deserves heavy attention when preparing written instructions. The subjects' reading ability was shown to be important only when the subjects had reading material to guide their execution of the procedure, however.

The results also support the hypothesis that there are other subject characteristics that are also very important to successful task performance. Coordination seems to be deserving of close attention, as it influenced both the without-instructions scores and the improvement scores. At the very least, it can be said that writing instructions with only the student's reading ability in mind may make for poor instructions.

Evidently, the task accompanying reading-to-do impacts the relative contribution to successful performance made by reading ability. Reading-to-do may well be much more different from reading-to-learn than reading researchers have recognized up to now. It may also be that simply improving the written portions of instructions cannot provide sufficient help to the reader to enable successful task completion. Subject characteristics may need greater attention than reading ability when the reading task is for performance. This would seem to complicate the improvement of written instructions. The subject characteristics that significantly influenced the sewing times were gender and coordination.

Metacognition

The three subject characteristics whose effects on the adjusting activities of the subjects that were significant were gender, prior experience, and sewing time for the with-instructions button. The significant effects of gender and prior experience have been demonstrated in previously discussed analyses of the button ratings. That these two variables would influence the adjusting activities of the subjects does not seem unusual in light of this. Gender related differences are apparently very strong influences on the execution of this task. Prior experience also seems to be very useful in explaining the subjects' behaviors in Study 2. The relationship of sewing time to adjusting is an obvious one, as well.

Evidently the adjusting activities of the subjects performing this task are influenced very much by the subject characteristics. The results continue to reinforce the suggestion that simply manipulating text features may not be sufficient to encourage adjusting, but that knowing something about the intended reader/performer is probably equally, if not more, important.

Guthrie and his colleagues did not suggest an optimal combination of information and cognition that might foster self-correcting. The results of Study 2 seem to suggest that subjects can be induced to make more adjustments by inserting particular features into instructional text, creating a new source of information about task performance. The results also appear to support the suggestion that the nature of the task also influences the adjusting activities of the reader/performer. This may complicate the task of improving written instructions.

Implications

Due to the widespread presence of reading-to-do tasks in job or job-related settings (Diehl & Mikulecky, 1980; Guthrie et al. 1989; Sticht, 1975), the results and conclusions of these two studies merit consideration by educators in many specialties. Adult, vocational, and community college educators, teachers who author their own procedural instructional materials, trainers in business or industry settings, or authors of on-the-job training materials are only a few of the many who could benefit from what was learned about procedural instructions.

The results of Study 1 seem to indicate that the three information sources tested (the exposition, the use of a procedural schema, and the representation of the outcome) should be included or improved in written instructions, and that the age/maturity of the reader should be considered. Representing both the textual material and the task itself is needed. It may be that older students have an advantage over younger students with tasks other than the one studied here. Testing instructions on potential users also seems helpful.

If effective exposition at the beginning of an instruction set contributes to successful execution of a procedural task, as the results of the first study seem to indicate, then attention to or improvement of that portion of the instructions would be time well invested. The Guthrie model specifies the necessary characteristics of effective exposition--an overview of the process including a statement of purpose and information about the outcome. Using this definition, an educator could examine existing or new procedural materials to improve or add the necessary relevant expository material.

The application of the suggestion that a procedural schema be used in written instructions is equally straightforward. A list, or set of numbered sentences is not difficult to create when the content is procedural. The conclusions drawn by Dixon (1987a-c) about the actual wording of the steps are also not difficult to apply, if the educator is aware of them. The removal of warnings or advice from the steps, but including them on the page in a separate location to make them available and to emphasize them is another apparently helpful step in creating useful instructions. This practice clarifies the step or action information, and also emphasizes important information related to the step.

Improving the third information source tested in Study 1, the representation of the outcome, is probably the most difficult to accomplish. The results of the ex   'ment appear to indicate that two photographs or drawings are not enough for successful execution of the button-sewing task. It is likely that such a small number of illustrations would complicate the execution of other tasks, too. At the least, authors of instructions should examine their written directions to determine if there are specific, clear sentences and clear, relevant graphics showing the positions of the equipment and hands.

Studying the metacognitive processing of the reader of procedural instructions was the object of Study 2. It appeared that interrupting the reader, to encourage self-testing and self-correcting, was not as

helpful to young readers as it was to older, and/or better readers. While slightly more adjustments were made by readers who were infrequently interrupted than by those who were frequently interrupted, the overall quality of the product was not increased. Also, fact learning did not occur, generally, during the execution of the procedure. Several subject characteristics (e.g., gender, prior experience, coordination) influenced performance.

The results of Study 2 seem to indicate that the metacognitive processing of a reader of a procedural document is complex, and influenced by many variables, both external (characteristics of the document) and internal (personal characteristics). Interruptions designed to foster self-testing and self-correcting seem to be differentially effective. Educators who wish to improve procedural materials, then, perhaps need to know more about their potential readers than their reading ability. The nature of the task also apparently impacts performance, and writers of procedural instructions need to carefully analyze the limitations imposed by the task. The ease with which elements can be separated, for example, may influence success. It may be, too, that the reader needs access to a list of possible adjustments.

Educators also need to recognize their purpose in writing instructions. That is, their goal should be well defined. Is performance the desired end product, or is fact learning? Apparently the cognitive demands of reading-to-do are heavy, especially for younger students, or poorer readers.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic reading processes in reading comprehension. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 255-291). New York: Longman.

Andre, M. E. D. A., & Anderson, T. H. (1978-79). The development and evaluation of a sel'-questioning study technique. *Reading Research Quarterly, 14,* 605-623.

Applebee, A. N., Langer, J. A., & Mullis, I. V. S. (1987). *Learning to be literate in America: Reading, writing, and reasoning. The nation's report card.* Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 281 162).

Baker, L. (1979a). *Comprehension monitoring: Identifying and coping with text confusions* (Tech. Rep. No. 145). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Baker, L., & Anderson, R. (1982). Effects of inconsistent information on text processing: Evidence for comprehension monitoring. *Reading Research Quarterly, 17,* 281-294.

Berlyne, D. E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology, 55,* 289-296.

Bieger, G. R., & Glock, M. D. (1984). The information content of picture-text instructions. *Journal of Experimental Education, 53,* 68-76.

Bieger, G. R., & Glock, M. D. (1986). Comprehending spatial and contextual information in picture-text instructions. *Journal of Experimental Education, 54,* 181-188.

Braby, R., Hamel, C. J., & Smode, A. F. (1982). *Handbook of format models for designers of technical training materials.* (ERIC Document Reproduction Service No. ED 226 135).

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717-726.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 103-152). New York: Academic Press.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77-165). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453-482). Hillsdale, NJ: Erlbaum.

Brown, A. L., & DeLoache, J. S. (1977). *Skills, plans, and self-regulation* (Tech. Rep. No. 48). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Carliner, S. (1987). Lists: The ultimate organizer for enginee.ing writing. *IEEE Transactions on Professional Communication, PC-30*(4), 218-221.

Chang, K. L. (1983). *An investigation of the curricular and occupational reading demands of the plumbing trade.* Unpublished master's thesis, University of Calgary, Canada.

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study of conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior, 14,* 56-72.

Collins, A., & Smith, E. E. (1980). *Teaching the process of reading comprehension* (Tech. Rep. No. 182). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Cranney, A. G., Rowley, M., & Stonehocker, L. (1984). *The reading skills of home economics: Problems and selected references.* (ERIC Document Reproduction Service No. ED 249 476).

Curran, T. E., & Mecherikoff, M. (1979). *Toward more comprehensible technical manual graphics. Final report.* (ERIC Document Reproduction Service No. ED 179 198).

Diehl, W. A., & Mikulecky, L. (1980). The nature of reading at work. *Journal of Reading, 24,* 221-227.

Diehl, W. A., Mills, C. B., Birkmire, E. P., & Mou, L. (1989, March). *Procedural texts: Predictions of judgments of importance from structural models.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Divesta, F. J., Haywood, K. G., & Orlando, V. P. (1979). Developmental trends in monitoring text for comprehension. *Child Development, 50,* 97-105.

Dixon, P. (1982). Plans and written directions for complex tasks. *Journal of Verbal Learning and Verbal Behavior, 21,* 70-84.

Dixon, P. (1987a). The processing of organizational and component step information in written directions. *Journal of Memory and Language, 26,* 24-35.

Dixon, P. (1987b). The structure of mental plans for following directions. *Journal of Experimental Psychology, 13,* 18-26.

Dixon, P. (1987c). Actions and procedural directions. In R. S. Tomlin (Ed.). *Coherence and grounding in discourse* (pp. 69-89). Philadelphia: John Benjamins.

Duffy, T. M., Post, T., & Smith, G. (1987). An analysis of the process of developing military technical manuals. *Technical Communication, 34,* 70-78.

Dwyer, F. M. (1971). Assessing students' perceptions of the instructional value of visual illustrations used to complement programmed instruction. *Programmed Learning and Educational Technology, 8,* 73-80.

Dwyer, F. M. (1972). The effect of overt responses in improving visually programmed science instruction. *Journal of Research in Science Teaching, 9,* 47-55.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34,* 906-911.

Fox, B. J., & Siedow, M. D. (1980). Written directions for content area classrooms--Do students understand the teachers? *Clearing House, 54,* 101-104.

Frase, L. T. (1981). Writing, text, and the reader. In C. H. Frederiksen & J. F. Dominic (Eds.), *Writing: The nature, development, and teaching of written communication, Volume 2 Writing: Process, development, and communication* (pp. 209-221). Hillsdale, NJ: Erlbaum.

Frase, L. T., & Schwartz, B. J. (1979). Typographical cues that facilitate comprehension. *Journal of Educational Psychology, 71*, 197-206.

Gagne, E. D. (1985). *The cognitive psychology of school learning*. Boston: Little, Brown & Co.

Garner, R., & Re. , R. (1981). Monitoring and resolving comprehension obstacles: An investigation of spontaneous text lookbacks among upper-grade good and poor comprehenders. *Reading Research Quarterly, 16*, 569-582.

Glock, M. D., Bieger, G., Crandell, T., Knowlton, M., Schorr, F., & Stone, D. (1984). *Understanding picture-text instructions. Final report* (Tech. Rep. No. 12, Series B). Ithaca: State University of New York, College of Agriculture and Life ⁵⁻ences at Cornell University. (ERIC Document Reproduction Service No. ED 241 911).

Glover, J. A., Harvey, A. L., & Corkill, A. J. (1988). Remembering written instructions: Tab A goes into slot C, or does it? *British Journal of Educational Psychology, 58*, 191-200.

Glover, J. A., Timme, V., Deyioff, D., Rogers, M., & Dinell, D. (1987). Oral directions: Remembering what to do when. *Journal of Educational Research, 81*, 33-40.

Graesser, A. C. (1978). How to catch a fish: The memory and representation of common procedures. *Discourse Processes, 1*, 72-89.

Guthrie, J. T. (1983). Teaching comprehension monitoring. *Journal of Reading, 27*, 190-191.

Guthrie, J. T., Bennett, S., & Weber, S. (1990). *Processing procedural documents: A transformational model*. Unpublished manuscript, University of Maryland.

Guthrie, J. T., Schafer, W. D., & Hutchinson, S. (1989, March). *Relationships of document literacy and prose literacy to occupational and societal characteristics of young black and white adults*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Guthrie, J. T., Schafer, W. D., & Wang, Y. (1990). *Document literacy and reading achievement among minority students at grades 3, 7, and 11*. Unpublished manuscript.

Haber, R. N. (1970). How we remember what we see. *Scientific American, 222*, 104-112.

Hare, V. C., & Pulliam, C. A. (1979). *College students' metacognitive awareness of reading behaviors*. ERIC Document Reproduction Service No. ED 182 726.

Harris, P. L., Kruithof, A., Terwogt, M. M., & Visser, T. (1981). Children's detection and awareness of textual anomaly. *Journal of Experimental Child Psychology, 31*, 212-230.

Hartley, J. (1981). Eighty ways of improving instructional text. *IEEE Transactions on Professional Communication, PC-24*(1), 17-27.

Hayes, D. A., & Henk, W. A. (1983, April). *Facilitative effects of analogies and illustrations on understanding and remembering written directions*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Hayes, D. A., & Henk, W. A. (1984, April). *Relative effects of analogies and pictures on understanding and remembering written directions*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Hayes, D. A., & Henk, W. A. (1986). Understanding and remembering complex prose augmented by analogic and pictorial illustration. *Journal of Reading Behavior, 18*, 63-78.

Henk, W. A., & Helfeldt, J. P. (1987). How to develop independence in following written directions. *Journal of Reading, 30*, 602-607.

Johnson, R. C. (1976). Technical order usability. In T. G. Sticht & D. W. Zapf (Eds.), *Reading and readability research in the armed services* (pp. 236-248). Alexandria, VA: Human Resources Research Organization. (ERIC Document Reproduction Service No. ED 130 242).

Karlsen, B., Madden, R., & Gardner, E. F. (1976). *Manual for administering and interpreting: Stanford diagnostic reading test*. New York: Harcourt Brace Jovanovich.

Kern, R. P., Sticht, T. G., Welty, D., & Hauke, R. N. (1975). *Guidebook for the development of Army training literature. Combat training techniques*. Alexandria, VA: Human Resources Research Organization. (ERIC Document Reproduction Service No. ED 140 099).

Kintsch. W. (1985). Text processing: A psychological model. In T. E. van Dijk (Ed.), *Handbook of discourse analysis* (Vol. 2, pp. 231-243). London: Academic Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

LeFebvre-Pinard, M., Bouffard-Bouchard, T., & Decary, D. (1983, April). *Reading to understand and remember: The role of metacognitive awareness about strategies*. Paper presented at the biennial meeting of the Society for Research in Child Development, Detroit, MI.

LeFevre, J., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction, 3*, 1-30.

Loftus, G. R., & Bell, S. M. (1975). Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 103-113.

Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis. *Reading Research Quarterly, 22*, 407-432.

Mackworth, N. H., & Bruner, J. S. (1970). How adults and children search and recognize pictures. *Human Development, 13*, 149-177.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2*, 547-552.

Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 529-540.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111-151.

Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 38-48.

Mark, A. W., & Bracewell, R. J. (1989, March). *The role of descriptive information in the comprehension and performance of written instructions.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Markman, E. M. (1979). Realizing that you don't understand: Elementary children's awareness of inconsistencies. *Child Development, 50,* 643-655.

Mayer, R. E. (1987). *Educational Psychology: A cognitive approach.* Boston: Little, Brown & Co.

Meyer, B. J. F. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16,* 72-103.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology, 19*(2), 155-160.

Norman, D. A. (1989). *The psychology of everyday things.* New York: Basic Books, Inc.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117-175.

Phifer, S. J., & Glover, J. A. (1982). Don't take students' word for what they do while reading. *Bulletin of Psychonomic Society, 19,* 194-196.

Pitts, M. M. (1983). Comprehension monitoring: Definition and practice. *Journal of Reading, 26,* 516-523.

Readence, J. A., & Moore, D. W. (1981). A meta-analytic review of the effect of adjunct pictures on reading comprehension. *Psychology in the Schools, 18,* 218-224.

Reeve, R. A., & Brown, A. L. (1984). *Metacognition reconsidered: Implications for intervention research* (Tech. Rep. No. 328). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Reynolds, H. N., & Booher, H. R. (1980). The effects of pictorial and verbal instructional materials on the operational performance of deaf subjects. *Journal of Special Education, 14,* 175-187.

Rigney, J. W., & Lutz, K. A. (1976). Effect of graphic analogies of concepts in chemistry on learning and attitude. *Journal of Educational Psychology, 68,* 305-311.

Schoff, G. H., & Robinson, P. A. (1984). *Writing and designing operator manuals.* Belmont, CA: Lifetime Learning Publications.

Schorr, F. (1982, March). *Comprehending procedural instructions: The influence of metacognitive strategies.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Schorr, F. L., & Glock, M. D. (1983). *Comprehending procedural instructions: The influence of comprehension monitoring strategies and instructional materials* (Tech. Rep. No. 10). Ithaca: State University of New York, College of Agriculture and Life Sciences at Cornell University. (ERIC Document Reproduction Service No. ED236 333).

Seymour, P. H. K. (1974). Generation of a pictorial code. *Memory and Cognition, 2,* 224-232.

Shepard, R. N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156-163.

Siegel, A. I., Lambert, J. V., & Burkett, J. R. (1974). *Techniques for making written material more readable/comprehensible.* Lowry, AFB, CO: Air Force Human Resources Laboratory. (ERIC Document Reproduction Service No. ED 097 629).

Simon, H. A., & Hayes, J. R. (1976). Understanding complex task instructions. In D. Klare (Ed.), *Cognition and instruction* (pp. 269-285). Hillsdale, NJ: Erlbaum.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. *Cognition and Instruction, 1,* 359-396.

Smith, H. K. (1967). The responses of good and poor readers when asked to read for different purposes. *Reading Research Quarterly, 3,* 53-83.

Smith, S. L. (1985). Comprehension and comprehension monitoring by experienced readers. *Journal of Reading, 28,* 292-300.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73-74.

Stein, J. (Ed.). (1980). *The Random House dictionary.* New York: Ballantine.

Sticht, T. G. (1975). *Reading for working: A functional literacy anthology.* Alexandria, VA: Human Resources Research Organization (HumRRO).

Sticht, T. G., Fox, L. C., Hauke, R. N., & Welty-Zapf, D. (1977, September). *The role of reading in the Navy* (NPRDC-TR-77-40). San Diego, CA: Navy Personnel Research and Development Center. (NTIS No. AD A044 228). (ERIC Document Reproduction Service No. ED149 300).

Stone, D. E. (1977a). *Comprehension of information in picture-text amalgams in procedural texts.* Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Stone, D. E. (1977b). *Text of picture-text amalgams in procedural texts.* (ERIC Document Reproduction Service No. ED 236 665). (Excerpted from author's doctoral dissertation, Cornell University, Ithaca, NY).

Stone, D. E., & Glock, M. D. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology, 73,* 419-426.

Terrell, W. R., Ewell, R. C., Scott, P., & Braby, R. (1983). *The design and production of a procedure training aid using the procedure learning format and the computer automated page layout (PLA) routine* (Tech. Note 12-83). Orlando, FL: Naval Training Analysis and Evaluation Group. (ERIC Document Reproduction Service No. ED 244 579).

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L., Stein, & T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, NJ: Erlbaum.

Trepanier, M. L. (1982, November). *How do I remember?: Young children's understanding of the memory process*. Paper presented at the annual meeting of the National Association for the Education of Young Children, Washington, DC.

Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly, 18*, 328-346.

Wilkinson, L. (1989). *SYSTAT: The system for statistics*. Evanston, IL: SYSTAT, Inc.

Winograd, P., & Johnston, P. (1980). *Comprehension monitoring and the error detection paradigm* (Tech. Rep. No 153). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Wright, P., & Wilcox, P. (1978). Following instructions: An exploratory trisection of imperatives. In W. J. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 129-153). New York: Wiley & Sons.

Figure Captions

Figure 1. Interaction between gender and reading ability without-instructions security ratings.

Figure 2. Interaction between gender and reading ability: Without-instructions appearance underside scores.
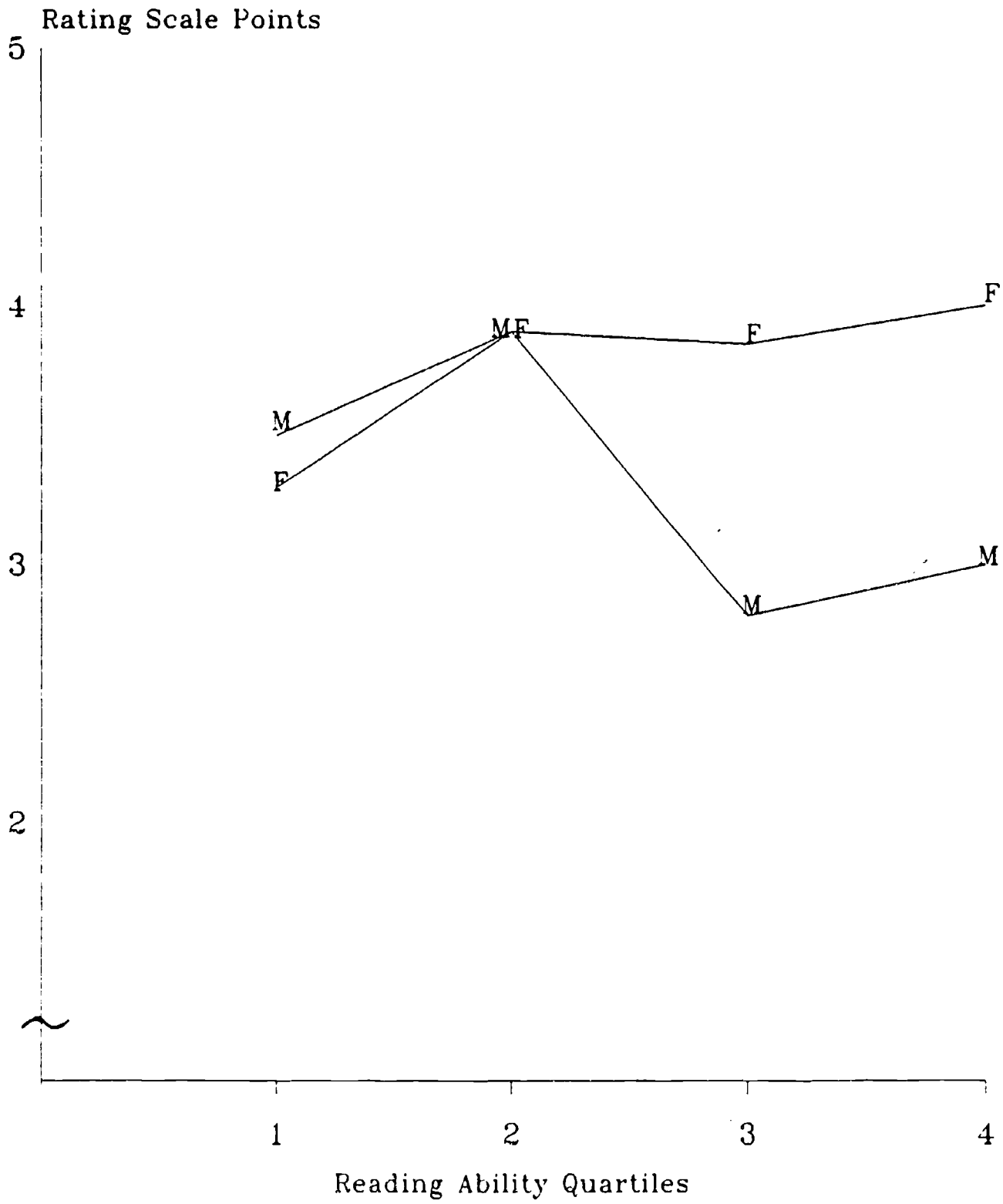
Figure 3. Interaction between gender and sewing time on the without-instructions security scores.

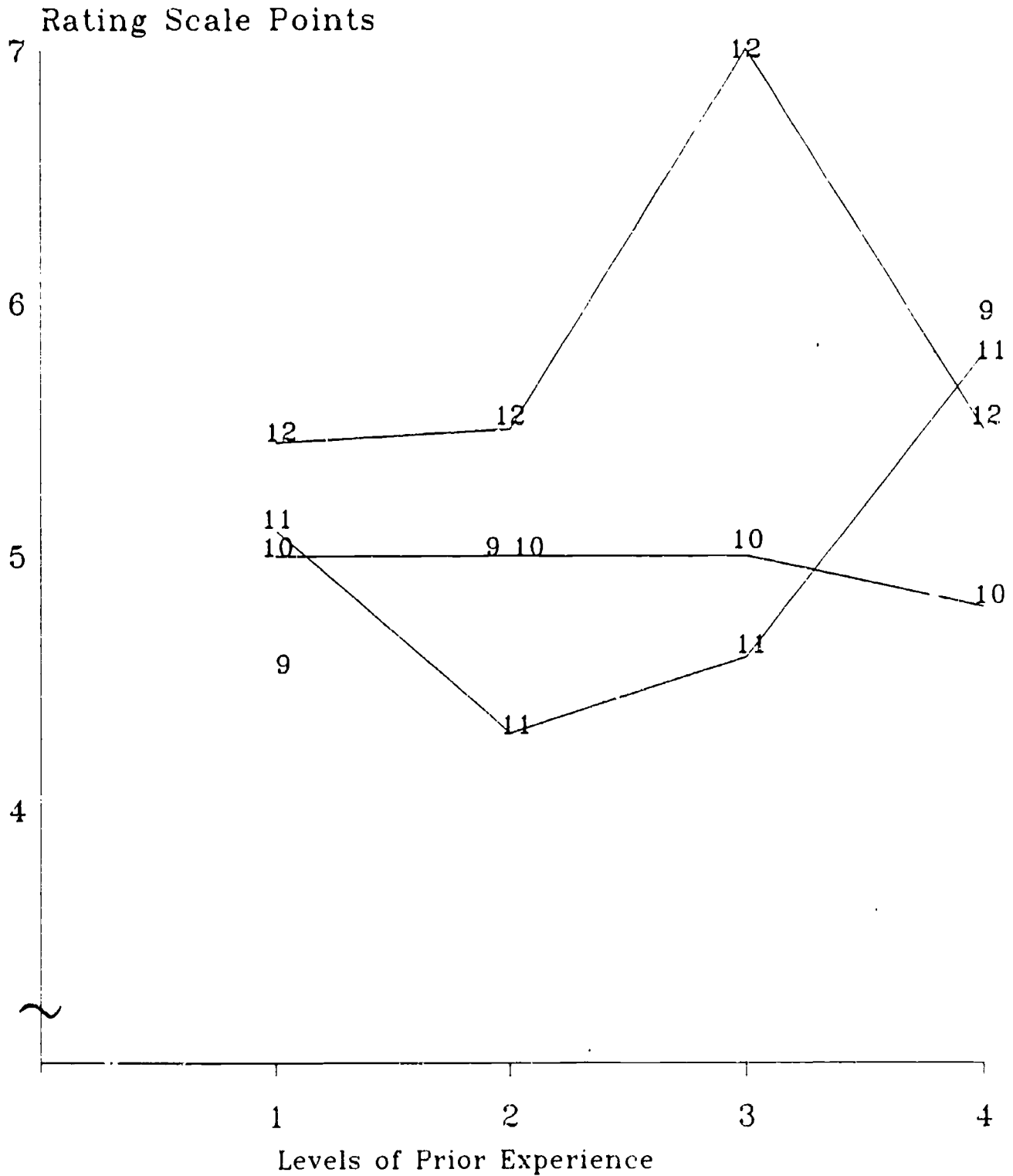**Figure 4.** Interaction between grade level, number of interruption, rating scales, and without-/with-instructions differences.

Figure 5. Interaction between grade level and prior experience: With-instructions shank scores.

Figure 1
Interaction Between Gender and Reading Ability
Without-Instructions Security Ratings

Rating Scale Points

Reading Ability Quartiles

Figure 2
Interaction Between Gender and Reading Ability:
Without—Instructions Appearance Underside Scores

Rating Scale Points

Reading Ability Quartiles

Figure 3
Interaction Between Gender and Sewing Time
on the Without—Instructions Security Scores

Rating Scale Points

Sewing Time by Quartiles

Figure 4

Interaction Between Grade Level, Number of Interruption, Rating Scales, and Without-/With-Instructions Differences

Figure 5

Interaction Between Grade Level and Prior Experience:
With-Instructions Shank Scores

Rating Scale Points

Levels of Prior Experience

*No 9th grade students were at Level 3
of prior experience.

Appendix A

$\mathcal{B}$uttons

Buttons come in all shapes and sizes, but actually, there are only two basic types—sew-through and shank. Whichever type you use, you can sew them on so they stay put.

Hand Sewing

Use a double strand of thread;

Sew-through button: Take a small stitch through fabric at button location. Place toothpick or wooden match on top of button; sew through holes of button, sewing over pick (C). Remove
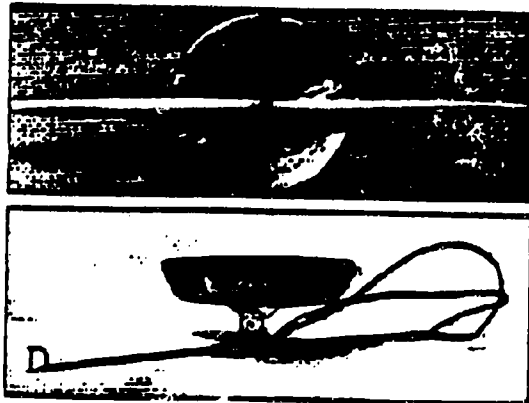
toothpick. Wrap thread tightly around the thread under button, creating a shank (D). Without a shank, stress from the second fabric layer might cause the button to pop off. Anchor thread with a few little stitches.

Appendix B

Attaching buttons

Sew-through buttons

A sew-through button has either two or four holes through which the button is sewed to the garment. When sewed flat, this button can be used as a closure only for very thin, lightweight fabrics, or as a decorative button. If a thread shank is added, the button can be used to close heavy or bulky fabrics as well. The shank permits the closure to fasten smoothly and will keep the fabric from pulling unevenly around the buttons.
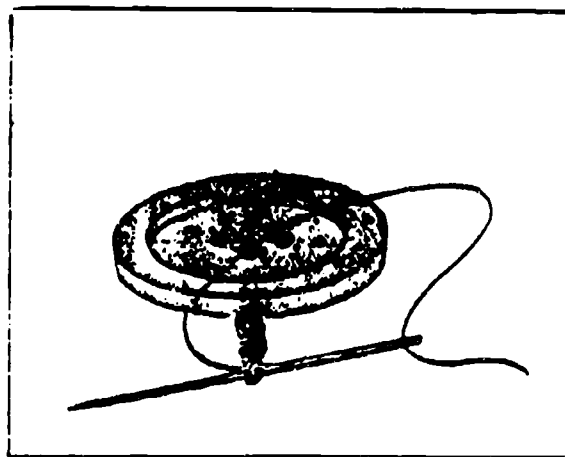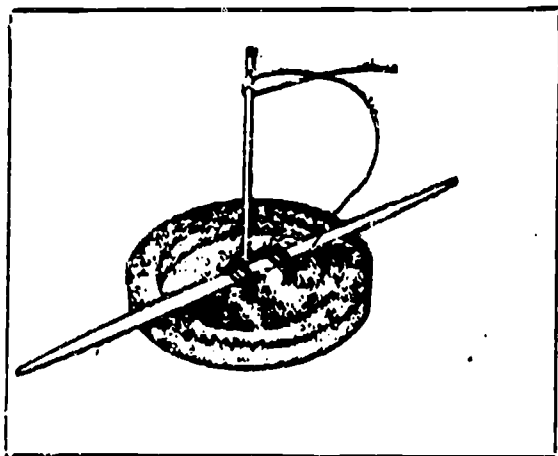
To make a thread shank, secure thread at button mark, then bring needle up through one hole in button. Lay a pin, toothpick, or matchstick across top of button. Take needle down through second hole (and up through third, then down through fourth, if a 4-hole button); make about six stitches. Remove pin or stick, lift button away from fabric so stitches are taut, and wind thread firmly around stitches to make shank. Backstitch into shank to secure.

Appendix C

Sewing on a Button with a Thread Shank

Many garments are fastened with a button and buttonhole closure.
A button needs a **shank** - a length of thread between the button and the
fabric - so that the layer with the buttonhole will fit smoothly under
the button. Then, when the garment is buttoned, the buttonhole will not
spread and cause unsightly wrinkles.
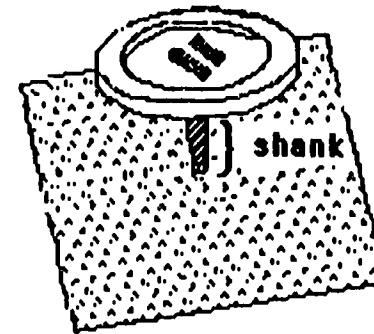
shank

To sew on a flat button with a thread shank, follow Steps 1 - 17.

Pointers

Positioning Button

1.

button location

button location

1. Mark the button location on both sides
   of the fabric using a sharp chalk pencil.

2.
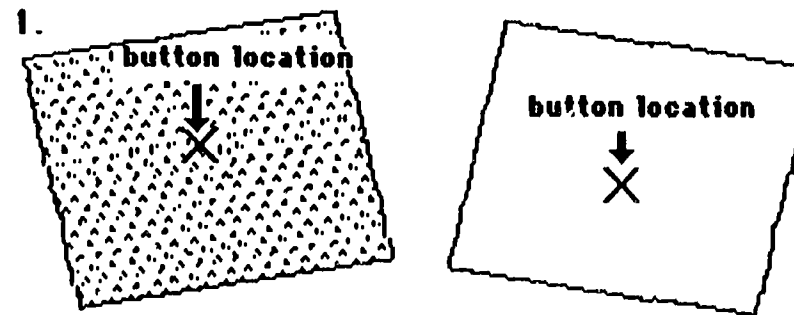
button location

2. Take a small stitch in the surface
   of the fabri   the button location
   marking using   needle threaded
   with a knotted double strand. Pull
   the thread through until the knot is
   on the surface of the fabric.
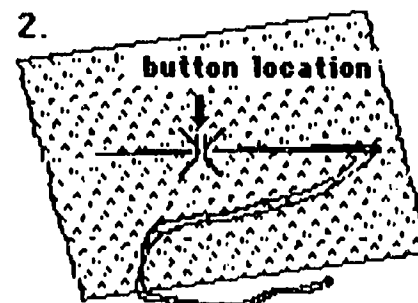
2. Notice that the needle starts and
   finishes on the same side of the
   fabric.

Positioning Button

3. Insert the tip of the needle into one of the holes in the button and pull the needle all the way through. Allow the button to slide down the thread until it rests on the surface of the fabric.

3.
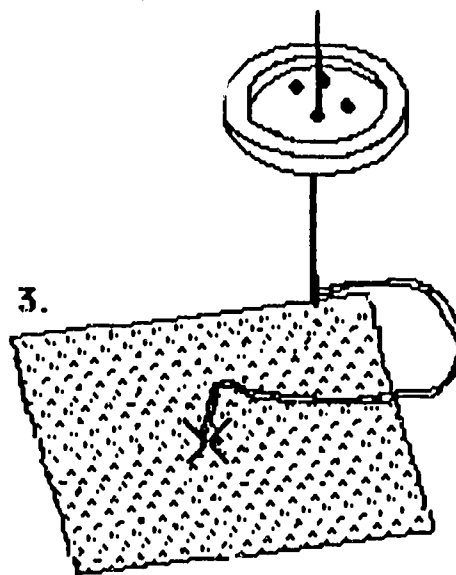
Forming Stitches

4. Place a round toothpick or matchstick on top of the button, so that it lies between the holes.

4. + 5.

Try to make your sample look like the "good" drawing.

Underside of Fabric

| X | ※ |
| good | bad |

5. To make the first stitch, place the tip of the needle in the hole directly across the toothpick from the first hole and push it through the center of the X marking. Pull the needle and thread all the way through. The thread should now for a loop around the toothpick.
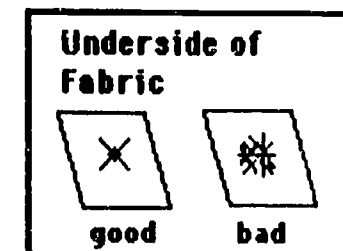
5. You will probably have to angle the needle to go through both the hole in the button and the center of the X.

Do not pull the thread so tight that the toothpick will not slide back and forth a little.
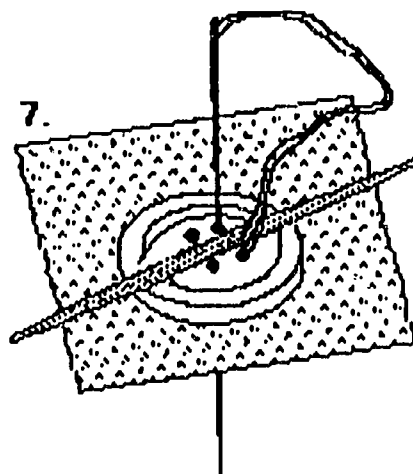
Forming Stitches

6. To begin the next stitch, push the needle through the center of the X marking on the underside of the fabric and through the first hole. Pull the needle all the way through, and pull the thread snug.

6.
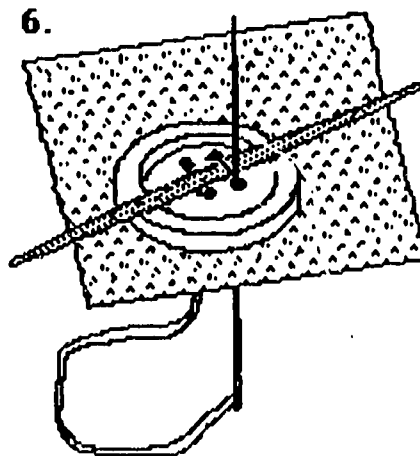
7. To finish the stitch, place the needle tip in the second hole again and push the needle through the button and fabric to the underside. Pull the thread to match the snugness of the first stitch.

7.

7. Do not pull the thread so tight that the toothpick will not slide back and forth a little.

Forming Stitches

8. Repeat Steps 5,6, and 7 using the same two holes until you have 3 or 4 stitches around the toothpick. End with the needle on the underside of the fabric.
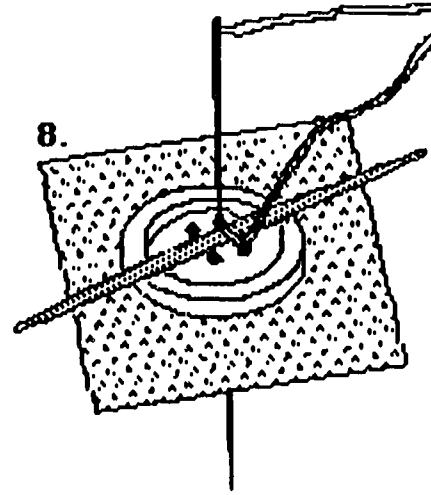
9. To finish attaching the button, use the remaining two holes. Push the needle tip through the fabric near the center of the X marking and through one of the two empty holes in the button. Pull the needle and thread all the way through.
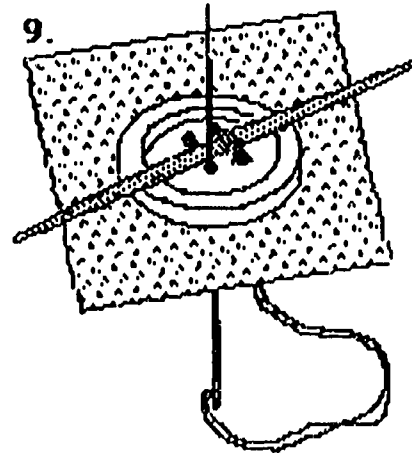
Forming Stitches

10. Repeat Steps 5,6, 7, and 8 using just the two remaining holes. **Stop** before pushing the needle down through the button and fabric on the last stitch.

10.
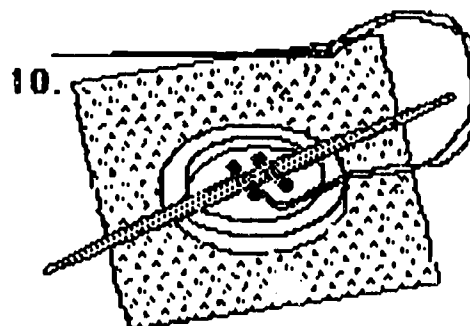
Making the Shank

11. Remove the toothpick.

11. If you cannot remove the toothpick you have pulled the thread too tight. Gently wiggle the toothpick. If it still will not come out you will have to cut your thread and begin again.

12. Complete the stitch by inserting the needle through the hole, but **not** through the fabric. Pull the needle and thread through the button.

12.

70

69
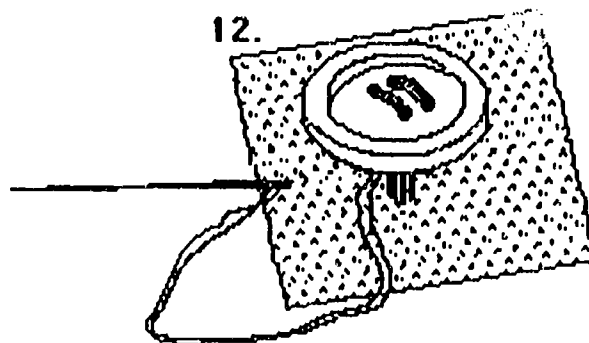
Making the Shank

13. Lay the fabric down on the table. If you are right-handed, hold the button between your left thumb and index finger. Hold the needle in your right hand. If you are left-handed, reverse the hand positions.

14. Holding the needle as directed in Step 13, and using a circular motion, wrap the supply thread snugly around the threads which join the button to the fabric. Start close to the button and wrap toward the surface of the fabric. Go around 3 or 4 times.

14a.

14b.
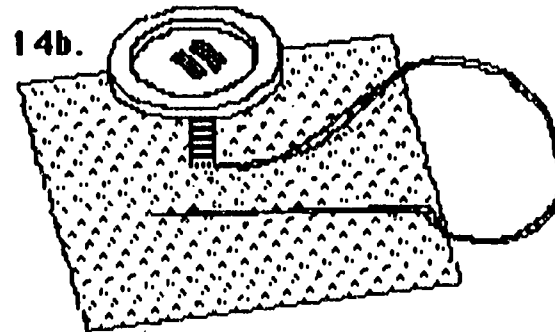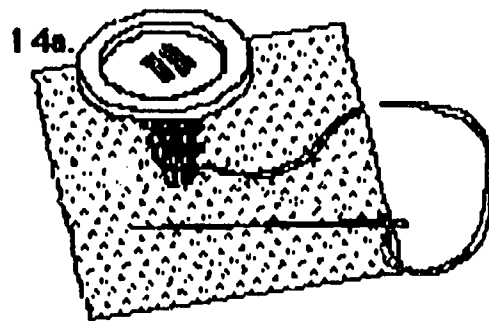
14. You will probably have to let go of the button with your left hand once on each wrap.

The shank should be uniform in width with no wrapped threads lying on top of each other.

Examine the shank after 2 wraps. If the wrapped threads lie on top of each other, unwrap them. Rewrap so each wrap lies below the preceding one.

71

72

Securing Thread

15. To secure the thread, take 2 or 3 small
stitches through the fabric at the
base of the shank.

15.
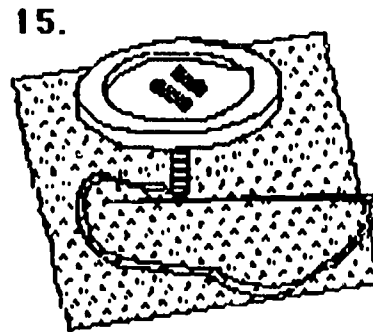
16. Cut the supply thread near the needle.
Separate the two strands and tie a square
knot.

17. Cut the supply thread as close to the surface
of the fabric as possible.

74

Appendix D

Shank Rating Scale

7	space allowed
	tightly wrapped vertically
	    - deduct 1 level on scale for tangles, or uneven thread
	        tension

_____

6	space allowed snugly wrapped, wraps spread outward **or** loosely wrapped vertically - deduct 1 level on scale for thread tangles

_____

5	space allowed wrapped loosely 3 - 4 times or snugly 1 - 2 times - deduct 1 level on scale for thread tangles

_____

4	space allowed wrapped loosely 1 - 2 times

_____

3	space allowed no wraps

_____

2	no space allowed loosely attached **or** no space but wrapped anyway

_____

1	no space allowed sewn flat and tight **or** too insecure to count as a shank

_____

Security Rating Scale

	initial knot on surface of fabric under button
	3 - 4 stitches per pair of holes
7	snug thread
	tied off with 2 - 3 stitches and knot tied
	initial knot on surface of fabric under button
	3 - 4 stitches per pair of holes
6	snug thread
	tied off with <2 stitches and / or no knot tied
	initial knot on surface of fabric under button
	2 - 3 stitches per pair of holes
5	snug thread
	tied off with 0 - 1 stitches and / or no knot

Scores of 5, 6, or 7 must be reduced as follows:

reduce 1 level on scale for :

too many stitches (bulky, holes of button filled with thread)

	thread knotted initially
	2 - 3 stitches per pair of holes
4	loose thread
	0 - 1 stitches and / or ineffective knot (may have thread tails)
	thread knotted initially
	1 - 2 stitches per pair of holes
3	thread loose
	evidence of tie-off - stitch attempted and knot near surface of fabric (may have thread tails)
	thread knotted to secure to fabric (may have thread tails)
	0 - 1 stitches per pair of holes
2	thread loose (button slides off X somewhat)
	tie-off attempted - some evidence of stitches
	knot in thread, not on surface of fabric
	no knot to secure thread (probably has thread tails)
	0 - 1 stitches per pair of holes
1	thread loose (button slides freely around X marking)
	no tie-off stitches or knot

Appearance Front Rating Scale

7
consistent stitch pattern (=)
thread tension consistent with smooth, untangled thread
    (deduct if loose)
      - deduct 1 level on scale for each of the following:
         too much thread (bulky appearance)
         thread tails or stitches around button edge
         puckers

6
consistent stitch pattern (X) **or** = pattern with loose stitches/
    variant tension
thread tension consistent
smooth, untangled thread, no loops
      - deduct 1 level on scale for each of the following:
         too much thread (bulky)
         thread tails, tangles, inconsistent tension, loops
         around button edge, or puckers

5
majority of stitches form consistent pattern (= or X)
minor unevenness in thread tension among stitches
fabric flat with only minor puckers, if any
      - deduct 1 level on scale for each of the following:
         tangles in thread
         if attachment stitches show
         needle attached

4
stitch pattern mixed (= or X)
thread tension slightly variant
minor thread tangles visible from surface of button
      - deduct 1 level on scale for each of the following:
         too much thread (bulky appearance)
         if attachment stitches show
         needle attached

3
more consistently patterned stitches than inconsistent ones
thread tension inconsistent among stitches and visible minor
thread tangles

2 stitch pattern missing or inconsistent
thread tension inconsistent among stitches
major thread tangles visible (loops, snarls, etc.)

or

0 - 1 stitch per pair of holes - too insecure to use

1 no apparent stitch pattern, or stitch placement inconsistent
thread tension uneven and inconsistent
thread tangles ignored (loops visible, thread tails on surface)

or

toothpick attached
needle attached and shows on surface
unwearable (stitched to fabric around edge of button)
major thread tails on surface

Appearance Underside Rating Scale

7
flat fabric
area covered by stitches very small (<3/16" across)
extremely focussed with short individual stitches
- deduct 1 level on scale for each of the following:
thread tail
thread tangles
puckers in fabric

6
flat fabric
area covered by stitches up to 1/4" across and quite dense
individual stitches up to @ 3/16"
- deduct 1 level on scale for each of the following:
thread tail
thread tangles
major puckers in fabric

5
area covered by stitches about 1/4" - 5/16" across
more focussed than stitches in level 4 but visible gaps between
stitches
- deduct 1 level on scale for each of the following:
thread tails
thread tangles
stitched in puckers in fabric

4
>1/4" in both horizontal and vertical dimensions with noticeable
gaps between stitches
slight puckers
- deduct 1 level on scale for each of the following:
thread tails
tangled thread
noticeable or stitched in puckers in fabric

3
3/8" - 1/2" across stitch pattern but more focussed than level 2
short thread tails on underside
slight puckers
small thread tangles

widespread but beginning to show some focus around a central
point (3/8" to 1/2" across)

2   will have major thread tangles, large loops, and / or snarls

probably will have thread tails on underside

major puckers in fabric

0 - 1 stitches; too few to judge

probably too insecure to show tangles

1 or

widespread (>1/2" in any direction)

thread tails present

major thread tangles, long loops or many snarls