National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

COMPARISON OF FEEDBACK GENERATED BY SUBJECT MATTER AND LEARNER EXPERTS DURING FORMATIVE EVALUATION

Diana Tremblay

Department of Educational and Counselling Psychology

McGill University

Montreal

1994

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Arts in Educational Psychology

ISBN 0-315-99942-X

Canada

Comparison of expert feedback generated during formative evaluation

ABSTRACT

The literature on formative evaluation advocates the use of both subject matter and learner experts to review instructional materials, yet there has been little research to support this recommendation. The present study investigates the distinguishing characteristics of these two expert reviewers, in particular the type and amount of oral feedback they produced and the cognitive processes they engaged in. The think-aloud method was employed to obtain feedback about a six page instructional unit from eight experts; four subject matter and four learner experts. Comments from the experts were transcribed, segmented and coded according to three coding systems. Results indicated that the two groups produced similar data. In addition, the findings showed that both groups referred to similar domains of knowledge, evoked or constructed similar plans and identified their task as detecting problems. These findings contradict the use of both types of experts during formative evaluation. Some practical recommendations for practitioners are offered.

RESUMÉ

La littérature sur l'évaluation formatif recommande l'utilisation d'experts des contenu et "learner experts" (les experts qui sont familiers avec les étudiants qui utilisent le matériel éducationel) pour le matériel éducationnel encore, il y a très peu de recherche pour supporte cette recommandation. La présente étude de recherche examine et caractérisé les deux experts réviseurs, en particulier le type d'opinion orale il produite et réfléchit leur engagement. La pensée a haute voix a été utilisée pour obtenir une évaluation du type d'opinion orale de directives de six pages de huit experts; quatre experts des contenu et quatre "learner experts". Les commentaires écrits des experts, segmenter et réglementer selon trois codes de système. Les résultats des deux groupes indique des similarités. En plus les résultats des deux groupes utilise le même connaissance et construise des plans d'organisation et identifier leur tache de trouve les problèmes. Les résultats ne soutiennent pas l'utilisation des deux types d'experts durant l'évaluation formatif. Des recommandations d'ordre pratique sont offerte pour le praticien.

ACKNOWLEDGEMENTS

I would like to extend my sincere appreciation to my supervisor Dr. Cynthia Weston for her patience, support and expertise.

Also, I would like to thank the members of the Formative Evaluation Research Team; Dr. Lynn McAlpine, Alanna Israeloff, Tino Bordonaro, Kate Le Maistre, Tim Rahilly and Malcolm Welch for their valuable contribution to this research project. Their opinions and ideas stimulated me to think in new directions.

Special thanks goes to Tino Bordonaro for creating my beautiful graphs and helping me through computer crises. Further thanks go to Richard Beland and Renee Champagne for the translation of my Resume.

TABLE OF CONTENTS

PAGE

PAGE

PAGE

PAGE

LIST OF TABLES

LIST OF FIGURES

PAGE

FIGURES

PAGE

CHAPTER ONE

Review of the Literature

Introduction

Prior to the 1960's, it was common practice for educational publishers to put unrevised materials into final production. These first drafts were distributed among the schools, frequently proving to be ineffective and unable to support classroom instruction. As curriculum development projects grew in the sixties, publishers attempted to improve instructional materials by comparing new products against existing materials. Like the earlier generation, many of the flaws in these materials continued to go unnoticed and as a result, a low level of learning carried on.

As a response to this difficulty, Cronbach (1963) and Scriven (1967) proposed that data identifying potential problems in the material be collected during their development. Following data collection, revisions should be made while the products were still in rough draft form. This process was termed formative evaluation and today it is a well established component of instructional materials or product development (Dick & Carey, 1990; Truett, 1984; Wager, 1983; Stolovitch, 1982; Andrews & Goodson, 1980).

Scriven (1967) identified two primary characteristics of formative evaluation when used for educational purposes. First, the aim of the

evaluation is to improve the effectiveness of instructional materials. Second, the feedback received from potential learners and experts form the basis for revision. An early and continual revision cycle is generally recommended (Geis, 1987; Weston, 1986; Komoski, 1983), although, studies have reported an improvement in student performance after only one revision of the prototype (Dick, 1980; Baker & Alkin, 1973).

In recent years there has been substantial support for formative evaluation. In an examination of the pertinent characteristics of 40 models of instructional design, Andrew and Goodson (1980) found that 38 of the models endorsed formative evaluation. More recent surveys conducted by Tessmer (1993) and Burt and Geis (1988) provide additional evidence that formative evaluation is highly recommended.

Research has shown that formative evaluation renders instructional materials more effective and as a result many studies endorse the recommendation to include formative evaluation as an important component of instructional design (Bordonaro, 1993; Schloss, Smith & Posluzsny, 1990; Wager, 1983; Ellis & Wulfeck, 1983; Baghdadi 1980). There are, however, several common approaches to formative evaluation. These methods are based on the principal source of feedback employed; either experts, learners or a combination of the two (Weston, 1986).

When an expert or a number of experts are used to critique instructional materials it is referred to as expert review. In this approach,

learner feedback is not collected and revisions are based solely on expert opinions.

The use of learners to evaluate educational products has been labelled; developmental testing, field testing or learner verification and revision. The common feature among all of these methods is that potential learners of the materials, either individually or in groups, are the primary source of feedback for revision.

The literature also advocates a third approach to formative evaluation, gathering feedback from multiple sources (Israeloff, 1992; Weston, 1987; Geis, 1986; Dick & Carey, 1990). This method proposes the use of a combination of experts and learners, with each group generating different kinds of information. For example, experts would be consulted in order to identify problems within their area of specialization, while learners would be asked to provide personal reactions to the materials. Weston (1987) points out that using multiple sources is advantageous as experts may recommend changes that may not be identified as problems by the learners.

All of these approaches have a common element, the recommendation that instructional materials be tried out and revised based on the feedback obtained. These methods can be seen as one continuous process consisting of two distinct stages; a data collection and a revision phase. Weston (1991) suggests there are several advantages for making

such a distinction. For example, this discrimination allows researchers to investigate a) the nature of different sources of feedback during data collection, b) the preference of revisors for certain kinds of feedback data, c) the impact of different types of data sources and d) the revisor's input on learning outcomes.

The focus of the current study is on the data collection stage of formative evaluation and one particular method of collecting information; the expert review process. The remainder of this chapter reviews contemporary literature regarding formative evaluation and expertise. Initial sections discuss the use of experts as a source of information and recent research on expert review. The last two sections examine the cognitive psychology literature on expertise and the characteristics of experts relevant to this research. This review concludes with a summary and critique of existing literature, followed by a discussion of the specific research questions suggested by the review and examined in this thesis. .

Experts as a Source of Data

Earlier on, it was explained that each of the approaches to formative evaluation is based on a principal source of feedback; either experts or learners. In this section experts will be described in more detail, as they are the focal point of this study. In particular, the types of experts most often recommended and advantages and disadvantages of using experts will be highlighted.

For centuries, practitioners have sought the advice of experts to evaluate their products. Today, expert review remains popular with publishers and developers of instructional materials as it is generally more cost-effective than the other approaches (Tyson & Woodward, 1989; Weston, 1986; O'Donnell et al., 1985; Nevo, 1985; Truett, 1984; Kline, 1984). Evidence of this is found in a survey conducted by Truett (1984) with educational software producers throughout the United States. She found that close to half of the producers used teacher evaluations to review programs and nearly 20% used other outside experts. Concerns about the higher cost of collecting learner data have led some researchers to promote the exclusive use of expert reviewers (Montague, Ellis, & Wulfeck, 1983; Macdonald-Ross, 1978). For example, Macdonald-Ross (1978) suggested that only experts be used to critique materials since revision data produced by learners was seldom richer or more meaningful than expert data.

During expert review, various kinds of experts can be consulted individually or as part of a team. Geis (1987) generated an extensive list of different types of experts used in formative evaluation and the unique function they perform. His categorization scheme includes: subject matter experts, instructional designers, pedagogical experts, learner experts, social and moral experts, and users of the materials. Weston and McAlpine (1990) added presentation, language, and cultural experts to the list. Since

each type of expert represents a specific area of expertise, they can be used to serve different purposes during formative evaluation (Stolovitch, 1982). For example, subject matter experts could review the instruction to assess its accuracy, comprehensiveness and recency. Pedagogical experts can judge the suitability of chosen instructional strategies, the ability of students to learn from the instruction or how effectively teachers can work with the materials. Learner experts, also known as audience specialists, can supply information about the relevance or acceptability of an instructional system to a particular target population. Instructional designers could use their skills in planning instruction to comment on how well the educational product reflects the principles of good instruction such as clarity of objectives or alignment of components.

A survey of current instructional design textbooks reveals that nearly half recommend the use of two particular outside specialists during formative evaluation; subject matter and learner experts. It is presumed that each type of expert offers a unique perspective and therefore, a wider range of information useful for revision is provided (Tessmer, 1993; Flagg, 1990; Dick & Carey, 1990).

There are several advantages to collecting feedback data from experts. Nevo (1985) reports four advantages of expert review: 1) experts' opinion is an inexpensive evaluation tool compared to other procedures of data collection, 2) obtaining feedback from experts can be accomplished in

a minimal amount of time, 3) generally experts' ideas are well respected and add credibility to a project and 4) experts' opinion can be used to evaluate new instruction before the project is implemented. This flexibility is important as it allows materials to be evaluated early in their development. In addition, experts may have an advantage over learners in detecting erroneous content or potential problems with the materials (Geis, 1987). Saroyan (1989) also suggests that another advantage is that experts' comments are generally more informative than learners, since experts tend to detect inaccurate and problematic content and often follow this with recommendations for improving the materials.

Nonetheless, expert review has been characterized as having certain shortcomings. Nevo (1985) admits that experts' opinions can be subjective and that experts are often hesitant to provide conclusive responses. He also points out that there are often shortages of experts in certain areas, making it difficult for developers to secure the appropriate experts for evaluation. Weston (1987) highlights another potential disadvantage; that experts cannot be relied upon to predict accurately how learners will respond to or learn from the materials. A further disadvantage is that using too many experts may produce idiosyncratic information (Thiagarajan, 1978).

Nevo (1985) suggests that when formative evaluation is structured in a systematic way many of these shortcomings can be minimized. He

presents five requirements developers need to incorporate into the expert review process. First, offer a detailed project description, then select an unbiased sample, present specific questions to experts, collect written and oral responses and lastly synthesize experts' opinions into a composite summary. Thiagarajan (1978) underscores the importance of providing checklists to the experts to focus their attention on specific areas, as well as, encouraging them to go beyond simply identifying problems. He recommends that reviewers suggest and implement solutions, so that nothing is lost in the translation from feedback to revision. Systematically incorporating these recommendations into the expert review process would overcome any inconveniences and enable developers to benefit from the significant advantages of this procedure.

The literature on formative evaluation indicates that both learners and experts produce meaningful data for revision, however, the central theme of this study is the role of expert reviewers during formative evaluation. In the upcoming section, research on the behaviors of expert reviewers will be examined.

Research on Expert Reviewers

There has been little research done on the characteristics of expert reviewers, however, two studies are relevant to this discussion. One is Saroyan's (1989) investigation of different types of expert reviewers and whether they produced distinctive feedback during formative evaluation. A

second is by Duy (1990), who examined similar expert reviewers to ascertain whether these experts stayed within their area of expertise and focused on corresponding problems in the materials.

In Saroyans' (1989) analysis of the expert review process, two different kinds of experts were selected. Microbiologists who taught introductory microbiology to health science students and professional instructional designers. Each subject was to review and revise a self-instructional module on microbiology. The results showed that the two groups differed in their representation of the task, the problems they focused on and the strategies they employed. Content experts consistently addressed problems with the inadequacy and irrelevance of the content. In addition, despite their experience teaching the intended audience, they refrained from commenting on the pedagogical aspects of the text. Instructional designers also focused on concerns with the content, but with less intensity and concentrated the rest of their attention on design and pedagogical issues. A second conclusion was that individuals within each expert group demonstrated a similar concern about gross inaccuracies. At an in-depth level of analysis, however, dissimilarities were found. Saroyan concluded that the unique training and work experience of the various experts led them to interpret their roles differently. The microbiologists, who routinely functioned as resource people, acted as specialists. On the other hand, the instructional designers behaved more like generalists.

Influenced by their background in formative evaluation and their familiarity with the types of opinions usually generated by different expert reviewers, the designers provided more diverse information. Based on these findings, Saroyan recommended that both types of experts be used to constitute a formative evaluation team, in order to reduce duplication in effort and increase efficiency.

Duy (1990) selected instructional designers, similar in terms of training and expertise, to investigate the uniformity of revisions made by each expert. She found that the experts were quite similar in addressing various instructional design attributes, placing their emphasis on rewriting the objectives or the need for introductions and improved transitions. Nonetheless, they were more idiosyncratic with respect to other attributes of the materials, such as presentation and subject matter. Duy concluded that the experts predominantly stayed within their area of expertise, however, distinct backgrounds might have contributed to their concentrating on different minor categories. She proposed that future studies examine the effects of focusing experts behavior through the use of a checklist.

These initial studies on expert reviewers suggest that different types of experts can be distinguished by the data they produce. They stay within their area of expertise and consistently recognize global problems. However, the samples used in each of studies was very small and

generalizations about the role of expert reviewers cannot be made.

While there is limited research on the behavior of expert reviewers during formative evaluation, the current cognitive science literature provides additional information on expert behavior.

Cognitive Psychology Literature on Expertise

The current investigation of cognitive science into expertise focuses on the collective talents of diverse experts and their ability to reason, understand, solve problems and learn. This section will summarize the vast literature on expertise by first discussing the acquisition of expertise and then describing the characteristics of three types of experts germane to the present study.

Acquisition of Expertise

Throughout the last decade cognitive psychologists have attempted to comprehend excellence by uncovering the learning processes experts go through. It has been estimated that 10,000 to 20,000 hours of concentrated learning and practical experience is needed to acquire a world-class level of expertise (Glaser, 1991; Chase & Simon, 1973). Looking at this from another perspective, 20,000 hours are roughly equivalent to spending more than 10 years of college and university building an expert knowledge base in a specific domain. According to Pylyshyn (1989), 10 years represents a ratio between the amount of knowledge and skill demanded by the external environment and the internal capabilities that influence the rate at

which people are able to learn this requisite knowledge and skill. These two components will be examined in more detail next.

Effects of External Demands on Expertise

The knowledge base acquired by experts is constantly being shaped by the demands of their external environment. Intelligent problem solving behavior is adaptive and can take on different forms in different milieus. Theorists speculate that exposure to distinctive external demands leads to diverse types of expertise or proficiency in a specific domain (Glaser, 1991; Lawrence, 1988). Two of the most profound external demands that influence expertise are the task environment and on-the-job experiences. As Lawrence (1988) explains, shared values and outlooks that are born from the social structure of the workplace "set certain constructions on reality for professional and cultural groups" (p.231).

Furthermore, the external environment can also act as a constraint on expertise. Typically, task instructions that are inadequately defined and restrict individuals to focusing on surface details can force experts to resort to novice-like behaviors (Glaser, 1991; Swanson, O'Connor & Cooney, 1990; VanLehn, 1989). Such ambiguous task instructions often lead to the construction of an ill-defined problem space, where experts are inhibited from isolating the specific principles around which their domain knowledge is organized. Instead they must retrieve large amounts of information and rely on inefficient general search heuristics. The outcome of such a vast

search generally evokes a highly variable set of responses.

In a study done by Voss, Greene, Post and Penner (1983) solutions elucidated by political science experts were compared with solutions obtained by physics experts. Political science faculty members and undergraduate students were presented with a typical ill-structured political science problem. Voss and colleagues found that there was considerably more within group variability in solutions derived from political science experts, when compared with solutions obtained from the physics experts. The researchers attributed this discrepancy to the ambiguous nature of the political science task that invariably elicits different translations of the problem. These varied interpretations in turn generate an array of corresponding solutions, many of which are unsuitable.

The external environment is an important feature in the development of expertise. It not only influences the style of cognitive functioning adopted by the expert, but also the breadth of their social and practical problem solving skills.

Effects of Internal Demands on Expertise

Due to internal capabilities, experts accumulate knowledge and skill at different rates. Nevertheless, there are certain common circumstances that are instrumental to the speed at which experts obtain this knowledge base. Two of these aspects are practice (Glaser, 1991; Neves & Anderson, 1981) and the development of a rich content knowledge (Chi & Ceci, 1987;

Carey, 1985).

The importance of practice is demonstrated in a study by Anzai and Simon (1979) that showed practice quickly improved subjects' ability to unravel a puzzle. The researchers proposed that once a subject discovers a correct solution path, it becomes a template for forming new productions capable of uncovering the solution more efficiently.

Several developmental studies illustrate the second aspect, the relationship between acquisition of prior content knowledge and new knowledge. One of these studies on cognition in young children, documents a change similar to the novice to expert shift (Carey, 1985). Carey based her theory upon her own observations of childrens' thinking. Younger children organize their knowledge structures around undifferentiated characteristics. As they gain more information about objects and ideas their knowledge structures change to reflect these new facts. As the child matures, detailed categories begin to develop around abstract principles. The general point that Carey makes is that as the learner acquires more specific concepts, their knowledge structures alter and these new structures facilitate advanced reasoning. Increased content knowledge, therefore, will in time promote expert thinking.

There are also constraints associated with the inner environment. Posner (1988) points out that people differ in abilities and interests and that these capabilities interact with their accumulation of knowledge. He

believes that the most important individual difference may be the willingness to undergo the long, rigorous training necessary for experts to become masters in their field. If this is the case, then the development of expertise may rely as much on motivation as on the capacity to learn.

Practice and a rich knowledge base all combine to place internal demands on experts. At the same time, the external task environment and on-the-job experiences influence the style adopted by the experts. Moreover, ill-structured tasks, individual abilities and a lack of motivation can diminish expert performance. Having examined the influences that contribute to the acquisition of expertise, the discussion will now turn to a description of the salient characteristics of three types of experts relevant to this research; domain or subject matter experts, literate experts and pedagogical experts.

Characteristics of Relevant Experts

In everyday life, people are encountered who have unique talents in particular areas of specialization. This high level of competence has become an intriguing subject for researchers and the list of expert characteristics continues to grow. In the current study, the characteristics of subject matter and learner experts as reviewers of written material are being examined. Presently, it is unknown what kind of expertise learner experts will bring to the formative evaluation task. A knowledge of teaching, as well as reading and writing, seems to be implied. Hence, this

section will review the universal traits of domain, literate and pedagogical experts as they are germane to the study of both subject matter and learner experts.

Domain Experts

Experts in various domains demonstrate superior skills in three main areas: 1) metacognitive skills, 2) memory and speed of performance and 3) knowledge structures. Each one of these competencies will be looked at in more detail.

The literature on expertise reveals that experts display self-regulatory or metacognitive skills that are absent in less experienced learners (Glaser, 1991; Swanson et al., 1990; Glaser & Chi, 1988). These capabilities include awareness of one's level of competence, allocation of attention to the analysis of problems, management of available resources and strong solution monitoring skills.

Two other features that distinguish experts is their accelerated speed of performance and superior memory. Initially experts, due to their extended analysis of a task, are slower than novices in the first phase of problem solving, but faster overall (Johnson, 1988; Chase & Ericsson, 1981). Also, experts' recall exceeds the usual limits of both short-term and long-term memory. Further investigation has determined that there is a link between this superior memory and accelerated speed of performance (Gentner, 1988; Chi et al., 1987; Neves et al., 1981).

An early investigation by Chase and Simon (1973) using chess players of varying strength revealed that master chess players exhibited greater speed and accurately recalled more positions. They proposed that experts' speed and increased memory comes from many hours of practice. Such intense rehearsal develops automaticity of skills and frees up resources for greater storage in memory. Moreover, they advanced that experts can arrive at a solution without conducting an extensive memory search. Building on this theory, Chase and Ericsson (1981) hypothesized that superior memory in experts is due to retrieval systems that chunk information together.

The third and perhaps most influential attribute of expertise is the possession of a well organized body of knowledge. This knowledge, which is the primary difference between experts and novices, is principle-based, domain-specific and readily accessed by excellent metacognitive skills. Experts' knowledge structures differ from novices' in several aspects; for example they are larger than novices, incorporating more core items and members. Experts also possess more superlinks or overlaps where distinct categories share common features (Chi et al., 1987). These superlinks enable experts to perceive large patterns of interrelated information.

Murphy and Wright (1984) examined differences in knowledge structures between experienced and beginning counsellors and found support for the claim that richness of categories increased with experience.

All subjects were asked to describe the characteristics of an aggressive child, a depressive child and a disorganized child. It was found that the skilled counsellors included more features in each category and that the classifications contained many items shared by two or more clusters, whereas novices' categories contained virtually no overlapping attributes.

This extended knowledge base is a critical component of expertise as it aids experts' in their interpretation of problem situations. A good problem representation is comprehensive and abstractly depicts problems in terms of goal statements or "what is to be done", rather than simply "how to do it". Experts confronted with new information examine the data in this goal-directed fashion and use their professional knowledge and other personal frames of reference to guide their search. This more coherent and principled representation results in a deeper understanding of the problem. In turn, such a comprehensive understanding leads to the selection of more appropriate procedures and ultimately, more effective solutions.

The characteristics of exceptional metacognitive skills, superior memory and intricate knowledge structures are common traits among experts in a variety of domains. More specific skills attributed to literate experts will be highlighted next.

Literate Experts

In the present study the experimental task requires subjects to read

through written materials and identify problems for revision. Also, for the purpose of this research, learner experts were all CEGEP or university professors. Since reading and writing are critical aspects of this profession, it is fitting to examine the expertise literature in this area; known as literate expertise.

Both reading and writing are highly goal oriented, intellectual tasks that include planning, implementation of the plan, and monitoring progress. The research on writing and revision indicates that skilled editors set themselves a more difficult task, are goal-directed, engage in more planning and make more meaningful revisions (Fitzgerald, 1987; Flower, Hayes, Carey, Schriver & Stratman, 1986). Skilled readers also set themselves a more difficult task, are goal directed and work at a deeper level of understanding to uncover the meaning behind the words (Scardamalia & Bereiter, 1991; Hare, 1981; Flower, 1989). Scardamalia and Bereiter (1991) reviewed the literature on strategic reading and compiled information on the reading styles of expert and inexpert readers. This search revealed that task execution was more laborious for skilled readers. They constantly monitored their comprehension of the materials, re-cycled to planning to consult their goals when the text became difficult, and integrated new information with prior knowledge. The end result is a deeper level of comprehension for the reader.

Another study examined the role task representation played when

students were reading for the purpose of writing (Flower, 1989). Flower found that the students' organizing plan was the dominant feature used to structure the task representation. She proposed that five organizing plans guided the process of reading to write and that skilled reader/writers were more likely to use two options; synthesizing ideas around a controlling concept and interpreting ideas for a particular purpose. The synthesizing plan requires readers to combine their own knowledge with information from the text and organize this information around a central topic. It is the mainstay of academic writing and permits reading for meaning not just details. The interpretation plan expands on the synthesizing plan by attempting to adapt the authors' knowledge into a format that addresses the needs of the audience and sets up an issue to make a definite claim. Although selecting an organizing plan is instrumental to reading and writing, the process is often carried out with little or no awareness on the part of the problem solver. Familiar problems are likely to be highly automated, with little conscious effort required to evoke them.

Research on revision of written materials indicates that experienced revisors also define the task differently and pursue different strategies (Flower et al., 1986). A study comparing novice and experienced editors found that experts interpreted their task as diagnosing problems and revising, while the novices viewed the task as detecting problems without elaboration and rewriting rather than making specific changes. Additional

expert characteristics identified were the use of well-developed and specific procedural knowledge, building complex problem representations and diagnosing more global problems that considered the audience and purpose for writing. The researchers did note however, that although most experts dealt with the same global deficiencies there was little consistency in specific problems. An interesting finding was that experts made surprisingly few revisions. The researchers concluded that although diagnosis was in the experts' repertoire, experts frequently used the less time consuming alternative of detection.

Ostensibly then one reason that experts avoid making revisions is an unwillingness to commit to the extra cognitive cost involved. In a review of literature on revision in writing, Fitzgerald (1987) presents six other potential roadblocks to revision. These obstacles include: 1) the lack of clear objectives for the task, 2) lack of knowledge of what is wrong, 3) lack of knowledge of how to fix the problem, 4) the difficulty of juggling presentation and content-related goals, 5) the inability to take the readers perspective and 6) difficulty in managing the process of revision.

In summary, literate experts engage in more arduous task execution by going beyond the surface level of facts to achieve critical literacy; interpreting facts for different uses. The plans they develop are goal directed and highly automated. Due to cognitive costs, experts may simply detect problems without making revisions or other obstacles may constrain

their performance. Next, a survey of pedagogical experts will complete the review of expert characteristics.

Pedagogical Experts

Learner experts in this study are all experienced teachers and consequently bring this specialized knowledge to the formative evaluation task. Since this background influences their performance, the key features of pedagogical expertise will be highlighted.

Master teachers are distinguished from novices by their more skilful planning, in-depth analysis of problems and goal-directed task representations. Like experts in other domains, their knowledge structures are more elaborate and detailed and each category is organized according to a general principle. A further distinction is that skilled teachers have a large quantity of knowledge about students available to them and they are able to use this knowledge to predict where difficulties might arise for students (Borko & Livingston, 1989; Berliner, 1986; Leinhardt & Greeno, 1986; Schulman, 1986; Leinhardt & Smith, 1985).

The skill of teaching rests on two fundamental systems of knowledge; subject matter and lesson structure (Leinhardt et al., 1986). The first is knowledge of the content to be taught. The second is the knowledge required to construct and conduct a lesson. This second category is made up of several distinctive components, including: 1) traditional lesson plans, 2) activity structures that include generic

methods of teaching, as well as, specific ways of teaching a particular content, 3) a repertoire of automatic routines that are associated with individual activities and 4) information schemata that enable teachers to make decisions about the instructional needs of students and subject matter coverage. Shulman (1986) introduced the term pedagogical content knowledge, that is a blending of content and pedagogy similar to Leinhardt and Smith's specific activity structures. This knowledge is unique to the teaching profession and enables experienced teachers to adapt topics to the interests and abilities of students (Borko et al., 1989; Berliner, 1986, Shulman, 1986).

A comparison study of mentor and novice teachers demonstrates that experts are better at predicting where in a course students are likely to have problems (Borko et al., 1989). Experts' more developed content knowledge, pedagogical content knowledge and knowledge of learners enable them to predict misconceptions the students may have. The researchers also found that when experienced teachers plan instruction, they think about "how" best to communicate the content and focus on explanations, examples, sequencing and students' understanding and involvement with the content.

The studies of pedagogical experts show that master teachers' more sophisticated content, lesson and student knowledge distinguishes them from novice teachers. This knowledge influences how the subject matter is

presented and ensures that the instruction is responsive to students' needs.

A summary of the cognitive psychology studies on expertise indicate that different training and professional experiences can result in diverse expertise and proficiency. Furthermore, ill-structured tasks and different levels of motivation can constrain an individuals' level of competence. A common characteristic among subject matter experts is that they spend more time in the initial phases of problem solving, analyzing problems and building goal-directed representations of the task. They also possess well-organized and complex knowledge structures. Traits specific to literate experts are setting for themselves the more demanding tasks of reading for meaning, and considering the audience and purpose of the task when writing. Pedagogical experts have a unique knowledge base that includes content, lesson structures and student characteristics. Skilful teachers use this extended knowledge to communicate the content in a manner that is relevant to their students' needs.

Summary

Formative evaluation refers to collecting feedback on rough drafts of educational products with the intention of improving the instructional quality of these materials. Since its inception, support for formative evaluation has increased. Research and experience have provided evidence that evaluation of instructional materials does improve quality and increase learning. Expert review is one of the common sources of

data collection used during formative evaluation. This source of data collection has several advantages that have increased its popularity with publishers and developers alike, however, there is a lack of research on the expert review process and the characteristics that distinguish each type of expert. A few studies indicate that subject matter experts and instructional designers are distinctive from one another. An additional conclusion is that individuals within each expert group consistently recognize global problems, but are more idiosyncratic about specific imperfections.

A further investigation into the cognitive psychology literature on expertise produced additional defining characteristics of experts. External and internal demands place different stresses on experts that constrain their performance and result in diverse specializations. Universally, experts are differentiated from novices by their superior metacognitive skills, goal-directed plans and well organized and complex body of domain knowledge. Literate experts, experienced in the revision of written text, can be distinguished by the difficulty of the task they set for themselves and their strategy of revising to make the text more meaningful to the audience. Likewise, pedagogical experts are unique due to their specialized knowledge of subject matter, lesson presentation and student characteristics. This combination of content and lesson structure knowledge guides experienced teachers in their translation of the content into instruction that is relevant to the reality of the classroom. This review

of the research sheds light on the characteristics of experts, however, it is insufficient to support the claims that various types of experts produce distinctive data when evaluating instructional materials. Neither does it provide empirically based guidelines to help the instructional developers structure expert review sessions.

Focus of this Study

This study was undertaken to address the paucity of research on the expert review process in formative evaluation. It attempts to identify distinguishing characteristics of two different types of expert reviewers; subject matter and learner experts. First, it examines the product produced by these two expert groups. Second, it investigates how they represent the task and the plans they construct. This information will aid instructional developers in their decision whether to use a single expert reviewer or a team of diverse experts. It will also help them to structure the task so that the experts will generate the data the developers are seeking.

Research Questions

The general question that guided this research was: What are the defining characteristics of the feedback generated by subject matter and learner expert reviewers during formative evaluation?

(a) What are the similarities and differences in the product generated by these experts both between and within the two groups?

(b) What are the similarities and differences in the cognitive processes employed by the experts both between and within the two groups?

CHAPTER TWO

Methodology

Overview

This study examines the defining characteristics of the oral feedback produced by two kinds of expert reviewers during formative evaluation. The types of specialists selected for the present study are subject matter and learner experts. This research also examines the cognitive processes of these two expert groups. Specifically, it compares the similarities and differences that exist in terms of the type and amount of data generated, as well as the strategies used by both groups.

The materials chosen for the reviewing task was a six page unit on the topic of diet and cancer. Three types of data were collected. First, the subjects completed a think-aloud procedure; second, they responded to specific questions from the administrators in a retrospective interview and third, they completed a debriefing questionnaire. The primary method of data collection was the think aloud, which required the subjects to verbalize all of their thoughts while reading the unit aloud. Verbal protocols were obtained by transcribing verbatim the comments made by the subjects. Initially, these protocols were segmented and coded using a coding scheme based on the human problem solving model; the inter-rater reliability of the coding scheme was established. Three additional levels of analysis were then applied to the coded protocol: 1) Product Attribute

analysis of essential characteristics of a well-designed instructional product, 2) Planning and Goal analysis of the formative evaluation plan constructed by the expert and 3) Task Representation analysis. From these analyses the similarities and differences between and within expert groups were compared. Demographic data was compiled with the use of a debriefing questionnaire.

Method of Data Collection

Subjects

Eight experts were selected for this study. Four were professional dieticians and four experts were university or CEGEP professors. All experts had practised their professions for a minimum of five years. This was judged to be the requisite amount of experience needed to acquire a basic level of expertise. Nevertheless, most subjects had significantly more years of professional experience than this minimal standard.

The experts in the current study were chosen because they represented two types of experts that are called upon to review instructional materials. The dieticians were selected as subject matter experts, since they were familiar with the issues pertaining to diet and cancer. The professors were designated as learner experts, as they were acquainted with the intended audience for these materials; undergraduate university arts and science students. The subjects were offered a small stipend for their participation in the study and informed consent was

ascertained by having each subject sign a consent form.

Materials

Stimulus materials. The stimulus materials used in this research was a six page self-instructional module entitled The Diet Cancer Relationship (Fenster, Harpp, & Schwarcz, 1990) (see Appendix A). These materials were written by university chemistry professors, who were experts in the content of the materials. The materials were created for use in an introductory level, undergraduate chemistry class that was open to both arts and science majors.

To be considered appropriate, the stimulus materials had to meet three standards. First, the materials had to be in a draft format, in other words without the benefit of undergoing any type of formative evaluation. This criterion was established so that the materials used in this study would have a similar composition to those usually tendered for formative evaluation. Next, the materials had to be written by content experts, as these are the most frequent authors of instructional materials in higher education. Lastly, the materials had to provide a minimum of 30 minutes of instruction. This time period was selected since it was long enough to be considered a learning task, yet short enough to elicit comments from subjects on all aspects of the materials.

Administrative materials. Subjects were provided with a folder containing the following documents: the stimulus materials, a summary of

activities for the session, a consent form, instructions for their task, a description of the warm-up exercise, and a hand-out explaining the think-aloud procedure (see Appendix B for a copy of all of these documents except for the stimulus materials, which are presented in Appendix A).

The administrator's folder contained the following information: a tracking sheet, which consisted of the stimulus materials in a smaller print with large margins for written comments, a script presenting some background on the research project and the purpose of this study, a script explaining the session's activities, a script explaining the task instructions (same as subjects'), a script explaining the warm-up exercise, a list of phrases used to prompt subjects to continue thinking-aloud during the session, a script for concluding the session and, a debriefing questionnaire (see Appendix C for a copy of administrator's materials except for the task instructions, which are presented in Appendix B).

Equipment and Setting. A Sony cassette recorder (model TCM-5000EV) with a separate microphone was used to record the session. As a precaution, another cassette recorder with a built-in microphone was used as a back-up. The room was arranged so that the tape recorders were placed on a separate table, with the microphone placed in front of the subject. One administrator sat next to the subject, while the other administrator sat where the tape recorders were located. This arrangement allowed the subjects to carry out the experimental task with ease, while the

researcher could change cassettes and take notes unobtrusively. The recorders began running after the warm-up activity and remained on until the end of the session.

Rationale for the Think-Aloud Method

The think-aloud is a data collection method for conducting in-depth analyses of cognitive processes. This research method is closely tied to the theory of human problem solving and its purpose is to provide insight into subjects' underlying thought processes (Ericsson & Simon, 1984). In the current study, the think-aloud method was used to gather detailed verbal information about what the experts were thinking while reviewing instructional materials. The subjects were to read the materials out loud, stopping to verbalize their thoughts as they came to mind. All thoughts were to be expressed no matter how trivial in nature they seemed to be.

There are two major assumptions behind this method. The first assumption is that during problem solving an individual processes available knowledge and possible operations into a sequence of intermediate steps that lead to an eventual solution. The theory also assumes that under normal circumstances, as intermediate steps come into focus they are available for verbalization and according to Ericsson & Simon (1984) articulating these steps does not interfere with the problem solving process. The result of this operation is a think-aloud protocol that provides a trace of all the erroneous sequences of steps, alternative operations considered

and the final selection of a solution (Breuleux, 1991).

A major benefit of using a think-aloud procedure is that it uncovers a rich data source of reflective thinking that is seldom accessible by other methods (Ericsson & Simon, 1984; McAlpine, 1987). An additional gain is that this method offers a means to reduce potential distortions and loss of information that occur when subjects recount their thoughts retrospectively (McAlpine, 1987).

The think-aloud method, however, is not without its critics. Ericsson and Simon (1984) respond to some of these criticisms in an article outlining the effectiveness of verbal reports as data. One of the main criticisms is that conducting a think-aloud will alter the subjects' performance. A second reproach is that this method produces incomplete verbal reports and inconsistent results.

Ericsson and Simon counter the first complaint by suggesting that evidence from concurrent reports contradicts this claim. In addition they report that even when subjects performed the more complex task of reading while verbalizing their thoughts, the subjects' performance, comprehension or number of inferences generated were comparable to subjects who did not think aloud while reading. Ericsson and Simon continue to refute think-aloud critics by pointing out that the potential disadvantages of incomplete and inconsistent data can be diminished by using clear probes in a retrospective interview that immediately follows the

last trial. They also emphasize that even though incompleteness of reports may make some information unavailable, this does not invalidate the information that is present (p.243).

Having weighed the advantages and disadvantages of the think-aloud method and its suitability for the experimental task, the think-aloud was determined to be the best method of data collection to use in this study while the subjects' worked through the materials. This information was supplemented with interviews.

Pilot Testing

Prior to conducting the study, a pilot test was conducted in order to identify problems with the data collection procedures and to determine the clarity of task instructions. The pilot test also provided an opportunity for the researchers to practice implementing the experimental procedures prior to the actual data collection. The methodology of the pilot test was identical to the proposed study, but limited to one subject. This subject was a high school, mathematics teacher with over 20 years of teaching experience. For the past two years, she had been conducting research on university course evaluations. During the pilot test no problems were uncovered, and the data collection methods remained unchanged.

Procedures

Preliminary procedures. Prior to data collection, the administrators contacted the potential subjects by phone. At this time, the nature of the

research was explained and a short description of the experimental task was given. The subjects were also informed that the task would take a maximum of three hours and that a $100.00 honorarium would be given at the end of the session. Upon agreeing to participate, dates for data collection were arranged with each subject. The initial phone call was followed by a reminder notice mailed to each subjects' home. This notice indicated the nature of the study, the time commitment required, the amount of the honorarium, the date and the location of the session. Upon arrival for their appointment, the subjects were greeted by the two administrators of the session and thanked for their participation. The subject was then seated at a table and provided with a folder containing a package of reading materials that they would use during the session. The participant was informed that the administrators would stop for a break midway through the session, however, a break could be taken any time the subject deemed necessary.

Procedures during the session. Each of the eight sessions consisted of: (a) an ice-breaker, (b) a summary of the activities for the session, (c) a description of the experts' task, (d) a warm-up exercise for the think-aloud, (e) the reviewing task, (f) a retrospective interview, and (g) the debriefing questionnaire.

The ice-breaker involved having the subjects read a script that thanked them for their participation and outlined the research project they

had volunteered to take part in. The general nature of the task was described and it was emphasized that the materials were being evaluated, not the subjects' performance. At the end of the script, the subjects were encouraged to articulate any concerns they might have.

For the summary of activities, the administrator read a list of actions that would be executed during the upcoming session. At the same time, the subjects read silently along with the administrator using their own copy of the list. After confirming that the subjects understood the task, they were asked to sign the participant consent form.

This was followed by the administrator reading the description of the task to the subjects. The script explained that the subjects were to give feedback as an expert on their observations about the content of the materials. They were also asked to comment on how well the intended audience could learn from the unit. The subjects were advised to highlight positive, as well as, negative features of the materials. The task description briefly outlined the think-aloud procedure and the subjects were informed that should they forget to think out loud, one of the researchers would prompt them to continue. Written comments were permitted, as long as the subjects also verbalized what was being written.

A warm-up exercise was included based on Newell and Simon's (1972) recommendation to allow subjects to familiarize themselves with the think-aloud procedure before executing the task. Through practice the

subjects become accustomed to verbalizing all of their thoughts, therefore, they will generate more comments during the experimental session. As with the previous activity, the administrator read an explanation of the warm-up exercise to the subjects while they read their own copy silently.

The warm-up activity required the subjects to form as many words as possible from 12 randomly chosen letters of the alphabet. Any of the letters could be used repeatedly. Three minutes were allotted for this task, and the subjects were told to verbalize their thoughts throughout the entire activity. While performing the task, if the subjects remained silent for more than ten seconds the administrator prompted them to continue thinking-aloud. After the warm-up activity, the subjects and the administrators appraised their performance. If the subjects still had questions concerning the think-aloud procedure, they were given additional information in a hand-out. This hand-out consisted of common questions and answers concerning think-aloud protocols, and was adapted by Saroyan (1989) from Hayes, Flower, Schriver, Stratman, & Carey (1987).

Prior to commencing the reviewing task, the definition of the task was repeated to the subjects and they were asked if they had any further questions. The subjects were reminded that the administrators would be unable to respond to their queries once the think-aloud procedure began.

During the reviewing task, the subjects read the stimulus materials out loud while the administrators sat quietly. One of the administrators

prompted the subjects to continue thinking out loud and wrote down on the tracking sheet any comments made by the subjects that were unclear or lacking in sufficient details. This was done so that further clarification of these comments could be sought following the think-aloud session. The other administrator was responsible for changing the cassettes and also tracked the subjects comments. If subjects stopped verbalizing for more than 10 seconds, the first administrator reminded them to think-aloud by using a list of prompting phrases (see Appendix C). These phrases were in a predetermined order, and if a subject remained silent after the first prompt was used, the administrator would go down the list of phrases until the subject began to verbalize again. The purpose for using designated prompting phrases was to increase consistency among subjects and reduce potential bias. For all eight subjects, the administrator rarely had to prompt and the first two phrases were the only ones used.

Following the reviewing task, the notes from the tracking' sheets were used to structure the retrospective interview. The objective of the retrospective interview was to obtain additional feedback from subjects. Ericsson and Simon (1984) recommend the use of a retrospective interview immediately after the think-aloud session, so that the information that is still stored in short-term memory can be directly accessed.

The interview consisted of specific questions about the feedback given by the subjects during the review task. General questions were

avoided, as they might not have elicited the information that was being sought (Ericsson & Simon, 1980). In general, the subjects were asked to clarify or elaborate on the comments the administrators had noted earlier. The length of the interview varied across subjects.

When the retrospective interview concluded, the subjects were asked to fill out a debriefing questionnaire (see Appendix C). This questionnaire consisted of questions on the subjects' definition of the task, their plan for reviewing the instructional materials, their domain of expertise and other background information on the subjects education and job experiences. There were also general questions that provided an opportunity for the subjects to express their opinion on the study or mention difficulties they had with the reviewing task. Once the questionnaire was completed, the subjects were asked to sign a receipt for the honorarium offered.

Method of Data Analysis

Transcribing

The recorded commentaries of the subjects were transcribed verbatim using a Sanyo Memo-Scriber TRC 9100 transcribing machine. The verbal reports or protocols were typed single space. The transition from the subjects' thinking aloud to reading directly from the text was designated by a double space. All segments representing reading aloud from the text were enclosed in quotation marks. Pauses in the subjects'

speech while thinking aloud were indicated by a comma.

Segmenting

The next step of analysis involved parsing the transcribed protocols into units or segments. This is known as segmenting and there are a number of ways to proceed with this operation. One way is to use clausal analysis, a system developed by Winograd (1983), and based on having a conjugated verb within each segment. Other possible segmenting procedures are based on speech bursts, temporal information, or repetitions (Saroyan, 1989).

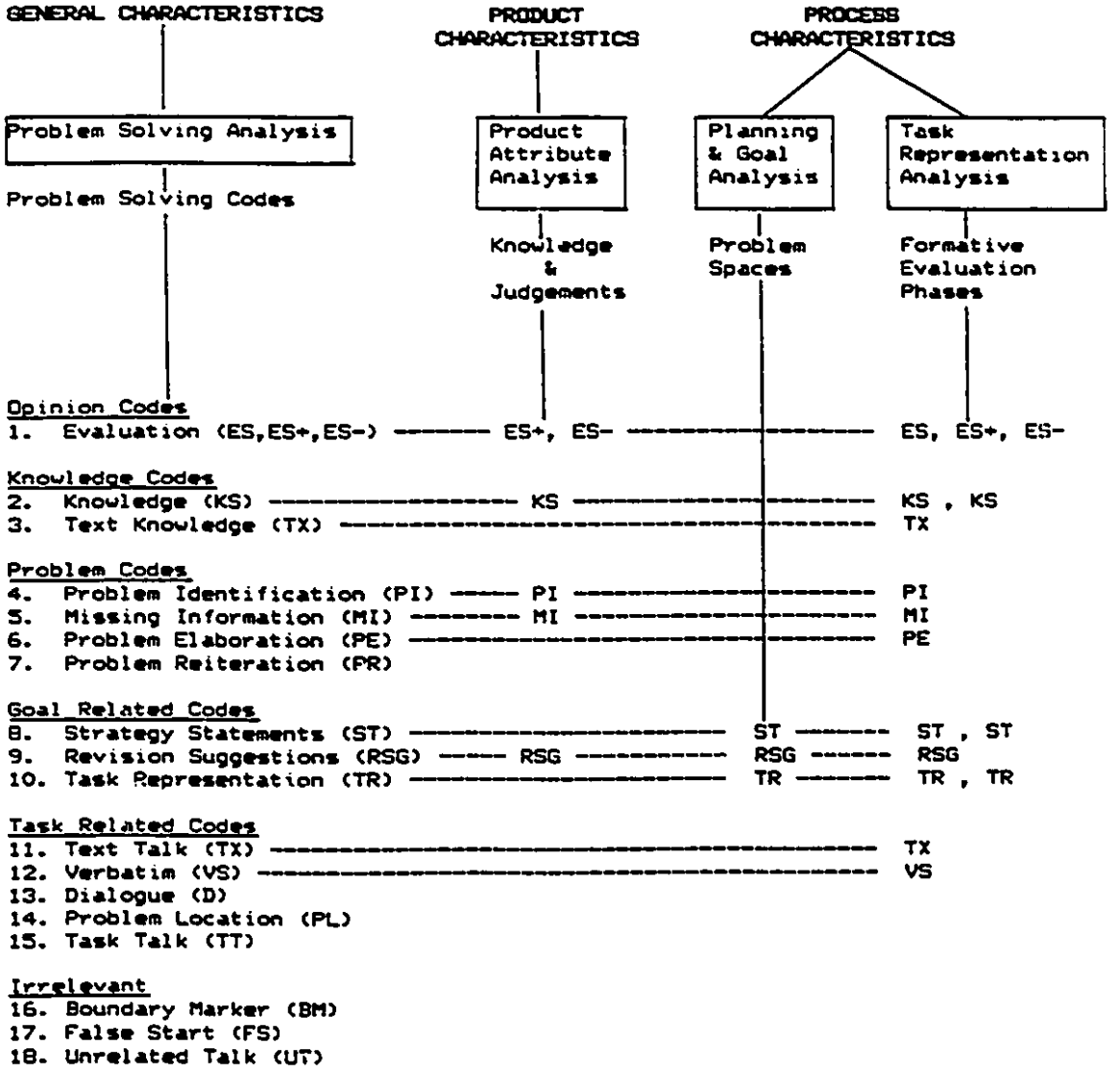
The segmenting system chosen for this study was adopted from Rahilly, Weston, and McAlpine (1991). The premise of this procedure is that a segment must be broad enough to reveal its meaning, yet narrow enough to ensure consistency in segmenting (Rahilly, 1991). In essence a segment is a "meaning unit" or a complete thought. Each unit might include a conjugated verb, a participle, an infinitive or implied verb, and a bound adjunct. Bound adjuncts are defined as phrases (subject with verb absent) or clauses (subject and predicate) that modify or add more meaning to the unit. These adjuncts are connected to the main clause by a binder such as "if", "because", "and", or "while" (see Appendix D for a sample of a segmented protocol).

Overview of Analyses

Table 1 provides an overview of the four types of analyses

Table 1

Overview of Analyses



performed. As indicated, each type of analysis investigated the data in different combinations in order to answer specific questions. The first level of analysis, Problem Solving Analysis, examines the general characteristics of the oral feedback according to the function it performs. The second analysis, Product Attribute Analysis, is associated with the product generated by the experts. For the purpose of this research, the product is defined as information that would be useful in future revisions of the instructional materials. The last two analyses, Planning and Task Representation, are process related and explore the cognitive operations the reviewers engage in.

General Characteristics: Problem Solving Analysis

Ericsson and Simon (1980) make two recommendations concerning the analysis of think-aloud protocols. First, the coding scheme should be established in advance and secondly, it should be based on the terminology of the theoretical constructs that support a particular study. The coding scheme used for this research was developed by Saroyan (1989) and based on the theory of human problem solving. The assumption behind this theory is that the search for a solution to a problem takes place in a problem space. A problem space is a symbolic work space made up of all the possible goals the solver might consider, all possible operations, strategies or moves that might be taken and all the possible conditions that might constrain the solver's actions (Flower, 1989).

Within this mental representation of the problem, an individual links, in a sequence, appropriate actions and their knowledge of the problem. This procedure is repeated until a path is constructed that moves the problem solver from the initial state to the goal state. In order to determine an encoding scheme that represents this process, all of the components of the problem space need to be taken into account and a code derived for each element (Breuleux, 1991).

Saroyan's coding scheme represents the path an individual takes as they search through the formative evaluation problem space. According to Saroyan, this problem solving model is comprised of four elements: 1) identifying a problem, 2) establishing criteria for an acceptable outcome, 3) implementing change, and 4) drawing upon knowledge sources to achieve the desired goal.

Although Saroyan's research investigated revisors rather than reviewers, it did examine experts who were formatively evaluating instructional materials and therefore, her coding scheme was deemed appropriate for this study. The fact that other previous studies on formative evaluation have also modeled their coding schemes on Saroyan's system (Rahilly, 1991; Duy, 1990), furnished additional support for this decision.

In total, 18 codes were used to sort the data; all 13 of Saroyan's codes, 4 codes adopted from Rahilly, and one new code. Some of the

codes underwent slight changes in definition. This coding scheme was applied to all think-aloud and retrospective interview segments.

The codes were grouped into six categories, with each category representing a specific formative evaluation function. This was done in order to facilitate analysis of similar codes. These six categories are: Opinion Related Codes, Knowledge Related Codes, Problem Related Codes, Goal Related Codes, Task Related Codes and Theoretically Irrelevant Codes. A description of each category, its codes and its source will be presented next. Also, an example will be provided for each code.

Opinion Related Codes. Opinion related codes are used to classify statements that express the subjects' feelings about the text without supplying specific details.

1) Evaluation Statements (ES, ES+, ES-): (Saroyan)

These statements represent opinions about the text and are the result of comparing the present status of the materials to the goal state. This category represents neutral, positive or negative comments that are expressions of preference, judgements, internal feelings or observations. The difference between a Problem Identification (PI) and an Evaluation Statement is that these statements do not explicitly state the source of the problem. The following statements are examples of the three different types of evaluation statements:

Neutral Evaluation Statement: "Unusual, now that surprises me."

Positive Evaluation Statement: "That looks OK, AH."

Negative Evaluation Statement: "I would not take it that easily."

Knowledge Related Codes. **Knowledge related codes are used to classify comments that reflect the subject's personal or declarative knowledge, as well as, knowledge gained through the text.**

2) **Knowledge Statement (KS):** **(Saroyan, revised by Tremblay)**

These statements represent references made to prior personal or declarative knowledge of the subject matter or intended audience. They may represent problem identifications when they provide a rationale for suggesting revisions. An example of a Knowledge Statement is:

" There's a lot of new information coming out in that area of research."

3) **Text Knowledge (TK):** **(Saroyan)**

These statements represent segments expressing knowledge acquired from the instructional materials. An example of a Text Knowledge statement is:

"I see that the word environmental refers to when you talk about things that are out of your control."

Problem Related Codes. **Problem related codes are used to classify remarks that refer to the identification of specific problems in the materials.**

4) **Problem Identification (PI):** **(Saroyan)**

These statements represent explicit reference to an observed problem in the instructional materials. An example of a Problem Identification statement is:

"I don't know what they mean by that."

5) **Missing Information (MI): (Rahilly)**

These statements represent a distinct type of Problem Identification (PI), which expresses desire for further information or an observation that information is missing in the text. An example of a Missing Information statement is:

"I would like to know who the study was done on."

6) **Problem Elaboration (PE): (Rahilly)**

These statements represent an expansion of a problem previously identified. This code is always preceded by a Problem Identification (PI) or Missing Information (MI) statement. It can be distinguished from a Knowledge Statement (KS) in that it precisely refers to a problem. The following are examples of a Problem Elaboration and its preceding segment:

Preceding Segment: "I didn't understand that clearly."

Problem Elaboration: "I had to read it a couple of times before it made sense."

7) **Problem Reiteration (PR): (Rahilly)**

These segments are restatements of a previously discussed Problem Identification (PI) or Missing Information (MI). The following are examples of a Problem Reiteration and its preceding segment:

Preceding Segment: "I didn't understand what they said."

Problem Reiteration: "It wasn't apparent to me what they meant by that."

<u>Goal Related Codes.</u> **Goal related codes represent the process of planning in which the subject's are engaged so as to carry out the formative evaluation task. A goal is defined as either the intention to perform an action, a potential action or a future action.**

8) **Strategy Statements (ST): (Saroyan)**

These statements represent references to a tactic usually used by the subject to review materials. A strategy statement typically includes words such as "would", "could", "should", "normally" and "usually". This code is similar to a Task Representation (TR), however, it is a potential action and is not tied to the current task. An example of a Strategy Statement is:

"I would have to go to the study and look it up."

9) **Revision Suggestion (RSG): (Saroyan)**

These statements represent references to proposed changes to the instructional materials. This code is future oriented. An example of a

Revision Suggestion is:

" I think the sentence can be rewritten in a clearer way."

10) Task Representations (TR):(adapted from Task Talk by Saroyan)

These statements represent references to an action or thought that the subject is currently undertaking, or has already taken (if it occurs in the retrospective interview). These statements also refer to the subject's setting up of a short term goal. This is an intended action. An example of a Task Representation statement is:

"Ok, consumer groups, I have to focus on that."

Task Related Codes. Task related codes group segments that consider the manner in which the subject becomes familiar with the task environment or accomplishes the reviewing task.

11) Text Talk (TX): (Saroyan)

These statements represent the subjects's paraphrases or summaries of the text. Segments that are in verbatim form are coded as (VS) Verbatim Statements). An example of a Text Talk statement is:

"It seems to imply a mixture of various foods."

12) Verbatim Statements (VS): (Saroyan)

These statements represent anything that the subject reads or repeats verbatim from the instructional materials.

13) Dialogue (D): (Saroyan)

These statements represent any comments or questions that are

addressed to the administrator. (If this exchange concerns the research task the segments are coded as (TT) Task Talk). The following statements are examples of a dialogue statement and its preceding segment:

Preceding segment: (Administrator) "Did you find the sentence difficult?"

Dialogue: "I don't recall."

14) Problem Location (PL): (Rahilly)

These statements represent references to a problem's location in the text. An example of a Problem Location statement is:

"It was the sentence before that."

15) Task Talk (TT): (Israeloff/Tremblay)

These statements represent comments or questions addressed to the administrator concerning the research task. An example of a Task Talk statement is:

"I can't quite focus on the materials as well as I would at home because I have to read it out loud."

Theoretically Irrelevant Codes. These are segments that represent idiosyncratic speech patterns or verbalizations that are not related to formative evaluation.

16) Boundary Markers (BM): (Saroyan)

These statements represent verbal utterances that sometimes represent a pause in the subject's verbalizations or are a link between

segments. An example of a Boundary Marker is:

"Um, uh, ok, hmm."

17) False Start (FS): (Saroyan)

These statements represent an incomplete thought. An example of a False Start is:

" But I don't know if its..."

18) Unrelated Talk (UT): (Saroyan)

These statement represent comments made by the subject that are not directly connected to the experimental task. An example of an Unrelated Talk statement is:

" I met him once at a conference."

Coding Reliability. Krippendorff (1980) suggests that the accuracy of the coding scheme be established through inter-rater reliability. For a coding scheme to be accurate, there must be general agreement on the definition of different categories. Without this understanding, codes would not be distinctive or recognizable and the coding system could not be applied to other related think-aloud protocols. Inter-rater reliability assesses the quality of this shared understanding by having independent raters attempt to reproduce the coding results.

Inter-rater reliability for this study was verified by having four independent raters code 10 percent of all think-aloud segments. These segments were selected randomly and represented equally both group of

experts. A practice session was provided for the four coders, so that they could familiarize themselves with the coding scheme and the nature of the verbalizations. The coders were supplied with a package that contained a copy of the code definitions and the segments they were to code. The percentage of agreement among all four coders was 75 percent. Irrelevant codes such as Boundary Markers, False Starts, Unrelated Talk and Dialogue were not included in the calculations.

Frequencies and Percentages of Problem Solving Codes. Once the reliability of the coding scheme was verified, the frequency of segments in each of the problem solving codes was calculated. Think aloud data and retrospective data were analyzed separately. Since think aloud data is believed to be a richer data source, only this data was converted into percentages while retrospective data was not analyzed further. Frequencies from think aloud data were converted into percentages based on the total number of segments. These percentages were then grouped according to related problem solving codes. Comparisons were made between the two expert groups, as well as, contrasting individuals within each group. Those segments classified as theoretically irrelevant; Boundary Markers, False Starts and Unrelated Dialogue, were excluded from further analysis

Product Characteristics: Product Attribute Analysis

Once the primary problem solving analysis was completed, a more detailed examination was performed on the data most likely to be used

during future revisions. This was done by applying a second categorization scheme to segments that reflected experts' knowledge and judgements. Using this secondary coding scheme, segments were sorted according to the attributes of a well-designed instructional product (McAlpine & Weston, 1994). This system consisted of four checklists originally designed to assist expert reviewers with the evaluation of instructional materials. Each checklist represented a major attribute of an educational product and consisted of several subcategories (Appendix E). A summary of the four major attributes and their subcategories are presented below, along with examples from the think aloud protocols.

Instructional Design. Included in this category are pedagogical attributes that should be evident in a well designed instructional product. The attribute category is based on the instructional design model and contains the following subcategories: (a) Justification of Need, (b) Target Audience, (c) Entry Level Prerequisites, (d) Objectives, (e) Motivation and Context for Learning, (f) Instructional Strategies, (g) Organization and Structure of Content, (h) Examples, (i) Practice, (j) Feedback, (k) Evaluation of Learning, (l) Internal Alignment and Integration and (m) External Alignment. An example of the Motivation subcategory is:

"It needs to be more interesting, if I were a student I would find it a bit boring."

Subject Matter. This checklist includes attributes relating to the

knowledge structure of a subject domain. The issues addressed by the category are: (a) Value of Content, (b) Content Accuracy, (c) Comprehensiveness, (d) Coherence, (e) Objective Presentation/Bias, and (f) Recency. An example of the Comprehensiveness subcategory is:

" The connection between saturated fats in margarine and cancer should be mentioned here."

Presentation. This checklist deals with the physical attributes of instruction materials. Among the characteristics found in this category are: (a) Space, (b) Typeface, (c) Titles, Headings and Sub-Headings, (d) Use of Numbers, (e) Graphics, Illustrations and Visuals, (f) Audio/Music, (g) Colour, (h) Page Size and Style, (i) Margins, (j) Columns, (k) Technical Quality, (l) Highlighting, and (m) Format and Layout. An example of the Titles, Headings and Sub-headings subcategory is:

" What do the numbers mean in this table? There are no headings."

Language. This category addresses how language is used to express ideas. Subcategories included are: (a) Choice of Vocabulary, (b) Complexity of Sentence Structure, (c) Verbs, (d) Redundancy, (e) Transitions, (f) Consistency, (g) Clarity, (h) Conciseness and (i) Appropriateness for Audience. An example of the Choice of Vocabulary subcategory is:

" These words are too heavy for first year students, this could be simplified".

Frequencies and Percentages of Knowledge Statements according to Product Attribute Analysis. The decision to obtain more details about Knowledge Statements (KS) was based on the fact that knowledge has a profound effect on the judgements made by the experts. The possession of a well organized body of knowledge has been shown to be the primary difference between experts and novices. Schemata for recurrent situations are stored in long term memory and one of their major functions is to construct interpretations of circumstances. Experts use and test this internal model whenever they try to impute meaning to a problem (Glaser, 1984).

Therefore, in order to get a sense of which knowledge bases experts were calling upon, Knowledge Statements in each of the major attribute categories were converted into percentages. These percentages were based on the total number of concurrent Knowledge Statements. Comparisons were made both between and within expert groups.

Frequencies and Percentages of Judgements according to Product Attribute Analysis. Segments coded as Positive and Negative Evaluation Statements (ES+, ES-), Missing Information (MI), Problem Identification (PI) and Revision Suggestions (RSG) were considered to be judgements and were chosen for additional coding as they represented conclusions about the product being formatively evaluated. This information would be useful data for revisors. In order to get a sense of what aspects of materials the choice to revise, Positive and Negative Evaluation Statements, Missing Information, Problem Identifications and Revision Suggestions in each of the attribute categories were converted into percentages. These percentages were based on the total number of concurrent judgement comments, excluding theoretically irrelevant codes. Comparisons were made between the two expert groups, as well as, contrasting individuals within each group.

Process Characteristics: Planning and Goal Analysis

Ill-structured tasks are activities that are not fully defined beforehand and normally evoke a highly variable set of responses (Glaser, 1991; VanLehn, 1989). These tasks are considered to be complex and knowledge-rich as they demand sifting through large amounts of problem related details. Formative evaluation matches this description and therefore, can be viewed as an ill-structured problem.

The standard problem solving approach to analyzing think-aloud

protocols is inappropriate for ill-structured tasks, as it focuses on knowledge structures and fails to examine planning (Lampert & Clark, 1990). With such ambiguous tasks, planning is often the distinguishing characteristic of expert problem solving, as experts are more likely to generate a working model of the problem prior to implementing a solution. By using a plan, which is defined as a set of goals, experts avoid the need to rely on time consuming random verification (Hayes, 1989). Due to experts' predilection towards planning, simply knowing how they structure their thinking about a problem provides little information on the application of this knowledge in practice.

Consequently, it was decided to follow Bracewell and Breuleux's recommendation (1992) to perform an in-depth analysis on goal and planning related comments evident in the think aloud protocoi. To analyze these strategic operations it is necessary to identify goals, which are the basic units of a plan. The outcome of this type of goal analysis is that direct identifications can be made of the process of planning in which the individual is engaged and the actual plan they constructed.
In order to provide a mechanism for investigating the planning process, all three Goal Related codes were further analyzed: Revision Suggestions (RSG), Strategy Statements (ST) and Task Representations (TR). After reviewing the goals the experts established, it appeared that they were operating in three distinctive problem spaces; Text Comprehension,

Review and Revision. A coding scheme, developed by the researcher to represent these three problem spaces, can be found in Appendix F. Each problem space contains a description of the identified goal and the strategic operations associated with this goal. This coding scheme was derived by first: sorting Goal Related Codes according to their related problem space and second: analyzing in more detail the various strategies within each category. A summary of the three problem spaces, as well as specific examples, are presented below.

Text Comprehension Problem Space. The goal of this problem space is to improve reading comprehension, and includes techniques used by strategic readers (Flood & Lapp, 1990). Goal related codes were sorted according to the strategic operations associated with this goal and include: (a) Define Objective, (b) Scan Text, (c) Build Background, (d) Concentrate, (e) Predict, (f) Monitor Comprehension and (h) Clarify. An example of the Scanning Text subcategory is:

"I'm going to get a sense of what this is about by reading the first and last sentence of every paragraph."

Review Problem Space. This problem space and associated strategies represent the goal of finding positive and negative features of the instructional materials and passing judgement on them. Goal related codes were sorted according to the following strategies: (a) Skim Text, (b) Read Thoroughly, (c) Identify Confusing Information, (d) Check

Original Sources, (e) Check Definitions, (f) Search for Ambiguous Words, (f) Rely on Prior Knowledge, (g) Challenge Content and the Quality and (h) Comment. An example of the Checking Original Sources subcategory is:

" I wonder about that, I'd have to go to the study and look up and see exactly what it says."

Revision Problem Space. This problem space represents the goal of making changes to the instructional materials for the purpose of improving them. It was found that the strategies mentioned by experts related closely to the product attribute they detected lacking in the text. For example, if a reviewer thought the text would not motivate the learner to continue reading, they might suggest some possible techniques for gaining the readers attention. Therefore, the product attribute checklists conceived by McAlpine and Weston (1994) was adapted for the purpose of analyzing this problem space. Goal related codes were sorted according to the same four headings: (a) Instructional Design, (b) Subject Matter, (c) Presentation and (d) Language. An example of a Presentation strategy is:

" That's what I do, I make a chart of these things listing the foods."

Frequencies and Percentages of Goal Related Codes according to Planning and Goal Analysis. Revision Suggestions (RSG), Strategy Statements (ST) and Task Representations (TR) codes, representing goals

and their subsequent plans, were sorted into each of the three planning problem spaces. Frequencies were converted into percentage scores based on the total number of concurrent goal related comments. Comparisons were made both between and within expert groups.

Process Characteristics: Task Representation Analysis

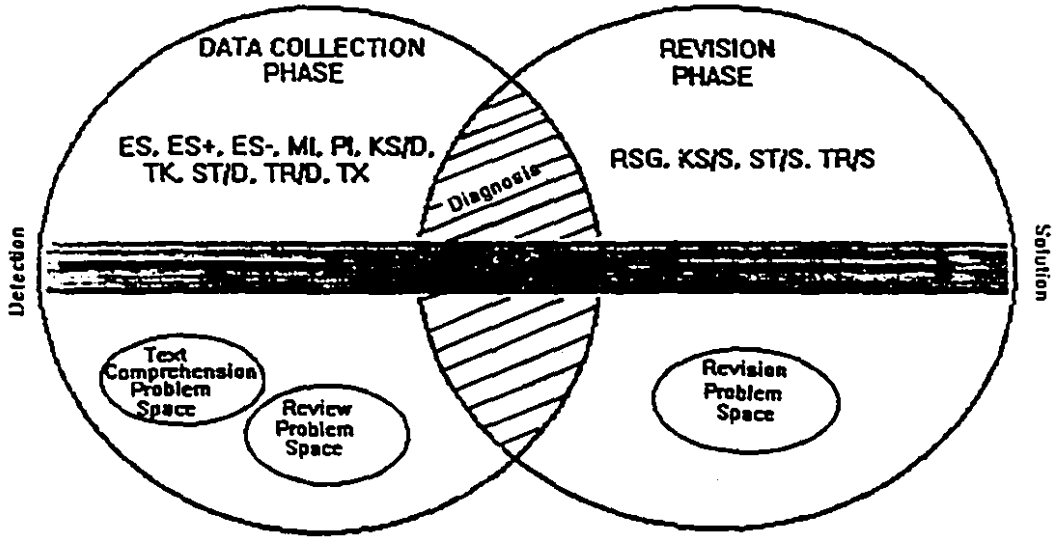
The final stage of analysis was to examine how the subjects represented the task for themselves, since task definition begins the process of problem solving and influences the operating plans developed. Flower, Carey and Hayes (1985) suggest that revisor's of text representation the task on a continuum from low information to high information. At the low end of the continuum subjects depict the task as only the simple detection of problems, however, at the high end subjects not only identify a problem but describe its symptoms and prescribe a solution. Thiagarajan (1978) describs expert feedback during formative evaluation as being on a similar continuum. The low end includes identifying problems and causes, while the high end; more useful since it allows for strategic revisions, includes suggesting and implementing revisions.

The low end, detection of problems, can be equated with the data collection phase of formative evaluation and the high end, development of solutions, can be compared with the revision phase of formative evaluation. To investigate how the experts in the current study represented their task,

problem solving codes were sorted according to whether they focused on detection (the data collection: low information) or on solutions (revision phase; high information). This examination made it possible to assess where the subjects concentrated their efforts.

Figure 1 was developed to illustrate the multiple problem solving nature of the formative evaluation task. This was done by demonstrating the relationship of the phases of formative evaluation to the previously described planning problem spaces. As can be seen, this ill defined task requires simultaneous attempts to resolve several related problem spaces. Both the Text Comprehension and the Review problem space occur during the data collection phase of formative evaluation; where judgements are made about the educational quality of the materials. In this phase, the subject searches through the materials in order to detect features they perceive to be either positive or negative. This is followed by the Revision problem space that occurs during the revision phase of formative evaluation; where changes and corrections are made to the materials. It appears that a subject's particular definition of the task influences whether they move on to the revision phase or remain in the data collection phase. If a subject perceives their task to be detecting, then they are likely to remain in the data collection phase and provide low level information. On the other hand, if a subject extends their representation of the task to include diagnosis and solution planning for identified

Figure 1. Task Representation in Relation to Phases of Formative Evaluation



problems, they are more apt to move along the continuum to revision and provide high level information.

Codes were sorted according to the formative evaluation phase to which they were linked. Three codes, Knowledge Statements, Strategy Statements and Task Representations, appear in both stages and are marked either D for detection or S for solution to symbolize the phase they are connected to. The following outline presents the two formative evaluation phases and the codes that are related to each phase.

Data Collection Phase - (ES, ES+, ES-, MI, PI, PE, KS/D, TK, TX, VS, ST/D, TR/D). This stage includes all of the codes that pertain to the data collection phase of formative evaluation. In this stage the reviewers compare the current materials against their standard for a well-designed instructional product and make judgements concerning its educational attributes. When in the text comprehension problem space, the subjects rely on their knowledge of strategic reading to establish short-term goals to develop appropriate solution plans for understanding the text. Whereas, when representing the review problem space they rely on their knowledge of evaluating educational text to construct their review plans. Hence, those statements that are linked to comprehending the text or identifying problems, such as Knowledge Statements (KS/D), Task Representation (TR/D) and Strategy Statements (ST/D), are included in the first phase. Another aspect of the text comprehension task is to read the text and

absorb the information in order to detect problems. Codes such as Verbatim Statements (VS), Text Talk (TX) and Text Knowledge (TK) represent this function. During the review task, when the expert recognizes either a positive or negative aspect of the materials, they might comment on this observation by making an Evaluation Statement (ES,ES+,ES-). On the other hand, they might choose to draw conclusions and pass judgement on the features they observed. These judgement codes include Missing Information (MI), Problem Identification (PI) and Problem Elaboration (PE).

Revision Phase - (KS/S, TR/S, ST/S, RSG). This stage includes only those codes that represent the revision phase of formative evaluation, which is typically carried out after data is collected. In order to represent the revision problem space, the reviewer apparently turns first to their declarative or factual knowledge to diagnose the type of problem they encountered and its symptoms. They then search through their knowledge of procedures to recall actions that have been successful previously in solving similar types of revision problems. If a ready-made solution is not available, new strategies might be devised and prescribed to correct problems within the text. The codes involved in this phase are Revision Suggestions (RSG) and those Knowledge Statements (KS/S), Task Representations (TR/S) and Strategy Statements (ST/S) that are linked to revisions.

Frequencies and percentages according to the Task Representation analysis. The frequency of occurrence for all think aloud coded segments in terms of task representation was determined. Frequencies of these coded segments were sorted into their corresponding problem solving phase and converted into percentages. Comparisons were made both between and within expert groups. Those codes considered to be redundant or outside of the formative evaluation problem solving model; Problem Reiteration (PR), Problem Location (PL), Task Talk (TT) and Dialogue (D) were excluded from this analysis.

Analysis of Debriefing Questionnaire

Responses on the debriefing questionnaires were summarized according to the experts' background information and their definition and planning of the review task. A synopsis of the reviewers' background information included their domain of expertise, educational credentials, prior experience with text evaluations and familiarity with the content and intended audience. This was followed by a further analysis of the experts' description of the formative evaluation task and the plans they made to accomplish this task.

CHAPTER THREE

Results

Introduction

This research is guided by two principal questions: 1) What are the defining characteristics of the product for revision generated by subject matter and learner experts during their review of instructional materials? 2) What are the characteristics of the cognitive processes that the two groups of reviewers engaged in? The following chapter focuses on the above questions by comparing the similarities and differences of the think aloud and retrospective data produced by the two expert groups. Comparisons have been made at two levels, both between and within the two groups.

This chapter is divided into four sections. The first section introduces preliminary results pertaining to the Problem Solving analysis, specifically the frequency in which each code is utilized and the percentages of theoretically relevant functional groupings. The second section provides an examination of those Problem Solving codes that reflect the product characteristics; Knowledge Statements and Judgement codes. These two codes, which would be of interest to a revisor of the instructional materials, are sorted according to Product Attribute analysis. The next section presents results pertaining to the processes employed by the various experts. Goal Related codes are studied using the Planning

and Goal analysis, while findings of the Task Representation analysis are presented in reference to two phases of formative evaluation; Data Collection and Revision. The final section offers qualitative comparisons of the subjects based on information obtained from the debriefing questionnaire.

General Characteristics: Problem Solving Analysis

Frequencies and Percentages of Think Aloud and Retrospective Data

The frequency of segments in each Problem Solving code are presented in Table 2 and Table 3. Table 2 presents think aloud data, and retrospective interview data is shown in Table 3. Both expert groups provide more think aloud than retrospective data, except for Dialogue and Problem Location codes. These are the only codes where the subjects make more comments during the retrospective interview than during the think aloud procedure. A major difference between the two group of experts, is the highly verbal nature of the learner experts. Combined totals of Table 2 and 3 show that these experts are twice as verbal in comparison to subject matter experts (M_{SME}=1415; M_{LE}=2737).

The remainder of the results section will discuss findings in terms of percentages only. As mentioned earlier, there is a large difference in total number of segments for both expert groups. In order to allow for meaningful comparisons between the two type of experts it is preferable to convert frequencies into percentage scores. Moreover, further results will

Table 2

Frequency of Think Aloud Segments in each Problem Solving Code

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
OPINION RELATED CODES:										
ES	1	13	3	23	40	12	-	56	9	77
ES+	7	31	22	60	120	14	17	35	17	83
ES-	1	1	1	6	9	14	2	45	62	123
KNOWLEDGE RELATED CODES:										
KS	31	44	39	156	270	74	77	83	64	298
TK	2	8	2	5	17	11	1	24	6	42
PROBLEM RELATED CODES:										
PI	30	47	12	28	117	90	25	64	85	264
MI	13	3	1	23	40	9	3	10	31	53
PE	10	6	-	17	33	23	2	19	32	76
PR	3	1	-	4	8	11	1	27	6	45
GOAL RELATED CODES:										
ST	2	-	-	3	5	16	3	11	1	31
RSG	16	18	1	22	57	37	55	24	19	135
TR	13	5	9	8	35	70	5	81	42	49
TASK RELATED CODES:										
D	5	-	1	-	6	2	1	49	4	56
TT	2	1	2	16	21	11	1	34	3	49
TX	7	20	15	20	62	30	3	58	85	176
VS	10	7	9	16	42	10	4	13	5	32
PL	-	-	-	-	0	-	1	1	-	2
IRRELEVANT CODES:										
BM	8	96	10	53	167	7	6	135	99	247
FS	11	2	6	4	23	29	20	24	15	88
UT	10	2	1	54	67	17	16	43	76	152
Total:	182	305	134	518	1144	487	243	836	661	2227

Table 3

Frequency of Retrospective Segments in each Problem Solving Code

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
OPINION RELATED CODES:										
ES	-	-	-	4	4	-	-	10	5	15
ES+	-	1	1	5	7	1	2	2	-	5
ES-	-	-	-	2	2	1	-	3	5	9
KNOWLEDGE RELATED CODES:										
KS	10	11	9	7	37	47	15	46	12	120
TK	1	-	-	4	5	1	-	1	-	2
PROBLEM RELATED CODES:										
PI	6	7	7	9	29	13	7	11	17	48
MI	-	-	-	3	3	-	-	-	2	2
PE	1	1	-	10	12	7	3	7	8	25
PR	1	2	1	1	5	4	1	5	5	15
GOAL RELATED CODES:										
ST	-	-	-	1	1	3	-	4	-	7
RSG	3	2	-	-	5	2	4	4	-	10
TR	15	4	1	8	28	29	3	27	1	60
TASK RELATED CODES:										
D	8	6	5	24	43	2	15	33	12	62
TT	5	-	1	-	6	-	-	-	-	0
TX	3	7	4	4	18	11	2	11	7	31
VS	3	-	4	5	12	4	2	3	7	16
PL	1	-	4	-	5	1	1	3	-	5
IRRELEVANT CODES:										
BM	1	6	1	7	15	-	2	15	7	24
FS	7	1	2	3	13	8	10	9	-	27
UT	1	-	2	18	21	-	18	66	2	86
Total:	66	48	42	115	271	134	85	260	90	570

focus solely on think aloud data as it is believed to be a richer data source. As mentioned previously in Chapter Two, using think aloud data reduces the potential loss or distortion of information that frequently occurs when individuals report retrospectively.

Problem Solving Codes Grouped by Function

In Table 4, percentages are calculated by excluding from the total number of segments those codes that are considered to be theoretically irrelevant; Boundary Markers, False Starts and Unrelated Talk. Both within group percentages and between group data are determined. Group mean data is calculated using the total number of segments across subjects. All subsequent tables, with the exception of Table 12 and 13, will be organized similarly.

Between Group. An overall comparison between the two expert groups is depicted in Figure 2. This graph shows that close to half of the comments made by the experts involve either identifying problems (Problem Related Codes: M_{SME}=23.0; M_{LE}=25.0) or referring to personal and factual knowledge about the materials or the intended audience (Knowledge Related Codes: M_{SME}=30.5; M_{LE}=21.5). One difference between the two groups is that subject matter experts have a considerably higher percentage of segments in the Knowledge Related Category (M_{SME}=30.5; M_{LE}=21.5). On the other hand, learner experts have a substantially higher percentage of Goal Related segments

Table 4

Percentage of Think Aloud Segments Grouped by Problem Solving Function

OPINION RELATED CODES: (ES, ES+, ES-)

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
6.0	22.0	22.0	22.0	18.0	9.0	9.5	21.5	19.0	15.0

KNOWLEDGE RELATED CODES: (KS, TK)

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
22.0	25.5	35.0	39.5	30.5	19.0	38.0	16.0	14.0	21.5

PROBLEM RELATED CODES: (PI, MI, PE, PR)

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
36.0	28.0	11.0	18.0	23.0	30.5	16.0	19.0	33.0	25.0

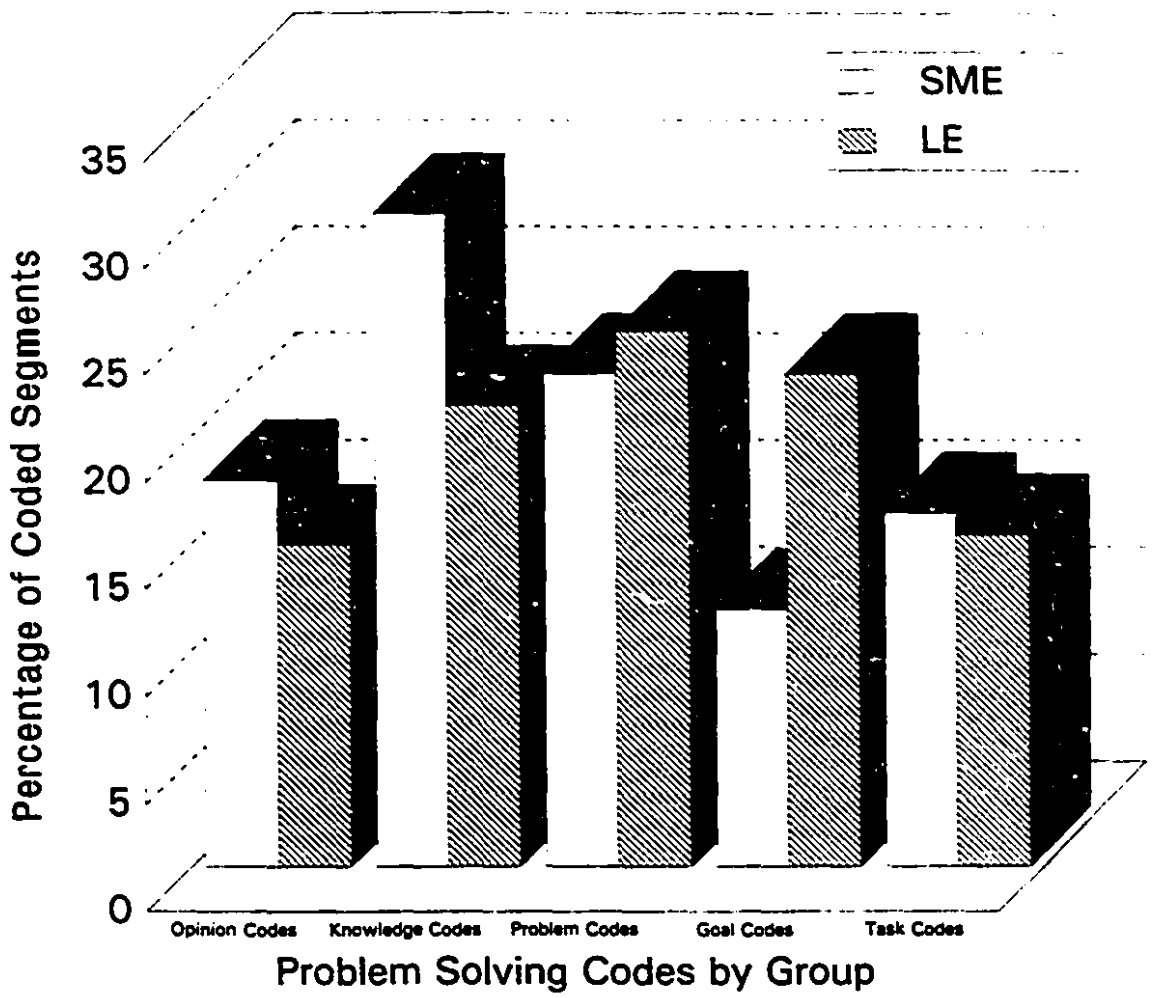
GOAL RELATED CODES: (ST, RSG, TR)

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
20.5	11.0	9.0	8.0	12.0	28.5	31.5	18.5	13.0	23.0

TASK RELATED CODES: (D, TT, TX, VS, PL)

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
16.0	14.0	23.0	13.0	16.5	12.0	4.5	24.0	20.5	15.5

Figure 2. Between Group Comparision: Percentage of Segments Grouped by Problem Solving Function



(M_{SME}=12.0; M_{LE}=23.0). A further dissimilarity is that the comments of these experts are more dispersed across categories than those of the subject matter experts.

Within Group. Within group percentages were examined for consistency among subjects with the same expertise. Figures 3 and 4 represent individual data by expertise and in general show that subject matter experts have a more congruous distribution of comments than do learner experts. One individual, Subject Matter Expert I, demonstrates a more unique pattern of comments and this discrepancy accounts for the wide range of scores found in the evaluation, problem and goal related classifications. In contrast, the learner experts are more idiosyncratic; and there are no coherent within group similarities.

Opinion Related Codes. Included in this category are Evaluation Statements that represent neutral, positive and negative comments. Group mean data for these statements are presented in Figure 5. Both groups of experts have the same percentage of Neutral Evaluation Statements (M_{SME}=4.0; M_{LE}=4.0), however, differences between the two groups are found with positive or negative evaluation comments. For example, subject matter experts have three times the percentage of Positive Evaluation Statements in comparison to learner experts (M_{SME}=13.0; M_{LE}=4.5), yet they have a considerably lower percentage of Negative Evaluation Statements (M_{SME}=1.0; M_{LE}=6.0).

Figure 3. Within Group Comparision: Percentage of SME Segments Grouped by Problem Solving Function

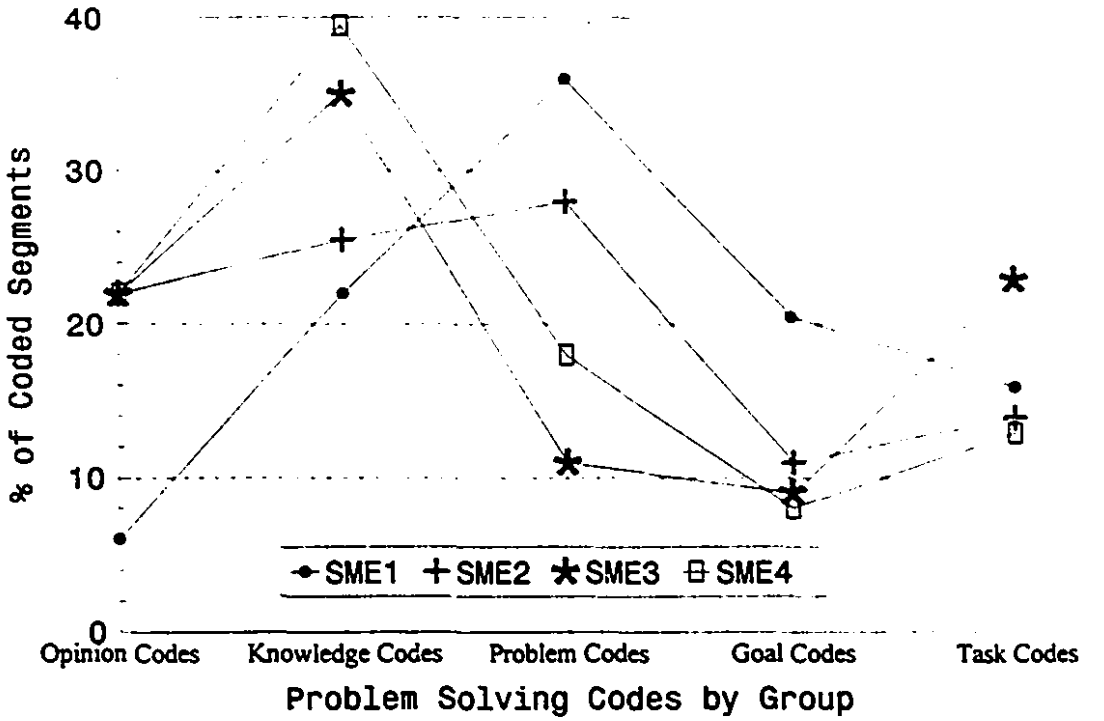


Figure 4. Within Group Comparision: Percentage of LE Segments Grouped by Problem Solving Function

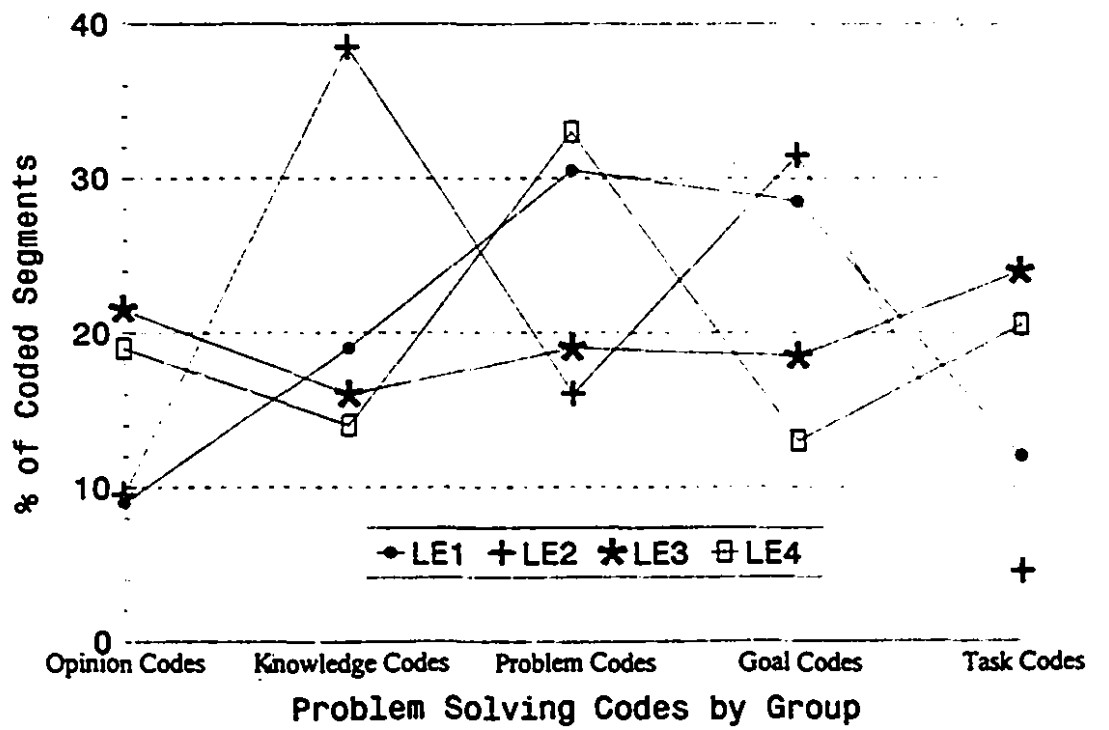
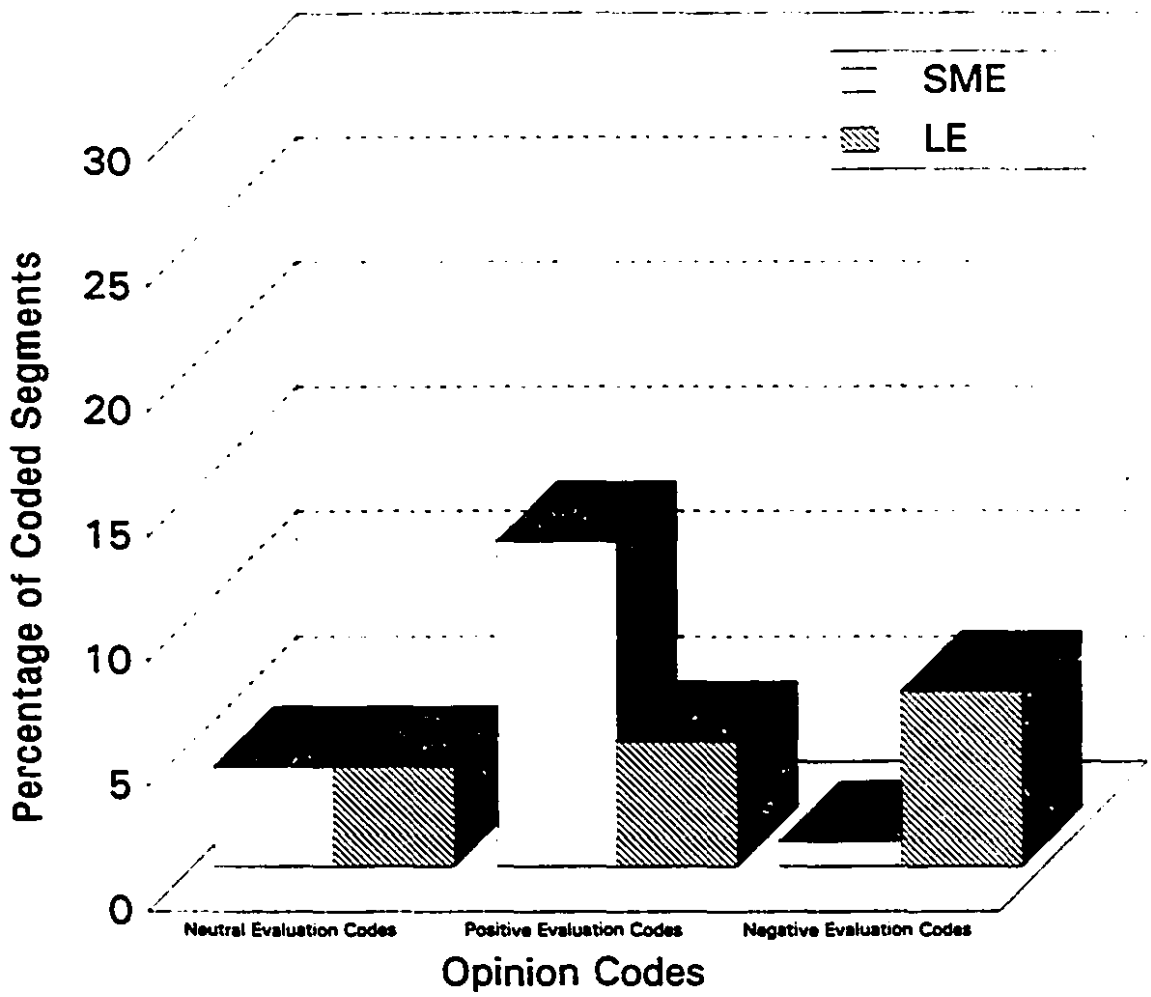


Figure 5. Percentage of Opinion Related Codes



Knowledge Related Codes. This category consists of Knowledge Statements and Text Knowledge. Figure 6 reveals that subject matter experts have a notably higher percentage of Knowledge Statements when compared to learner experts (M_{SME}=29.0; M_{LE}=18.5). Both groups have approximately the same percentage of segments for Text Knowledge statements (M_{SME}=1.5; M_{LE}=3.0).

Problem Related Codes. This category includes Missing Information, Problem Identification, Problem Elaboration and Problem Reiteration codes. As indicated by Figure 7, there is only a slight difference in the percentage of comments uttered by both groups of experts for any of the above codes.

Goal Related Codes. Contained in this category are Strategy Statements, Task Representations and Revision Suggestions. Figure 8 presents the mean data for this category and establishes that when compared to subject matter experts, learner experts generate close to three times the percentage of Task Representation comments (M_{SME}=4.0; M_{LE}=11.5). The percentages of Strategy Statements (M_{SME}=1.0; M_{LE}=2.0) and Revision Suggestions (M_{SME}=7.0; M_{LE}=8.0) are essentially the same for both groups.

Task Related Codes. Included in this category are Dialogue, Task Talk, Text Talk, Verbatim Statements and Problem Location. As Figure 9 demonstrates, subject matter experts when compared to learner experts,

Figure 6. Percentage of Knowledge Related Codes

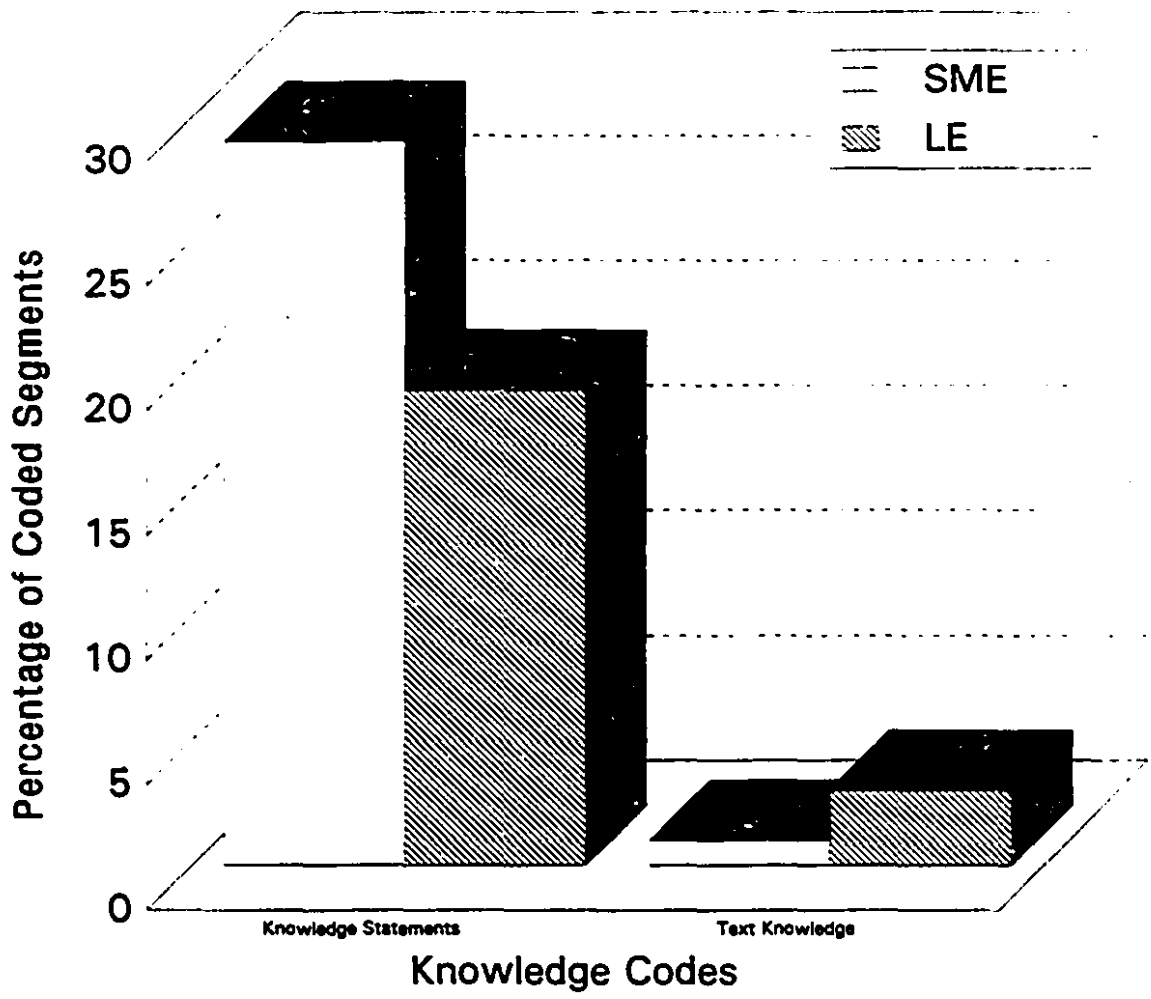


Figure 7. Percentage of Problem Related Codes

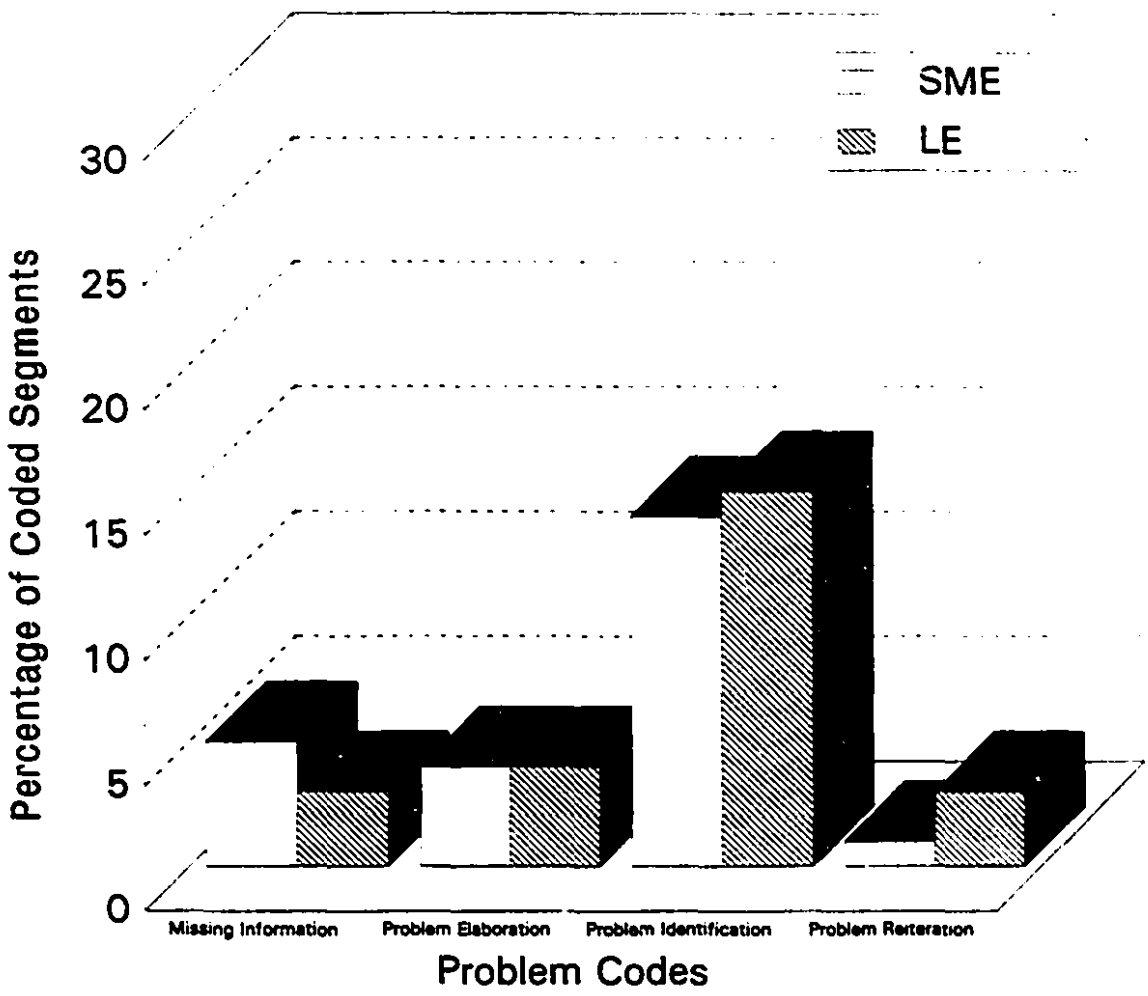
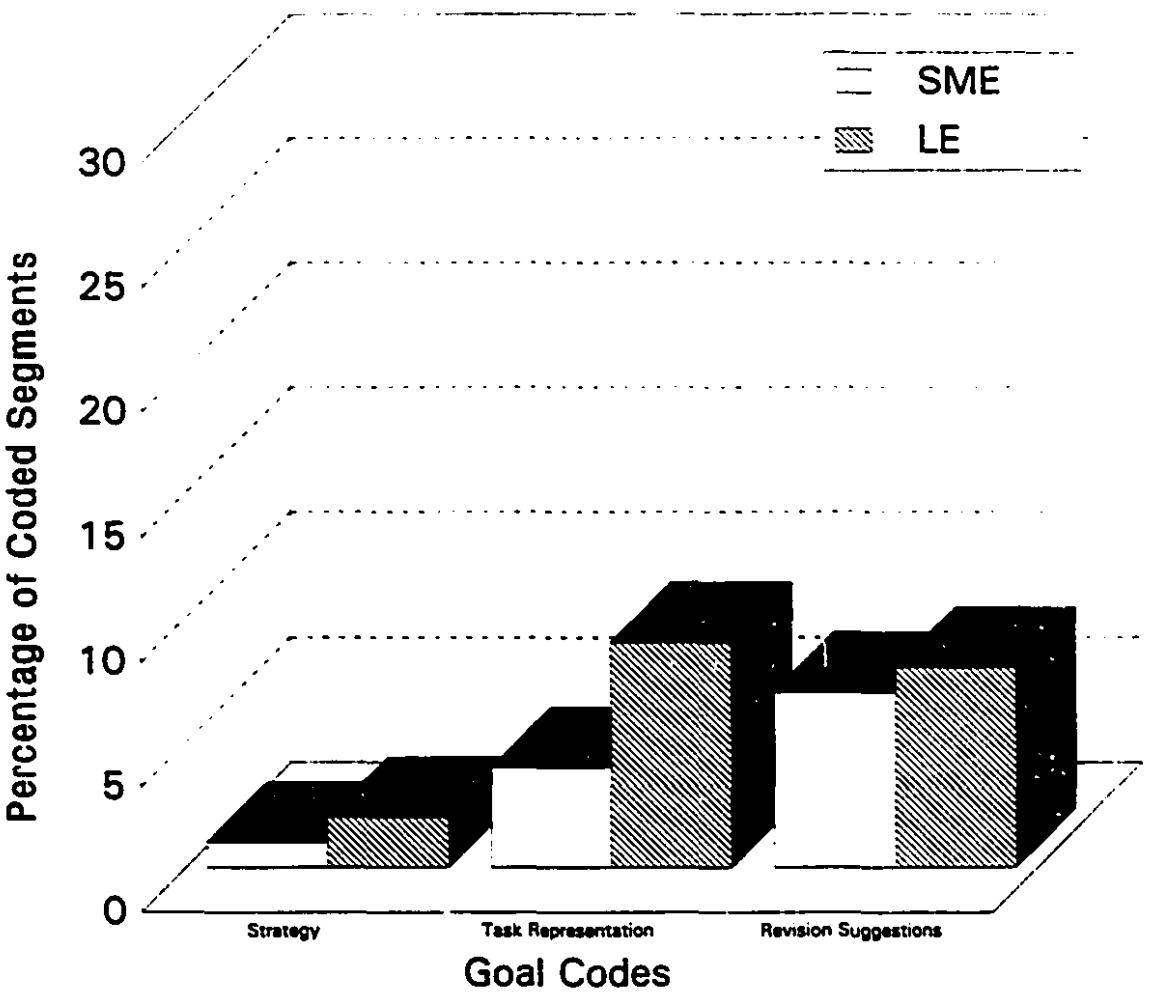


Figure 8. Percentage of Goal Related Codes



have a considerably higher percentage of Verbatim Statements (M_{SME}=5.5; M_{LE}=1.5). For the other task related codes the percentages are approximately the same.

Product Characteristics: Product Attribute Analysis

Product data is information provided by reviewers that is deemed to be the most useful for revision. The Problem Solving codes discerned as representative of this data are Knowledge Statements and Judgement codes. All of these codes are sorted according to Product Attributes.

Product Attribute Analysis of Knowledge Statements

The decision to obtain more details about Knowledge Statements was based on the fact that available knowledge structures have a profound effect on the type and quality of the judgements made by experts. Furthermore, in the earlier analysis of related problem solving codes (see Table 4), the expert groups showed a marked difference in the knowledge statements they uttered. It was found that subject matter experts refer to their personal and declarative knowledge considerably more than do learner experts (M_{SME}=29.0; M_{LE}=18.5).

Percentages based on the total number of Knowledge Statements and sorted into the four main attribute categories are presented in Table 5 between and within groups.

Between Group. Figure 10 shows that both group of experts have a substantially higher percentage of knowledge related comments in the

Figure 9. Percentage of Task Related Codes

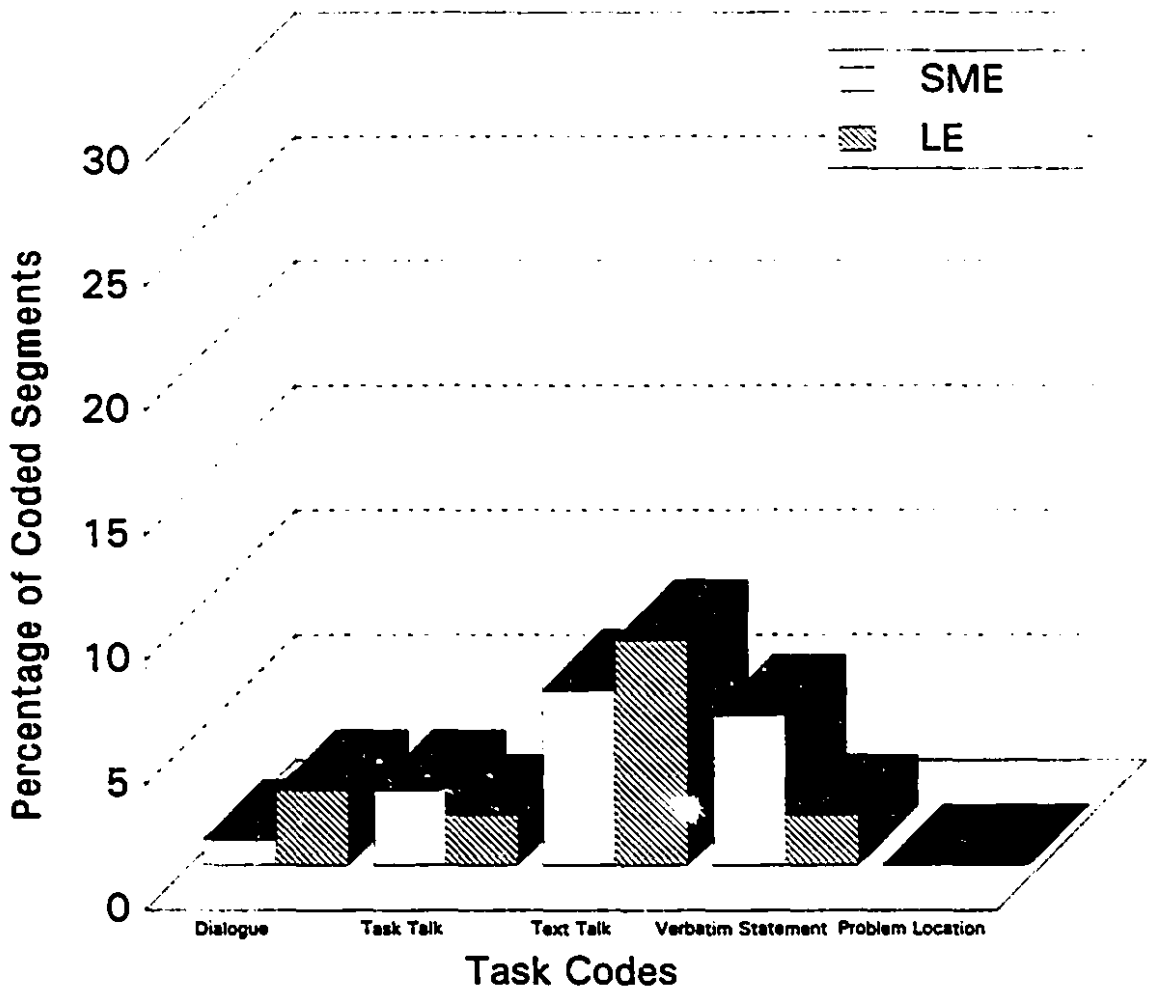


Table 5

Percentage of Knowledge Statements (KS) according to Product Attributes

I - INSTRUCTIONAL DESIGN ATTRIBUTES:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
16.0	30.0	-	2.5	12.0	28.0	4.0	17.0	11.5	15.0

II - SUBJECT MATTER ATTRIBUTES:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
67.0	65.0	100.0	93.5	81.5	65.0	90.0	65.0	83.0	76.0

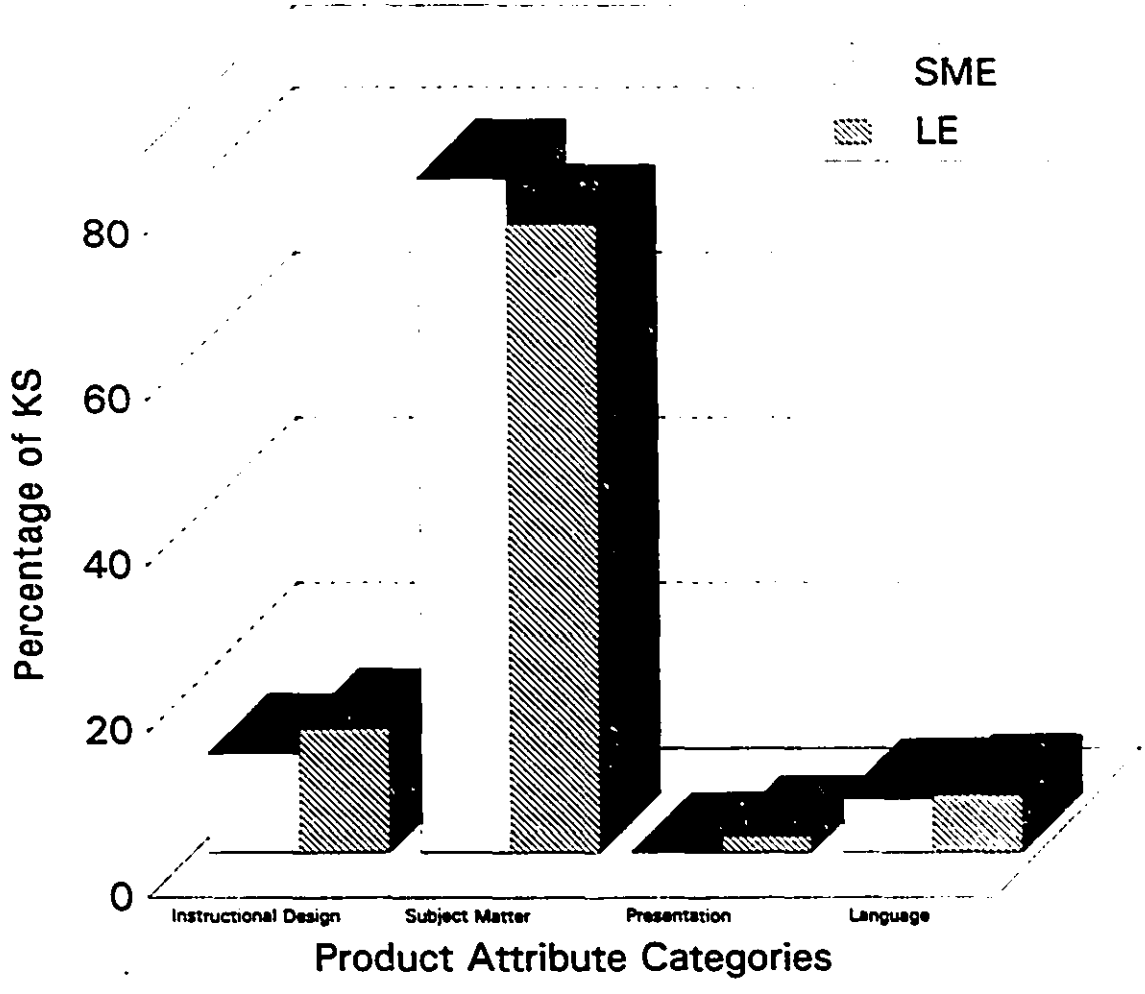
III - PRESENTATION ATTRIBUTES:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
-	-	-	-	0	3.0	5.0	1.0	-	2.0

IV - LANGUAGE ATTRIBUTES:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
17.0	5.0	-	4.0	6.5	4.0	1.0	17.0	6.0	7.0

Figure 10. Between Group Comparison: Percentage of Knowledge Statements according to Product Attributes



subject matter category (M_{SME}=81.5; M_{LE}=76.0), with the lowest percentage of comments in the presentation category (M_{SME}=0; M_{LE}=2.0).

The two groups differ in terms of the amount of attention they give to each of the attribute categories. Learner experts make fewer references to their content knowledge than do subject matter experts, and place slightly more emphasis on the knowledge of instructional design, language and physical presentation attributes of the text.

Within Group. As Figure 11 demonstrates, all subject matter experts focus most of their attention on content issues and ignore presentation problems. Similarly, Figure 12 demonstrates that learner experts also place more emphasis on content issues than the other attributes.

Product Attribute Analysis of Judgement Codes

Judgements about the materials; Positive and Negative Evaluation Statements, Problem Identifications, Missing Information and Revision Suggestions have been coded according to product attributes. These codes were chosen for additional analysis since they represent judgements about the text and provide information helpful to making revisions. Percentages are based on the total number of think aloud data segments and subsequently divided into the four attribute categories: instructional design, subject matter, presentation and language. Table 6 presents these results both between and within groups.

Figure 11. Within Group Comparison: Percentage of SME Knowledge Statements according to Product Attributes

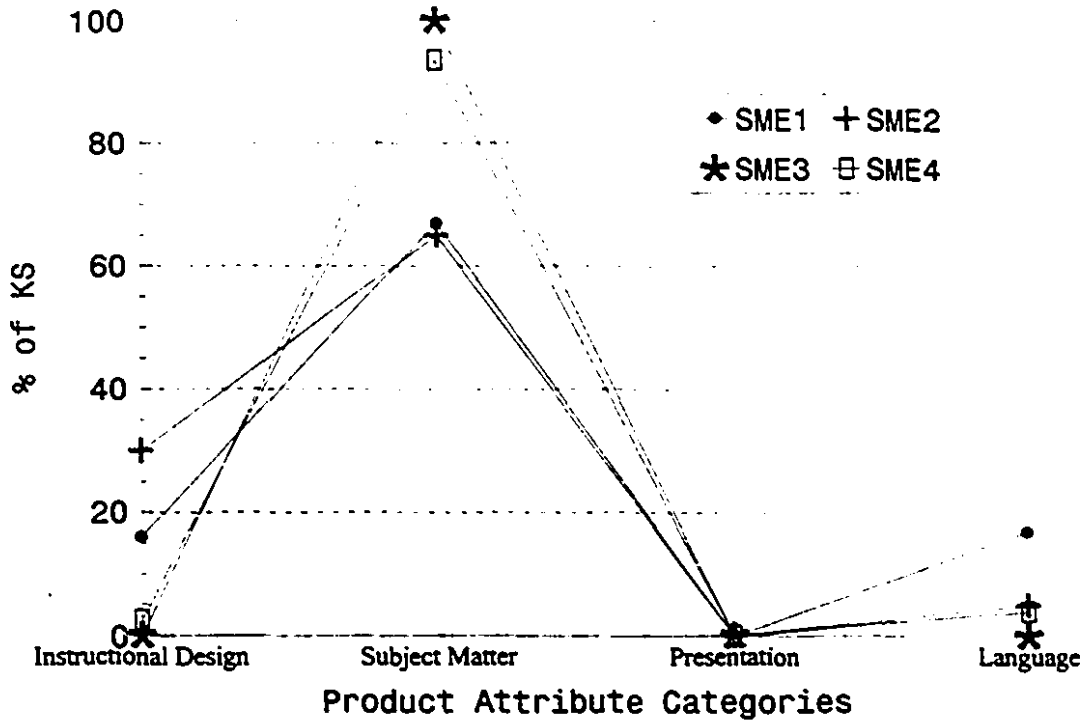


Figure 12. Within Group Comparision: Percentage of LE Knowledge Statements according to Product Attributes

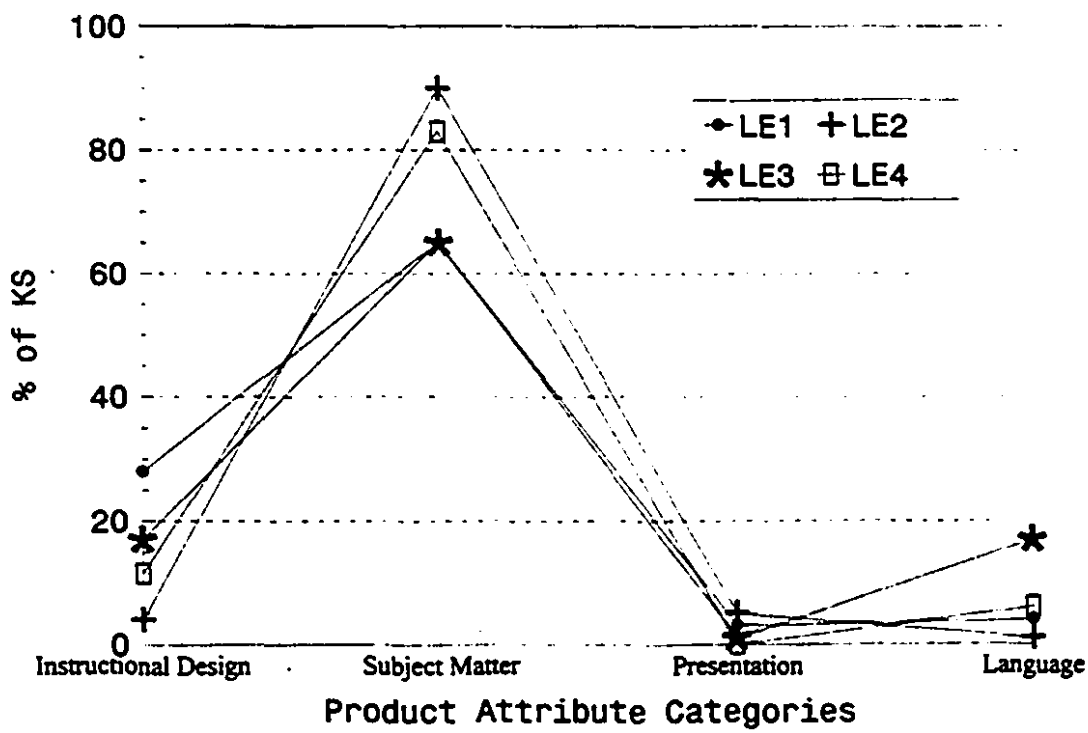


Table 6

Percentage of Judgements Codes (ES+,ES-,MI,PI,RSG) according to Product Attributes

I - INSTRUCTIONAL DESIGN ATTRIBUTES

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
ES+	1.0	.5	-	.5	.5	1.0	-	2.0	-	1.0
ES-	-	-	-	-	0	-	-	2.5	2.0	1.0
MI	1.0	-	-	-	.5	-	-	-	-	0
PI	-	11.0	-	2.0	3.0	10.0	1.0	2.0	6.0	5.0
RSG	3.0	2.0	1.0	0	1.5	3.5	2.0	0	1.0	2.0
Total	5.0	12.5	1.0	2.5	5.5	14.5	3.0	6.5	9.0	9.0

II - SUBJECT MATTER ATTRIBUTES

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
ES+	3.0	14.0	19.0	14.0	12.5	2.0	7.5	3.5	3.5	4.0
ES-	1.0	.5	1.0	1.0	1.0	1.5	2.0	3.0	10.5	4.0
MI	6.5	1.5	1.0	6.0	4.0	2.0	1.5	1.5	6.0	3.0
PI	13.0	9.0	9.0	3.0	8.5	6.0	8.0	3.0	9.0	6.5
RSG	7.0	5.0	1.0	4.5	4.5	3.0	19.0	.5	1.0	5.5
Total	30.5	30.0	31.0	28.5	30.5	14.5	38.0	11.5	30.0	23.0

III - PRESENTATION ATTRIBUTES

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
ES+	-	-	-	-	0	.5	-	-	-	0
ES-	-	-	-	-	0	1.0	-	.5	-	.5
MI	-	-	-	-	0	-	-	-	-	0
PI	3.0	-	-	-	1.0	3.0	.5	1.0	.5	1.5
RSG	2.5	-	-	-	0	1.0	4.5	0	0	1.5
Total	5.5	0	0	0	1.5	5.5	5.0	1.5	.5	3.5

IV - LANGUAGE ATTRIBUTES

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
ES+	-	.5	-	1.0	.5	-	2.0	.5	-	.5
ES-	-	-	-	-	0	.5	-	1.0	-	.5
MI	.5	-	-	-	0	-	-	-	.5	0
PI	3.0	2.5	1.0	2.0	2.0	1.5	3.0	2.0	3.0	2.5
RSG	0	0	-	.5	0	1.0	3.0	2.0	1.0	2.0
Total	3.5	3.0	1.0	3.5	2.5	3.0	8.0	5.5	4.5	5.5

Between Group. Figure 13 shows that when judgement codes are sorted according to product attributes, the pattern of comments are parallel to the pattern of knowledge statements made by the experts (see Figure 10). The overall distribution of results is the same, but the amount of attention given to each attribute differs somewhat. Learner experts focus slightly less on content and more on other attributes, while subject matter experts focus slightly more on content and less on other aspects. Both groups have the highest percentage of segments in the Subject Matter category (M_{SME}=30.5; M_{LE}=23.0) and the lowest percentage in the Presentation Attribute category (M_{SME}=1.5; M_{LE}=3.5).

Within Group. As Figure 14 demonstrates, subject matter experts are very consistent in their concern about content related issues. They are, however, less consistent in the amount of attention they pay to the other For example, Subject Matter Expert I emphasizes presentation attributes considerably more than the other subjects, while Subject Matter Expert II is notably more attentive to instructional design issues.

In the case of learner experts they show more variability in their judgement comments. As depicted in Figure 15, two subjects focus primarily on content issues, whereas, the other two subjects present a more even distribution of comments across categories.

Process Characteristics: Cognitive Process Analysis

This study has defined the cognitive processes the experts' engaged

Figure 13. Between Group Comparison: Percentage of Judgement Codes according to Product Attributes

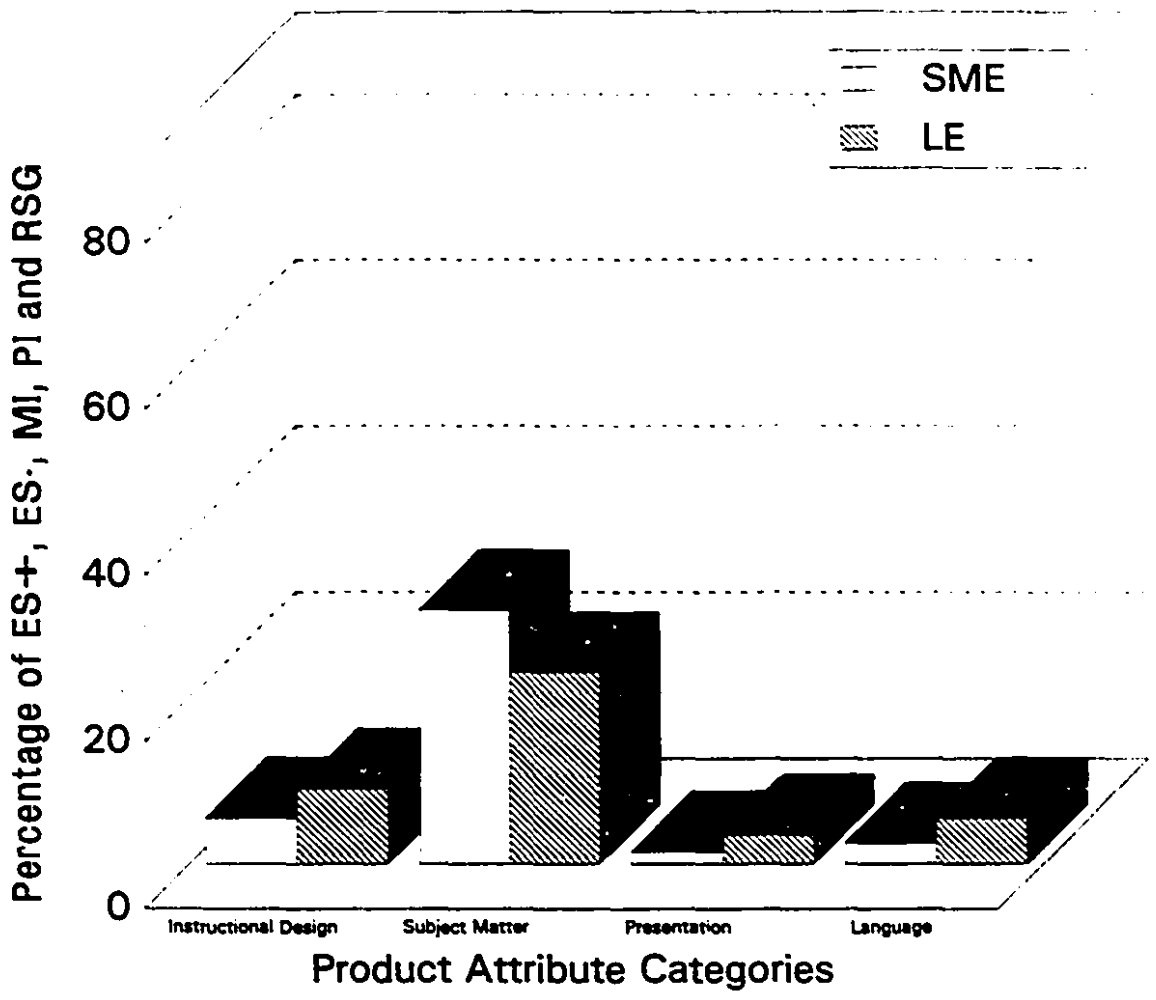


Figure 14. Within Group Comparision: Percentage of SME Judgement Codes according to Product Attributes

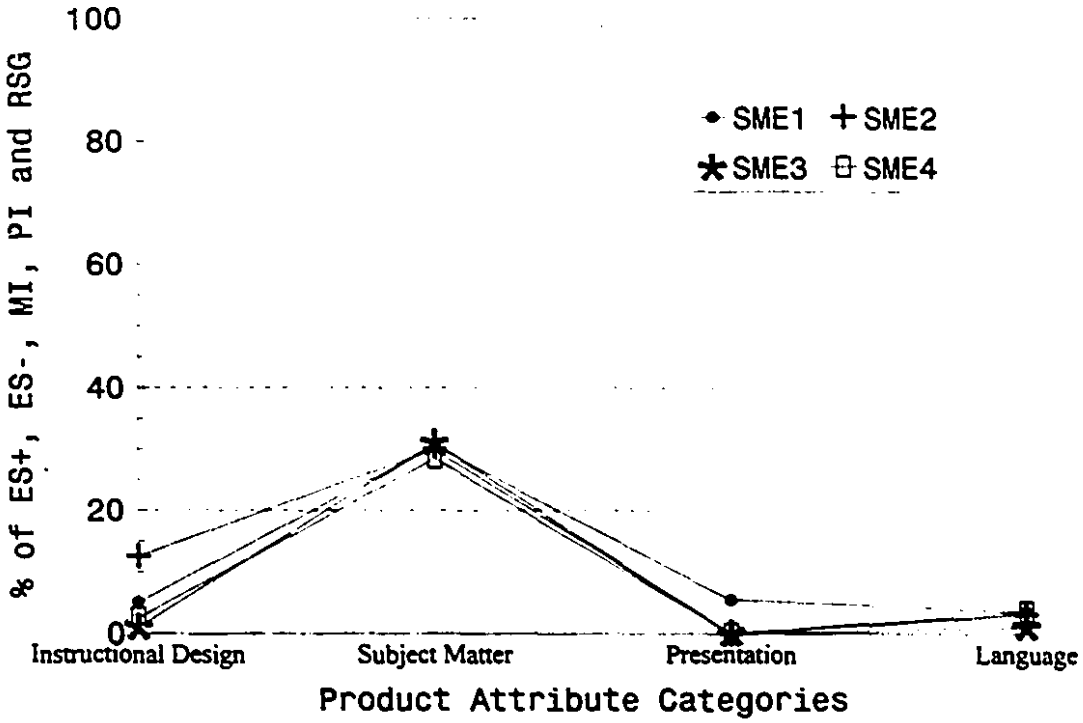
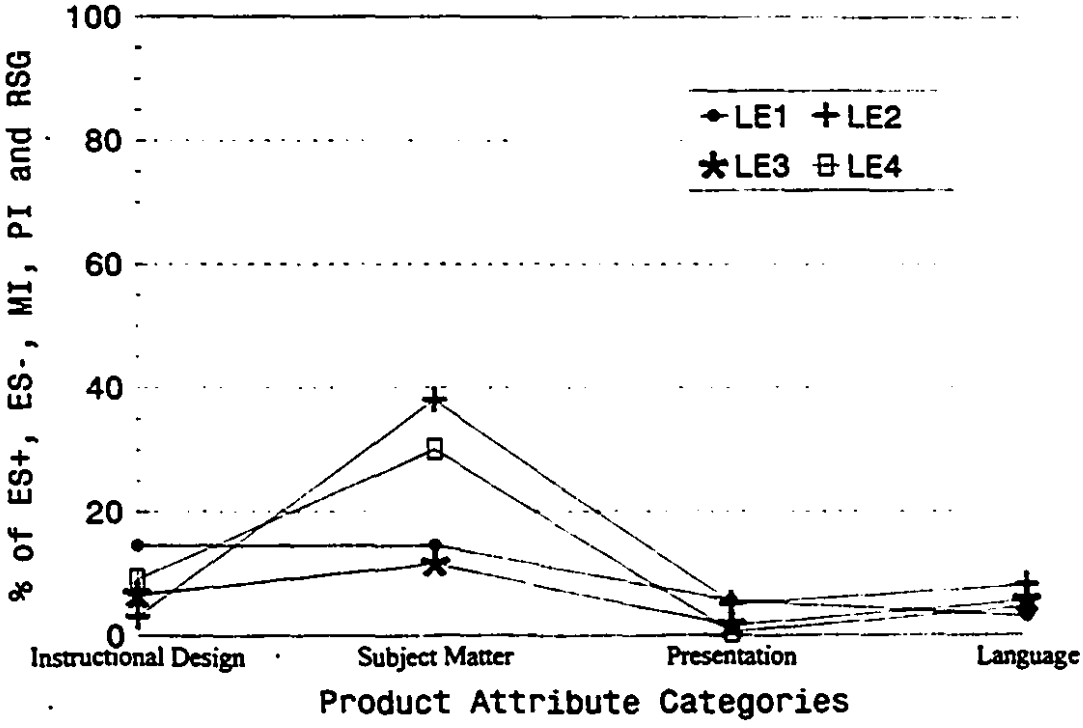


Figure 15. Within Group Comparision: Percentage of LE Judgement Codes according to Product Attributes



in during formative evaluation as 1) the plans they evoked or built and 2) the task representations they constructed. Experts' organizing plans are examined by classifying Goal Related codes according to three problem spaces. Whereas, an analysis of the reviewers' task representations are analyzed by sorting applicable Problem Solving codes according to the two phases of formative evaluation (see Figure 1).

Planning and Goal Analysis

Earlier results establish that learner experts are considerably more engaged in the process of planning than are subject matter experts, as they spend more time in goal related activities (see Figure 2). This finding supports the recommendation of Breuleux and Bracewell (1992) to analyze Goal Related codes in more detail. Using the Problem Space Strategies checklist (Appendix F), goal statements were sorted into three problem spaces: Text Comprehension, Review and Revision. After this initial coding was completed, a subsequent analysis was done of the various strategies associated with each category.

Percentages based on the total number of Goal Related codes and sorted into the three problem spaces are presented in Table 7.

Between Group. Figure 16 shows that when Goal Related codes are sorted according to problem spaces, both group of experts have a similar pattern of comments but the intensity is different. Both groups have a higher percentage of goal statements pertaining to the development of

Table 7

Percentage of Goal Related Codes (RSG,ST,TR) across Problem Spaces

I - TEXT COMPREHENSION PROBLEM SPACE:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
3.0	0	9.0	22.0	8.5	30.0	0	13.5	39.5	21.0

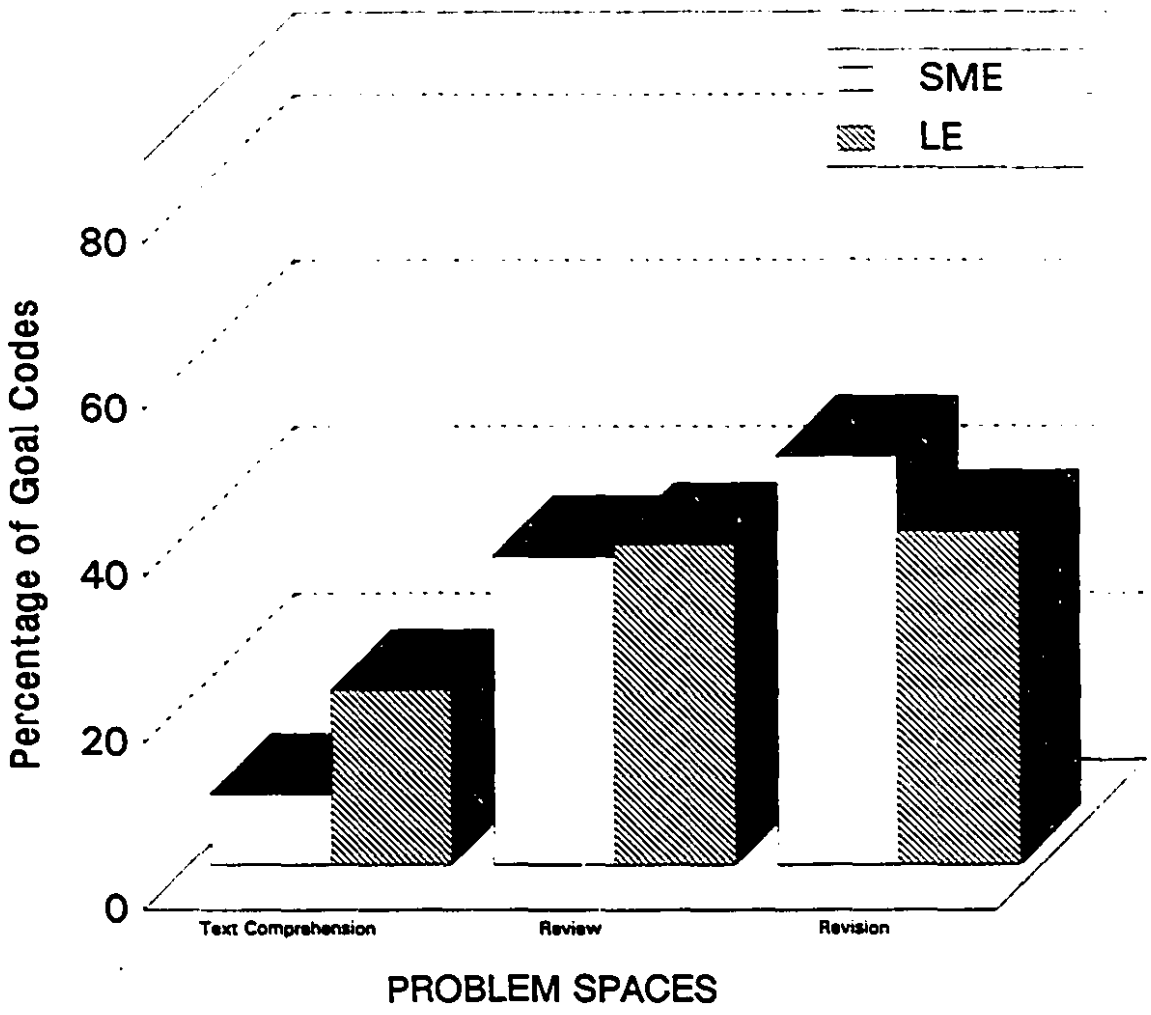
II - REVIEW PROBLEM SPACE:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
42.0	21.0	73.0	12.5	37.0	37.0	6.5	66.0	44.0	38.5

III - REVISION PROBLEM SPACE:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
55.0	67.0	9.0	65.0	49.0	29.0	93.5	20.0	16.5	40.0

Figure 16. Between Group Comparison: Percentage of Goal Related Codes across Problem Spaces



revision plans (M_{SME}=49.0; M_{LE}=40.0) and the lowest percentage relating to text comprehension (M_{SME}=8.5; M_{LE}=21.0). The percentage of review planning comments are approximately the same for both type of experts (M_{SME}=37.0; M_{LE}=38.5).

Although the configuration of results is similar, learner and subject matter experts differ in the attention they give to the various problem spaces. When compared to subject matter experts, learner experts make more than double the percentage of comments in the Text Comprehension Problem Space, but considerably fewer comments in the Revision Problem Space.

Within Group. When within group percentages are examined, the two expert groups demonstrate a contrasting pattern of comments. Figures 17 and 18 illustrate that most subject matter experts make more statements about their revision plans, while in contrast most learner experts are more verbal about their review plans. In each group, however, there is one subject with an aberrant pattern of comments. Contrary to the other content experts, Subject Matter Expert III utters a higher percentage of review planning statements. In reverse, Learner Expert II differs from the others by verbalizing a considerably higher percentage of revision planning comments. This discrepancy results in an increased variability of results, especially for learner experts.

Figure 17. Within Group Comparision: Percentage of SME Goal Related Codes across Problem Spaces

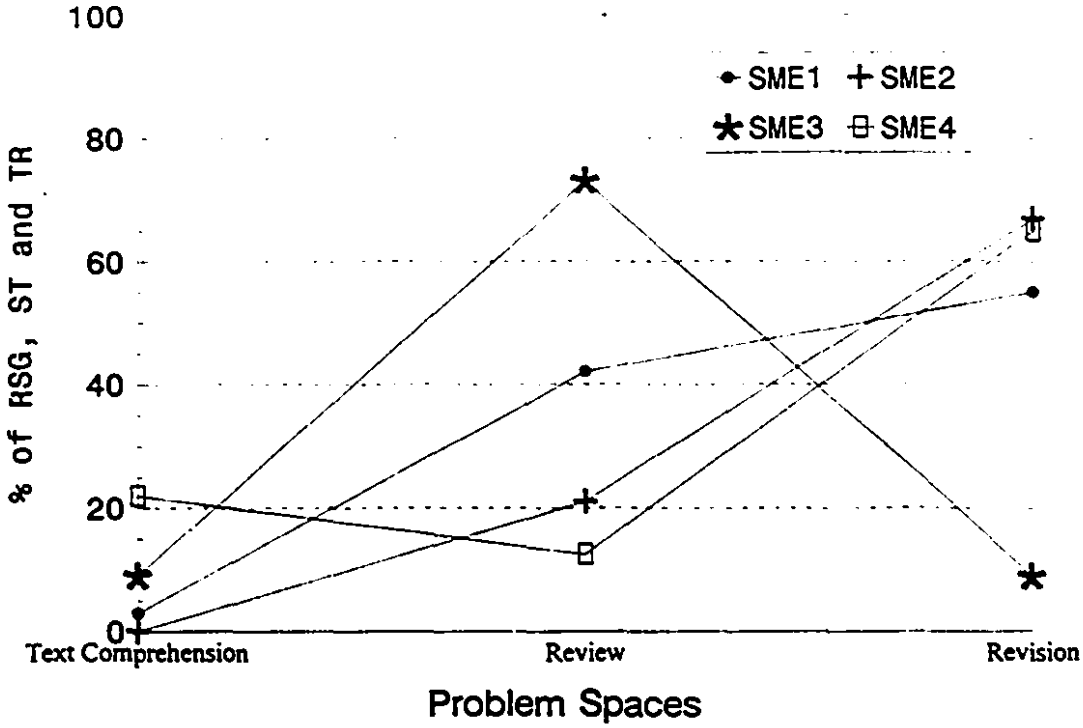
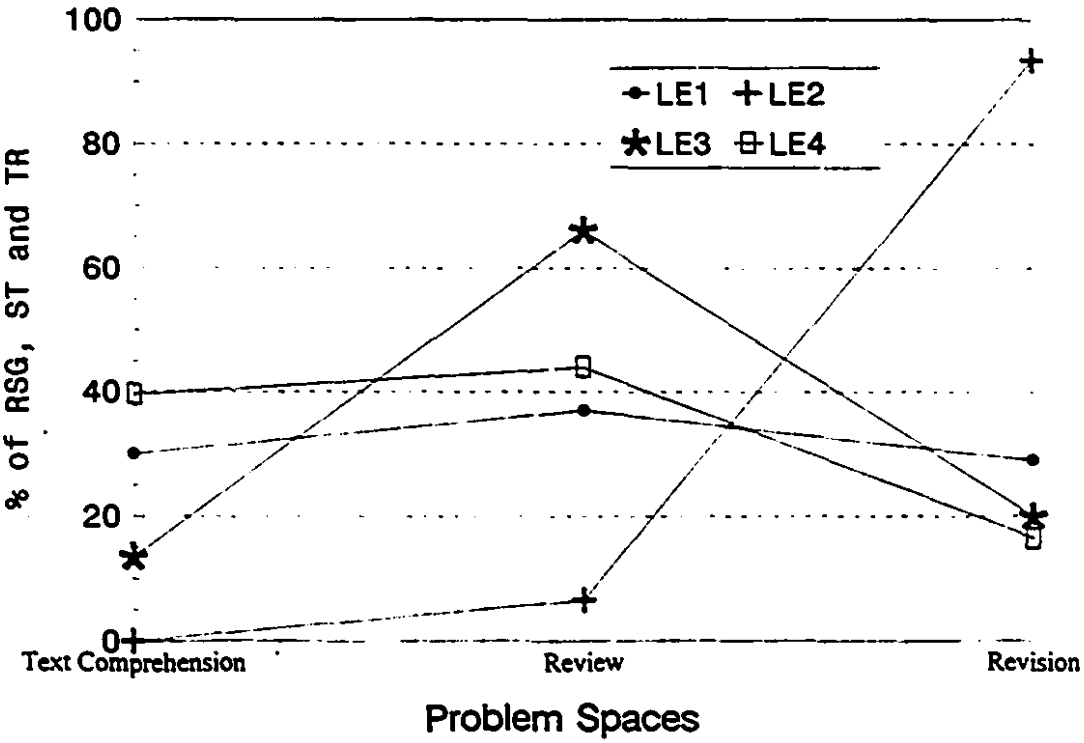


Figure 18. Within Group Comparision: Percentage of LE Goal Related Codes across Problem Spaces



Goal Related Codes: Text Comprehension Strategies. The percentages of Goal Related codes sorted into the various strategies that comprise the text comprehension problem space are presented in Table 8. This table reveals that the two expert groups are different in their verbalization of procedural knowledge pertaining to text comprehension. Overall subject matter experts articulate fewer strategies for understanding the materials, while learner experts utilize more diverse tactics. In particular, learner experts focus considerably more on monitoring their comprehension (M_{SME}=7.5; M_{LE}=1.0) by using their metacognitive knowledge; personal knowledge about their own cognition. Also, whencompared to subject matter experts, these learner experts evoke a higher percentage of previewing strategies (M_{SME}=1.5; M_{LE}=3.5) by scanning the text for key information.

When comparing only between group means, both groups appear to equally mention the need to define objectives (M_{SME}=4.0; M_{LE}=4.0). In fact, since the mean for content experts is based on the comments of only one subject; Subject Matter Expert IV, it is learner experts who use this strategy most frequently.

Within group analysis shows that individual subject matter experts do not present a steady pattern in their use of text comprehension strategies. In contrast, learner experts consistently mention the necessity of clarifying the objectives of the text during reading.

Table 8

Percentage of Goal Related Codes across Text Comprehension Problem Space

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
Define Objective	-	-	-	15.5	4.0	2.5	6.0	2.5	4.0	4.0
Scan Text	-	-	-	6.0	1.5	6.0	-	-	8.0	3.5
Build Background	-	-	-	-	0	4.0	-	-	3.0	2.0
Concentrate	-	-	-	-	0	2.0	-	4.0	3.0	2.0
Predict	-	-	-	-	0	2.0	-	-	2.0	1.0
Monitor Comprehension	3.0	-	-	-	1.0	9.5	-	7.0	14.0	7.5
Clarify	-	-	9.0	-	2.0	2.0	-	-	3.0	1.0

Goal Related Codes: Review Strategies. Table 9 presents the percentages of comments that can be categorized as possible reviewing tactics. The major similarity between both groups is that the strategy used most often is challenging the text by contemplating its quality and questioning the issues covered (M_{SME}=21.0; M_{LE}=9.0). Also, the two type of experts make approximately equal use of strategies such as thoroughly reading the text (M_{SME}=9.5;M_{LE}=7.5) and commenting to the administrators about particular aspects of the materials (M_{SME}=3.5;M_{LE}=3.5).

Both expert groups are dissimilar in the variety of procedures they proposed. Subject matter experts again restrict themselves to the number of strategies they use to review the text. In contrast, learner experts articulate several possible review tactics. In addition to employing the same strategies as the subject matter experts, these experts also make considerable use of approaches such as: searching for confusing information (M_{SME}=0; M_{LE}=6.5), relying on their prior knowledge to evaluate the text (M_{SME}=1; M_{LE}=4.5) and returning to the original sources of information to verify content (M_{SME}=1,M_{LE}=4.0).

An analysis of individual group members within each group show that subject matter experts were not coherent in the tactics they use to review the materials. Learner experts, however, do consistently select the strategies of thoroughly reading the text and commenting to the

Table 9

Percentage of Goal Related Codes across Review Strategies

	SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
	I	II	III	IV	Mean	I	II	III	IV	Mean
Skim Text	-	-	-	3.0	1.0	1.0	-	2.5	1.0	1.0
Read Thoroughly	6.0	8.0	27.5	-	9.5	6.0	1.5	15.0	7.5	7.5
Identify Confusing Information	-	-	-	-	0	8.0	-	9.5	8.5	6.5
Check Original Source	3.0	-	-	-	1.0	5.0	-	4.0	6.0	4.0
Check Definitions	-	-	-	-	0	-	-	4.0	-	1.0
Search for Ambiguous Words	-	-	-	-	0	-	-	6.0	-	1.5
Rely on Prior Knowledge	3.0	-	-	-	1.0	6.0	-	5.0	7.5	4.5
Challenge	29.0	8.5	45.5	-	21.0	10.5	-	13.5	12.5	9.0
Comment	-	4.0	-	9.5	3.5	1.0	5.0	6.0	1.0	3.5

administrators.

Goal Related Codes: Revision Strategies. For both expert groups most of their planning statements concern revision strategies to improve the educational materials (M_{SME}=49.0; M_{LE}=40.0). These comments were sorted according to the four principal strategy classifications of the Revision Problem Space; Instructional Design, Subject Matter, Presentation and Language, and their subcategories (see Table 10).

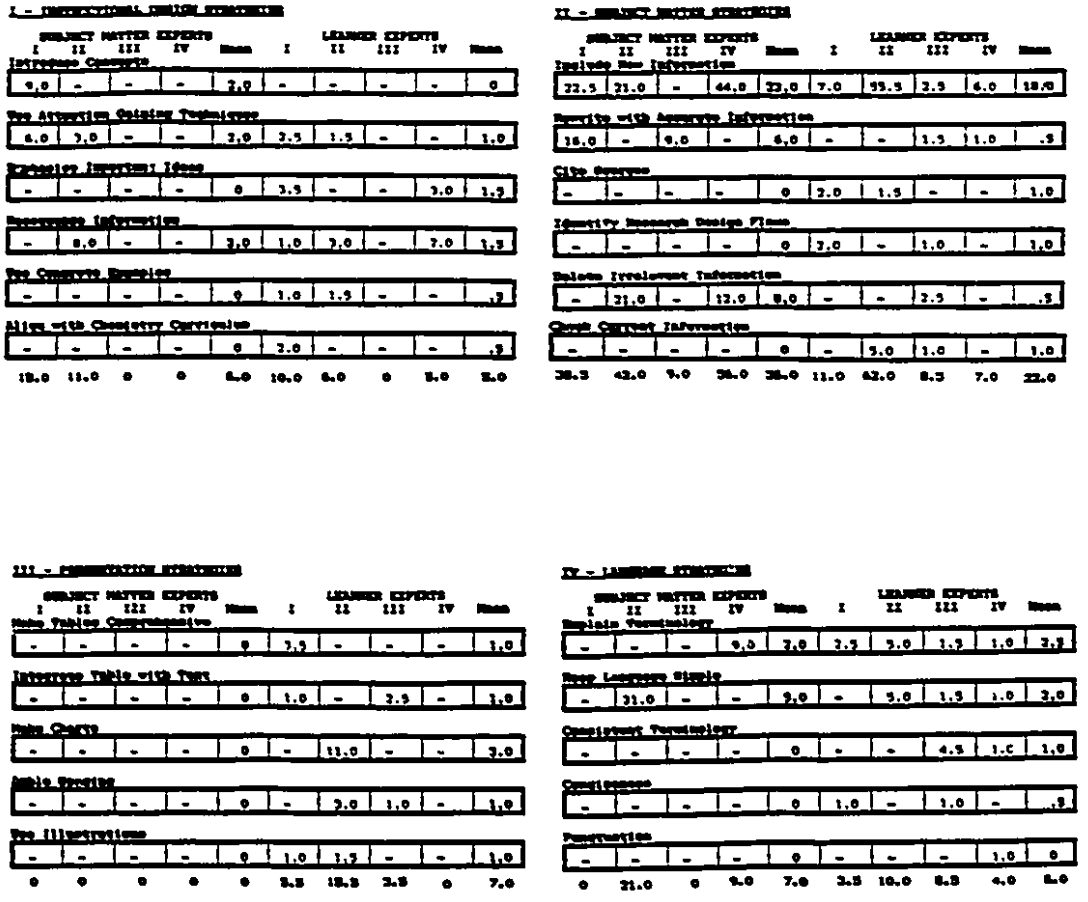
An analysis of the principal strategy classifications, show that the two groups mainly suggest the use of Subject Matter strategies (M_{SME}=39.0; M_{LE}=22.0) with considerably less emphasis on the other three categories. Both groups propose a similar number of Instructional Design and Language tactics, however, learner experts recommend more Presentation strategies than do subject matter experts (M_{SME}=0; M_{LE}=7.0).

A comparison of the subcategories within each main classification, show that both groups are similar in their dominant use of the Subject Matter strategy of adding new information (M_{SME}=22.0; M_{LE}=18.0). Both groups also make approximately the same percentage of comments concerning the instructional design strategy of resequencing information (M_{SME}=2.0;M_{LE}=1.5) and the language strategy of explaining terminology (M_{SME}=2.0;M_{LE}=2.5).

A further comparison between the two groups is the subject matter experts' bias in favour of two other Subject Matter strategies; rewriting the

Table 10

Percentage of Goal Related Codes across Revision Strategies



text to increase the accuracy of the content (M_{SME}=6.0; M_{LE}=.5) and deleting irrelevant information (M_{SME}=8.0; M_{LE}=.5). They also emphasize the language strategy of maintaining an appropriate reading level (M_{SME}=5.0;M_{LE}=2.0). On the other hand, learner experts suggest more frequently the need for the presentation strategy of using charts and diagrams (M_{SME}=0; M_{LE}=3.0).

Again subject matter experts do not display a coherent pattern of revision strategies. Learner experts are somewhat more consistent as they all suggest revision techniques such as the need to incorporate new information and explain terminology.

Task Representation Analysis

As mentioned previously in the Methodology section, individual codes of the Problem Solving Coding Scheme were assigned to the two Phases of Formative Evaluation. In Table 11, percentages were calculated by excluding from the total number of segments those codes that were considered to be redundant or outside of this model; Dialogue, Problem Location, Problem Reiteration and Task Talk. All other segments were categorized according to either the data collection phase or the revision phase of formative evaluation.

Between Group. As shown in Figure 19, there is a strong similarity between the two group of experts in their definition of the task assigned to them. Both groups are predominately involved with the first data

Table 11

Percentage of Segments according to the Phases of Formative Evaluation

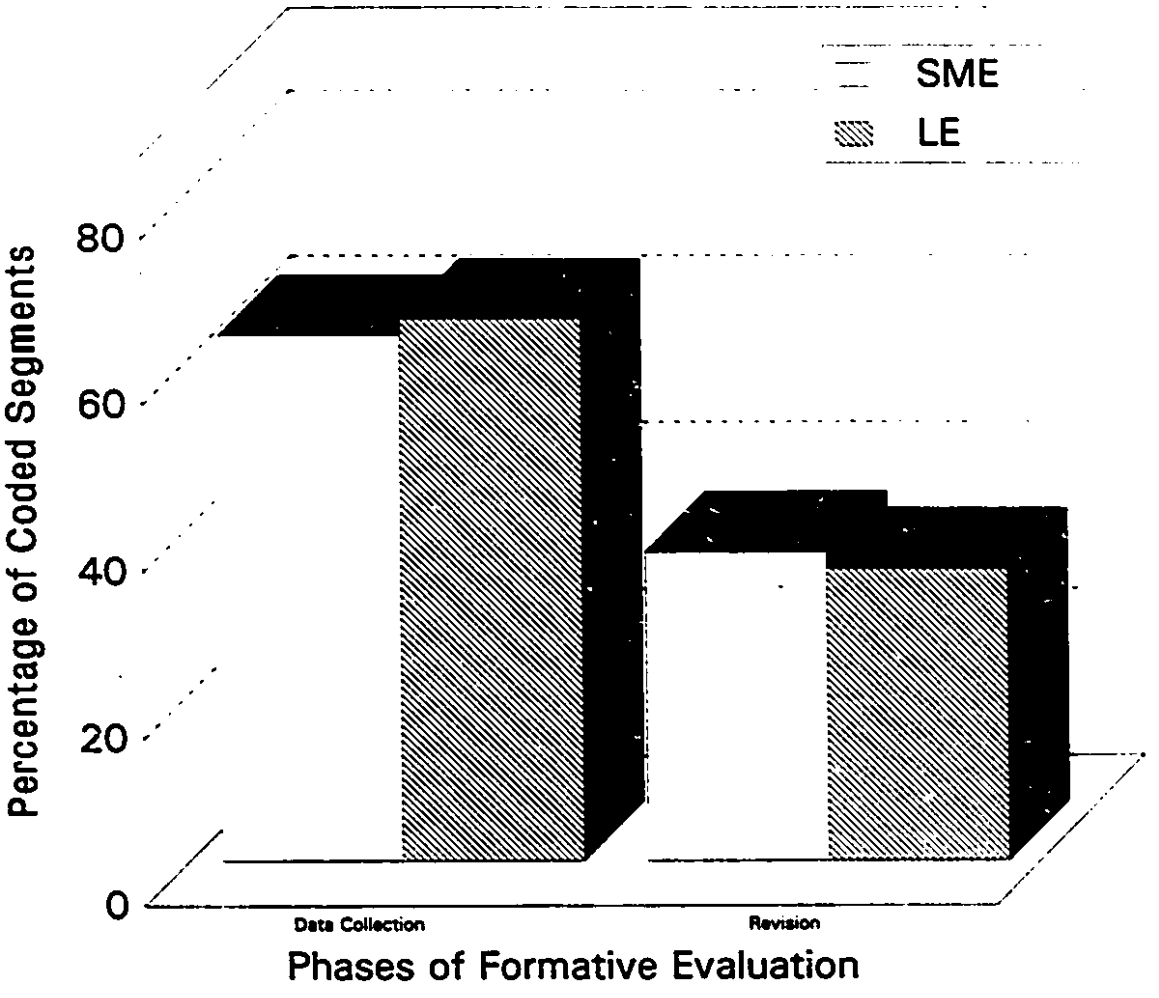
DATA COLLECTION PHASE:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
66.5	68.0	66.0	53.5	63.0	71.5	31.5	78.0	81.0	65.0

REVISION PHASE:

SUBJECT MATTER EXPERTS					LEARNER EXPERTS				
I	II	III	IV	Mean	I	II	III	IV	Mean
33.5	32.0	34.0	46.5	37.0	28.5	69.0	22.0	19.0	35.0

Figure 19. Between Group Comparision: Percentage of Segments according to Phases of Formative Evaluation



collection phase (M_{SME}=63.0; M_{LE}=65.0) and define their task as giving opinions about the text and identifying problems. Only a third of their remarks relate to the Revision Phase (M_{SME}=37.0; M_{LE}=35.0). Therefore, the experts are more concerned about detection and less concerned with diagnosing the type and symptom of the problems identified or making specific changes to eradicate these flaws.

Within Group. Subject matter experts have a relatively coherent pattern of comments, as is demonstrated in Figure 20. All subjects emphasize data collection; Phase 1, rather than revision; Phase 2. For learner experts, the within group pattern is less consistent (see Figure 21). Most of this variation is owing to the atypical comments of one subject, Learner Expert II, who has a considerably higher percentage of segments in the Phase 2 rather than the Phase 1.

Analysis of Debriefing Questionnaire

Following data collection, all subjects were asked to complete a debriefing questionnaire that consisted of nine questions in total. Two questions were designed to examine the education and job experiences of each subject and three were devised to gather information on the subjects' definition of the task and the plans they constructed.

Background Information

A summary of the responses pertaining to training and relevant job experiences are presented in Table 12.

Figure 20. Within Group Comparison: Percentage of SME Segments according to Phases of Formative Evaluation

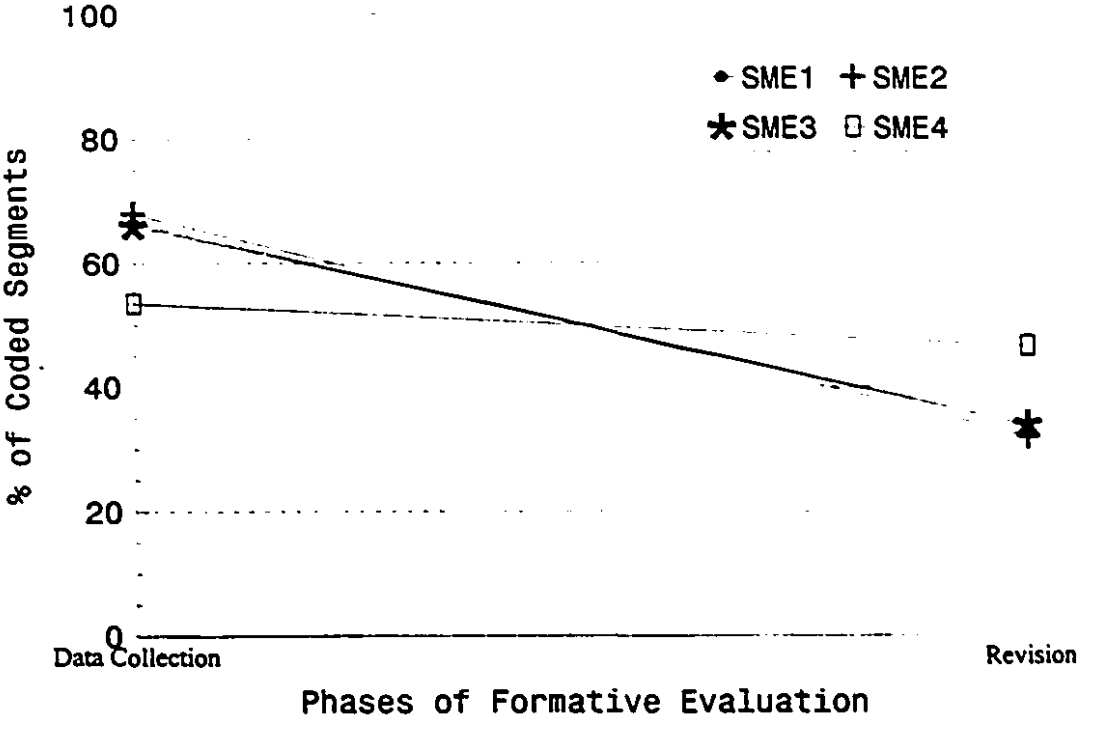


Figure 21. Within Group Comparision: Percentage of LE Segments according to Phases of Formative Evaluation

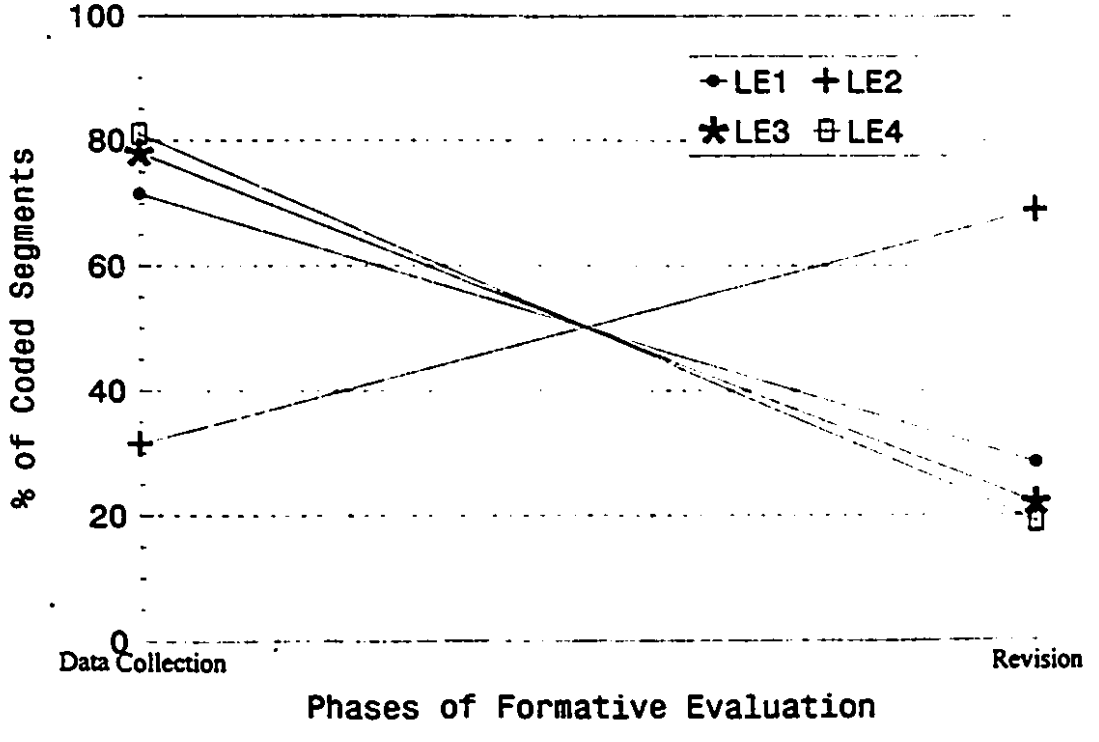


Table 12

Background Information on Expert Reviewers

	SME I	SME II	SME III	SME IV
Domain Expertise	Dietetics/ Pedagogy	Dietetics	not an expert	Dietetics
Degree	MSC Dietetics	MSc Nutrition	BSc Nutrition	BSc Food Science
Prior Text Evaluation	YES	NO	NO	NO
Years Experience in Field	11 Dietician 8 Teacher	8 Dietician	13 Dietician	5 Dietician
Familiarity with Content	somewhat	somewhat	somewhat	moderate
Familiarity with Audience	somewhat	somewhat	completely unfamiliar	moderate

	LE I	LE II	LE III	LE IV
Domain Expertise	Chemistry/ Pedagogy	Environmental Issues/ Chemistry	Evaluation	Psychology Empirical Research
Degree	Phd Organic Chemistry	MSc MEd	MA Counselling	Phd Clinical Psychology
Prior Text Evaluation	YES	YES	YES	YES
Prior Experience in Field	16 Teacher	20 Teacher 15 Environ.	20 Counselling	30 Teacher
Familiarity with Content	somewhat	very familiar	somewhat unfamiliar	moderate
Familiarity with Audience	somewhat	very familiar	somewhat unfamiliar	very familiar

When subject matter experts were asked to define their expertise, most of them acknowledged that their domain of expertise was dietetics, however, one subject claimed to be a generalist rather than an expert. Subject Matter Expert I also claimed to be an expert in pedagogy. Next the experts were asked to rate their familiarity with the content and the audience. All subject matter experts said they were at least somewhat familiar with the content, however, they added that cancer and diet was not their area of specialization. On average subject matter experts felt they were somewhat familiar with the audience, although their definition of the audience was the general public, which was not the population targeted by this research. In addition, the experts were questioned about any prior experiences they had evaluating materials. Only one subject, Subject Matter Expert I, had previous experience evaluating textbooks.

The second group of experts had very different backgrounds in comparison to the subject matter experts. First, learner experts offered a wider range of responses to the question of domain expertise that included empirical research, pedagogy, and content expertise; either psychology or science. Although all subjects had experience as teachers, only one expert mentioned pedagogy as an area of proficiency.

Concerning the content of the materials, in general learner experts perceived themselves as informed laymen who were somewhat familiar with the topic and therefore qualified to evaluate certain aspects. As

Learner Expert IV expressed, " In strict terms I'm not an expert on the subject that was presented, but if the subject is scientific presentation and ideas, then I am a content expert". Moreover, Learner Expert II's rating of familiarity with the content was higher than any of the subject matter experts in this study. For the past fifteen years the subject had been teaching environmental issues. In addition, learner experts typically rated themselves as being better acquainted with the intended audience; undergraduate arts and science students.

In response to the question of previous experience reviewing instructional materials, all subjects stated that as teachers they were occasionally required to evaluate textbooks.

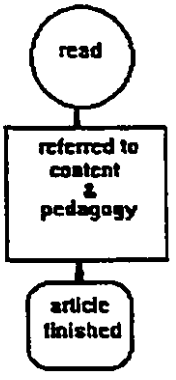
Definition and Planning of the Review Task

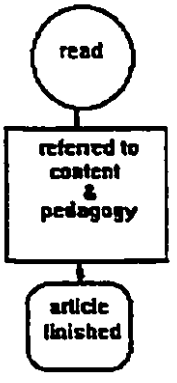
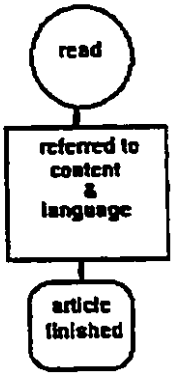
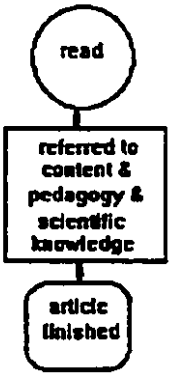
The various experts were requested to summarize the task they had just completed and the methods they chose to accomplish this task. Table 13 summarizes this information and shows that all experts, from both groups, described the task as evaluating or commenting on the materials. More specifically they mentioned that they read the text, analyzed it and made comments.

The two group of experts, however, were not as cohesive in their explanation of how they accomplished this task. The majority of subject matter experts asserted that they "just responded naturally" and did not make a plan other than to read the text. Subject Matter Expert I is the

Table 13

Task Descriptions and Planning of Experts Reviewers

	SME I	SME II	SME III	SME IV
Task Description	reviewed content	reaction to content	comment on materials	comment on materials
Plan		no plan	no plan	no plan

	LE I	LE II	LE III	LE IV
Task Description	evaluated materials	evaluated materials	evaluate & comment	evaluate & comment
Plan			no plan	

only subject in this group that articulated a fairly elaborate plan for evaluation. This subject stated that while reading the text, they referred to both their knowledge of pedagogy and the subject matter. Their decision to terminate the task coincided with the completion of the article.

On the other hand, Learner Expert III was the only audience expert who claimed to be without a plan. Most learner experts verbalized a more detailed plan than subject matter experts, which appeared to reflect the learner experts' expanded domains of expertise. In the plan constructed by Learner Expert I, two domains of knowledge were referred to while reading the text; pedagogical and content knowledge. A second subject, Learner Expert II relied on content and literate expertise. While Learner Expert IV related the text to three knowledge bases; scientific inquiry, content and pedagogical knowledge. As with all the other subjects, the plans began with reading the text and ended with the completion of the article.

CHAPTER FOUR

Discussion and Conclusion

The formative evaluation procedure is seen as being divided into two distinct stages; data collection and revision. The focus of this research was the data collection phase. Comparisons were made between both the output (product analysis) and organizing plans (process analysis) of two types of expert reviewers commonly asked to provide feedback during formative evaluation. The think aloud data generated by subject matter and learner experts were analyzed in terms of the similarities and differences in the type and amount of feedback generated.

The results of this study revealed that the products of these experts were similar, but the processes they engaged in were more distinctive. The focal point for both groups of experts were content issues, however, each group took a somewhat different approach in accomplishing the evaluation task. In particular, subject matter experts made more references to their knowledge, while learner experts focused on planning activities. A possible explanation for this discrepancy is that subject matter experts had more knowledge of the content available to them. Like domain experts found in the cognitive literature on expertise, subject matter experts might have used this well-organized body of knowledge to aid them in their review of the text. Conversely, learner experts who in general were less familiar with the subject matter, elaborated more on their plans and potential

solution strategies. This extended goal-related activity might have compensated for any deficiencies in content knowledge that these experts were experiencing.

Overall, it appears that both group of experts provided similar data, drew upon the same types of knowledge and represented the task in the same manner. This refutes the recommendations in the literature to use both subject matter and learner experts during formative evaluation (Tessmer, 1993; Dick & Carey, 1990; Flagg, 1990). The remainder of this chapter will discuss these findings further and present some recommendations for future research.

Similarities and Differences in Feedback

General Characteristics

An examination of the think aloud data according to the Problem Solving codes, found that the information provided by the two groups of subjects had a highly similar pattern. For both groups approximately one quarter of all verbalizations centred on the identification of problems. Few revision suggestions were made. It is possible that these findings reflect the subjects' interpretation of their role in formative evaluation. In this study, task instructions did not offer a definition of the review task, nor did it specify procedures for conducting such a task. This was done intentionally, so that the manner in which expert reviewers interpreted their task could be studied. Apparently these experts perceived their role

to be evaluators of the text and not revisors, which helps us to see and describe how a reviewer in data collection operates.

In addition, both group of experts made a high percentage of references to their knowledge base. Again this finding could be related to the subjects' task definition, since as evaluators the experts would use their declarative knowledge to provide insight into the type and nature of the problems found.

Product Characteristics

One of the principal objectives of formative evaluation is to collect data that will aid in revision. To be effective, this information should focus on diverse product attributes and provide a maximum number of identified problems. Furthermore, it would be helpful if this data included elaborations of the problem's symptoms and possible revisions. Both Knowledge Statements and those Problem Solving codes that reflect judgements about the instructional materials represent this information.

This study revealed that for both group of experts over 75% of Knowledge Statements were related to content issues. It is perplexing that the learner experts, who were either university or CEGEP teachers, refrained from referring to their pedagogical knowledge. Although during the debriefing interview several experts claimed to rely on this knowledge, in fact less than 15% of Knowledge Statements involved pedagogy. Perhaps these experts could not access their own knowledge of pedagogy,

as this information is believed to be specifically linked to content knowledge (Leinhardt & Greeno, 1986). Since the learner experts in this study, who were selected due to their knowledge and experience with the audience, taught chemistry and psychology their pedagogical knowledge would be associated with these two fields. As most learner experts had never been called upon to teach nutrition and environmental science; the subject matter of the text under review, they could not comment on its pedagogy.

A further similarity between the two types of reviewers was found in the judgements they made on the materials. Once the data was sorted according to product attributes, it became apparent that both groups were inclined to mainly draw conclusions about the subject matter of the instructional materials. Close to 25% of all judgements fell into the subject matter category, which was considerably higher than judgements made in the other attributes categories. This finding corroborates Saroyan's (1989) initial conclusions that experts from diverse backgrounds report problems related to content, even when subject matter is not their area of specialization.

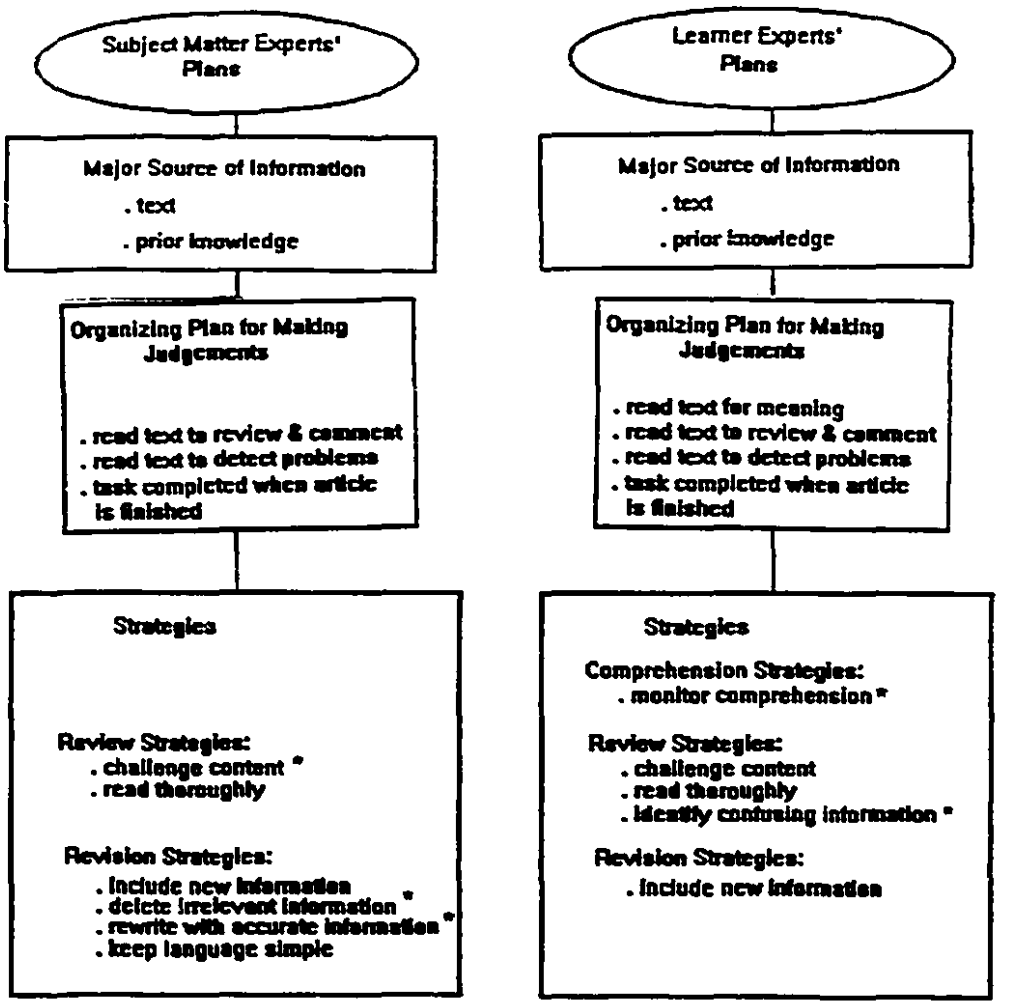
Process Characteristics

Another aim of this study is to examine the cognitive processes that the expert groups engaged in during formative evaluation. One aspect of this process is to assess the goals set by the reviewers. This was done by

analyzing the plans the experts built or evoked to accomplish the task of evaluating the text. An examination of Goal Related codes found that the two group of subjects appeared to be somewhat similar in their planning. For both groups, nearly half of all goal related comments fell within the Revision Problem Space. Nonetheless, two dissimilarities were apparent between the two groups. First, learner experts operated more within the Text Comprehension Problem Space and second, subject matter experts made substantially more comments within the Revision Problem Space.

A more detailed examination of the planning subcategories revealed that these differences were due to the learner experts' greater tendency to monitor their comprehension while reading, and the subject matter experts' inclination to engage in more types of revision planning. Figure 22 presents an overview of the plans developed by each expert group and illustrates this fundamental distinction. Knowledge of the content may be the critical factor involved in this discrepancy. Since learner experts in general were less familiar with the topic of diet and cancer, they presumably approached the reading task in a manner typical of learners. As they could not depend on their own declarative knowledge to interpret the content, they had to continually monitor their comprehension while attempting to determine the text's principal message and major points. Ostensibly, it was only after the learner experts understood the text that they were able to turn to the review task. In contrast, subject matter

Figure 22. Overview of Expert Reviewers' Task Representations



Note. All segments representing less than 5% of Goal Related Codes omitted. * = notable between group difference of 5% or more.

experts' greater content knowledge evidently enabled them to quickly move on to reviewing the instructional materials. Moreover, it allowed them to consider a greater variety of revision strategies, such as rewriting the text with more accurate facts and deleting irrelevant details.

A final feature of the experts' cognitive processes to be considered here is how the subjects' represent the task assigned to them. The results of this study show that both expert groups similarly interpreted their task as simply detecting problems and provided a low level of revision information, with more than 60% of their comments falling within the Data Collection Phase of the formative evaluation model. Since designers have been encouraged to obtain a higher level of information from reviewers; including identifying the causes of problems and suggesting solutions (Thiagarajan, 1978), this finding provides some insight into why expert data has sometimes been seen as inconclusive and incomplete. Possibly, the developer of the materials and the reviewer having different expectations for the task, resulting in this impasse.

This difference may be due to the organizing plan the experts chose when structuring their task. Flower (1989) describes several types of structures for reading and writing that require different levels of task representation. As Figure 24 demonstrates, both group of experts acted more like novice writers by evoking the less demanding organizing plan of reviewing and commenting. With this plan the subjects carry on a dialogue

with the text, in which they alternated between reviewing, summarizing and adding their own comments, criticisms or associations. The fact that these reviewers represented their task more like inexperienced writers than expert writers is understandable when the backgrounds of the subjects are considered. These subjects were considered experts in the domains of dietetics or teaching, not revision and writing.

Other possible reasons for the experts emphasis on the identification of problems and their limited involvement with revision are a lack of revision procedural knowledge and the cognitive costs involved. Fitzgerald (1987) noted that one obstacle to revision is a deficiency in the knowledge of how to rectify problems observed. Reviewers either might not have acquired this knowledge, or have the requisite knowledge, but had difficulty recalling and/or representing it. Also, the experts might have focused on review because the task description gave them the option to do the easier task of commenting and reviewing, instead of the harder task of revision. Other studies have shown that even experts designated as revisors proposed very few text changes (Flower et al., 1985; Nevo, 1985). Since diagnosing problems and making revisions is costly in terms of attention and requisite knowledge, the strategy of choice is often the simpler procedure of detection.

Within Group Characteristics

Resemblance among individuals group members was observable in

both the plans they elicited and the judgements they made. Subject matter experts were a highly cohesive group, with all members uttering comments that were mostly content related and defining their task as detecting problems. Also, the majority of individual members sought out a higher percentage of revision strategies.

In general, learner experts resembled one another as well. As a rule, the majority of members commented primarily on content related issues, probed their procedural knowledge for review and revision strategies and defined their task as problem identification. It is one subject in particular, Learner Expert II, who demonstrated a unique profile and increased the range of variability within the learner expert group. This noteworthy subject will be discussed in more detail later.

When Goal Related statements were analyzed in more depth, disparities among individual members were more evident for both expert groups. Only learner experts showed a slight group likeness in the strategies they employed.

It is common for experts to not fit a single "expert" profile when a more specific level of analysis is done (Duy, 1990; Saroyan, 1989; Flower et al, 1985; Leinhardt & Smith, 1985). One possible explanation for this variability is that even within specific domains, distinctive backgrounds can lead to diversity of expertise or proficiency (Glaser, 1991; Duy, 1990; Saroyan, 1989; Lawrence, 1988). Furthermore, ill-structured problems like

formative evaluation, demand sifting through a large amount of problem related details and will normally evoke a conflicting set of responses (Voss, Greene, Post and Penner, 1983).

As mentioned earlier, Learner Expert II presented a unique profile. In contrast to all the other experts, this reviewer defined the task as diagnosing the symptoms of problems and proposing solutions, with the majority of comments falling within the Revision phase of formative evaluation. This subject also provided more data to guide the revision process, which included a higher percentage of judgement comments and revision suggestions. Although the primary focus was on subject matter attributes, a comparatively higher percentage of language and presentation comments were made as well. This reviewer's greatest concern was to ensure that concepts were clarified, so as to facilitate the learners comprehension of the materials (see Table 9).

The high level of information provided by this subject is probably due to background experience as a CEGEP teacher of this particular content. For the last fifteen years this expert had been teaching a course on the environment and acknowledged being very familiar with the subject matter and the intended audience. Possibly this experience fostered the confidence and knowledge necessary to locate problems and suggest proposed changes to the material. These results support the findings of Borko and Livingston (1989), Berliner (1986) and Shulman (1986). They

all found that expert teachers who had well-developed and interconnected knowledge structures for subject matter, pedagogical content and learner characteristics were better predictors of where in a course students were likely to have problems. These teachers could foresee misconceptions the students might have and areas of learning these misconceptions were likely to affect. Nonetheless, this hypothesis is based on only one subject and more research is necessary for validation of these results.

Implications for Practitioners

The literature on formative evaluation suggests that a team consisting of different types of experts be used to review instructional materials under development. One specific recommendation has been to use a combination of subject matter and learner experts. The current study refutes this advice. Since both subject matter and learner experts generally comment on similar aspects of the text, their joint data does not yield a wider variety of information to guide the revision process. Furthermore, this study did not validate that learner experts, as reviewers, are a distinctive group. Although this group was selected due to their proficiency as teachers, they did not focus more on pedagogical knowledge or learner issues in the product attributes. Therefore, this research cannot justify the use of both expert groups during formative evaluation.

The differences in planning between the two expert groups indicates that reviewers who are unfamiliar with the content of the materials being

evaluated are at a disadvantage. Since these reviewers must continuously monitor their comprehension in an attempt to isolate the main idea and major points of the text, they should be allotted extra reading time. In addition, these experts are likely to compensate for the increased amount of text comprehension planning required by reducing their planning of possible revisions.

The finding that reviewers tend to remain in the data collection phase of formative evaluation implies that when experts are left to define the review task, they choose the simpler procedure of detection. Even if the reviewers had wished to go beyond the present task and offer revision suggestions, for some of them their lack of formative evaluation experience could have been a hinderance.

If a high level of information is required certain guidelines could eliminate these obstacles. First, where possible select reviewers who are familiar with the subject matter of the instructional materials. In addition, use clear and comprehensive task instructions that will turn formative evaluation into a more well-structured task, where the desired outcome would be mutually understood by both developer and expert. Finally, a reviewer whose background is similar to Learner Expert II might be the better choice. An individual who teaches the subject matter to the intended audience appears to be an excellent combination of both types of experts. The ability of this expert to rely on pedagogical content

knowledge could make for good predictions of the learners success and allow foresight of areas where misconceptions might arise. At the same time, the capacity to assess accuracy of content and recommend recent material is enhanced. This hypothesis, although it is beyond the scope of the present study could be an interesting question for future research.

Limitations of Study

Despite the benefits of using the think-aloud procedure, it did impose an artificial task environment that might have constrained the subjects and altered their performance. This was borne out by the concerns of one subject who was distressed by the fact that she would not normally evaluate materials in this manner.

A second limitation of this procedure is that it can be intimidating and exhausting, even for highly verbal subjects (McAlpine, 1987). In fact, several experts mentioned that reading out loud and commenting simultaneously was a difficult task for them. Some subjects also questioned whether they had verbalized all their thoughts. This disadvantage may limit the amount of data obtained from experts, however, using other sources of data to triangulate results will reduce this liability. In particular, a well-structured retrospective interview is crucial.

One final limitation is that due to the length of time required for each session it was necessary to keep the sample size small. As a result the data collected may not be completely representative of each expert

group.

Recommendations for Future Research

This study highlighted the defining characteristics of two kinds of expert reviewers; subject matter experts and learner experts. It would be interesting to extend this research to other types of reviewers such as instructional designers, language experts or cultural experts.

Also, it would be useful to look at the medium of instruction used as stimulus materials. For this investigation written materials were tested, however, other sources of instruction such as film, video or computer programs could be explored.

Since Learner Expert II generated such interesting results, future research using reviewers who teach the subject matter to the intended audience seems warranted. This study might focus on not only the characteristics of the type of data generated, but the quality of the product produced in terms of its value to revisors.

In relationship to the primary coding scheme, future studies might include codes that represent various classifications of knowledge. For example, different categorizations could be included such as: personal, declarative, procedural, diagnostic, metacognitive; personal knowledge about one's own cognition and conditional knowledge; when and where procedures are applicable.

Finally, future researchers might wish to design aids for reviewers

that are based on the knowledge schemas of the various experts used in formative evaluation. These checklists should be comprised of the types of problems detected, symptoms that distinguish these problems and potential solutions.

Concluding Remarks

This study analyzed the key features of formative evaluation feedback generated by subject matter and learner experts. Results indicate that the two expert groups produce similar data. In addition, both groups refer to the same knowledge domains and evoke or construct similar plans. These findings refute the literature advocating that during formative evaluation feedback data or expert review, be collected from both subject matter and learner experts.

The results also strengthen the notion that formative evaluation is an ill-defined problem and therefore, the experts and the designer of the materials might interpret the role of the reviewer differently. In the present study, the experts identify their task as detecting problems, rather than proposing changes to the materials. On the other hand, the designer might expect the reviewer to go beyond simply identifying problems by elaborating more on the nature of the problem, and occasionally providing solutions. Several guidelines are offered to practitioners who might require a higher level of information from the reviewers. In particular, the designer is advised to provide specific task instructions to the experts when

they are conducting formative evaluation and to select a reviewer who teaches the subject matter to the intended audience.

REFERENCES

Andrews, D.H., and Goodson, L.A. (1980). A Comparative Analysis of Journal of Instructional Development, 3(4), 2-16.

Anzai, Y., and Simon, H.A. (1979). The theory of learning by doing. Psychological Review, 86, 124-140.

Arnold, D. (1989, November). A discipline sensitive model of textbook selection criteria in the higher education faculty. Paper presented at the meeting of the Association for the Study of Higher Education, Atlanta, GA.

Baghdadi, A. (1980). A comparison between two formative evaluation methods. Dissertation Abstracts International, 41, 3387-A.

Baker, E.L. and Alkin, M.C. (1973). Formative evaluation of instructional development. AV Communication Review, 21, 389-418.Berliner, D.C. (1986, August). In Pursuit of the Expert Pedagogue. Educational Researcher, 5-13.

Borko, H., and Livingston, C. (1989). Cognition and Improvisation: Differences in Mathematics Instruction by Expert and Novice Teachers. American Educational Research Journal, 26(4), 473-498.

Bordonaro, T. (1993). A Comparision of the Effectiveness, Cost and Efficiency of Four Formative Evaluation Conditions. Unpublished master's thesis, McGill University, Montreal.

Bowler, M. (1978). The making of a textbook. Learning, 6, 38-42.

Bracewell, R.J., and Breuleux, A. (1992). Substance and Romance in analyzing think-aloud protocols. Technical Report 92.2, Laboratory of Applied Cognitive Science, McGill University, Montreal.

Breuleux, A. (1991). The analysis of writers' think-aloud protocols: Developing a principled coding scheme for ill-structured tasks. In G. Denhiere and J.P. Rossi (Eds.), Text and Text Processing, (pp. 333-362). North-Holland: Elsevier Science Publishers.

Broyton, R. (1989). Yes, but... English Journal, 78(6), 14-21.

Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman, J. Segal and R. Glaser (Eds.), Thinking and Learning Skills: Current research and open questions. (Vol.2, pp. 485-517). Hillsdale, NJ:Lawrence Erlbaum Associates.

Chase, W.G. , and Ericsson, K.A. (1981). Skilled memory. In J.R. Anderson (Ed.), Cognitive Skills and Their Acquisition, (pp. 141-189). Hillsdale, NJ: Lawrence Erlbaum Associates.Chase, W.G., and Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Chi, M.T.H., and Ceci, S.J. (1987). Content knowledge: Its role, representation and restructuring in memory development. Advances in Child Development and Behavior, 20, 91-142.

Cronbach, L.J. (1963). Course improvement through evaluation. Teachers College Record, 64, 672-683.

Dick, W. (1980). Formative evaluation in instructional development. Journal of Instructional Development, 3(3), 3-6.

Dick, W. (1987). Instructional Design and the Curriculum Development Process. Educational Leadership, 44(4), 54-56.

Dick, W. and Carey, L. (1990). The Systematic Design of Instruction. (3rd ed.) Glensview, Il.: Scott, Foresman Company.

Duy, N. (1990). Expert revisors use of student data. Unpublished master's thesis, McGill University, Montreal.

Ericsson, K.A., and Simon, H.A. (1980). Verbal reports as data. Psychological Review, 87(3), 215-251.

Ericsson, K.A., and Simon, H.A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Fenster, A.E., Harpp, D.N., and Schwarz, J.A. (1990). The diet cancer relationship. In A.E. Fenster, D.N. Harpp & J.A. Schwarz (Eds.) The world of chemistry. Part 1 (pp. 50-55). (Available from McGill University Bookstore, Montreal, Quebec)

Fitzgerald, J. (1987). Research on Revision in Writing. Review of Educational Research, 57(4), 481-506.

Flagg, Barbara N. (1990). Formative Evaluation for Educational Technologies, Hillsdale: NJ: Lawrence Erlbaum.

Flood, J., and Lapp, D. (1990, April). Reading comprehension instruction for at-risk students: Research-based practices that can make a difference. Journal of Reading, 490-496.

Flower, L. (July, 1989). The Role of Task Representation in Reading-To-Write. Technical Report No.6, CSW, Carnegie-Mellon University, Pittsburgh.

Flower, L., Carey, L., and Hayes, J.R. (1985). Diagnosis in Revision: The Experts' Option. Technical Report No. 27, Communications Design Centre, Carnegie-Mellon University, Pittsburgh.

Flower, L., Hayes, J.R., Carey, L., Schriver, K., and Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. College Composistion and Communication, 37(1), 16-55.

Geis, G.L. (1988, April). Profiles of the actors in formative evaluation. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Geis, G. L. (1987). Formative Evaluation: Developmental Testing and Expert Review. Performance & Instruction, 26(4), 1-8.

Geis, G.L. (1986). Student input on materials development. Paper presented at the annual meeting of the American Educational Research

Association, San Fransisco, CA.

Gentner, D.R. (1988). Expertise in typewriting. In M.T.H. Chi, R.Glaser, and M.J. Farr (Eds.), The Nature of Expertise, (pp. 1-22). Hillsdale, NJ: Lawrence Erlbaum Associates.

Glaser, R. (1991). The Nature of Expertise. (Occasional paper No. 107). The Ohio State University, The National Center for Research in Vocational Education.

Glaser, R. (1984). Education and thinking: The role of knowledge. American Psychologist, 39 (2), 93-104.

Glaser, R., and Chi, M.T.H. (1988). Overview. In M.T.H. Chi, R. Glaser and M.J. Farr (Eds.), The Nature of Expertise, (pp. xv-xxviii). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gooler, D.D. (1980). Formative Evaluation Strategies for Major Instructional Development Projects, Journal of Instructional Development, 3(3), 7-11.

Hare, V.C. (1981). Readers' problem identification and problem-solving strategies for high and low-knowledge comprehenders. Journal of Reading Behavior, 13, 359-365.

Hayes, J.R., Flower, L., Schriver, K., Stratman, J., and Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), Advances in applied psycholinguistics: Vol. 2 (pp. 176-240). Cambridge, England:

Press Syndicate of the University of Cambridge.

Israeloff, A. (1992). Comparison of Feedback Generated by Experts and Learners during Formative Evaluation. Unpublished masters thesis, McGill University, Montreal.

Johnson, E.J. (1988). Expertise and decision under uncertainty: performance and process. In M.T.H. Chi, R. Glaser and M.J. Farr (Eds.), The Nature of Expertise, (pp. 209 -228). Hillsdale, NJ: Lawrence Erlbaum Association.

Kline, M.J. (1984). Social influences in textbook publishing. The Educational Forum, 48, 222-234.

Komoski, K. (1983). Formative evaluation: the empirical improvement of learning material. Performance and Instruction Journal, 22(5), 3-4.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Beverly Hill, CA: Sage.

Lampert, M. and Clark, C.M. (1990, July). Expert Knowledge and Expert Thinking in Teaching: A Response to Floden and Klinzing. Educational Researcher, 21-23.

Lawrence, J.A. (1988). Expertise on the bench: Magistrates' judicial decision-making. In M.T.H. Chi, R. Glaser and M.J. Farr (Eds.), The Nature of Expertise, (pp. 229-259). Hillsdale, NJ: Lawrence Erlbaum Associates.

Leinhardt, G. and Greeno, J.G. (1986). The Cognitive Skill of Teaching. Journal of Educational Psychology, 78(2), 75-95.

Leinhardt, G. and Smith, D.A. (1985). Expertise in Mathematics Instruction: Subject Matter Knowledge. Journal of Educational Psychology, 77(3), 247-271.

McAlpine, L. (1987). The think-aloud protocol: A description of its use in the formative evaluation of learning materials. Instruction Science, 17, 3-27.

McAlpine, L. and Weston, C.B. (1994). The Attributes of Instructional Materials, Performance Improvement Quarterly, 7(1), 19-30.

McCutcheon, G. (1980). How Do Elementary School Teachers Plan? The Nature of Planning and Influences on It. The Elementary School Journal, 81(1), 4-23.

McDonald-Ross, M. (1978). Language in texts. In L.S. Schulman (Ed.), Review of Research in Education, 6, 229-275.

Montague, W.E., Ellis, J.A. and Wulfeck II, W.H. (1983). Instructional Quality Inventory: A Formative Evaluation Tool for Instructional Development. Performance and Instruction Journal, 22(5), 11-14

Murphy, G.L., and Wright, J.C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. Journal of Experimental Psychology: Learning, Memory and Cognition,

10(1), 144-155.

Nevo, D. (1985, March). Experts' Opinion: A powerful evaluation tool. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Neves, D.M., and Anderson, J.R. (1981). Knowledge compiliation: Mechanisms for the automatization of cognitive skills. In J.R. Anderson (Ed.), Cognitive Skills and Their Acquision, (pp. 57-84).

Newell, A., and Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice Hall.

O'Donnell, H. (1985). Improving textbooks - who is responsible? Journal of Reading, 29, 268-270.

Pfleigher, J., Chomienne, M., Bordeleau, P., and Stolovitch, H. (1979). Enquete sur l'education des documents audiovisuels dans la Province du Quebec. GERDAVE (Rapport No.8). Montreal, Canada: Universite de Montreal, Section Technologie Educationelle.

Posner, M.I. (1988). Introduction: What is it to be an expert? In M.T.H. Chi, R. Glaser and M.J. Farr (Eds.), The Nature of Expertise, (pp. xxix-xxxvi).

Rahilly, T.J. (1991). An analysis of three learner-based formative evaluation conditions. Unpublished masters thesis, McGill University, Montreal.

Rahilly, T.J., Weston, C.B., and McAlpine, L. (1991). Issues in Segmenting. Unpublished manuscript, McGill University, Montreal.

Saroyan, A. (1989). The review process of formative evaluation: The role of the content expert and the instructional designer. Unpublished doctoral dissertation, McGill University, Montreal.

Scardamalia, M. and Bereiter, C. (1991). Literate Expertise. In K. A. Ericsson and J. Smith (Eds.), Toward a General Theory of Expertise, (pp. 172-193). Cambridge, Massachusetts: MIT Press. Schloss, P.J., Smith, M.A. and Posluzany, M. (1990). The impact of formative and summative assessment upon test performance of special education majors. Teacher Education and Special Education, 13(1), 3-8.

Scriven, M. (1967). The methodology of evaluation. In R.W. Tyler, R.M. Gagner and M. Scriven (Eds.). Perspective of curriculum evaluation, Chicago, Ill: Rand McNally.

Shulman, L.S. (1986). Those Who Understand: Knowledge Growth in Teaching. Educational Researcher, 5(14-15), 4-14.

Stolovitch, H.D. (1982). Application of the intermediate technology of learner verification and revision for adapting international instructional resources to meet local needs. Performance and Instruction Journal, 21(7), 16-22.

Swanson, H.L., O'Connor, J.E., and Cooney, J.B. (1990). An information processing analysis of expert and novice teachers' problem solving, American Educational Research Journal, 27(3), 533-556.

Tessmer, (1993). Planning and Conducting Formative Evaluations. London, England: Kogan Page Limited.

Thiagarajan, S. (1978). Instructional Product Verification and Revision: 20 Questions and 200 Speculations. Educational Communication and Technology, 26(2), 133-142.

Truett, C. (1984). Field testing educational software: are publishers making the effort? Educational Technology, 24(5), 7-12.

Tully, M.A. (1985). A descriptive study of the intents of state level textbook adoption procedures. Educational Evaluation and Policy Analysis, 7(3), 289-308.

Tyson, H. and Woodward, A. (1989). Why students aren't learning very much from textbooks. Educational Leadership, 47(3), 14-17.

VanLehn, K. (1989). Problem solving and cognitive skill acquistion. In M. Posner (Ed.), Foundation of Cognitive Science, (pp. 527-579). Cambridge, Massachusetts: MIT Press.

Voss, J.F., Greene, T.R., Post, T.A., and Penner, B.C. (1983). Problem solving skills in the social sciences. In G.H. Bower (Ed.), The Psychology of Learning and Motivation: Advances in Research Theory,

(Vol. 17, pp. 165-213). New York, NY: Academic Press.

Wager, J.C. (1983). One to one and small group formative evaluation: an examination of two basic formative evaluation procedures. Performance and Instructional Journal, 22(5), 5-7.

Weston, C.B. (1986). Formative evaluation of instructional materials: an overview of approaches. Canadian Journal of Educational Communication, 15(1), 5-17.

Weston, C.B. (1991, April). Separating data collection from revision: A framework for conducting formative evaluation research. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Weston, C.B. (1987). The Importance of Involving Experts and Learners in Formative Evaluation. Canadian Journal of Educational Communication, 16(1), 45-58.

Weston, C.B., and McAlpine, L. (1990). How to get formative evaluation feedback from experts. Paper presented at the annual conference of the National Society for Performance and Instruction, Toronto, Ontario.

Weston, C.B., and McAlpine, L. (1990). Product attribute categorization system. Unpublished manuscript, McGill University, Montreal.

Winograd, T. (1983). Language as a cognitive process: Vol 1. New York: Addison-Wesley.

APPENDIX A

Stimulus Materials

"Sound nutrition is not a panacea. Good food that provides appropriate proportions of nutrients should not be regarded as a poison, a medicine, or a talisman It should be eaten and enjoyed." This statement by the Food and Nutrition Board of the National Research Council in the U.S in a 1980 publication called "Toward Healthful Diets" raised more than a few eyebrows. Reaction from consumer groups was furiously negative. These groups along with many individuals objected to the conclusion that no specific dietary advice was appropriate for all citizens. The recommendation of a balanced diet with moderation in consumption did not sit well with people who were convinced that a great many of the ills of North American society are related to improper nutrition. A document detailing the evils of food additives, the benefits of vitamin supplementation and the virtues of "organic" foods would undoubtedly have received more favorable reaction. Science however cannot deal with emotions, beliefs or anecdotal evidence; it must be based on facts stemming from well controlled and reproducible experiments. Unfortunately in the area of nutrition it is very difficult to design and carry out studies which lead to conclusive results. Accordingly many reports of results are speckled with phrases like "may cause", "is consistent with", "is associated with"; all of which imply uncertainty. The difficulty of providing "proof" one way or another in the areas of food science and nutrition leaves the door open to a variety of opinions not only among the alarmists and self styled authorities but among nutritional experts as well.

Indeed, just two years after the above mentioned report the National Research Council issued a new document entitled "Diet, Nutrition and Cancer" with more specific recommendations reflecting the state of knowledge and information pertinent to the diet and the incidence of cancer. The guidelines now recommended a reduction of fat intake from about 40% to 30% of total calories, a reduction in the consumption of cured, pickled and smoked foods and an increase in the consumption of whole grain cereal products as well as fruits and vegetables, especially those rich in carotene. Vegetables belonging to the cabbage family were highly recommended but vitamin supplementation was not advised. The new report was in turn also criticized. Many scientists believe that not enough is known about the diet-disease connection to warrant specific guidelines for the population as a whole and furthermore the suggestion was made that if the guidelines were improperly applied they could lead to nutritional deficiencies. In light of the ongoing controversy it is appropriate to examine the studies and the kind of data that have lead to the debated recommendations. An examination of this controversy also serves to underline the need for a basic scientific understanding of chemical and nutritional concepts. Familiarity with terms like "minerals", "vitamins", "fat", "fiber", "carotene" etc. is essential for an objective and critical discussion of the relationship between diet and cancer.

There appears to be little doubt that many cancers are environmentally related. Epidemiological studies have clearly shown large differences in cancer rates between countries. For example, breast and colon cancer rates in many areas of the world are less than one fifth that in North America. The Japanese in turn have the highest incidence of stomach cancer in the world. Immigrants from other countries to the U.S. and Canada however experience the local cancer rates, suggesting an environmental influence.

Perhaps the best demonstration of this environmental effect comes from a study made public in 1984 by the National Cancer Research Institute of Japan. An epidemiological study spanning 16 years and involving over 100,000 men clearly showed that the incidence of cancer was greatest among those who smoked, drank alcohol, ate meat regularly and did not consume vegetables daily. Indeed the absence of vegetables from the diet appeared to increase the risk of a wide variety of cancers. The results of the survey are summarized below:

RELATIVE MORTALITY RATES

Smoking	Drinking	Meat	Vegetables	
No	No	No	Yes	1.0
No	Yes	Yes	No	1.1
Yes	Yes	Yes	Yes	1.7
Yes	Yes	No	No	1.8
Yes	No	Yes	No	1.8
Yes	Yes	Yes	No	2.5

The protective effects of vegetable consumption are dramatically illustrated by the above data; in fact even in the high risk group (smokers, drinkers and meat eaters) the risk of cancer can be reduced by one third if vegetables are regularly eaten. This protective effect may be manifested through the fiber, Vitamin C or carotene components of the vegetables as discussed below.

Accordingly many cancer experts now estimate that as much as 90% of North American cancers are environmentally determined and that a large fraction of these should therefore be avoidable. "Environmental" must not be confused with "man made"; in the present context the word is used to differentiate from "genetic" factors. Cigarette smoking and toxic wastes are environmental and obviously "man made", but exposure to sunlight and the consumption of naturally occurring carcinogens can also be termed "environmental". In fact, Bruce Ames of the University of California (Berkeley) has concluded after a survey of the scientific literature that most of the carcinogens that non-smokers encounter in their daily life come from natural foods and cooking methods. For example celery and parsley contain a carcinogen which becomes activated by light; mushrooms, beans and even alfalfa sprouts contain compounds which may increase the risk of cancer. Cooking, especially when food is browned or burned adds carcinogens to the diet. On the other hand, suggests Ames, food also appears to contain natural anti- carcinogens like Vitamins C and E, selenium and carotene which may decrease the risk of the dreaded disease. The fact that cancer rates aside from those related to smoking have remained almost constant over the years appears to imply that the "natural" components of the environment may be more important than the "man made" factors in inducing cancer.

In a controversial article in Science, 221, 1256 (1983), Ames produced summarized the many natural foods (above) which contained various carcinogens. In this same article, he also indicated that there were many foods which were also anti-carcinogens. The main idea here was that a minimum of the questionable foods coupled with a reasonable amount of the "good" ones (vide infra) would provide as good a balance of risk/benefit as could be achieved in this very complex area. Ames was criticized by a group of 18 academics, union officials and environmentalists in a 1984 letter to Science for "trivializing" cancer risks. Ames recently published a summary of relative risk factors for cancer by a careful (but controversial) examination of the literature. The resulting index called HERP (Human Exposure dose/Rodent Potency dose). This index considers two questions: How much of the material causes considerable rates of cancer in lab animals, and how much of it might an average person be exposed to over a lifetime? The rankings do not predict a person's actual chances of developing cancer, but show comparisons. If the relative ranking of tap water is 1.0, then peanut butter (2 tablespoons/day) is 30

(aflatoxin risk) as is comfrey tea (1 cup/day) (symphytine, a natural pesticide is present). One pack /day of cigarettes is rated at 12,000 while the risk of cancer from PCBs (once used in electrical transformers) is 0.2. Needless to say, such a detailed list has created concern and discussion and will stimulate research in the future.

Since the second World War some 50,000 synthetic chemicals have been introduced into the environment with about 500 new ones coming into use every year. Many of these are mutagenic or carcinogenic in lab tests yet the cancer epidemic that many scientists (even Bruce Ames at one time) have predicted has not materialized. Accordingly there is widespread, though certainly not universal, belief that most cancers are caused by natural carcinogens. Many of these carcinogens are produced by plants as natural pesticides to ward off insects. Ironically the current practice of breeding insect resistant plants in order to minimize the use of synthetic pesticides may actually be introducing new carcinogens into the diet. It is also a fact of course that not everyone gets cancer even though everyone consumes natural carcinogens.The explanation for this apparent inconsistency may lie in the possibility that whereas chemicals isolated from food can cause cancer, the whole food does not. Mutagens and "anti- carcinogens" are often present in the same food. For example the potentially harmful effects of the psoralens in parsley and celery may be counteracted by the carotene and vitamin components of these foods. It appears then that attention to a scientifically balanced diet may be more important in warding off cancer than worrying about the trace amounts of synthetic carcinogens in the environment. The following summarizes the current state of knowledge in this important area.

The Dietary Fat-Cancer Relationship

The above mentioned recommendation to reduce the fat content of the diet stems mostly from correlations noted by epidemiologists. A strong correlation exists between per capita fat intake and breast cancer mortality in women as well as between fat intake and mortality from colon cancer. It must be pointed out however that such associations do not imply cause. For example a similar correlation exists between gross national product and breast cancer. Although the "per capita" correlation of dietary fats with cancer is strong, there appears to be no conclusive correlation of individual fat consumption and cancer. There may be other variables in the relationship as well. Hormones like estrogen have been linked with cancer. Does the fact that women are having fewer and later pregnancies influence the average estrogen levels ? Could it be the added calories and not the fat per se which is instrumental? The human feeding studies which would be needed to clarify the situation can never be ethically done but studies in animals do suggest that higher levels of fat intake cause mammary tumors. Theoretically the argument can be put forward that fats cause cancer by undergoing oxidation in cells leading to the production of cancer causing reactive species called free radicals. These free radicals then damage the DNA of the cell, leading to improper replication. If this mechanism is correct, unsaturated fats may pose a greater risk since they are more easily oxidized. Some studies have indeed shown an association between cancer and "trans" fatty acids which are produced when vegetable oils are converted into margarine. Adequate Vitamin E , beta carotene and selenium consumption may prevent the oxidation of fats.

Dr. Keith Ingold at the National Research Council in Ottawa has in fact shown that Vitamin E is the major "free-radical trapping" anti-oxidant in human blood. Beta carotene can also act as an anti-oxidant, especially at low oxygen concentrations such as are found in cells. It is noteworthy that this important research started out as an investigation into why engine oils break down upon exposure to oxygen in the car's engine; a nice demonstration of how important results can come from seemingly "unimportant" research. Similarly the antioxidants BHT and BHA which had originally been developed to prevent fats in cereals from going rancid (and incidentally have been much maligned) may turn out to have an important role in not only the prevention of cancer but in actually slowing down the aging process.

Colon cancer has also been associated with high fat, high cholesterol diets. Once again though, epidemiological studies in individuals have yielded inconsistent results. Animal feeding studies in turn have shown that dietary fat promotes colon cancer. Furthermore, populations with high rates of colon

cancer have increased levels of bile acids in the feces; these have been associated with cancer and are known to be formed in larger amounts in high fat, high cholesterol diets. In summary, the evidence may appear to be somewhat circumstantial, but the recommendation to reduce fat content by 25% does not represent a risk as long as a balanced diet is maintained.

The Cured Foods-Diet Association

Once again population studies have shown that cancers of the stomach and esophagus are more common in countries such as China, Japan and Iceland where the diet is high in foods that are salt cured and smoked. There is no doubt that smoke contains cancer causing compounds and salt has been reported to promote gastric cancer in rats. Sodium nitrite, a pickling agent and preservative used in cold cuts, hot dogs, ham, etc. has been linked with the potential formation of nitrosamines, known carcinogens, in the body. Based upon these observations, limiting the intake of such cured or smoked foods would appear to be wise. Yet, even this recommendation has been challenged. It has been pointed out that the death rate from stomach cancer has been declining in North America while the consumption of processed meats has been rising. Furthermore, nitrite addition is so strictly regulated now that only minimal amounts are used; in fact the amount of nitrite now added can only prevent growth of the clostridium Botulinum organism if it is used in conjunction with salt. It is also true that most of the "smoked" foods presently marketed are smoked with liquid smoke. This is made by passing smoke through water; since the carcinogenic compounds do not dissolve in water foods "smoked" by this process are safer than "naturally" smoked foods. Although credence can be given to these criticisms, it must also be pointed out that foods high in smoke flavor and nitrites are generally high in fat and thus in calories-perhaps enough of a reason to minimize consumption.

The Selenium-Cancer Association

Selenium is a mineral required by the body in "trace" amounts. It plays a role in the activity of the enzyme glutathione peroxidase, an enzyme which protects cells from damage by oxidation. Consistent with this activity is the observation that mammary cancer in rats fed a high polyunsaturated fat diet can be inhibited by selenium. Selenium is found in the soil and is absorbed by crops. High soil selenium areas correlate inversely with cancer but these areas are also less populated and differ from low soil selenium areas in several respects. Indeed lung cancer rates are lower in countries where tobacco contains more selenium. Mexican and Colombian tobaccos have three times as much selenium as American and British tobaccos. Some correlations between blood selenium levels and cancer have also been noted and preliminary research has shown that the selenium content of hair and nails may reflect blood levels. High intake of selenium can be toxic and the presently available information does not warrant the recommendation of supplements.

The Cancer- Vitamin C and E Connection

The evidence for this association is essentially anecdotal although both of these vitamins are anti-oxidants and therefore could behave as anti- carcinogens. Vitamin E has been reported to reduce mutations in some bacterial systems and Vitamin C does block the conversion of nitrites to nitrosamines. For the latter reason Vitamin C is added to hot dogs. Similarly since both tomatoes and lettuce contain Vitamin C they can conceivably do more than just dress up the appearance and flavor of a bacon sandwich. Indeed a BLT may be the best way to consume bacon. There is however no evidence that either Vitamin E or C can prevent cancer.

The Cancer-Vitamin A Connection

Remember the stories about eating carrots to see better? This may be stretching the point, but the vitamin A in carrots does play an essential role in the chemistry of vision. Furthermore, the vitamin and its precursor compound (beta- carotene) may also protect the body against cancer. The rationale for this belief lies in the fact that vitamin A plays an important role in the control of cell differentiation and in that both vitamin A and especially beta-carotene are efficient scavengers of chemical species called free radicals. Since loss of cell differentiation is a basic feature of cancerous cells and since free

radicals are unstable, highly reactive chemicals which can damage our genetic materials (DNA and RNA) there is good reason to suspect that these two nutrients may have a protective effect against cancer.

Vitamin A itself can be obtained from animal products such as liver, eggs and meat or it can by synthesized by the body from. beta carotene. Many green vegetables produce this bright orange compound but the richest sources are pumpkins, spinach and of course carrots.

In 1975 a major epidemiological study showed that Norwegian men consuming more than the average amount of vitamin A had less than half the rate of lung cancer as compared with men having below average consumption of the vitamin. Similar findings were also reported in the following 5 years from scientists in Japan, Singapore and the United States.

A further study (Nov. 1981) published in the British medical journal Lancet supported the hypothesis that the pro-vitamin A (beta carotene) and not the vitamin itself was the beneficial factor. The study showed that there was an inverse relationship between intake of dietary beta-carotene and lung cancer in 1,954 middle aged male smokers over a period of 19 years. Intake of preformed vitamin A did not show a significant effect.

Unfortunately, studies on vitamin A are often limited due to its toxicity. High levels of vitamin A lead to liver damage, headaches, lack of appetite, hair loss, menstrual problems and retarded growth in children -- problems sometimes seen among vitamin and health food faddists. On the other hand, optimal investigative approaches are possible with beta carotene since there are no known serious side effects, even with doses so high as to cause and obvious orange skin coloration. In recent years synthetic analogs of vitamin A have been prepared in an effort to reduce its toxicity. These safer compounds are now being tested with high risk groups to determine if other forms of cancer can be prevented. One such group consists of albino children in Africa who have a 100% risk of developing skin cancer. In addition, at the present time the U.S. National Institute of Health has invited all male physicians between the ages of 40 and 85 to participate as subjects in a placebo-controlled general study of beta- carotene and cancer.

A major report on this issue published in the New England Journal of Medicine, March 1984 (by the Harvard School of Public Health) explained that although the protective effect against lung cancer of beta-carotene is strongly supported by many studies, there are indications that these effects may not apply to other types of cancer.

In conclusion, it should be noted that the main cause of lung cancer, smoking, also increases one's risk of several other serious diseases, including atherosclerosis -- a primary cause of death in North America. However, there is no evidence that either vitamin A or beta carotene affects this condition in any way.

The Cancer-Fiber Connection

Roughage? Unappetizing, tasteless, completely indigestible but... it fights cancer!

It all started with Dr. Dennis Burkitt's 20-year observation of diets and incidence of colorectal cancer in rural Africa. The British surgeon noted that although cancer of the lowest five to six feet of the intestine is very prevalent in the western world it is almost nonexistent among people in Africa consuming a high fiber diet. In Canada, about 100,000 people get colon cancer every year, half of whom die within the same year. The same high frequency of this malignancy has been found in the U.S., Scotland, Denmark and especially New Zealand, countries which consume the highest amounts of meat and animal fat around the world.

The incidence of this type of cancer appears to be 100 times more prevalent in the lowest 1% of the small intestine. This leads scientist to believe that carcinogens are not swallowed with our food but are produced in the colon from material in the feces. It has been suggested that bile acids (biomolecules

naturally released into the gut in response to the presence of fat in the diet) are chemically altered by bacteria to produce carcinogens. High colon cancer areas have been found to be much more abundant in colorectal cancer patients than in control groups. In a recent study conducted by Dr. Tracy Wilkins, a microbiologist at the Virginia Polytechnic Institute in Blacksburg, a chemical mutagen, named faecapentaene, was isolated from the feces of about 20 per cent of the white residents of Johannesburg. The same compound was detected in less than 2 per cent of the rural population. The diet of the urban community is very similar to ours (high in refined carbohydrates and fat), whereas that of the rural population is low in meat and fat and high in fruits and vegetables. Although most carcinogens are mutagens not all mutagens are carcinogens, and therefore the presence of faecapentaene does not necessarily mean that it is the cause of cancer. Dr. David Kingston, a chemist at the Virginia Polytechnic Institute, has synthesized this compound and its cancer-causing potential will now be investigated in laboratory animals.

These findings certainly support the theory that fiber, which increases the rate of feces elimination, should lower one's chances of developing cancer of the colon. . However, there are some inconsistencies in the findings related to the effects of fiber. For instance, in a Canadian study published in 1980 higher consumption of dietary fiber was shown not to have any significant effect on cancer whereas in Puerto Rico high consumption was associated with higher incidence of colon cancer. Such discrepancies may be related to the extremely heterogeneous nature of dietary fiber. Dietary fiber is a mixture of indigestible chemicals: cellulose, hemicellulose, lignin and pectin. Preliminary studies have shown that wheat bran and fiber from citrus fruits protect laboratory animals against chemically-induced colon cancer. Since citrus fruits are also an excellent source of vitamin C (a scavenger of carcinogenic free radicals) an orange a day, or even the traditional apple a day, may not be such a bad idea.

GUIDELINES FOR AN ANTICANCER MENU

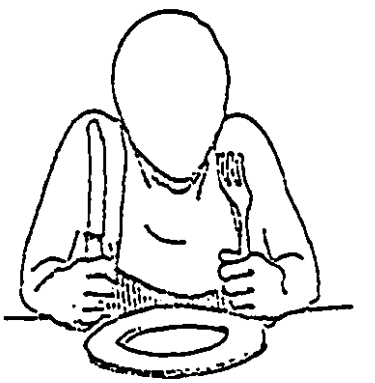
-decrease consumption of fats, nitrite-cured meats, smoked or charcoal-broiled meats and large amounts of alcohol.

-increase consumption of foods rich in dietary fiber,. beta carotene, vitamins A, E and C and the mineral selenium (megadoses of dietary supplements are presently not recommended).

-consume often, cruciferous vegetables such as cabbage, broccoli, Brussels sprouts and cauliflower.

RECENT REVIEW

A recent summary which gives a balanced report is from Scientific American, November, 1987, p. 42.



APPENDIX B

Subjects' Materials

Summary of Activities

1. I'll ask you to read a page of instructions, then check that you understand the task.

2. The technique you'll be using is called a Think-Aloud. I'll give you some information on that next, and give you time to read it.

3. I'll give you a short warm-up task to allow you to practice thinking aloud.

4. I'll give you the module, ask you to work through it and tape record your comments as you think aloud.

5. Finally, I'll ask you for some brief demographic information an we'll discuss the materials more generally.

CONSENT FORM

This is to state that I agree to participate in a programme of research conducted by Alanna Israeloff and Diana Tremblay under the supervision of Dr. Cynthia Weston of the department of Educational Psychology at McGill University.

I understand that the purpose of the research is to evaluate educational materials and that my performance or ability is not being judged.

I understand that my participation in the study is totally anonymous and confidential.

I understand that I can have a full description of the results of the study after its completion.

I understand and give my consent to have the session audiotaped.

I understand that the data from this study may be published.

I understand that I will be paid one hundred dollars ($100) for my participation upon completion of the session.

I HAVE CAREFULLY STUDIED AND UNDERSTOOD THIS AGREEMENT AND THEREFORE I FREELY CONSENT AND AGREE TO PARTICIPATE IN THE STUDY.

NAME: _____

SIGNATURE: _____

DATE: _____

Description of your task

PLEASE TAKE SOME TIME TO READ THIS CAREFULLY

This unit, "The diet-Cancer Relationship", is being developed for use in a first-year chemistry course for both Arts and Science students who are not majoring in chemistry. Please give me your comments and feedback as an expert on anything you observe with regard to the content of the material and how well learners could use this unit. Feel free to highlight anything that is positive, as well as negative, about the materials.

While you are reviewing, we want you to speak your thoughts out loud so that we can record what you are saying. This technique is called Think-aloud, which means saying out loud all the thoughts, questions, comments and strategies that go through your mind while you are performing a task. This does not mean that you should analyze what you are doing. Just report your thoughts. Since the Think-aloud process is unfamiliar to most people, we shall be having a short practice session in a few moments. You may wish to write comments directly on to the materials while you are thinking out loud, but please remember to say what you are writing. Every comment you make is valuable to this research.

At times you may forget to think out loud, so my role will be to prompt you to continue. While you are speaking, I shall also be keeping track of your comments. In the Think-aloud procedure, I can't answer any questions that you may ask, but there'll be a chance to do so at the end of the session. Please take your time on the task. I shall suggest times for taking breaks, but whenever you feel you want a break, please do not hesitate to say so.

Is there anything unclear about these instructions?

Warm Up Exercise

The following task is a warm up exercise to help you to become familiar with the Think-Aloud procedure. You will be provided with 12 letters and your task will be to combine these letters to form as many words as possible. You may use any of these letters as many times as you wish.

While performing this task, it is important that you verbalize all your thoughts as they occur to you. Do not be concerned about speaking correctly. What is more important is that you continue to think aloud during the entire exercise. If at any time I feel you are not talking often enough, I will prompt you to continue verbalizing your thoughts. You will have three minutes to complete this exercise.

QUESTIONS SUBJECTS FREQUENTLY ASK ABOUT TALK-ALOUD PROTOCOLS OF THE WRITING AND REVISION PROCESS

WHAT IS A TALK-ALOUD PROTOCOL?

A protocol is a sequential recording of a person's attempt to perform a task. Developed early this century, protocol analysis is a powerful tool in educational research. In particular, the information captured in "talk aloud" protocols enables the educational researcher to construct detailed models of human thinking processes, and construct detailed models of human thinking processes, and in some cases to stimulate these processes in a computer program. In short, protocols give the researcher a "window" through which to look at otherwise invisible mental processes that occur from moment to moment. We are concerned in this short explanation with talking-aloud protocols of writers revising a text. Listed below are some more questions that subjects of protocol experiments ask.

HOW IS A TALKING-ALOUD PROTOCOL MADE?

The procedure is really very simple. The researcher will ask you to talk-aloud while you are revising or rewriting a particular document. You are to say out loud what you are thinking. You are not to worry about speaking correctly, stopping in the middle of thoughts or sentences, etc., but you should try to verbalize as continually as you can during the entire time you are at work. Pauses in your talk will naturally occur, but try to avoid them. If the researcher feels you are not talking often enough, he/she may prompt you.

SHOULD I TRY TO EXPLAIN HOW MY WRITING PROCESSES WORK, OR HOW I WOULD USUALLY DO THIS TASK?

Subjects who ask this question are usually trying to do the researcher's work themselves, at the same time they are revising or rewriting. You are not to describe what you "would" do, but only what you are actually thinking about at the time you are working. In fact, you are not expected to "analyze" your writing habits or creative processes at all. You are not being asked to "introspective", or to give an explanation or interpretation of your writing. You need only say what is on your mind at the moment. <u>Concentrate on the task you have been given, and simply say aloud whatever occurs to you</u>.

HOW CAN A PROTOCOL CAPTURE MY THINKING PROCESSES IF A CAN'T SAY ALOUD EVERYTHING I AM THINKING?

Of course, you will not be able to say everything you are thinking when you are completing even a simple writing or revision task. A portion of your thinking is lost and falls between the cracks. But almost everything you do say is valuable to the researcher. The amount of information retrieved from the talk-aloud method probably exceeds the amount to be g=gained by any other research method currently employed for the study of how people write and revise. Moreover, your transcribed protocol is also studied by the researcher; your finished or revised text is also studied, and compared with the talk-aloud transcript. By itself, your finished text tells the researcher very little about the processes you used to create the text. However, when your finished text is "matched" with the protocol transcript, the researcher has a much more detailed picture of how your writing and revising unfolded. Again: concentrate on the task and on whatever you are conscious of as you work. Say aloud everything that comes to mind.

DOESN'T TALKING-ALOUD INTERFERE WITH MY THINKING, SO THAT I AM NOT THINKING AND WORKING AS I NORMALLY WOULD?

This question is often asked, and rightfully so. It's a very important question for researchers to deal with. At the present time, no one knows for certain if talking-aloud does interfere with your thinking during problem-solving. A lot of research is presently being conducted to find out. So far, researchers have been unable to find any strong evidence that talking interferes with thinking. Some research has even shown that, with very little practice, you can solve the same problem in the same amount of time whether you are talking aloud or not. Talking loud can also improve decision-making, and many people talk to themselves when they write anyway. The first few minutes of a protocol may feel awkward, but with a little practice this feeling will disappear and you will feel more comfortable.

SHOULD I WRITE AND TALK AT THE SAME TIME, OR ONLY BEFORE OR AFTER I WRITE SOMETHING DOWN?

You should talk as continuously as possible, whether you are writing or not. Sometimes you will find yourself only able to say exactly what you are writing on paper. This is perfectly fine, so long as you don't pause too long between words. If you do, the experimenter will prompt you.

SHOULD I TALK-ALOUD EVEN IF I AM JUST RE-READING WHAT I'VE WRITTEN?

Yes, you should. Avoid the temptation to mumble if, and when, you re-read your text. Even if you are skimming rapidly, and not re-reading sentences in their entirety, talk-aloud and make sure your voice is audible and clear.

APPENDIX C

Administrator's Materials

Tracking Sheet

- 1 -

THE DIET CANCER RELATIONSHIP

"Sound nutrition is not a panacea. Good food that provides appropriate proportions of nutrients should not be regarded as a poison, a medicine, or a talisman. It should be eaten and enjoyed." This statement by the Food and Nutrition Board of the National Research Council in the U.S. in a 1980 publication called "Toward Healthful Diets" raised more than a few eyebrows. Reaction from consumer groups was furiously negative. These groups along with many individuals objected to the conclusion that no specific dietary advice was appropriate for all citizens. The recommendation of a balanced diet with moderation in consumption did not sit well with people who were convinced that a great many of the ills of North American society are related to improper nutrition. A document detailing the evils of food additives, the benefits of vitamin supplementation and the virtues of "organic" foods would undoubtedly have received more favorable reaction. Science however cannot deal with emotions, beliefs or anecdotal evidence; it must be based on facts stemming from well controlled and reproducible experiments. Unfortunately in the area of nutrition it is very difficult to design and carry out studies which lead to conclusive results. Accordingly many reports of results are speckled with phrases like "may cause", "is consistent with", "is associated with"; all of which imply uncertainty. The difficulty of providing "proof" one way or another in the areas of food science and nutrition leaves the door open to a variety of opinions not only among the alarmists and self styled authorities but among nutritional experts as well.

Indeed, just two years after the above mentioned report the National Research Council issued a new document entitled "Diet, Nutrition and Cancer" with more specific recommendations reflecting the state of knowledge and information pertinent to the diet and the incidence of cancer. The guidelines now recommended a reduction of fat intake from about 40% to 30% of total calories, a reduction in the consumption of cured, pickled and smoked foods and an increase in the consumption of whole grain cereal products as well as fruits and vegetables, especially those rich in carotene. Vegetables belonging to the cabbage family were highly recommended but vitamin supplementation was not advised. The new report was in turn also criticized. Many scientists believe that not enough is known about the diet-disease connection to warrant specific guidelines for the population as a whole and furthermore the suggestion was made that if the guidelines were improperly applied they could lead to nutritional deficiencies. In light of the ongoing controversy it is appropriate to examine the studies and the kind of data that have lead to the debated recommendations. An examination of this controversy also serves to underline the need for a basic scientific understanding of chemical and nutritional concepts. Familiarity with terms like "minerals", "vitamins", "fat", "fiber", "carotene" etc. is essential for an objective and critical discussion of the relationship between diet and cancer.

There appears to be little doubt that many cancers are environmentally related. Epidemiological studies have clearly shown large differences in cancer rates between countries. For example, breast and colon cancer rates in many areas of the world are less than one fifth that in North America. The Japanese in turn have the highest incidence of stomach cancer in the world. Immigrants from other countries to the U.S. and Canada however experience the local cancer rates, suggesting an environmental influence.

Perhaps the best demonstration of this environmental effect comes from a study made public in 1984 by the National Cancer Research Institute of Japan. An epidemiological study spanning 16 years and involving over 100,000 men clearly showed that the incidence of cancer was greatest among those who smoked, drank alcohol, ate meat regularly and did not consume vegetables daily. Indeed the absence of vegetables from the diet appeared to increase the risk of a wide variety of cancers. The results of the survey are summarized below:

Script for Ice-Breaker

Thank you for your participation in this study. I'd like to start by explaining what we're going to be doing today, and if you have any questions, feel free to interrupt me.

A bit of background on our research: This study is part of an ongoing research project concerned with the improvement of instructional materials. We get feedback about instructional materials from learners and experts, then investigate how to use this information to revise and improve the materials.

We'd like to get your feedback about a module from an introductory chemistry course for non-chemistry university students from both Arts and Science. We tell experts and learners alike that we're evaluating the materials, not you!

The procedure for the session is outlined for you on this sheet:

(Hand out Summary of Activities sheet)

Script for Summary of Activities

1. I'll ask you to read a page of instructions, then check that you understand the task.

2. The technique you'll be using is called a Think-Aloud. I'll give you some information on that next, and give you time to read it.

3. I'll give you a short warm-up task to allow you to practice thinking aloud.

4. I'll give you the module, ask ycu to work through it and tape record your comments as you think aloud.

5. Finally, I'll ask you for some brief demographic information an we'll discuss the materials more generally.

Is that short description of the procedure clear?

Before beginning, I have a consent form which I'd like you to read and sign, if you agree with what it says;

(Give consent form)

If you agree to the statements, would you please sign at the bottom of the form?

Script for Warm Up Exercise

The following task is a warm up exercise to help you to become familiar with the Think-Aloud procedure. You will be provided with 12 letters and your task will be to combine these letters to form as many words as possible. You may use any of these letters as many times as you wish.

While performing this task, it is important that you verbalize all your thoughts as they occur to you. Do not be concerned about speaking correctly. What is more important is that you continue to think aloud during the entire exercise. If at any time I feel you are not talking often enough, I will prompt you to continue verbalizing your thoughts. You will have three minutes to complete this exercise.

(subject performs warm-up exercise)

ARE YOU CLEAR ABOUT HOW TO DO A THINK-ALOUD??
YES - move on
NO - think-aloud handout to read

BEFORE WE BEGIN REVIEWING THE STIMULUS MATERIALS, I WOULD LIKE TO REMIND YOU THAT YOUR TASK IS TO COMMENT AND GIVE FEEDBACK AS AN EXPERT ON ANYTHING YOU OBSERVE WITH REGARD TO CONTENT OF THE MATERIAL AS WELL AS, HOW WELL LEARNERS COULD USE THIS UNIT. HIGHLIGHT ANYTHING THAT IS POSITIVE AS WELL AS NEGATIVE ABOUT THESE MATERIALS.

Possible prompting words during Think-Aloud

1. Keep thinking aloud please
2. Keep talking please
3. Tell me what you're thinking please
4. What are you thinking please?
5. Please remember to keep talking
6. Tell me more about it
7. Is there something you want to say about that?
8. Tell me more

Last resort:

9. Is something wrong?

Script for Questionnaire for Retrospective Interview

(Tape recorder should be running)

Thank you for taking part in this study. As you were working through the materials, I took some notes. I wasn't permitted to ask or answer any questions then, but I'd like to check some specifics now, then ask a few more general questions.

1. (Ask questions generated by tracking sheet)

Thank you for clarifying these points. Now I have some more general questions to ask.

2. Aside from the think aloud, how would you describe the task we asked you to do?

   IF NEEDED: a) Describe what you did while you were working with the materials.

3. You just described the task as __(use subjects definition of task from question #2)__. How did you go about accomplishing this task?

   IF NEEDED: a) Did you make any kind of plan?

4. How did you decide that you had completed your task?

5. When you were reviewing these materials, how much attention did you give to the following areas?

a) SUBJECT MATTER: Examples of subject matter are the value of the content, how related items are grouped together, if the content is unbiased, and if the content represents current trends

1_____2_____3_____4
very little a lot

b) PRESENTATION: Examples of presentation are spacing, typeface, headings, visuals, highlighting, format, and layout

1_____2_____3_____4

c) LANGUAGE: Examples of language are the choice of vocabulary, sentence structure, redundancy, clarity, concciseness, and coherence of ideas

1_____2_____3_____4

d) INSTRUCTIONAL DESIGN Examples of instructional design are concerns about the target audience, objectives, motivational elements, instructional strategies and coherence of instruction.

1_____2_____3_____4

6. How would you define your area of expertise?
IF NEEDED: a) We asked you here today because we considered you to be an expert. What kind of expert would you consider yourself to be ?
b) Would you be a content or audience expert?

7. a) Did you experience any problems with this reviewing task (i.e., the time involved, having to verbalize all your thoughts etc.)?

b) Do you have any other comments you would like to make?

8. Are there any questions that I could answer for you?

9. Now I'm going to ask you a few questions on your background:

a) Would you give me a full list of your professional qualifications (i.e., degrees, professional affiliations etc.)?

b) Have you any prior experience with <u>(use subjects definition of the task from question #2)</u>?

c) **IF YES:** Is this how would you normally go about it?

d) How many years of experience do you have in your field?

e) How do you rate your familiarity with the content of "The Diet Cancer Relationship"?

1_____2_____3_____4
completely very
unfamiliar familiar

d) How would you rate your familiarity with undergraduate arts and science students for whom these materials are intended.

1_____2_____3_____4
completely very
unfamiliar familiar

Goodbyes

That's all!

I feel we have a lot of very useful data here, and I appreciate the effort that you put into it. I know that doing a Think-aloud is a demanding process and I hope it hasn't been too tiring for you.

Once again, thank you for taking the time to help us with this research. Please accept this as a token of our thanks. I'll need one more signature from you to acknowledge receipt of the cheque.

If later on you should have questions you wish to ask us, we can be reached through the Centre for University Teaching and Learning at McGill University (398-8063).

(S should be escored from room or give directions for way out. Offer of results should be made in a follow-up thank you letter.)

APPENDIX D

Sample of Segmented and Coded Protocol

		TEXT:	"The diet cancer relationship"
1	S2:	ES+/SM1	Sounds like an interesting topic./
		TEXT:	"Sound nutrition is not a panacea."
2	S2:	BM	Um,/
3		ES+/ID	that certainly sounds like a good um introduction./
		TEXT:	"Good food that provides appropriate proportions of nutrients should not be regarded as a poison, a medicine"
4	S2:	ES/SM5	Oh, sounds pretty negative there/
5		BM	um/
		TEXT:	"or a talisman"
6	S2:	PI/L	I don't know what they mean by that so/
7		BM	um/
		TEXT:	"It should be eaten and enjoyed."
8	S2:	PI/SM5	Right there it almost sounds like it's presented as a medicine, a little bit preaching./
		TEXT:	"This statement by the food and nutrition board of the national research council in the U.S. in a 1980 publication called toward healthful diets raised more than a few eyebrows."
9	S2:	ES+/SM2	Ya, I can understand why./
		TEXT:	"Reaction from consumer groups was furiously negative. These groups along with many individuals objected to the conclusion that no specific dietary advice was appropriate for all citizens."
10	S2:	PI/SM2	Well, I I don't know if I'd uh necessarily go along with um their reasons for being uh furious with that./

Appendix E

Product Attribute Checklist

REVIEW CHECKLIST: INSTRUCTIONAL DESIGN ATTRIBUTES

Included are pedagogical attributes that should be evident in a well designed instructional product. Questions under each attribute are not inclusive, but are only representative examples. This checklist might best be used by an instructional designer.

1. Justification of need:
   - Do the materials meet a definite need?

2. Target Audience:
   - Is the audience described?
   - Is there evidence that ability level, readiness, attitudes culture of the target audience have been considered?
   - Have individual differences been taken into consideration, if appropriate?

3. Entry level prerequisites:
   - Have prerequisites for students (e.g., necessary skills) and teachers (e.g., qualifications) been specifically stated?

4. Objectives:
   - Are objectives explicitly or implicitly evident?
   - Are objectives attainable and suitable?

5. Motivation and context for learning:
   - Are student experiences considered in the way mate rials are presented?
   - Are there attention gaining techniques within the instructional content that require students to link new learning to prior learning?

6. Instructional strategies:
   - Are instructional methods and media appropriate for objectives, audience and subject matter?

7. Organization and structure of content:
   - Have important ideas been repeated and emphasized to draw attention and enhance learning?
   - Have structural features such as outlines, overviews, advance organizers, transitions, review s and summaries been used where appropriate?
   - Has information been chunked and sequenced appropriately?
   - Has a clearly explained verbal or visual cueing system been used to emphasize important concepts and maintain logical presentation? (N.B. differentiated from 'Format and Layout' in Presentation Checklist which focuses on visual

consistency as opposed to conceptual or logical function implied here)

8. Examples:
   - Have appropriate examples and non-examples been provided where necessary?

9. Practice:
   - Has the opportunity been provided for the learner to practice the desired learning outcomes(s) (e.g., discussion, exercises, labs, tutorials, case studies, questions)?

10. Feedback:
    - Is prompt and appropriate feedback provided where necessary?

11. Evaluation of learning:
    - Are desired learning outcomes assessed by means of valid and reliable measures?
    - Can evaluation measures be used to diagnose difficulties?

12. Internal alignment and integration:
    - Has a match been maintained between various instructional components such as goals, objectives, content, practice, feedback, and evaluation measures?
    - Is there a complete description of the product including purpose and all relevant components?
    - Are all relevant components of instruction included as appropriate (e.g., table of contents, glossary, answer key, index, teacher manual, supplementary material)?
    - Is there information on the instructional environment (e.g., required facilities, student/teacher ratio, recommended methodology, group size, and allotted time)?

13. External Alignment:
    - Do the materials have educational value?
    - Is instruction coherent with audience, curriculum, needs and environment?
    - Is there any information on the reliability of materials and their effectiveness with representative learners?
    - Is the purpose and rational for instruction evident?

REVIEW CHECKLIST: SUBJECT MATTER ATTRIBUTES

1. Value of content:
   - Ascertain that content is relevant, important, appropriate, and necessary.

2. Content Accuracy:
   - Ascertain that content is accurate and that integrity of the subject matter is maintained.
   - Verify the source of the subject matter
   - Specify whether the content is documented and research based, and whether it has been reviewed by scholars in the field.
   - Use credible authors who are known in the field.
   - Do not include misleading content or misinformation.

3. Comprehensiveness:
   - Ensure that instructional content is comprehensive in terms of both quality and quantity.
   - Ensure that the rationale/philosophy are in harmony with the educational goals of that particular area of education.

4. Coherence:
   - Ascertain that content elements are properly integrated.
   - Group related items together.

5. Objective presentation / bias:
   - Remain objective and unbiased in the presentation of content.
   - Do not use stereotypes.

6. Recency:
   - Present content which is `state of the art', that represents current trends in the area.

REVIEW CHECKLIST: PRESENTATION ATTRIBUTES

This checklist deals with the physical attributes of instructional materials. It might best be used by a specialist in the particular medium of presentation (e.g., text, video).

1. Space:
   - Provide ample space where written answers are elicited.
   - Use a consistent method for allocating space between headings, sub-headings, paragraphs, words and lines.

2. Typeface:
   - Use legible typeface (e.g., use simple serif or sans serif type styles).
   - Maintain consistency in typeface.
   - Avoid crowding the text or the screen in order to make reading easier.
   - Use upper case type for initial letters and proper nouns; otherwise use lower case which facilitates reading.
   - Avoid extensive use of italics as they reduce reading speed.
   - Avoid using a string of capital letters, in particular whole paragraphs.
   - Use bold or extra bold typeface only where emphasis is needed.

3. Titles, headings, and sub-headings:
   - Use to clarify and guide
   - Make them as short as possible.
   - Print in the same fashion throughout the text.

4. Use of numbers:
   - The use of numbers is encouraged for a sequence of steps or in lieu of sub-headings, and for displaying nested content.
   - The use of the number symbol rather than prose is preferable in instructional text in particular when presenting a series of items.

5. Graphics, illustrations, visuals:
   - Are they appropriately used?
   - Are they supportive of content and accomplish something that the narrative cannot?
   - Are they closely integrated with the meaning of the narrative?
   - Illustrations should be appropriate for the intended audience.

- Simplicity or complexity matches needs of content.
- Make tables as comprehensive as possible, displaying patterns and trends clearly.
- Use flowcharts when complex information has to be sorted or choices have to be made.
- Provide the reader with ample information on using flowcharts.

6. Audio/Music:
   - Are they appropriately used?
   - Are they supportive of content and accomplish something that visuals/text alone cannot?
   - Music is integrated.

7. Colour:
   - Use colour sparingly and with a purpose which is clearly explained.
   - Use colour to enhance or highlight a display and to promote discrimination between elements.

8. Page size and style:
   - Use standard page size.
   - Maintain a consistent structure, especially in page length and visual balance to make presentation aesthetically pleasing.
   - Avoid using dark coloured paper.

9. Margins:
   - Use unjustified right margins. Right justification impairs reading and causes awkward word spacing and hyphenation.

10. Columns:
    - Use a two column structure instead of a one or three column structure for straightforward prose text.
    - Use a single column text for content which is interrupted by charts and tables.

11. Technical Quality:
    - Ensure technical quality of visuals, audio, text (e.g., clarity of graphics, exposure, no typographic errors).

12. Highlighting:
    - Use various techniques to emphasize important concepts (e.g., colour, typeface, typestyle, graphic conventions such as boxes).
    - Prompts, visuals, narrative displays, colour, and sound are used to support instructional plan (e.g., a new term is identified by a visual cue such as underlining, a different typeface, or bold).

13. **Format and layout:**
    - Maintain a consistent format.
    - Use various aids such as numbering systems, headings, indentation, and spacing to promote a consistent presentation.

REVIEW CHECKLIST: LANGUAGE ATTRIBUTES

Included are attributes that deal with how language is used to express ideas. This checklist might best be used by a language arts specialist who is familiar with the target audience.

1. Choice of vocabulary:
   - Avoid jargon, acronyms, abbreviations, technical or complex terms.
   - Use appropriate words for reading level of population.
   - Define new terms.
   - Use prose (e.g., half of the group) instead of numbers (e.g., 50% or 16 out of 32) tc facilitate retention of general concepts.

2. Complexity of sentence structure:
   - Limit the number of clauses contained in a sentence.
   - Avoid noun strings.

3. Verbs:
   - Use active voice rather than passive.
   - Avoid using negative except when a particular emphasis is to be made.

4. Redundancy:
   - Provide references for reader to link subjects within and between sentences.
   - Use relative pronouns and other `function words' to promote comprehension.

5. Transitions:
   - Indicate relationship between sentences (e.g., therefore, however, in contrast).

6. Consistency:
   - Use parallelism when appropriate (e.g., consistent use of infinitive form in a list).
   - Use standard English conventions.

7. Clarity:
   - Use elaboration, example, restatement.

8. Conciseness:
   - Be brief, to the point.

9. Coherence of ideas:
   - Ideas are closely tied.

10. Appropriateness for audience:
    - Maintain a balance between novel and familiar content. Novel content adds to complexity.
    - Use familiar terms especially when relationships are described.
    - Rewrite abstractions into concrete ideas so that readers can perceive them with ease.
    - Provide comparisons and contrasts when introducing new concepts.

APPENDIX F

Problem Space Strategies

TEXT COMPREHENSION PROBLEM SPACE

Included in this problem space are techniques used by strategic readers (Flood & Lapp, 1990) to improve their comprehension of the materials. At the initial stage of the plan, the major concern is to prepare for the upcoming critical reading episode by specifying the purpose for reading the text and then determining the appropriate reading method. The next step is to identify the principal message and major points of the text. The last step requires effective, analytical reading in order to comprehend what the author is attempting to communicate.

1. **Define objective:**
   - Set purpose for reading by asking what is to be achieved during the reading episode.
   - Choose a reading method that is suitable for the purpose that has been specified.

2. **Scan text:**
   - Preview text by scanning title, pictures, headings, summary, references, study questions and highlighted information.
   - Assess the credibility of the author.

3. **Build background:**
   - Activate prior knowledge through self-questioning about the topic, vocabulary and stylistic form of text.

4. **Concentrate:**
   - Focus attention.
   - Pause to reflect.

5. **Predict:**
   - By using context clues predict what will come next.

6. **Monitor comprehension:**
   - Use metacognitive knowledge to monitor comprehension.

7. **Clarify:**
   - Re-read text to clarify the main points.

REVIEW PROBLEM SPACE

Included in this problem space are strategies the reviewers in this study used to identifying problem areas or positive aspects of the instructional materials.

1. **Skim text:**
   - Skim through materials to quickly identify positive and negative aspects.

2. **Read the text thoroughly:**
   - Fully read the text to isolate positive and negative areas.
   - Re-read to distinguish positive and negative aspects.

3. **Identify confusing information:**
   - Mark confusing details.

4. **Check original source:**
   - Assess credibility of author or publisher.
   - Recheck data.

5. **Check definitions:**
   - Define unknown terminology to judge whether these terms have been used appropriately.

6. **Search for ambiguous words:**
   - Look for imprecise terminology that requires clarification.

7. **Rely on prior knowledge:**
   - Use expert knowledge domain to evaluate the materials.

8. **Challenge content and the quality:**
   - Question issues covered in the materials.
   - Contemplate the quality of the materials.

9. **Comment:**
   - Criticize or comment on aspects of the materials.

REVISION PROBLEM SPACE

This problem space incorporates strategies suggested by the reviewers in this study to improve the instructional materials. This category is based on the product attribute checklists conceived by Weston and McAlpine (1990) and the strategies involved fall under the same four major categories, which are: (a) Instructional Design, (b) Subject Matter, (c) Presentation and (d) Language.

I - INSTRUCTIONAL DESIGN STRATEGIES

1. **Introduce concepts:**
   - Present an overview of a concept to introduce the learner to this new idea.

2. **Use attention gaining techniques:**
   - Use attention gaining techniques to motivate the learner to continue and to help the learner link new learning with prior knowledge.

3. **Emphasize important ideas:**
   - Emphasize and repeat important ideas in order to draw attention and enhance learning.

4. **Resequence information:**
   - Chunk and sequence information in a logical manner that will enhance learning.

5. **Use concrete examples:**
   - Use appropriate examples and non-examples that the learner can relate to.

6. **Align with chemistry curriculum:**
   - Make instruction coherent with the requirements of the curriculum.

II - SUBJECT MATTER STRATEGIES

1. **Include new information:**
   - Include certain information that will ensure the content is more comprehensive in terms or quantity and quality.

2. **Rewrite with accurate information:**
   - Rewrite text to include more accurate information.

3. **Cite sources:**
   - Include appropriate references.

4. **Identify research design flaws:**
   - When citing empirical studies identify the obvious research design in order to train learners to critique scientific work.
5. **Delete irrelevant information:**
   - Delete certain information that is irrelevant or unimportant.

6. **Include current information:**
   - Rewrite text to include more recent information.

III - PRESENTATION STRATEGIES

1. **Make tables more comprehensive:**
   - Make tables comprehensive by displaying patterns and trends clearly.

2. **Integrate table with text:**
   - Table should be situated so that it is closely integrated with the meaning of the narrative.

3. **Make charts:**
   - Where information is complex use a chart to summarize quickly.

4. **Amble spacing:**
   - Use double space and amble margins to make materials easier to read.
5. **Use illustrations:**
   - Include illustrations to make text more interesting.

IV - LANGUAGE STRATEGIES

1 **Explain terminology:**
   - Define all new terms.

2. **Keep language simple:**
   - Avoid jargon, technical or complex terms .
   - Reading level should be appropriate for learners.

3. **Consistent terminology:**
   - Use the same terms throughout the materials.

4. **Conciseness:**
   - Keep information brief and to the point.

5. **Punctutation:**
   - Correct punctuation errors.