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.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

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.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

THE SOLO TAXONOMY APPLIED TO UNDERGRADUATE INSTRUCTION

BY

BRUCE DOUGLAS GALENZA ©

A thesis submitted to the Faculty of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

IN

THE DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1993

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

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-82168-X

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Bruce Douglas Galenza

NAME OF THESIS: The SOLO Taxonomy Applied to Undergraduate Instruction

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1993

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific purposes only.

The author reserves all other publication and other rights in association with the copywrite in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Bruce Douglas Galenza,
44 Sonora Drive,
St. Albert, Alberta.
T8N OL4

February 8, 1993

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "The SOLO Taxonomy Applied To Undergraduate Instruction" submitted by Bruce Douglas Galenza in partial requirements for the degree of Doctor of Philosophy.

Dr. L.L. Stewin, Supervisor

Dr. T.O. Maguire

Dr. R.H. Short

Dr. R.F. Mulcahy

Dr. T.C. Montgomerie

Dr. J.B. Biggs

Dedication

To Louise, Paul, Jon and Anna

Abstract

This study was designed to examine the SOLO Taxonomy of Biggs and Collis in an application to university and college level undergraduate instruction. Its purpose is to examine the SOLO's possible contribution to the field, and to discuss how this new approach fits into the present theoretical perspectives of cognitive psychology and academic instruction. Undergraduate education is fully defined in terms of the goals of the liberal arts curriculum, higher order thought, theories of learning and the students' experience, instructional techniques, and assessment practices. This paper continues with a theoretical evaluation of the SOLO as a valid instrument for instructional design and assessment for undergraduate education. Following this is a presentation of two studies evaluating the effectiveness of this approach compared to traditional methods, in terms of the quality and quantity of learning, as well as its durability of retention. Three college classes of undergraduate psychology were compared, one class receiving instruction with multiple choice examinations only, one with short essay examinations devised according to the SOLO partial credit model, and one class with the same short essay examinations paired with higher levels of explanation, examples and feedback concerning the SOLO. The general conclusion was that use of the SOLO taxonomy results in better student performance on examinations, especially for students of mid-range academic ability, and most markedly on long term retention of course material. This paper concludes with an indepth analysis of the levels of responses that students had attained, and an overall evaluation of the SOLO as an educational tool in undergraduate education.

Table of Contents

List of Tables

List of Figures

I. INTRODUCTION

Biggs and Collis (1982) outline the SOLO Taxonomy (Structure of the Observed Learning Outcome), a model for evaluating student learning in cognitive based academic education. This evaluative attempt comes at a time when the educational psychology field of assessment is seriously questioning the purposes, uses and validity of academic achievement tests, and journals are devoting entire issues to the problems (Educational Researcher, 1989). The central question concerns a valid operational definition of academic learning (Frederickson & Collins, 1989; Snow, 1989); including the goals of education, methods of instruction to fulfill these goals, and most importantly, the testing of students as an evaluation of both their performance and the efficacy of curriculum and instructional practices. The SOLO Taxonomy promises a workable and valid tool to answer these concerns in its attempt to define learning in terms of the cognitive structuring of knowledge manifested in the verbal structure of essay examination answers.

Preliminary results on the validity and reliability of this model (Biggs & Collis, 1982; Biggs, Holbrook, Ki, Lam, Pong, & Stimpson, 1989; Collis & Davey, 1986) are promising in its application to elementary and secondary school levels. Its initial success suggests the model may be applicable to college and university undergraduate classes as well. Undergraduate education is the "forgotten area" of educational research, with little or no work being done on goal and curriculum development, instructional, or assessment practices. Little is known of the extent to which post-secondary education in the liberal arts is fulfilling its mandate, and what, indeed, that mandate is.

This thesis discusses and evaluates the SOLO Taxonomy in the context of undergraduate education to examine its possible contribution to the field. The SOLO Taxonomy is evaluated from three perspectives: how this new approach fits into the present theoretical perspectives concerning cognitive psychology, its contribution to and agreement with models of instructional design and classroom instruction, and finally from an empirical perspective of its efficacy in the classroom. An overview of the field of undergraduate education is first presented in order to define the terms and concepts that will be used in the analysis of the SOLO Taxonomy, including the goals of a liberal arts education, definitions of higher order thought, learning theory, instructional methods and assessment techniques. Following this is a brief explanation of the SOLO Taxonomy itself, and a theoretical analysis of how the SOLO Taxonomy's function is based in cognitive theory. Finally, an empirical study that applies the SOLO taxonomy's technique of assessment and secondarily, instruction, to undergraduate level psychology courses is presented. The purposes of this study are three-fold: to discover if the SOLO Taxonomy is applicable to a higher level of cognitive-based education than is currently being tested, to further evaluate the taxonomy as a model for academic learning, and to provide a tool by which post-secondary education can be investigated, evaluated and improved. In light of the empirical results, an analysis and interpretation of the SOLO Taxonomy's function as an educational tool is discussed.

II. LIBERAL ARTS EDUCATION

A. GOALS OF THE LIBERAL ARTS CURRICULUM

The goals of a liberal arts education have historically been the development of intellectual powers and the transmission of a broad, deep and organized body of knowledge concerning mankind, the world, our cultures, our societies, and ourselves (Elias & Merriam, 1980; Langenbach, 1988). The liberal arts curriculum traces its roots to the first educational programs of ancient Greece which have formed the basis of undergraduate liberal arts education until the early part of the twentieth century (Hutchins, 1952). While it is emphatically not training for a particular job or career, it emphasizes personal development of the young in terms of intellectual excellence, knowledge and wisdom basic to the fulfillment of occupations and citizenship. It has traditionally formed the basis for later training specific to law, medicine and graduate work in specialized areas.

The liberal arts curriculum seeks to begin the process of lifelong learning in its students, and to transmit the values and provide the impetus for this learning. It seeks to develop critical thought and rational judgement in its students through exposure to the humanities: philosophy, literature, history and the social and natural sciences. Through rigorous intellectual training in that which is true, wise, virtuous and good, the liberal arts curriculum seeks to sustain and perpetuate the highest cultural values and traditions of that which is best in our societies. This knowledge is embodied in classic works such as the Great Books of the Western World reading program (Hutchins, 1952) and the Paediea model (Adler, 1982; Adler & Van Doren, 1984).

At the basis of these high-minded goals is the development of "intellectual powers" that are generally articulated in broad, general terms that have a great deal of intuitive appeal but little specific definition. A loose and comprehensive definition by Norris (1989) states "critical thinkers are disposed to seek reasons, try to be well informed, use credible sources and mention them, look for alternatives, consider seriously points of view other than their own, withhold judgement when evidence and reasons are insufficient, [and] seek as much precision as the subject permits (p 22)". Percy and Salter (1976) define "excellence of thought" as objectivity, reaching conclusions based on data rather than personal feelings, the abilities of independent analysis and criticism, the abilities to differentiate the relevant from the irrelevant and the self from the subject matter, and the abilites of logical thought inferences and skeptical approaches to accepted answers and conclusions. Wales and Stager (1977) see the educated person as one who acquires knowledge, thinks for oneself, gathers information for intelligent decisions and is capable of value judgements, among others.

It is, however, still far from clear what "critical thought," "rational judgement" or "intellectual excellence" are; hence the many problems with how to teach this and how to evaluate whether or not it is being transmitted to and learned by the students. It is difficult to grasp what exactly the educated person is doing when thinking, and how these processes differ from one who has not been educated. If the goal of a liberal education is to develop these intellectual abilities, currently termed "higher order thought" by cognitive psychology, it is necessary for the educator to operationally define what that capability is, so that the definition can direct the planning, delivery and assessment of the educational program.

B. HIGHER ORDER THOUGHT

The purposes of the following discussion of higher order thought are two-fold: to articulate the goals of undergraduate liberal education in terms of cognitive theory, and to define concepts of higher order knowledge such that the SOLO Taxonomy's contribution to student learning can be fully evaluated.

From the above definitions of the excellence of thought of an educated, critical and rational thinker, it is possible to differentiate four basic elements of higher order thought as it applies to academic education; these are: the declarative, procedural, metacognitive and attitudinal elements of higher order thinking. Briefly, declarative knowledge is organized and structured world knowledge, knowledge of facts, relationships, principles, theories and perpectives. Procedural knowledge is the activity or skill of thought, the use of cognitive operations and strategies involving and manipulating the information of declarative knowledge. It includes such procedures as logical thought, critical thinking, problem solving and decision making. Metacognitive thought is the knowledge and executive control of these procedural operations, the planful use and goal-directed application to the appropriate situation or problem. The attitudinal element is the motivation to use the appropriate strategies, the desire for truth or adequate solutions, the search for precision and wider perspectives. In other words, excellence of thought involves the what, how, when and why of higher order thinking (McKeachie, Pintrich, Lin & Smith, 1986). This is not to suggest that these are separate areas of thought, as in all cases the four elements are intrinsically intertwined in their applications. This paper will, however, look at each of the areas in turn.

1. DECLARATIVE KNOWLEDGE

As defined above, declarative knowledge is organized and structured world knowledge such as data, facts, and information; the "what" of any thought processes. It is generally recognized in the field of the analysis of thinking that the process of thought always involves content, and the higher order procedural operations involve declarative knowledge specific to the appropriate domain (Bransford, Sherwood, Vye & Rieser, 1986; Ennis, 1989; Glaser, 1984; Greeno, 1980; Greeno, 1989; Perkins & Salomon, 1989). It is further apparent that the efficacy of higher order thought is based upon the structuring of knowledge and its retrievability; as will be discussed later, the differential abilities of experts over novices in a particular field depends primarily on their organizational abilites of that field's knowledge (Bransford et al, 1986; de Groot, 1965). It is appropriate, therefore, to begin by defining models of knowledge structures before defining procedural thinking operations.

World knowledge is conceptualized by cognitive researchers not as accumulated facts but as organized in a variety of hypothetical mental structures. Rosch (1973) and Smith, Rips and Shoben (1974) depict comprehension of category membership and relationships as similarities to prototypes or point by point matching of features of incoming stimuli. For example, we recognize a never-before-seen pet as a "dog" by comparing incoming perceptions to a prototypic idea of what a dog is, on the basis of the necessary characteristics of four legs and a snout, while ignoring optional features such as size and fur length. Collins and Quillian (1969) represent category membership and features as hierarchical structures connected by relational links. For example, the subordinate "robin" would be connected to the superordinate "bird" by an "isa" link, while "bird" has "hasa" links connected to it specifying that a bird "hasa" beak and feathers. Later versions (Collins & Loftus, 1975) exchange the hierarchical arrangement for network models, with concepts or "nodes" connected by links of belongingness, similarity, temporal and spatial proximity, and so forth. Bats and birds, in spite of being in different animal domains, are still linked on the basis of shared flight abilities. Anderson and Bower's (1973) ACT networks consist not of interrelated concepts but of propositional statements linked by relationships between agents, actions, objects and so forth. For example, one could hold the knowledge that Anderson and Bower (agents) have developed (action) a model of declarative knowledge (object).

Others see world knowledge represented in terms of schemas, structures that specify in general terms the relationships between concepts, attributes of concepts, category membership and prescribed actions for dealing with everyday objects and events (Bobrow & Norman, 1975; Bransford & Johnson, 1972; Mandler, 1984; Neisser, 1976; Rumelhart & Ortony, 1977). Knowledge is built up by schematic instantiation or assimilation of new information into existing schemas, or the accomodation of the schemas into more comprehensive and differentiated structures. Models of schemas will be more fully developed in the last chapter of this paper as a theoretical interpretation of the educational use of the SOLO Taxonomy.

All cognitive models stress the point that learning and comprehension will occur to the extent that incoming information matches or fits into existing knowledge structures (Bartlett, 1958; Bransford & Johnson, 1972; Mandler, 1984). Stimuli that are processed to the point of semantic understanding, that is, related to currently held knowledge, are more likely to be encoded and retrieved than stimuli processed only at surface levels (Craik & Lockhart, 1972). It is also apparent that the extent to which world knowledge is interrelated, that is, the amount of detailed knowledge that is "chunked" into structures, affects procedural abilities. The more that knowledge is interrelated, the more information can be activated in working memory at one time, and the higher the levels of memory and comprehension (Atkinson & Shiffrin, 1968; Collins & Loftus, 1975).

From this brief summary of memory models, several principles can be extracted. Knowledge is not a collection of facts or details, such that the educated person can be conceived as simply having "more" information. The knowledgable person is one who has organized and interrelated information, hypothetically structured in terms of hierarchies, networks, and schemas. These structures organize knowledge around central general themes, specifying essential components and features, category memberships, relations and similarities, and allowable instantiations. Multiple access routes to these structures provide retrievability information to the point where the structures are easily and automatically accessed and applied to novel situations. Efficacy of the procedural aspects of thought depends both on the amount of knowledege to which the operations have access, and the retrievability of that knowledge, which is a function of its organization and structure. The efficacy of the SOLO Taxonomy as an educational tool will partially depend on the extent to which its use induces and promotes students' active structuring of academic course material and the formation of cohesive, organized, and interrelated knowledge bases to which the procedures of higher order thought have easy access.

2. PROCEDURAL THOUGHT

Procedural thought, the "how" of higher order thinking, is most generally defined as a series of cognitive procedures or operations in terms of what it accomplishes; that is, it is seen as a purposeful manipulation of internally represented information in the solving of a problem, completing a task, or arriving at a veridical or useful representation of reality.

The emphasis in the literature is on the functional and pragmatic nature of procedural thought (Perkins & Salomon, 1989) in the movement from an initial problem or ignorance state to a desired goal state of a solution or understanding (Anderson, 1990). This perspective often equates procedural thought with problem solving, although other writers (Beyer,

1987) separate procedural thought into different strategies such as critical thought, conceptualizing, decision making and so forth. Although the divisions are quite arbitrary at best, this paper will define different modes of cognitive operations in terms of trends within the literature, after first defining thought in its most general terms.

Beyer (1987) sees thinking as the mental processes by which individuals make sense out of experience, the internal operations of a search for meaning. He initially separates thinking into operations (cognitive and metacognitive), knowledge of subject areas, and dispositions for using thought, which roughly correspond to the four elements of thought outlined above. Cognitive operations, which have been called procedural above, are separated into skills and strategies; skills being the elementary, low level operations like recalling, comparing and analyzing, and strategies being combinations of these skills in their application for pragmatic uses such as problem solving or decision making. Some definitions of these two levels of thought will be examined shortly.

Another distinction that is receiving a great deal of attention in the field is that of domain-general and domain-specific skills and strategies; that is, whether thinking operations can be learned independently of subject matter and the extent to which these operations will generalize within and across knowledge domains. This argument has its roots in the "faculty psychology" of traditional liberal arts education (Lehman, Lempert & Nisbett, 1988) with the notion that simply exercising thought with difficult learning and thinking tasks like Latin and mathematics would ultimately "build up" a student's "thinking powers" such that thought in any domain would be improved. Thorndike and Woodworth (1901) later demonstrated that learned skills for dealing with one task did not generalize to other tasks, which, for a while, signalled the end of that philosophy of education. The cognitive field first started dealing with the problem of generalizability with Newell and Simon's (1972) attempt to develop an artificial intelligence general problem solver. It quickly became apparent that a broad knowledge base was necessary for problem solving and that general heuristics could not operate effectively as isolated processes. The more generality the program offered, the less its power as a problem solver, and vice versa (Perkins & Salomon, 1989). Glaser (1984) concluded that learning and reasoning skills do not develop as abstract mechanisms of information processing, but as functions intimately tied in with the knowledge base with which they operate. Thinking began to be seen as always an interaction between knowledge structures and cognitive processes (Bransford et al, 1986; Glaser, 1984). Thinking is always "about" something; one does not simply "think" (Ennis, 1989). As such, a knowledge base is always necessary and therefore thinking operations were generally considered to be context-bound. Further research pointed out that strategies learned in one knowledge domain very rarely spontaneously transfer to other domains (Belmont, 1989), and would only do so under the most rigid conditions of being guided, primed, encouraged or specifically ordered by teachers or experimenters.

Further work, however, showed that some skills and strategies can be identified that will generalize across domains, which Bransford et al (1986) call "general strategic" thought. Lehman, Lempert and Nisbett (1988) discovered that students formally trained in application of strategies such as contractual schemas, causal schemas and statistical rules will, to some extent, generalize these strategies to "real life" problems. Nickerson, Perkins and Smith (1985) and Palincsar and Brown (1984) report similar findings, and Herrnstein, Nickerson, de Sanchez and Swerts (1986) report the teaching of some generalizable cognitive skills independent of subject matter. These general strategies, however, are more issues of metacognitive thought and attitudinal dispositions, and will be discussed later in those sections. But the work of Herrnstein and others highlights Beyer's (1987) separation of skills and strategies outlined above. Skills, the elements of thought that make up the more complex strategies, appear to be more independent of a subject domain and more generalizable across domains than strategies like critical thinking or problem solving. Although skills and strategies seem to be the extremes of a continuum rather than distinctive processes, it is /orthwhile to look at these two levels separately.

a. Skills

Beyer (1987) defines skills as lower level thought processes that can be used in any subject domain. At the most basic, this includes processes like associating, recalling, categorizing, perceptual comparing and so forth that require little in the way of training and are used in most content areas. At a more complex level, skills consist of processes of distinguishing, identifying, discriminating, analyzing, synthesizing and evaluating.

Gagne, Briggs and Wager (1988) separate the internal components of learning into three elements: verbal information, intellectual skills, and cognitive strategies for the application of t' ›se skills. While they use these terms differently from how they are defined here, their terms correspond closely to the above defined declarative, procedural and metacognitive elements of thought respectively. Gagne et al emphasize the behavioral aspects of the cognitive skills they purport to be teaching. In both the definition of the goals of education and the assessment of its quality, Gagne goes to great lengths to operationally define the skills that the students will be required to learn. Students will "state" or "identify" the facts and principles of the knowledge domain under consideration, "discriminate" category members from non-members, "abstract" general principles from data, and "apply" rules or algorithms to problems. Other equally well defined skills are identification, demonstration, generation, and choice.

A slightly different perspective at a more complex level is offered by Bloom's (1956) taxonomy of objectives of education. His first level, knowledge, corresponds to declarative or verbal information, consisting not only of subject matter facts but knowledge of principles, abstractions, classifications, categories, rules and methods of inquiry as well; in sum, an organized and structured knowledge base. His other five levels of objectives all have to do with the procedural use of that knowledge, arranged in a hierarchical and cumulative fashion. Comprehension is the ability to interpret or translate information into one's own words and to understand the details of the subject matter in other contexts. Application is the ability to abstract out general principles from course material and to apply them to solve problems in novel situations. Analysis is the ability to break down subject material to its component parts, its relationships and its organization, while discriminating the essential from the irrelevant arguments in support of a thesis. Synthesis is the ability to creatively produce unique ideas, patterns, views and perspectives from the analyzed elements, in effect forming a novel productive rather than reproductive communication. Evaluation is the ability to make judgements about the ideas based on cognitive and empirical reasoning, constrained by the criteria of the particular field of the subject matter.

While Gagne et al (1988) and Bloom (1956) do not take the position that these skills can be taught independent of subject matter, both writers consider these skills to be relatively consistent across subject domains. Students can identify critical issues or perform analyses in any academic area, and the procedures will be, at least, similar. Higher level strategies of critical thought and problem solving are less general, and these strategies are considered to be more particular to their own domain. Further, strategies such as problem solving are made up of the lower level skills. One must first identify a problem as such, discriminate relevant from irrelevant information concerning its solution, and so forth (Beyer, 1987). On this basis, strategies will be considered as a separate topic, again in terms of areas within the field of cognitive research. Later, the SOLO Taxonomy's contribution will be examined in terms of both the lower level skills and the more elaborate procedures of higher level strategies.

b. Strategies

i. Problem Solving

The term "problem solving" is almost interchangeable with "higher order thinking" as used by some writers (Anderson, 1990; Best, 1986; Sternberg, 1982). Congruent with the perspective that higher order thought is purposeful and goal directed and can only be defined in terms of what it accomplishes pragmatically (Glaser, 1984; Perkins & Salomon, 1989), problem solving may indeed be the best operational definition of higher order thought we have, and has probably been the most investigated.

Anderson (1990) defines problem solving as a procedural strategy that consists of a sequence of operators or actions that must be found that will transform some initial problem state to a desired goal state. Glaser (1984) characterizes problem solving as the matching of a well structured problem with extant knowledge structures. Beyer (1987) summarizes the steps as follows. First, one must be aware of a problem. It must be clearly represented, along with factors that may inhibit or augment the solution. A plan must be selected for its solution such as an equation, formula or method, a simpler trial-and-error procedure, or working backwards from an end state. The plan must be implemented and its results evaluated. And last, the procedure must be abstracted and stored in memory for future reference. Bransford and Stein (1984) have developed a similar approach to the definition of problem solving represented by the acronym IDEAL: identify, define, explore, act, look and learn.

Due to the complexities of academic subject domains, most testing of problem solving models have been done with puzzles or simple every-day problems. Anderson (1990) summarizes the field of laboratory results dealing with Hanoi towers, broken checkerboards, orcs and hobbits, pouring water between different size jugs and so forth, and the generalizability of the conclusions to undergraduate subject matter must be questioned. Attempts at analyses of academic subject matter, for example, Greeno's (1976) work with high school geometry, have been unable to define students cognitive processes much beyond the earlier Gestalt conception of unexplainable sudden insight. Glaser (1984) concludes that, at present, our understanding of problem solving procedures is too rudimentary to tackle such a complex area.

Two more fruitful areas of problem solving appear to be the novice-to-expert research and artificial intelligence representations of problem solving. The former tradition began with de Groot (1965) and Chase and Simon (1973) in their investigations of cognitive differences between chess beginners and masters. These differences can be summarized by two major points, one structural and the other procedural. Experts were better at perceiving and comprehending chess piece configurations primarily by virtue of their organization or "chunking" of the relations between pieces; that is, more information can be held in working memory for use by cognitive processes. Further, procedural knowledge of what to do with these configurations differed in numbers of available procedures and in automaticity. The novices had to work through possible moves and consequences piece by piece, while experts could "intuitively" follow sequences with automatic procedures that were no longer open to conscious introspection. It is this work that contributed so much to our current understanding of higher order thought being so dependent on both structured knowledge and trained processes.

The dominant area of problem solving research is that of artificial intelligence, first developed by Newell and Simon (1972). Their original problem solver was designed as domain-free. By using think-aloud protocols of subjects, they were able to identify and simulate heuristic-style searches for solutions to problems. They also eventually concluded that general problem solvers could not be effective without knowledge bases, both for properly representing the problem to be solved and for designing and evaluating effective solutions. Again, their problems were lab puzzles with suspect generalizability. Later models emphasized schematically organized knowledge of the domain as necessary for directing searches for effective problem solving procedures (Hinsley, Hayes & Simon, 1978), which proved to be more successful (Bransford et al, 1986; Glaser, 1984).

ii. Decision Making

Many writers group problem solving and decision making as essentially the same procedural strategy (Wales & Nardi, 1984). Beyer (1987) does draw a distinction, pointing out that problem solving involves arriving at one "correct" solution whereas decision making consists of choosing the "best" alternative from several. Further, problem solving is serial, while decision making is a simultaneous or parallel procedure; problem solving is valueless in the sense that answers are right or wrong rather than good or bad, while decision making is a procedure in which results are qualitatively evaluated. Other than that, the principles of decision making appear to be identical to problem solving. Wales and Stager's (1977) steps to decision making are practically identical, and the effectiveness of the procedure again is seen to be based on an adequate and organized knowledge base specific to the field in which the decisions are being made. The areas of research, however, differ somewhat; rather than artificial puzzles, real world problems are used as material for investigation. This area also has more of an applied slant to it, concentrating its efforts in clinical, medical, business and political applications, and is therefore outside the basic academic scope of this paper.

iii. Critical Thinking and Reasoning

As with most terms in the area of cognitive strategies, "critical thinking" is often used synonomously with higher order thought in general, defining it in very general terms as reasonable, reflective thinking focussed on deciding what to do or to believe (Norris, 1985). Ennis (1989) sees critical thinking as the propensity or skill to engage in a cognitive activity with reflective skepticism. This emphasis on skepticism may occasionally equate critical thought with over-skepticism, giving it a negative connotation similar to fault-finding or carping (Beyer, 1987). More precisely, "critical thinking" is used to describe evaluative processes, cognitive procedures that ask the truth, worth or authenticity of an idea, statement, argument, conclusion, or possibly the end result of a problem solving procedure (Ennis, 1962).

Beyer (1987) does not consider critical thought to be a strategy in the sense that it is not a defined sequence so much as a collection of operations. These operations primarily involve discriminatory processes; distinguishing fact from opinion, truth from falsehood, relevant from irrelevant information, and logical from illogical reasoning. They are also made up of the lower level skills defined above; the analysis of the parts of an argument or hypothesis, the synthesis of these parts to abstract out principles and unstated assumptions, and the evaluation of the internal or external consistency of the argument.

Critical thinking is the ability to produce, follow or evaluate an argument; the making of a claim, the reasoning through of supporting and contradicting evidence, and the denial of alternative claims (Beyer, 1987). It is knowing not only how to question an assertion and what sorts of questions to ask, but also knowing when to question it. A healthy skepticism is required, plus a judgement of what needs to be questioned and when. It is apparent that the proper use of critical thought requires metacognitive judgement, as well as the motivation or desire to achieve precision and truth (Norris, 1989), both of which will be discussed later.

One clear prerequisite for critical thought is the discrimination of the self from the argument; that is, the ability to objectively analyse reasons in the light of evidence or logic rather than on the effect the argument would have on the person who is arguing. Another prerequisite, as in problem solving and decision making, is a strong interrelated knowledge base of the subject area one is critiquing. Ennis (1989) points out that background knowledge is essential for critical thinking in a given domain, for in different fields, different sorts of things count as good reasons. Critical thinking, therefore, does not generalize from domain to domain as it is a different process in each field. The reasoning of philosophy as a deductive method of obtaining truth is very different from the inductive scientific method of psychology, and the deterministic perspective of the natural sciences differs from the probabilistic approach of the social sciences and the contractual approach of law. As Lehman et al (1988) showed, specific training in one reasoning method does not transfer to novel situations where another method is required.

While critical thinking and reasoning may be considered different cognitive strategies, they are intertwined enough to consider them a single type of cognitive procedure, or at least to see reasoning as a special sub-category of critical thought. Logical thought and the scientific method of gathering information follow prescribed orders of procedures which may be considered one special approach to, or an element of, critical thinking's main purpose of arriving at truth or veridical reality representations. Logic and science are both designed as procedures to discriminate evidence from opinion, both provide safeguards against involving the truth seeker subjectively in the process. Most reasoning methods are designed to produce conclusions based on evidence, to define exactly what is supporting and contradictory evidence, and by what criteria this evidence is to be accepted or rejected. If critical thought is to be seen as a collection of operations as Beyer states, methods of reasoning are essential members of that collection.

iv. Algorithms, Models, Analogues, Learning Skills and Heuristics

This section is a catch-all of all other cognitive strategies representative of higher order thought that can properly be called the province of liberal arts education. Algorithms are similar to the methods of logic and science as articulated above in that they are defined, step-by-step procedures for arriving at a definitive answer. The most often mentioned examples of algorithms are arithmetic operations, and therefore could be seen as a subclass of problem solving. Due to the restricted nature of these procedures, they are generally not seen as modes of higher order thought in that they require little in the way of a knowledge base beyond the ability to repeat the procedure in novel situations.

Model building and the formulation of analogies could be considered to be higher order strategies in the sense that they are efficient ways of structuring knowledge in many domains. Model building involves the recognition of similarities between a relatively unknown complex phenomenon and one that is simpler and well known. Through the process of drawing analogies, model builders attempt to describe the unknown phenomenon in the simpler terms of the known one (West, Farmer & Wolff, 1991). The natural sciences make use of models in describing atomic structure and gravity; the social sciences use animals to model human behavior (Domjen, 1987) and computers to model human intelligence (Atkinson & Shiffrin, 1968). Further to understanding complex phenomena, models can provide a common language for communication of ideas, and also generate testable hypotheses. The student who is well versed in modeling new course material in familiar terms and is capable of generating new questions from the model has mastered a valuable learning skill.

The collection of strategies known as learning skills are those cognitive operations involved in acquiring new knowledge. The essence of learning skills, including modeling, is that they associate new subject material with currently held world knowledge; that is, new knowledge is assimilated into previously structured material. This can be done in a variety of ways. Elaborative rehearsal, imaging, semantic processing, pegwords, concept mapping and mnemonics are just a few (Baine, 1986; West et al, 1991). These will be discussed more fully in the section on instruction. An essential part of the definition of what an educated person is is the emphasis on the person's self-educating nature, and familiarity with these strategies contributes to education within and beyond academia.

Heuristics are similar to algorithms in that they are procedural methods of arriving at solutions, decisions or judgements, but they are "rules-of-thumb" rather than prescribed, step-by-step operations. Anderson (1990) points out that there is a progression of stages as one masters a new field from simple declarative knowledge of details to associative relationships and step-by-step procedures to autonomous and automatic functioning. Writers in the novice-to expert field remark that experts eventually become unable to articulate reasons for why they did what they did (de Groot, 1965), other than it simply "felt right" at the time. While this sort of intuitive judgement appears to be something other than cognitive processes, more in the line of affective judgement, some writers have analysed these judgements as schema-based cognitive heuristics that are essentially short-cuts of more complete reasoning skills, and have identified their strengths and weaknesses (Kahneman & Tversky, 1972; Tversky & Kahneman, 1973). The representative heuristic is a "quick-and-dirty" method of reducing a complex judgement to a simple judgement of similarity. For example, if one goes to a new restaurant and identifies it as such from a few similiarites it has in common with all restaurants (i.e.: tables, waiters and menus), one can comfortably assume it is indeed a restaurant and get on with the business of ordering and eating, rather than subjecting the restaurant to a long, drawn out and precise critical analysis concerning its identification. Any person with social knowledge will be familiar with restaurants and the required appropriate behavior. Similarly, the chess expert can draw upon a vast knowledge base of similar chessboard configurations and make an instantaneous and non-critical judgement that is generally correct. The second major heuristic, the availability heuristic, is used for decisions of frequency or probability of events occurring based on the availability to memory retrieval of earlier events. If one has caught a bus on a particular street corner every day for the last week, one can comfortably predict that a bus will arrive at approximately the same time today. If a chess master is playing a duffer who has made a common major error in a previous game, the master can safely look for similar errors to exploit in later games.

Kahneman and Tversky point out, however, that the use of heuristics is only effective in areas of familiar knowledge domains, and their use in novel areas often leads to misuse and errors. The process of stereotyping exemplifies both the heuristics discussed above. A member of a visible minority is identified on the basis of a few similarities to others of the same culture, and to the extent that examples of that minority's behavior is available to memory, this new member quickly has attributed to him or her a list of unwarranted traits. If a social scientist is making non-researched estimates of rates of crime or unemployment, the availability of memories of being mugged or suffering unemployment as a student will color those estimates.

The misuse of heuristics provides a counterpoint to the definition of higher order thought; in order to define what an educated person is, it may be helpful to define what an educated person is not. Glaser (1984) lists commonly observed errors in problem solving and critical thought. They are the failing to observe and use all relevant facts of a problem, failing to proceed in a systematic or step-by-step manner, jumping to conclusions and not checking them, failure to construct a representation of the problem, not defining problem goals, unsystematic information intake, and so on. Quite simply, what he is describing is the misuse of heuristics in a lesser known subject domain, the shortcut methods of critical thought and judgement where a more thorough procedure is called for. An expert in the field could make all these procedural errors and still arrive at a correct solution. To the extent that higher order thought has been defined as proper and systematic procedures for arriving at truth or solutions, the absence of higher order thought is characterized by the over use of nonapplicable heuristics.

It is clear that procedures associated with higher order procedural thought do not arise spontaneously but must be actively and purposefully taught to students (Belmont, 1989), as the natural tendency of thought is to schematically assume rather than systematically examine incoming information (Beyer, 1987). Greeno's (1989) scathing put-down of those untrained in higher order thought sees them as viewing knowledge as "something whose validity depends only on an unanalysed affective response." One of the goals of higher education is to move students encountering novel academic knowledge domains from a heuristically based level of procedural thought to a perspective of employing more proper and exacting systematic procedures. The evaluation of the SOLO Taxonomy will examine to what extent it produces this change.

v. Communication

The last strategy to be discussed here is that of the communication of thought processes, either verbally or in written form. While not generally viewed as a strategy by cognitive theorists, it is certainly one of the goals of the liberal arts curriculum in that students are being trained to continue the transmission of knowledge to the next generation. Communication can be considered as part of the process of analysis and synthesis of ideas in Bloom's terms, as it is the formation of novel, productive ideas. Communication may also be seen as the final product or application of decision making or problem solving strategies. Communication may be seen, alternatively, as a separate strategy in that it is a reflection of all other cognitive skills and strategies; a verbal-behavioral indication of the efficacy of these processes. Higher order thought may be structured by language (Whorf, 1956), or may structure language (Piaget & Inhelder, 1969); it is probably the former in cognitive, language based higher education. But it is conceivable that a student may be capable of higher order thought by any of the above definitions, but is poorly able to communicate either the process or product of this thought. The actual process of organizing, sequencing and translating ideas into a verbal or written communication to define, explain, analyse, synthesize or evaluate ideas for another reader or listener is therefore a cognitive strategy beyond the thought processes that have been discussed to this point, and is therefore considered a complimentary goal to the above cognitive and liberal arts goals.

In summary, procedural thought is the functional and pragmatic application of structured knowledge. Procedural thought covers the range of cognitive operations from the lower level skills like simple association and categorization, through the intermediate procedures such as analysis and synthesis, to the most complex strategies of problem solving, decision making and communication. Simply having knowledge is essentially useless without the procedural abilities of application; only then does education fulfill itself in accomplishment. Procedural thought is also essential in the academic assessment of knowledge. It is only through the use of assessment instruments demanding application of knowledge in procedures of skills, strategies and communications is it possible to assess the levels of structured knowledge and the generalizability of that knowledge that students have attained. Assessment instruments, including the SOLO taxonomy, must therefore be evaluated in terms of how well they demand and reveal structural and procedural knowledge.

3. METACOGNITIVE PROCESSES

Having attained a knowledge base and the cognitive strategies for using that information is one thing, knowing when and where to use them and having the motivation to go through with the extra effort of their use is something else entirely. It has been confirmed that simply having the declarative and procedural knowledge required for complex problem solving does not guarantee that this knowledge will be applied; it must be activated from its inert condition at the appropriate time and place (Bransford et al, 1986). This point is the generalizability or transfer problem that has been discussed above, the difficulty involved in getting learned skills and strategies to be used in novel situations (Belmont, 1989; Ennis, 1989; Lehman et al, 1988; Pintrich Cross, Kozman & McKeachie, 1986). Education in general, and the SOLO Taxonomy specifically, must be evaluated in terms of the extent to which it induces students to systematically transfer procedural strategies to new and appropriate situations.

Metacognition, or the executive control functions (Sternberg, 1985), is defined as knowledge about knowledge (Glass & Holyoak, 1986) or thinking about thinking (Beyer, 1987); knowing what one knows and the limitations of one's knowledge, knowing about the process of learning and what it involves, knowing which strategies are appropriate to which tasks, and so forth. Metacognition is considered to be the operations involved in the direction and control of cognitive processes; the planning, monitoring and assessment of those processes, the selection, sequencing and executing of cognitive operations (Beyer, 1987). Beyond differences in levels of knowledge structures and processes, differences in metacognitive knowledge most of all define how and why experts differ from novices, and what constitutes an educated person.

The processes of metacognition are often conceptually merged with problem solving or critical thinking strategies by many writers. The first step in problem solving is recognizing that a problem exists and that a solution is possible, given the correct selection of a strategy (Beyer, 1987). This awareness could be considered either part of the cognitive problem solving strategy or a separate executive analysis. The same can be said of the evaluation of a solution, or of planful problem representation, or even in the analysis of whether data are supportive of, or contradictory to, a given thesis (Bransford & Stein, 1984). Other writers separate cognitive and metacognitive functions as domain-specific and domain-general processes respectively, or equate metacognitive processes with any systematic critical or higher order thought in general (Greeno, 1989). Others (Glaser, 1984), see all active, goal directed cognitive processes as metacognitive, as opposed to reflexive, unreflected cognitive functions. There is no one clear definition of what this modern-day deus ex machina might be, or whether its differences from cognitive processes are qualitative or simply quantitative. Indeed, some writers simply admit that metacognitive processes are too general to be defined empirically (Greeno, 1989).

One area of research that makes somewhat of a distinction is that of learning strategies. One can have knowledge of the world or one can have knowledge of one's knowledge; that is, have some understanding of what it means to know and to learn. Bruner (1985), for example, identifies a series of models of what a learner might be, from tabula rasa to hypothesis tester, and this might properly be defined as metacognitive knowledge on his part. One can hold a personal epistemological position, or one can define or articulate that position (Greeno, 1989). The "mature learner" (Glaser, 1984) is utilizing metacognitive processes when thinking about what he or she knows and does not know, and has some understanding of what efforts are necessary before one can learn a new subject area, or can accurately predict the outcome of one's performance. The classic example in this field is the study of giving a phone number to a child to be recalled at a later time (Flavell & Wellman, 1977). Younger children will be convinced they will always remember it, but older children, knowing their limitations, will deliberately employ strategies like rehearsal or will simply write it down.

The learning strategy area, therefore, sees metacognitive knowledge as the deliberate application of memory skills to the learning of a new subject, the deliberate learning of problem solving and critical thinking skills with the purpose of having them available for future applications (Belmont, 1989). It is learning to learn, or perhaps learning how to learn (McKeachie et al, 1986). It is planfully reading and writing, not simply at the sentence level of comprehension but with a constant monitoring of whether a passage makes sense or is in agreement with the rest of the text (Pintrich et al, 1986). It is actively reading by questioning main points, clarifying any difficulties, summarizing for the essence and predicting what will come next (Perkins & Salomon, 1989). It is the planful attacking of new material using understood learning strategies; prereading, summarizing, organizing and interrelating. It is noticing when a passage gives contradictory or incomplete information (Bransford et al, 1986), and looking beyond the expressed ideas for the writer's assumptions and biases. It is the systematic and efficient assignment of one's time and cognitive resources. It is monitoring one's performance with questions like "Is this getting me anywhere?", and "Could I be doing it better?" (Perkins & Salomon, 1989). It is the planful use of communication to serve a purpose.

It follows, then, that one of the main goals of an undergraduate education must be the development of metacognitive operations, the self-monitoring, reflective processes that exemplify the higher levels of problem solving, critical thinking and self-educating abilities. If cognitive strategies are to have any generalizability to novel or everyday contexts at all, it will be to the extent that executive processes are developed to the point that adequate judgements can be made as to their applicability. Metacognitive processes can be equated, at their most general level, to judgement abilities as defined by the liberal arts tradition. It is the necessary conclusion to the acquisition of knowledge structures and processes; the ability to judge where and when this knowledge is to be applied, how one is to do so, and why it should be applied.

4. ATTITUDINAL CONSIDERATIONS

As cognitive psychology moves further away from the deterministic perspective that marked its earlier years, more attention is being paid to the attitudinal aspects of the student. It is being recognized that even if the student possesses a large body of structured knowledge along with the attendant cognitive processes to manipulate this knowledge and the metacognitive judgement of when it is to be applied, it still does not follow that the processes will be applied optimally. This is another perspective on the generalizability problem; in this case, processes do not transfer not because the student lacks knowledge of when strategies are to be applied, but because the student simply doesn't want to go to all the extra effort. The fourth cognitive goal of undergraduate education is therefore the end result of producing students that are not satisfied with quick and easy heuristic-style answers, but value and seek as much precision and exactitude as a situation will allow (Norris, 1989). Attitude is defined here in terms of Schachter and Singer's (1962) two-factor theory of emotion, which includes an undifferentiated arousal state as well as emphasizing the cognitive components of understanding that emotional state and its causes. In other words, attitudes toward the use of higher order thought in appropriate situations are comprised of a cognitive understanding of its practicality and a value judgement in favor of its use.

The issue of student motivation has been addressed on two different levels. The first, and more common, concerns motivating the student to learn specific class material. To this end, instructional designers have dealt with motivational concepts such as intrinsic-extrinsic motivation, self-worth concepts and cognitive goal formation (McKeachie et al, 1986), as well as various ways to present course material such that it will be adaptable to various student learning styles, termed "aptitude-instructional (or treatment) interactions" (Corno & Snow, 1986; Cronbach & Snow, 1977). These issues concern instructional methods more than educational goal analyses and will not be addressed here. The second level seen in the literature concerns the larger goal of producing students that are motivated not simply to learn what the class has to offer, but are disposed in general to learn in order to effectively apply their knowledge and processes to daily and academic problems. It concerns producing attitudes in students concerning the value of education and higher order thought as an adopted perspective and approach to life.

Returning to the classical goals of a liberal arts education, one of the more important end results is the production of students who are "disposed" to seek reasons, consider other perspectives beyond their own, to withhold judgement until all available information is collected, and to seek alternatives even when satisfactory solutions have been reached (Norris, 1989). Its highest goal is to produce students who are self-educating within and beyond academia (Gow & Kember, 1990). This attitudinal approach is seen by Beyer (1987) as a series of personal characteristics; flexibility of judgement, respect for truth, skepticism of approach to current answers and solutions, curiosity for new information, willingness to examine different approaches and conflicting evidence, and a desire for the widest, most encompassing perspective one can attain. Excellence of thought, therefore, is best characterized not so much as the ability to employ higher order thought processes, but as the demonstrated use of these processes in many situations and the articulated acceptance and valuing of these processes as one's own perspective. The apex of educated thought is not in the attainment of knowledge and cognitive strategies, but in the judgement of when to apply this knowledge and the desire to do so.

C. STUDENT LEARNING

Given the above goals of undergraduate liberal arts education, instructional methods must be designed and employed that optimally contribute to the development of students' cognitive structures and processes, more than a lecture style of simply presenting course material facts and details. How the student is perceived by the instructor dictates to a great degree how instructional practices will proceed; notions of how students learn determine how course material will be presented. Several possible views of students will be examined in this section, as well as theories of classroom learning, both behavioral and cognitive, in order to set the groundwork for an analysis of instructional techniques. The aim of this section is to define this area of research such that the SOLO Taxonomy's place within the field as an instructional tool can be determined.

Writers in the area of educational research (Reigeluth, 1983; Van Patten, Chao, Chun-I & Reigeluth, 1986) continually place emphasis on the structuring of course material such that it can be cognitively restructured by the students, on the relating of course material to knowledge structures already attained by the students, and on the necessity of active processing of course material on the part of the students. Concepts concerning the way course material is best structured begin with Bruner's (1985) survey of historical conceptualizations of what a learner is, and how the learning process proceeds. The first of Bruner's epistemologies is the student as "tabula rasa," the blank slate as seen by writers from Aristotle to Watson. All knowledge is built up from sensory impressions, from simple to complex ideas, reflecting the order in the real world. Implications for educational practices would involve a careful ordering and interrelating of course material from simple to complex concepts, and allowing the student to experience the association of related ideas. A more active, participatory epistemology is that of the student as "hypothesis generator;" students are seen as naive researchers who hold particular "theories" based on partial evidence from limited experience, a procedure reminiscient of the heuristic processes discussed in the last chapter. This view, held by writers from Acquinas to Tolman and Dewey, sees the student as an active and intentional searcher for reasons and knowledge, and would be particularly amenable to instruction directed at the gathering of empirical evidence and the construction of more and more elaborate and comprehensive theories. The "nativist" view of Plato through Chomsky and the Information Processing theorists sees the students' minds and behavior shaped as evolutionarily determined structures with natural tendencies toward categorization and organization. An instructor adopting this view would give students the opportunity to use and exercise these innate abilities through the structuring of categories, schemas and scripts. The "constructionist" view of Piaget and the contextualists (Gergen, 1985; Jaeger & Rosnow, 1988; Rosnow & Geogoudi, 1986) holds that the world is not found and reflected so much as constructed; knowledge is created according to a set of rules imposed by social experience. As such, students learn by representing their world at more and more detailed, discriminated and inclusive perspectives. The "novice-to-expert" view shared by artificial intelligence theorists stresses the task performance differences as students master new skills, analyses and rules pertaining to the subject matter. Implications for instruction would involve the exposure to course material and the direction, modelling and practice of abstracting out principles for the creation of rule governed processes. Given the above perspectives, the task for the instructional designer is to choose which model, or models, best describes the learner in the knowledge domain in question. In concert with this chosen view, a perspective on learning must be developed.

Cognitive theories and educational practices have clearly demonstrated that passively reading or listening to course material is a poor way to retain information. Craik and Lockhart's (1972) seminal study had subjects either counting vowels in presented target words (visual encoding of the words), forming words that rhymned with the target words (acoustic encoding) or fitting target words into sentences according to their meaning (semantic encoding). Visual and acoustic encoding can be considered a definition of "surface" or sentence level comprehension; semantic encoding is seen as "deep" or meaning level understanding. Subjects were then surprised with an incidental memory test in which they were required to recognize or recall the target words they had seen earlier. For recall, memory was 15%, 17% and 28% respectively for visual, acoustic and semantic encoded words, discrimination memory for recognition was 16%, 48% and 80% respectively. Craik and Lockhart summarized their findings with their Levels of Processing model: information will be retained to the extent that it is actively encoded according to its semantic value, interpreted in terms of extant world knowledge.

Extensive research in schema theory provides similar results (Bobrow & Norman, 1975; Bower, Black & Turner, 1979). New information is retained to the point that it can be related to existing knowledge structures. Information that cannot be so related is poorly encoded or inaccessible to retrieval processes. Bransford and Johnson (1972) present similar findings; information that is organized and sequenced such that it can be related to existing knowledge is retained, unorganized material is encoded only at a surface structure level. Network theories of knowledge structure (Anderson & Bower, 1973; Collins & Loftus, 1975) contain the same premise; knowledge is accessible to the extent that it is related to existing concepts in an organized fashion. The implications of these findings and theories to educational practices are clear: if students are to retain course information at a comprehension level, it must be actively encoded, organized and structured by the student, and related to what the student already knows (Reigeluth, 1983).

Behavioral learning theories have thoroughly laid out the environmental conditions that facilitate learning. Learning is defined as behavioral change rather than in terms of "latent learning," consequently responses must be elicited and differentially reinforced. Simply reading and studying are seen as essentially becoming familiar with the discriminative stimuli, and little learning occurs in terms of behavioral change. Examinations must be given frequently to encourage distributed rather than massed studying behavior, a more efficient way to facilitate learning (Dempster, 1987). Further, it is understood that more frequent examinations generally raise performance levels (Bangert-Drowns, Kulik & Kulik, 1988, cited in Crooks, 1988; Hopkins, Stanley & Hopkins, 1991). Examinations must test deeper comprehension rather than recognition of factual detail, so recall or comprehension studying will be reinforced while recognition studying is on extinction. Feedback from examinations should be extensive so as to allow students to discover what they do and don't know (Kulhavy, 1977). Examination marks should be posted quickly to maximize the reinforcement value (Kulik & Kulik, 1988).

All of the above perspectives share the idea of the necessity of student activity and corrective feedback. Whether cognitively restructuring course material or practicing responses in terms of assignments or examinations, the emphasis is on the work of the student, quantitatively and qualitatively. Instructional methods must be chosen that direct, guide and evaluate students in these efforts.

Another field of inquiry concerning the student suggests an extensive analysis of learner characteristics (Cranton, 1989; Gagne et al, 1988; Knowles, 1984) in terms of age, prior knowledge, past experiences, attention span, IQ, and visual and acoustic acuity, to name a few. This paper's analysis of university or college level education assumes a more homogeneous student population than the above models were designed for, and does not consider these analyses necessary. Further, these models are directed at instruction for adult learners whose individual needs are of primary importance. For undergraduate education, the standards of the institutions and industries generally take precedence over those of the student. Therefore, the emphasis here is not on giving the individual student only what he or she wants or can handle, so much as demanding that the student meet university or industry standards. There is also a great deal of literature dealing with differences in learning "styles" (Dunn, Dunn & Price, 1981; Gregorc, 1982) and brain hemispheric differences (Rennels, 1976; Samples, 1975) which suggest that instructors tailor their lectures to satisfy as many different students as possible. This paper will not deal with these approaches for several reasons. First, the postulated personality differences are generally speculative and poorly defined, and there is no clear evidence showing students with different learning styles or preferences learn more efficiently when taught with different teaching methods (Corno & Snow, 1986; Cronbach & Snow, 1977; Tiedemann, 1989; Tobias, 1989). Second, there is no clear evidence that these differences in students exist in any innate, determined sense (Hiscock & MacKay, 1985), so much as learned cognitive strategies that are modifiable, especially in higher level students (Marton & Saljo, 1976a, 1976b; McKeachie et al, 1986). And again, the goals of higher education are not to present course work to students' particular cognitive strategies and preferences but to develop those strategies to the critical thinking standards of the liberal arts tradition. The goals are not to build up one preferred learning style but to develop several, in order to give the student a wider range of choices of skills based on their applicability.

This is not to say, however, that students are to be ignored by instructors and left to their own devices in the development of those skills. As will be developed later, teaching students how to learn, along with the teaching of the course material, is recommended (McKeachie et al, 1986). Instructors must also thoroughly understand the level of thought at which their students are currently functioning so as to direct their early lecture material at that level to best facilitate cognitive development. Aiming at too high or too low a cognitive level will either frustrate or bore students, at the expense of not meeting the course objectives.

It is also a common sense idea to vary one's teaching styles and methods to keep up student interest and to present the course material in a variety of contexts to facilitate transfer. It is generally suggested that instructors become familiar with their incoming students' modal level of thought when dealing with the subject domain. This can be done with pre-tests early in the course.

It must be understood that, in a new knowledge domain, students will have little structured knowledge at entry, and students will have few reference points to understand new material. As such, they will be "surface processors" of new information (Entwhistle & Ramsden, 1983; Marton & Saljo, 1976a, 1976b), reading and listening at the level of the meaning of the words and factual details, and not yet at the deeper idea comprehension level. This understanding will directly affect the initial level of instructors' lectures and will dictate the amount of structure the instructor must provide (McKeachie et al, 1986); that is, the less the students know of the area, the more structure that must be presented for them. In summary, some writers suggest an analysis of learner characteristics not in terms of "who they are" so much as "what they know;" an analysis of the skills, strategies and background knowledge they possess as they enter the course (Gagne et al, 1988; McKeachie et al, 1986). This analysis can be tied directly to Bruner's models of the learner as a guideline for how the instructor will structure and sequence the course. Assessment methods to fulfill this instructional function, for which the SOLO Taxonomy is a possible candidate, will be discussed in a later section.

D. INSTRUCTIONAL METHODS

The primary concern of instructional methods is how best to present course material to create the desired knowledge structures and processes in students in ways most amenable to principles of student learning. It is a problem of the "sequencing and synthesizing" of course material (Van Patten et al, 1986); sequencing in terms of in what order the material is to be presented, and synthesizing in terms of how the material is to be related both within itself and to knowledge structures and skills students already possess. Course material must be analyzed so as to identify the elements that make up the principles to be taught, the elements must be ordered according to some defining scheme, and they must be tied together by some encompassing and organizing principle. The primary purpose of the instructional designer is to create an identifiable structure that students can emulate and which will guide lecture presentation such that it becomes readily assimilated into that structure.

The most obvious sequence to use is that employed by the course text, but most writers suggest course presentation using a different organizational scheme, as the more contexts in which the student sees the material, the more memorable it becomes (Light & Carter-Sobell, 1970). Cranton (1989) suggests several ways that information can be sequenced, such as historically, traditionally or practically. Van Patten et al (1986) suggest logical, chronological or psychological orderings; the latter referring to sequences of moving from the known to the unknown, the simple to the complex, concrete examples to abstract principles, and from observation to reasoning. They also suggest organizing course material either from the part to the whole, which Gagne (1968) also advocates, or from the whole to the detailed elements as suggested by Bruner (1966). It is apparent from their analysis of the order of sequence suggested by other writers that there is little consensus or evidence supporting any conclusion of how information should be ordered, as long as it is indeed ordered according to some strongly apparent underlying plan. It may be that more than one organizing and sequencing scheme is more effective in that it gives students multiple ways to structure and therefore understand the material. This multi-structural approach is also advocated by Bruner (1966) with his "spiral" sequence, a periodic recycling of topics with a progressive complexity of presentation. Ausubel (1960) and Gagne (1968) both suggest a hierarchical structuring of course material to match students' natural tendencies of knowledge organization, but Ausubel advocates teaching in a top-down fashion from general superordinate to detailed subordinate concepts, whereas Gagne would sequence material from the bottom up, from the parts to the whole. Reigeluth (1979) also prefers a simple to complex sequencing, but borrows Ausubel's "advance organizer" idea of first presenting the overall concept's most critical ideas in concrete, applicable terms and then rebuilding back up to them in a bottom-up manner. There is some evidence (Donald, 1983) on the effectiveness of hierarchical representations as an organizing plan, with natural sciences being more amenable to causal relationships while the social sciences and humanities are better representatively linked by similarities. The general lack of agreement in this area, however, suggests that there are far too many variables at work in the classroom to discover or comfortably predict which sequencing scheme is most effective. But it is understood that ordering material according to some identifiable principle is necessary, and the more ways students can interrelate the material within itself and to known information, the more comprehensive and memorable it will be (Bransford, 1979; Van Patten et al, 1986). It appears to be the case that the sequencing and synthesizing of material may also be specific to the particular knowledge domain (Schwab, 1962) and may be expected to interact with the instructor's preferences and understanding of the field. In the face of the conflicting ideas in this area, it is probably best to suggest instructors discover the sequence and relating principles with which they are most comfortable, and which, according to assessment instruments, the students find most conducive to learning.

Another view of instructional presentation is conceptualized by Gagne (1977), emphasizing procedural knowledge development as opposed to the above structural views. Knowledge is seen to be built up in a hierarchical and sequential fashion in a series of steps: 1. Simple stimulus-response and response-consequence routines, both behavioral and verbal. 2. The chaining together of multiple responses, what might be called the rote memory of intellectual skills. 3. Multiple discriminations and generalizations of the appropriate situations in which to use the response chains. 4. Conceptual learning, the abstracting out of the essential elements of the concepts. 5. Rule learning, the formalization of the concept or principle. 6. Problem solving, the application of the rule. 7. Signal learning, the metacognitive control of choice of rules to solve problems. Kameenui and Simmons (1990) have adapted Gagne's behavioral hierarchy to the cognitive domain by specifying the steps of learning as simple facts, verbal chains, discriminations, concepts, rule relationships and cognitive strategies. Each step is operationally defined in terms of the operations students can perform at each step in the hierarchy; for example, rule learning is defined as the ability to verbally respond with a statement that specifies the relationship between two or more facts, discriminations or concepts.

A third direction of instruction toward procedural knowledge development would be the above-mentioned taxonomy of Bloom (1956), proceeding from factual detail knowledge through analysis and synthesis of principles to the evaluation of theories. In each of these conceptualizations, complex intellectual behavior is seen as being built up from simpler procedures, and instruction would proceed hierarchically from the simple to the complex.

While the above views of instructional methods emphasize the instructors' roles in contributing to learning, it is never suggested that this is the entire story. Students are currently seen as active processors of information, and these processes must be stimulated before learning can occur. Concurrent with the teaching to students' structures and processes, instructors must also teach students how to learn, both in lecture and in self study. McKeachie et al (1986) consider it a necessity to teach learning strategies prior to and concurrent with class material delivery, as do Gagne et al (1988). They see these strategies primarily in terms of metacognitve abilities, teaching students the principles, natural progressions and limitations of the learning process so that students can planfully apply these to their studying. They specifically point out the metacognitive control processes of active attention, semantic encoding, organization and interrelationships, and the active retrieval of information as being crucial elements students must be taught before effective learning of class material can occur. Weinstein and Mayer (1986) echo these elements in their essential processes of selection, acquisition, construction and integration of course material, as do Pintrich and Johnson (1990). In all cases, the perspective of the student is as an active participant in the learning process, rather than a passive vessel to be filled with information.

Further to the efforts of encouraging students to take a more active role in learning, instructional designers can plan ways to assist students to more deeply encode course material. Wittrock (1990) suggests learning must be a generative process; reading and listening generally become passive, reactive processes which are not conducive to the active construction of knowledge representations. He suggests students make the reading process more like writing, the active, generative constructions of relations of class material. Pintrich and Johnson (1990) encourage instructors to teach their students to paraphrase, summarize, elaborate and question class material, to discuss it in class or among themselves and to suggest possible exam questions. Harrison (1990) would have students translate material into their own words, give analogies, draw distinctions, define the contexts, and ask further questions of the material. Baine (1986) and West, Farmer and Wolff (1991) advocate the use of learning heuristics such as anagrams, imaging, elaborative rehearsal strategies, pegwords, and so forth. In all cases, the point is to encourage the students to be active processors of information and creators of knowledge structures, to interrelate the elements of their knowledge bases and to practice the cognitive skills and intellectual abilities of organizing and applying this information. It is the intent of this paper to investigate the use of the SOLO Taxonomy as an instructional tool, to discover if demanding examination answers in terms of the SOLO Taxonomy's structures throughout the term will encourage students to more actively structure course material for themselves on a final examination.

E. STUDENT ASSESSMENT

The most obvious purpose of student assessment is performance evaluation in the sense of differentiating which students can and cannot continue with further instruction or training. The primary goal as demanded by the educational institution and later graduate schools is stratification of students, separating those who have satisfactorily performed from those who haven't (Heartal, Ferrara, Korpi & Prescott, 1984, as cited in Stiggins, Conklin & Bridgeford, 1986). But assessment has many other crucial functions as well. Hopkins, Stanley and Hopkins (1991) envision testing as a necessary component of a three-part system consisting of educational objectives, learning experiences and evaluative procedures. Each component contributes to, and provides feedback for, the other two components. This systems approach is very much in line with current models of instructional design (Briggs, 1977; Gagne, Briggs & Wager, 1988), who stress that all components of the design process must be considered in relation to each other, with ongoing evaluative feedback between each component. Specifically, this means that examinations must assess whether the stated goals are being met, and feedback from examinations must redefine the goals as to their feasibility. Instruction must be designed to fulfill the stated goals and goals must be restated in light of what is possible through instruction. And examinations must test only principles developed in instruction, while instructional practices must be directed at the examinations. This last approach, sometimes termed "teaching to the test," is suspect and will result in inflated estimates of performance when teaching and examinations deal only with factual information, but it is the recommended approach when goals, instruction and assessment techniques deal with higher cognitive skills of application, analysis, synthesis, and evaluation (McKeachie et al, 1986).

This approach considers the instructor not only as designer but as researcher as well, in terms of the continuing testing and evaluation of the effectiveness of the instructional design through constant monitoring and adaptability of approach in the light of evaluative feedback (Angelo, 1990; Winn, 1990). The field of instructional design is currently placing great emphasis on the development of a theoretical approach to designing educational programs, and we are seeing the emergence of a "science of instruction" (Reigeluth, 1984) that relies on an empirical base of tested and effective delivery systems (Andrews & Goodson, 1980; Dick, 1981; Gagne et al, 1988).

From the liberal arts perspective, the main objective of assessment would be to evaluate and provide feedback to the instructional design process as to whether the course is accomplishing its goals, both at the level of a general liberal arts education and at the level of the cognitive goals specific to the course content. Examinations must be derived directly from the stated course goals in order to assess to what extent students are experiencing the development of a basis of structured knowledge and processes in the particular knowledge domain. Examinations are therefore the measurable behavior or operational definition of learning; as knowledge is not directly observable or measurable it becomes a critical question of what behaviors shall be choosen to test as indicators of that knowledge. These measurements must be valid in the sense that they do indeed reflect higher cognitive structures and processes within the knowledge domain, and must be reliable in the sense that they are accurate and repeatable measures (Messick, 1989).

Tests can also be used as an educational tool; that is, as a further technique for the transference and consolidation of subject content. The purpose of an examination is to direct the students' attention to the important points of the course, to provide a second exposure to the material, to require active processing of the information beyond simple exposure, and to differentially reinforce student responses such that incorrect answers can be replaced by correct ones. Tests can allow students to practice higher cognitive skills. In short, assessment techniques can directly reflect and facilitate the development of those cognitive abilities they are measuring (Frederickson & Collins, 1989). Tests can act as motivators for learning, not only in terms of providing the expectation of evaluation for the student, but in terms of providing the structure and goals to direct student studying. It has long been understood also that examinations are an effective tool in that they enhance long term retention of the course material through overlearning (Hopkins et al, 1991; Jones, 1923).

Assessment is an area that suffers from a great deal of misunderstanding in the instructional field, and as such is seldom optimally used as an educational or evaluative tool. University and college instructors, while generally concerned about their evaluative profficiency, generally have little or no formal training in evaluation (Gullickson, 1984; Stiggins & Bridgeford, 1985). Tests of any sort, essay or objective, are generally of poor quality when developed by teachers without specific training in test construction (Hopkins et al, 1991). Crooks and Collins (1986), on the basis of judges' agreement, found that about 80% of test questions test the lowest form of knowledge according to Bloom's taxonomy; that is, recognition memory for detailed factual knowledge. As Nickerson (1989) concludes, higher educational assessment restricts its testing to declarative or low level procedural knowledge, and does not test quality of thought.

Higher education examinations are often computer scored multiple choice examinations. This is not to say that multiple choice examinations can't test higher levels of cognitive understanding; Hopkins et al (1991) present an impressive selection of multiple choice, matching and short answer questions that assess all levels of cognitive skills. The point is that multiple choice questions can, but generally don't, get at higher order thought, according to Nickerson (1989) and Norris (1989), and the form of communication strategies that they test is certainly impoverished. Instructors often encounter difficulties in writing higher level questions, whereas questions of factual detail are easily constructed. The same is true of essay questions in that marking according to the number of included facts is easy, whereas the analysis of a student's level of cognitive strategies and of qualitative understanding of course content is very difficult. Other reasons (Haertal, 1986) for employing factual questions are the instructors' concerns for the reliability of their tests. Marking is subjective, and people are poor judges of others' knowledge due to ubiquitous heuristics and biases in their judgemental processes (Kahnemann & Tversky, 1972; Tversky & Kahneman, 1973). Hopkins et al (1991) detail the reliability problems of essay style examinations: different standards between markers, different distributions of grades over the rating scales, halo effects, carry-over effects, order effects and so forth. Stiggins and his collegues (Stiggins & Bridgeford, 1985; Stiggins, et al, 1986) present research showing informal observation and mental recordkeeping are the primary assessment techniques used in lower grades, with the result that the assessment of intellectual competence correlates higher with social competence and skills than with academic performance (Erickson, 1977). Rosenthal's (1966) classic study of experimenter bias and demand characteristics suggested that teachers' understanding of student abilities may change both their behavior toward the students and the students' performance. To the extent that higher education instructors interact with their students and employ non-objective assessment techniques, these biases can be expected to be present. Instructors are also often wary of asking difficult questions in that some percentage of the class will experience the failure of being at a loss for any answer at all (Haertal, 1986). And last, the marking time necessary for large classes makes essay examinations impossible, or at least very aversive for the instructor. For whatever reasons, most tests require simple factual "surface" knowledge rather than a "deep" knowledge of actively searching for meaning and underlying principles (Marton & Saljo, 1976a; 1976b). The stated objectives of a course generally deal with goals of analytic and evaluative understanding of course material, but students quickly pick up on the "hidden" curriculum (Elton & Laurillard, 1979), and study accordingly, successfully relying on rote memory to pass the examinations. White and Horwitz (1987, cited in Masters & Misley, 1991) contend that students can succeed in college courses in this way while still entertaining common misconceptions about the subject matter and never really acquiring indepth knowledge of the underlying principles. Entwistle and Ramsden (1983) summarize research showing that the expected type of exam has an effect upon strategies of studying, and it is also apparent that the nature of study strategies affects the nature of what is learned and how it is retained and accessed (Marton & Saljo, 1976a; 1976b). Nickerson (1989) concludes that the type of test sets the tone for instructional methods and student behavior; if low level declarative facts will be tested, that's what teachers teach and students study. Students generally prepare and perform better when expected examinations are of the recall, essay type than if they are recognition style multiple choice (d'Ydewall, Swerts & De Corte, 1983). It is the case, however, that the expected cognitive level of the test, that is, factual versus analytic-evaluative, has more of an effect on preparation and performance than does test format. Hunkins (1969) showed that students who were used to higher order thought questions throughout the year outperformed controls who were used to knowledge level questions when both were tested on a common final exam. All questions were of the multiple choice variety.

To sum up, essay style examinations are generally unreliable, while multiple choice examinations and essay questions asking for lists of facts or details are suspect in terms of their validity. While multiple choice examinations do provide a reliably objective and easy measurement tool, they are seldom demonstratively valid in their evaluative function (Frederickson & Collins, 1989). Frederickson and Collins conclude that objective examinations generally emphasize low level skills, factual knowledge, memorization of procedures and isolated skills. As such, they lack systematic validity in that they do not foster the development of the cognitive skills claimed and desired by higher education.

As factual examinations demand little in the way of complex responding and feedback is delayed, they can't be considered much of a learning experience for the students either. About the best that can be expected of such a test is that its threat is so aversive that it can act as a negative reinforcer for encouraging students to study for it. As such, though, we can expect the emotional response associated with punishment to occur, and we must recognize that the exam is reinforcing other than optimal studying behavior.

We lack adequate tools for assessment, and as such, we do not know if higher education is succeeding in its stated goals. Masters and Mislevy (1991) assert that current testing practices are not in line with current cognitive theory. Knowledge is not a case of either quantitatively "having" or "not having" information, so much as a qualitative broadening of perspectives or progressively developing more and more complete and comprehensive representations of the world. Yet traditional test items are scored on a "right or wrong" criterion, reflecting a concept of knowledge being either present or absent. What is called for, in light of recent cognitive theories, are testing methods that allow for an inference of students' current perspectives and conceptions. The underlying concept held by Masters and Mislevy is of identifiable developmental stages of academic knowledge that may be inferred from gradients of sophistication in student responses. Recognition of these progressive developments will provide feedback and reinforcement for sudents' partial knowledge, and will also serve as a diagnostic indicating in what areas students must be helped. These stages can be analyzed with multiple choice examinations by presenting distractors of common misunderstandings that are representative of students with partial understanding, for which part marks may be awarded. More effective, however, are essay style answers that do not constrain students to given answers, but allow students to provide their own particular levels of understanding for the instructor's analysis. Biggs and Collis' (1982) SOLO Taxonomy provides a possible solution to the testing problems outlined in this section, in line with the recommendations of Masters and Mislevy.

This chapter has presented an overview of the field of undergraduate education in the liberal arts, delineating the field into areas of the goals of higher education, definitions of higher order thought in education, theories of student learning, and perspectives concerning instruction and assess ment. Given this view of undergraduate education, the SOLO Taxonomy will now be evaluated as to its position and possible contribution to the field. The taxonomy itself will first be introduced and discussed, followed by a theoretical analysis of its possible contribution to each of the above defined areas; as a template for the formulation and communication to students of academic goals, as a suggested definition of higher order thought, as an example of the application of cognitive learning theory, and as a tool for both instructional methods and assessment practices. Following will be an experimental analysis of the effectiveness of the SOLO Taxonomy's application to undergraduate psychology classes, and an analysis of student responses to the employment of the SOLO Taxonomy as an educational tool.

III. THE SOLO TAXONOMY

A. INTRODUCTION TO THE SOLO TAXONOMY

Biggs and Collis' (1982) SOLO Taxonomy is purported to quantitatively measure student comprehension of course material by inferring hypothetical cognitive structures from essay type examination responses. Biggs and Collis postulate stages of comprehension, similar to the gradients of sophistication of knowledge and progressive academic development suggested by Masters and Mislevy (1991). These stages are topographically similar to Piaget's stages of cognitive development (Inhelder & Piaget, 1958), but are not age related. That is, any learner of declarative knowledge of any age will show a stage-like progression as he or she acquires information, in terms of how knowledge is cognitively organized and structured, inferred from how essay answers are organized and structured.

The postulated taxonomic stages are expressed in terms of how examination responses are structured; that is, not only in terms of how many facts are included, but how the information in the response is organized, interrelated, and applied. The stages are as follows: 1. Prestructural; answers are non-existent, wrong in the sense of being off topic, or circular in the sense of repeating the question while adding no new information. 2. Unistructural; one single idea is developed with points or details supporting it. 3. Multistructural; several ideas with supporting details are developed separately. 4. Relational; several ideas are integrated, compared and contrasted. 5. Extended abstract; principles are abstracted from related ideas such that they can be criticized, applied to situations beyond the learned data, and can provide the basis for hypothesis generation. Examples illustrating these stages, drawn from introductory psychology material, are presented below.

Question: Why are rats used in psychological research?

Prestructural: Circular: They're used to experiment on.

Wrong: Because they're very intelligent.

Unistructural (one idea):

...used to study learning, for example, operant conditioning where a rat is reinforced for pushing a lever...

Multistructural (several ideas):

...operant conditioning (as above)...or classical conditioning where rats will associate a neutral environmental stimulus with an unconditioned stimulus...

Relational (integrated):

...operant...classical conditioning (as above)... rats and people show similar patterns of operant and classical conditioned behavior change. Therefore rats' behavio[illegible] be generalized to human learning.

Extended Abstract (critical analysis):

> ...generalizability. However, large differences in the number of trials necessary to show learning to criterion may indicate qualitative differences in learning, so generalizability is limited.

It can be seen that not only do the better answers contain more information but they are organized in a progressively and qualitatively better manner. The taxonomy therefore measures both the subject matter content acquired by the student and the process of acquisition as demonstrated by the resulting hypothetical cognitive structure of that content inferred from the structure of the response. The assumption is that answers are not "right" or "wrong," but reflect the progressive development of more and more complete understanding.

A primary question being asked in this evaluative paper concerns the validity of the SOLO Taxonomy as an assessment technique and as a measurable definition of academic learning. Biggs and Collis provide some information on the validity and reliability of their taxonomy. Validity, in terms of measuring quality of learning, is difficult to assess as there are no external criteria to which SOLO Taxonomy measurements can be compared other than subjectively evaluated essay answer examinations or objective multiple choice results. In high school English classes, SOLO Taxonomy measurements correlate with subjective evaluations anywhere from $\underline{r}$ = .4 to .7. However, different examinations within the course correlate at about .6 when both are marked with the SOLO Taxonomy, while the same examinations subjectively evaluated show a correlation of about .45. In other words, the SOLO Taxonomy appears to be the more consistent evaluative tool for quality of learning. Preliminary results in English literature, poetry and creative writing examinations show evaluations according to the SOLO Taxonomy correlate modestly with test battery results measuring IQ, school performance, school achievement, interest, motivation, learning strategies and cognitive abilities, while correlating negatively with rote learning abilities and strategies. In summary, Biggs and Colis conclude that the SOLO Taxonomy is measuring some aspect of school achievement, congruent with teachers' subjective evaluations and objective measures of intelligence, motivation, achievement and organizational abilities. Students who score the higher levels on the SOLO Taxonomy tend to be highly intrinsically motivated students who search for meaning in their academic work and avoid rote learning of facts and details. What Biggs and Collis do not develop is the extent to which the SOLO Taxonomy is grounded in cognitive theory, or the SOLO Taxonomy's function in operationalizing curriculum and instructional goals, task analyses, specifying adequate performances or determining instructional processes. This paper attempts to establish the SOLO Taxonomy's construct validity at these levels.

In terms of reliability, quality of learning outcomes is not a trait but a response, and is therefore sensitive to instruction, motivation and exposure to class material. Consequently, across item and test-retest measures show poor correlations. But the more critical measure of interjudge reliability shows a correlation of about .8, and has been as high as .95 for the English examinations, far higher than subjective quality grading procedures.

Later versions of the SOLO Taxonomy (Biggs et al, 1989; Collis & Davey, 1986) moved away from the single question examinations in an effort to differentiate cognitive levels more objectively. Instead, a "Partial Credit" model was introduced, a hierarchically and cumulatively arranged set of four interrelated questions were asked. The first required only unistructural knowledge to answer, the second multistructural and so on. Analyses revealed high construct validity and internal consistency can be obtained with questions of this sort. It is this Partial Credit Model; also suggested by Masters (1982) and Masters and Mislevy (1991) that will be empirically tested below.

B. THEORETICAL ANALYSIS OF THE SOLO TAXONOMY

As mentioned earlier, the field of instructional design currently defines the goals of academic instruction in cognitive-behavioral terms (Case & Bereiter, 1984). This is the perspective that emphasises the active nature of the students' internal cognitive processes, specifically the organization and structuring of knowledge and the procedural skills of use and application of that information. Yet it is recognized that knowledge is not a question of "having" something, so much as "doing" something one could not without this knowledge (Kidd, 1988). Hence, the emphasis of student evaluation is on measures of behavioral change that reflect cognitive changes, or from which one can infer the development of cognitive structures and processes.

The SOLO Taxonomy begins from this perspective, as its analysis is of the observable behavior of structuring essay answers. Biggs and Collis (1982) suggests the written structure may be used to infer "hypothetical cognitive structures;" yet go no further in theorizing what form these structures may take, nor do they examine how studying for and writing SOLO Taxonomy answers may contribute to the creation and development of these structures. The SOLO Taxonomy is, however, entirely amenable to cognitive models of schematic memory organization, and as such, suggests it has a strong theoretical base. Its construct validity lies in its assumptions, drawn from cognitive research psychology, that knowledge attainment is a progressive development of widening perspectives, built upon structured knowledge representations with attendant procedural abilities of use.

Biggs and Collis' (1982) experimental results show the

SOLO Taxonomy is effective when used as an instructional tool, but an interpretation is necessary to model what the SOLO Taxonomy is doing, in terms of cognitive theory. The following perspective is offered. Schemas are conceptualized as being knowledge structures of generic information that specify the presence and relationship of objects, events, people and concepts in well-learned domains (Bobrow & Norman, 1975; Brandsford & Johnson, 1972; Mandler, 1984; Neisser, 1976, Rumelhart & Ortony, 1977). As such, they are considered to be the basis of perception and understanding (Neisser, 1976; Rummelhart, 1980), action decisions (Bower, Black & Turner, 1979; Schank & Abelson, 1977), and of memory for past experiences (Mandler, 1978; 1984). Schemas consist of concepts and procedures that have been perceived through repeated experience as spatially and temporally contiguous.

The most popular example for illustrating the concept of a schema is the "restaurant script" (Bower, Black & Turner, 1979; Graesser, Gordon & Sawyer, 1979; Minsky, 1975; Schank & Abelson, 1977). A person visiting a restaurant for the first time with no prior knowledge of restaurants will have no organized knowledge or expectations upon first entering. Such a person will be faced with the cognitive task of an entirely "bottom-up" or data-driven nature; the processing of all details, events and objects in a non-discriminatory fashion. Little will be understood, except to the extent that objects, events or people match similar experiences in different settings. The person will not have ideas of appropriate behavior or reactions to others' behavior. When asked to recall this experience at a later time, the person may recite series of unrelated occurrences, without communicating the essence of what a restaurant is or its purpose.

Upon visiting several different restaurants, the naive person will begin to abstract out the essential characteristics and typical occurances that are present in all restaurants. One always exchanges money for food, one generally sits and reads the menu, one is usually approached by a waiter or waitress, and it is optional to leave a tip when leaving. This generic information is ordered sequentially, along with scripted behaviors of how one is to behave, in general, at each juncture; that is, one must order, but the details of what is to be ordered are not specified. A range of possible or acceptable specifics is joined to the generic information slots depending on their typicality of occurance in other settings. The table may be metal, wood or plastic but never paper; tips may range from 10 - 15% or even higher depending on the service; one orders hamburgers at a fast food outlet but never at an expensive restaurant. Once this schema is developed, the person will have expectations of, and ready behavior for, any new restaurant excursion. The perception during a new visit will be primarily "top-down" or hypothesis driven, and behavior will become more automatic. The detailed experiences of the visit, termed the instantiations of the schema's generic slots, will be stored relative to the schema's organization, in effect organizing the autobiographical memory. When asked to define the purposes or essential characteristics of restaurants, the person is able to do so from the stored sequence of generic information. When asked to recall a particular visit, the schema will guide recall and the organization of specific occuring details, even to the point where details can be chosen or created from the ranges of possible instantiations to fill in, by default, details that were not specifically retrieved. As the person becomes more familiar with restaurants and behavior becomes more automatic, the schema dealing with restaurants may become embedded in more encompassing schemas, for example, a schema for "going out on a date," in which case the restaurant script becomes another slot for an event during a date.

The concept of schematic processing has been applied to comprehension and recall of stories (Bartlett, 1932), pictures of scenes (Brewer & Teyens, 1981; Mandler & Johnson, 1976), people and behavior (Fiske & Taylor, 1984, Hastie, 1981), written text (Bower, Black & Turner, 1979; Schank & Abelson, 1977) and other domains, and the robustness of the concept suggests it may be useful in conceptualizing learning in an academic knowledge domain as well. A student of introductory psychology, when first exposed to the topic of the scientific method applied to the investigation of human behavior, will perceive, understand and recall this information as a disjoint mass of details. Processing is again bottom-up or data-driven as the student has no expectations and can make no inferences. As much detail as possible is processed, with no discrimination as to what is essential to the method or what is simply optional. Students must cognitively process this information through rote rehearsal memorization or simple elaborating heuristics. To the extent that texts or lectures point out the essential characteristics and their relations to each other, the student will begin to construct a schematic representation of the scientific method's critical points and procedures. Continual exposure to examples of their use will allow the student to abstract out the commonalities and defining characteristics; a hypothesis is always generated, ways of collecting data are determined, controlled procedures are designed, the hypothesis is tested, results are analyzed and a conclusion is derived. The student will arrange the concepts along a continuum of essential, usual, optional and arbitrary occurences concerning data collection and hypothesis testing. Generic informatinis ordered sequentially and the student will construct a range of possible and allowable variations and instantiations in each step; for example, which statistical tests are amenable to describing data and inferring effects. Once the schema is adequately represented, it becomes available for understanding novel examples of behavioral studies and for recalling and analysing earlier studies, or for inferring elements from studies that were not specifically stated. Students will develop expectations of, and analytic procedures for, exposure to any new research or empirical problem. Perceptions of empirical problems become top-down, and analysis becomes more automatic. Writing about the scientific method and having feedback provided will serve as extensions to the experience of reading about other studies, not only expanding and elaborating the scientific method schema but adding the procedural or scripted behavior elements to it. As the schema becomes more automaticized, it can become embedded in larger, more encompassing schemas of the variety of methods used in measuring, describing, predicting and controlling human behavior.

This brief description of the development of schematic representation coincides directly with the SOLO Taxonomy's definition of levels of learning. A prestructured answer of unrelated details will be offered by the pre-schematic student who has yet to coalesce the details into a related sequence or to abstract out the essential characteristics. One organized schema concerning the course material would offer a unistructural answer, several schemas would allow a multistructural answer of applying knowledge of the scientific method to two or more examples, and then on to a relational answer of comparing and contrasting their use in different settings. As schemas become embedded as instantiations of more elaborate schemas, an extended abstract answer becomes possible, as the student would now be capable of generating the appropriate steps of the method to answer specific questions about investigating human behavior, or for generating new hypotheses.

IV. EMPIRICAL ASSESSMENT OF THE SOLO TAXONOMY

A. PILOT STUDY

While the SOLO Taxonomy has a great deal of intuitive appeal for designing classroom instruction and evaluating student learning, it has still not been adequately tested and no results are available on its use in university or college undergraduate classes. It was proposed, then, that its efficiency be evaluated in this area, specifically that of introductory psychology courses. The first question to be answered was whether the SOLO Taxonomy was amenable to this domain. It was piloted at Blue Quill College in St. Paul, a college under the auspices of Grant MacEwan Community College in Edmonton, Alberta. The program at Blue Quill is designed for a diploma in social work for adult students. The student sample consisted of 10 native Indian students enrolled in the second half of an introductory psychology course. Several had failed the first half but had been allowed to continue. Few of the students had completed high school. Unstructured questions had indicated that all but two students had practically no understanding or memory of any course material from the first half. If a course based on SOLO Taxonomy instruction and evaluation could be effective here, it would suggest it is a very powerful teaching tool.

Students received lecture instruction in a casual seminar style in a three hour block each week and were assigned complementary text readings. Each following week, they were given an hour quiz of basic concepts. Students had been given explanations of the SOLO Taxonomy criteria for marking, and each exam was taken up in class immediately following the writing period, continually re-emphasizing the organization of the material and providing them feedback of where they stood in terms of class demands. Throughout the term, marks according to the SOLO Taxonomy scheme of marking went up dramatically for many of the students. While no formal analyses of results were conducted, the subjective impression was that students, while upset at the amount of work required of them, took away a significant amount of structured knowledge of basic psychological principles.

It was proposed that a formal testing and evaluation of the SOLO Taxonomy be conducted with a more representative sample of undergraduate students. The question being asked is whether the use of the SOLO Taxonomy as an instructional technique in terms of exam demands encourages better comprehension and retention of undergraduate psychology course material.

B. STUDY #1

1. Introduction

Two sections of PSYCO 201 from Camrose Lutheran University College, fall term, 1990, provided the students. This course is the first half of introductory psychology, dealing with scientific methods, statistics, neurophysiology, sensation, perception, consciousness, learning, memory and cognition. Enrollment in the two sections was not random so the design was quasi-experimental. The college administration did, however, assign students to the two classes concurrently in an effort to keep the enrollment numbers equal. The classes met at 8:00 and 9:00 A.M. on Mondays, Wednesdays and Fridays; a coin toss determined the earlier class (N=50) to be the experimental group and the later class (N=59) to be the control. The larger number in the later class suggests some bias may be present in terms of the more motivated students signing up for a class earlier in the registration period and being spared the earlier start. Drop out rates were high for the earlier class as well, other students were dropped from the final analysis for reasons such as missed examinations, suspicion of cheating on examinations and deaths in the family during exam week, leaving an experimental group of N=36 and a control of N=48.

The independent variable was method of instruction and evaluation. Both sections received traditional lecture-seminar instruction for 3 hours per week for 14 weeks, with four, hour-long examinations throughout the term. The control group received only multiple choice examinations while the experimental group had examinations made up of multiple choice questions randomly chosen from the comparison groups's exam plus sets of four partial credit questions as suggested by the SOLO Taxonomy. The questions from the first exam dealing with the introductory chapter are listed below:

1. List the five basic models or perspectives of psychology.

2. List and explain the four goals of research psychology.

3. Give three basic assumptions of the cognitive model that make it different from all the other models.

4. Why are there different models or perspectives in the field of psychology?

This partial credit series purposefully leads the student through the hierarchy of structures as defined by the SOLO Taxonomy. The first question is unistructural; it asks for a single idea or structure concerning psychological perspectives with the attendant details of simply naming those perspectives. The second question demands a multistructural answer of up to four separate ideas to be listed, each with attendent facts that will be developed in the "explanation" request of the question. The third is a relational question; not only must the student present multistructural ideas of the cognitive and other models, but the ideas must be compared and contrasted to each other. The fourth question asks for extended abstract information that goes beyond the student's text material, asking for a synthesis of that material in order to abstract out the epistemological position of the science of psychology in its attempt to formulate its knowledge of human behavior in terms of models. It was expected that the use of this partial credit approach in all of the four quizzes, plus full and ongoing explanations of what was expected of the students, would encourage them to approach the course material in that manner; that is, by organizing ideas, comparing and contrasting them to each other and attempting to abstract out the basic principles of psychology's study of human behavior. It was hypothesized that these cognitive structures will be the basis of a better, more indepth understanding of course material that would be reflected in the course assessment measures, outlined below.

The identical multiple choice questions of the first test served as a "pre-test" for both to equate the groups before the experimental manipulation, although it was not a pre-test in the sense that the same test was given before and afte treatment. The dependent variable of the final exam was identical for both sections; that is, half multiple choice and half essay type questions covering the entire course content. All multiple choice questions were selected from an exam file of questions with demonstrated validity and reliability in terms of item analyses, selected by the instructor on the basis of intuitive judgement that they require not simple memorization of detail but some synthesis of information in order to draw conclusions. The essay type format for the final exam consisted of three single, broad questions that were marked according to the SOLO Taxonomy's definition of structure. These questions, and sample student answers, will be presented in the last chapter. It was expected that the experimental group, due to their training in organization and comparison of material ideas, would produce higher level answers on the essay questions and would be able to retrieve more factual details for the multiple choice section of the final exam.

The second dependent variable was performance on a "post-test" of course material, administered six months after the end of class. Other measures of interest of lesser importance were the number of students volunteering for the post test signifying differences in interest in psychological experimentation, and the traditional student evaluation of the instructor and course that follows the end of each class.

2. Procedure

Both classes received a lecture seminar instruction of course material, a structured, interactive style with which the instructor is most comfortable. Efforts were made to keep lecture material consistent across the two sections; if students brought up a peripheral topic for discussion in one class, the instructor would introduce the same topic in the other class. Both classes wrote four hour-long quizzes throughout the term and the following class period was used entirely for discussion and feedback on exam material and performance. The only difference between classes was that the experimental group also received instruction on the SOLO Taxonomy marking scheme during that hour, as articulated above. The cognitive strategies necessary for answering the questions and the criteria by which they would be marked were laid out for the students. Students from both classes were encouraged to spend as much time as they wished going over the examinations during the feedback hour; examinations were not kept by the students beyond the feedback hour as the multiple choice question file was used by other instructors in the college and would be used in following years as well. Few students ever made use of the entire hour in the control group, many of the experimental group used the entire hour.

Before administration of the examinations, the partial credit questions were rated by four judges as to the level of cognitive structure needed to satisfactorily answer them. All questions from the four examinations were constructed by the instructor according to the SOLO Taxonomy, then all questions were randomized and given to three other judges who rated each as unistructural, multistructural, relational or abstract. Percentage agreement between each pair of judges is presented in Table I. Also presented is the percentage agreement of each level of question across judges. At least 3 judges agreed on the designation of 81% of the exam items, although a 4-way agreement was rather hard to obtain at only 36%. A further analysis of these questions is presented in the last chapter.

TABLE I

PERCENTAGE AGREEMENT BETWEEN JUDGES: ALL QUESTIONS

Judges	1	2	3	4
1		69	53	64
2			58	69
3				61

PERCENTAGE AGREEMENT AMONG JUDGES: EACH QUESTION TYPE

Question Type	3-Way Agreement	4-Way Agreement
Uni	100	67
Multi	44	11
Relat	78	44
Ex Abs	58	22
Totals	81	36

3. Results

A marking scheme was devised that would both satisfy the requirements of student performance discrimination for the assignment of marks and provide a measure of performance according to the SOLO Taxonomy stages. Out of a possible 20 marks, answers judged to be unistructural were assigned marks between 1 and 5, multistructural between 6 and 10, relational between 11 and 15, and extended abstract between 16 and 20. Once essay answers were categorized according to levels, each was compared to others of the same category on the basis of number of details or facts presented. This was neces s, for example, two answers both judged to be relational may show a great difference not in the structure of the answer but in the nummber of details supporting that structure.

The pre-test, the common multiple choice portion of the first quiz, showed a significant difference between the two classes. The experimental group averaged 56.6%, the control group averaged 61.5% (F (1,78) = 11.20, p = 0.001). On the final exam, neither the multiple choice (M.C.) nor the essay answer (E.A.) portion showed significant differences between classes. Results are presented in Table II. As can be seen, the groups were not equated on the pre-test, and although the experimental group did marginally better on the final exam, this effect cannot be differentiated from regression to the mean due to sampling error of exam questions on the pre-test.

TABLE II

STUDENT PERFORMANCE IN PERCENTAGE ON PRE-TEST, M.C. AND E.A. PORTIONS OF THE FINAL, AND FINAL EXAM PERCENTAGES, N = 84.

	Pre-test	M.C.	E.A.	Final	N
CON	61.5	59.2	40.3	49.7	48
EXP	56.6	58.5	44.2	51.3	36

It was suspected, however, that the students were not assigned to the groups in an unbiased fashion and consequently the groups were not equal in academic ability. The experimental group showed a high rate of absenteeism and a greater reluctance to join in class discussion. There were no apparent differences in the number of students enrolled in each of the faculties, so it is unknown if the differences were due to a bias in assignment or simply the 8:00 A.M. start of the class. The absenteeism would be expected to decrease the effect of the SOLO Taxonomy instruction. In any case, efforts were made to statistically equate the two classes before analysis. First, the lowest seven experimental students on the pre-test were dropped from the analysis, resulting in equal pre-test means of 61.5%. This had little effect on the final multiple choice portion, but a significant difference was found on the essays, with the experimental group averaging 48.4% compared to the control's 40.3% (F (1,71) = 3.958, p = 0.05). Table III presents the new group means after the removal of these subjects.

TABLE III

PERCENTAGES ON PRE-TEST, M.C. AND E.A. PORTIONS OF THE FINAL, AND FINAL EXAM PERCENTAGES AFTER EQUATING PRE-TEST MEANS, N = 77.

	Pre-test	M.C.	E.A.	Final	N
CON	61.5	59.2	40.3	49.7	48
EXP	61.5	60.5	48.4	54.4	29

The second method of equating the groups was to look not at their absolute scores but at the degree to which students changed in their performance throughout the term in relation to their pre-test marks; that is, the differences between their marks on the pre-test and the final.

Again, as the pre-test and final examination were not the same test, these two tests do not share the same scale, so one would not expect to see increases from pre-test to final that reflect increases in structures of knowledge. A "difference score" is therefore an arbitrary metric in itself, but comparisons of difference scores between groups is indicative of the relative performance of the groups as a result of the different methods of teaching.

It is apparent from Table III above that both classes dropped in terms of absolute percentages, indicating a more difficult final exam, especially in the more stringent marking criteria of the essay portion. Differences in marks are presented in Table IV. The control group dropped by 11.8%, greater than the experimental's drop of 5.2%, a significant difference of F (1,78) = 7.425, p = 0.008. Significant differences were found in both the multiple choice portion (F (1,78) = 4.081, p = .047), and the essay answer portion (F (1,78) = 5.609, p = 0.02); in both cases the experimental group performed better than the controls relative to their entry level abilities as measured by the pre-test.

TABLE IV

PERCENTAGE DIFFERENCES BETWEEN THE PRETEST AND EACH OF M.C., E.A. AND TOTAL SCORES OF THE FINAL, N = 84.

	M.C.	E.A.	FINAL	N
CON	-2.4	-21.2	-11.8	48
EXP	1.9	-12.4	-5.2	36

Further analyses showed the differences between groups were not uniform within the groups. Botn classes were blocked by thirds according to their position within their own group on the pre-test, labelled Top, Middle, and Bottom on Figures 1 through 3. This resulted in three groups of twelve students in the experimental group and three groups of sixteen in the control group. Means for blocks were not equal between groups. However, the experimental group was only about 5% lower than the control group at each level, which would minimize any unequal regression tendencies.

Analyzing the differences between pre-test and final as a 2 X 3 ANOVA showed a non-significant interaction (F (1,78) = 2.476, p = 0.09), and post-hoc tests showed no differences between groups for either the top or bottom thirds of the classes. The middle third of the two classes differed greatly (F (1,78) = 11.211, p = 0.001). This same pattern of mid-range differences was repeated in pre-test to final multiple choice differences (F (1,78) = 12.324, p = 0.001) and pre-test to final essay differences (F (1,78) = 5.408, p = 0.023). As can be seen from the figures, it appears that the difference is due not to the experimental group doing relatively better so much as the control group doing relatively poorly.

These differences must be seen in the context of the top and bottom groups regressing toward their means between the first and last examinations of the course. As these three groups were formed on the basis of their pre-test performance, the regression principle would predict the top and bottom groups would approach the class means, while the middle group should remain relatively constant. This effect is seen in the experimental group pre-test to final exam differences as shown in Table V; all differences between thirds are significant for the entire final exam and for the multiple choice portion, and the same non-significant trend is apparent in the short answer portion. The control group shows the top and bottom thirds regressing at nearly the same levels as the experimental group, slightly less as would be expected from the unequal pre-test means. But the difference scores of the middle thirds shows the experimental group outperforming the control group on all three measures.

TABLE V

F - RATIOS AND PROBABILITIES OF COMPARISONS OF BLOCKED THIRDS DIFFERENCES WITHIN GROUPS

		EXP	CON
	Top vs Mid	7.40	0.60
Pre-test - M.C.		(0.01)	(0.44)
	Mid vs Bot	3.66	34.70
		(0.06)	(0.001)
	Top vs Mid	2.17	0.06
Pre-test - E.A.		(0.14)	(0.80)
	Mid vs Bot	2.32	8.47
		(0.13)	(0.005)
	Top vs Mid	5.48	0.29
Pre-test - Final		(0.02)	(0.59)
	Mid vs Bot	4.09	23.51
		(0.05)	(0.001)

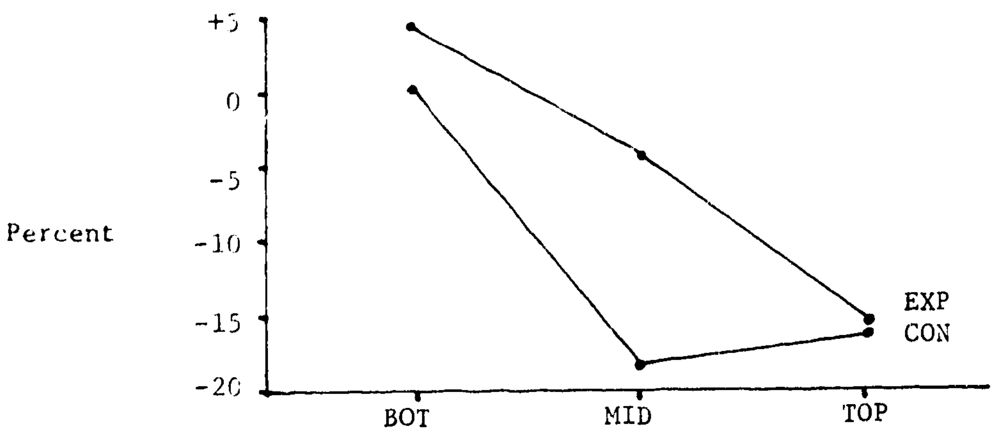

FIGURE 1. Differences: First Exam to Final Total

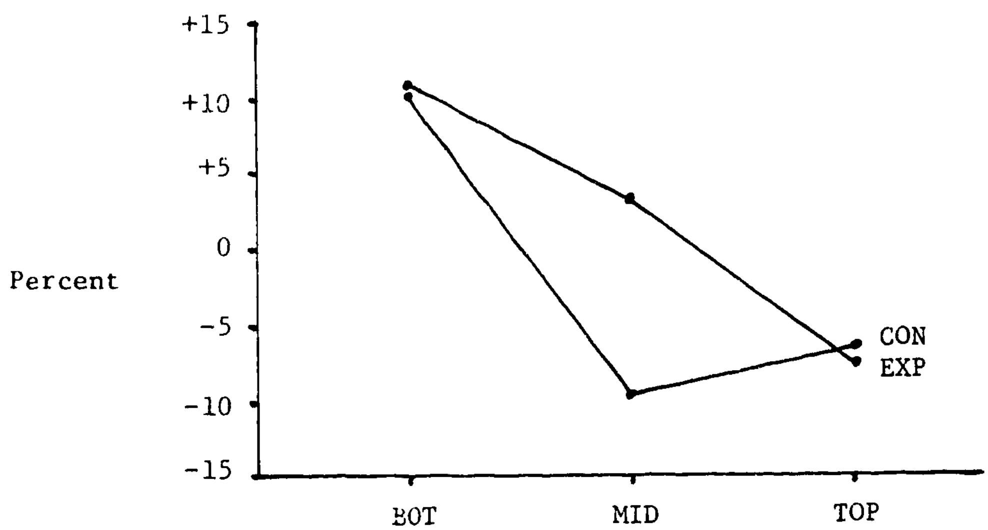

FIGURE 2. Differences: First Exam to Final Multiple Choice

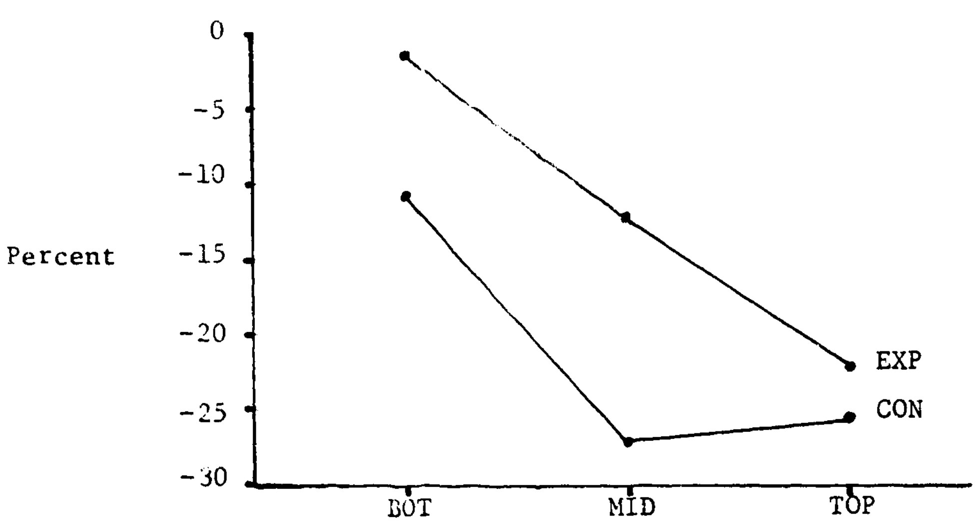

FIGURE 3. Differences: First Exam to Final Essay Answer

4. Post-Test

A post-test was conducted six months after the end of the class to examine differences in long term memory of course material between the two teaching methods. Volunteers were solicited for an undefined "experiment," for which they would be paid $15. A significantly larger percentage of the experimental group volunteered, 38.5% versus 20.8% of the control group ($X\ (1)=5.29$, $p < 0.05$). This higher level of interest may have reflected students' feelings of learning more and therefore valuing the subject matter more highly. It was interesting to note that all volunteers from both classes scored between the 45th and 80th percentiles of their respective classes on their final course mark, roughly corresponding to the mid-range groups in Study #1 that showed the greatest differences. All volunteers had continued with the second half of introductory psychology in the interim, all with a common instructor.

Students were contacted by telephone, asked if they were still interested in participating, and a shortened form of the final exam was mailed to them. This exam was subjectively judged by the experimenter to be similar but somewhat easier than the final as the more difficult multiple choice were dropped out while the same essay questions were asked in broader terms to give the students more leeway in their answers. Of those students that could be located and were still willing, 7 of the experimental group and 6 of the control group returned the examinations. Results were tabulated and sent back to the students, with a cheque for $15 and a full debriefing, including results of the study.

The final examinations of the course were averaged within each group for those who returned the post-tests, resulting in a control group mean of 61.5% versus an experimental group mean of 60.1%, a roughly equivalent pairing of volunteers. On the post-test, the control group averaged 35.2%, the experimental group 56.6% ($t(11) = 2.58$, $p = 0.03$). There were no significant differences in the multiple choice portion alone, but the essay answers averaged 23.9% versus 59.5% respectively, significantly different at the 0.05 level.

5. Discussion

Concerning the final exam, after the two groups are equated by either dropping the lower portion of the experimental group or by analyzing the change in exam performance over the term, the experimental group was seen to perform significantly better than the control group. This conclusion, however, is suspect for several reasons. The first is that almost the entire difference is accounted for by the middle third of the two groups. It might be predicted that the top third of both classes scored at that level because they had already mastered some analytical cognitive strategies and the SOLO Taxonomy instruction could not add to the experimental group's abilities beyond that level.

Conversely, the bottom third of the experimental was too ill-prepared or unmotivated to take advantage of the SOLO Taxonomy instruction and didn't benefit either. This suggests that the SOLO Taxonomy training was perfectly tailored to the average student who most needed training in analytic skills and could most benefit from it. This overly optimistic conclusion is tempered by several problems. First, although the differences are statistically significant, there is only a difference of about 5 - 6 % on the final exam between the two groups; not an effect to warrant the changing of instruction techniques. The second problem is that it is apparent that there was a great deal of movement within the rankings of both classes from the pre-test to the final examinations. In both classes, several of the initial top third students scored in the bottom third at the final, and vice versa. Enough of the control group's bottom two thirds switched places such that the bottom third on the first exam averaged a higher score on the final than the middle third did; 47.6% versus 40.9% respectively. The movement in rankings is reflected in the significant but mid-range correlations between the pre-test and the three final exam measurements presented in Table VI.

It is still the case, however, that of the middle thirds of the two classes, half the experimental group moved up to the final top third of their class, while more than half of the control group moved down into the final bottom third. If the movement in relative standings in the bottom two thirds of the control group is normative for this population, it must be concluded that the middle third of the experimental group was able to benefit from the SOLO Taxonomy instruction.

TABLE VI

CORRELATIONS BETWEEN PRE-TESTS AND THE THREE FINAL EXAM MEASURES, WITHIN GROUPS

(All are significant at 0.01 or beyond)

	EXPERIMENTAL GROUP (N = 36)		
	M.C.	E.A.	FINAL
PRE-TEST	.69	.48	.65
M.C.		.49	.77
E.A.			.91

	CONTROL GROUP (N = 48)		
	M.C.	E.A.	FINAL
PRE-TEST	.51	.40	.49
M.C.		.60	.87
E.A.			.90

Over a six month period, the differences in course material retention became more pronounced, with the SOLO Taxonomy group scoring 21.4% higher. The post-test is perhaps the more important dependent measure from an education standpoint, for it is hoped that course material will be retained by students long enough to at least provide a basis for advanced courses, if not to effect a permanent change in students' knowledge and ways of thinking. It is on this measure that the SOLO Taxonomy manipulation had its greatest effect. All students in the post-test study indicated they had gone on to take the second term of introductory psychology with the same professor. It may be surmised that the stronger knowledge base the SOLO Taxonomy students constructed during the first term allowed the second term material to consolidate and elaborate that base. For example, students covered the nature-nurture issue in the first half that was strongly emphasised in the developmental psychology areas in the second term. Also, learning and cognition principles were developed in the first term that are the basis for abnormal and social psychology in the second. A stronger knowledge base of these and other principles from the first term would make second term material more comprehensible, which would in turn consolidate and provide examples for the general principles learned earlier.

C. STUDY #2

1. Introduction

If the SOLO Taxonomy was an effective variable in raising student marks above a normally taught control, it may follow that an increased emphasis on the SOLO Taxonomy may increase marks even more. It was also hypothesized that the experience of teaching the Camrose class may have served to further familiarize the instructor with the SOLO Taxonomy, such that it could be explained more clearly to later classes. An introductory psychology class from Concordia College, spring session 1991, provided a second experimental group. This group of 28 students was treated the same as the first experimental group from Camrose, with a few exceptions. More emphasis was placed on explaining and providing examples of the SOLO Taxonomy's levels of answers. Sample examinations, presented in Appendix A, were given out and thoroughly discussed in an early class period, and more emphasis was placed on presenting course material in terms of structures, relational ties, and evaluations. Lectures periodically made reference to the idea of structuring, relating and evaluating course material in terms of the SOLO Taxonomy. A major difference with this class was that it was held over a three week period at three hours per day, an intensive program that did not allow students much reading and study time. The students were slightly older and possibly more motivated; on the other hand, many had full time jobs and families that interfered with their class work. Six dropped out during the class's duration, leaving 22 subjects in the group.

The same course material and examinations were presented as with the first experimental group. There were a few minor changes such as the wording of exam questions that had been shown to be unclear in the Camrose class. Examinations were the same multiple choice and partial credit combination, the final exam was indentical to the Camrose final.

2. Results

The pre-test of the first exam's multiple choice portion averaged 62.7%; dropping the top student equated the Concordia group to the two Camrose groups at 61.5%. An ANOVA was performed on the final examination scores of the 3 groups crossed with the 3 levels as indicated by the pre-test, resulting in a significant difference between groups of F (2,89) = 6.95, p = 0.002. Means are presented in Table VII, while difference scores for each of M.C., E.A. and final examinations, blocked by thirds, are plotted in Figures 4 through 6. The second experimental group outperformed both Camrose groups on the final examination at all three pre-test levels, but again only the middle level differences were significant. The control versus the second experimental group was F (1,89) = 14.88, $p < 0.001$; the first experimental group versus the second experimental group was F (1,89) = 4.31, p = 0.04. This again supports the suggestion from Study #1 that the mid-range students were most likely to benefit from the SOLO Taxonomy manipulations, and also from the added emphasis on the SOLO Taxonomy in the second experimental group, such that they outperformed the middle level of the first experimental group.

TABLE VII

FINAL EXAM MEANS, ALL THREE GROUPS

	M.C.	E.A.	FINAL
CONTROL	59.2	40.3	49.7
EXP1	60.5	48.4	54.4
EXP2	75.0	47.8	61.6

Further analysis showed the differences in the final examinations were primarily due to the multiple choice portion, where the second experimental group outperformed the two Camrose groups at all three pre-test levels. In the essay portion, the second experimental group showed a non-significant improvement over both Camrose groups at the top and middle levels, but were non-significantly lower than the Camrose experimental group at the bottom level.

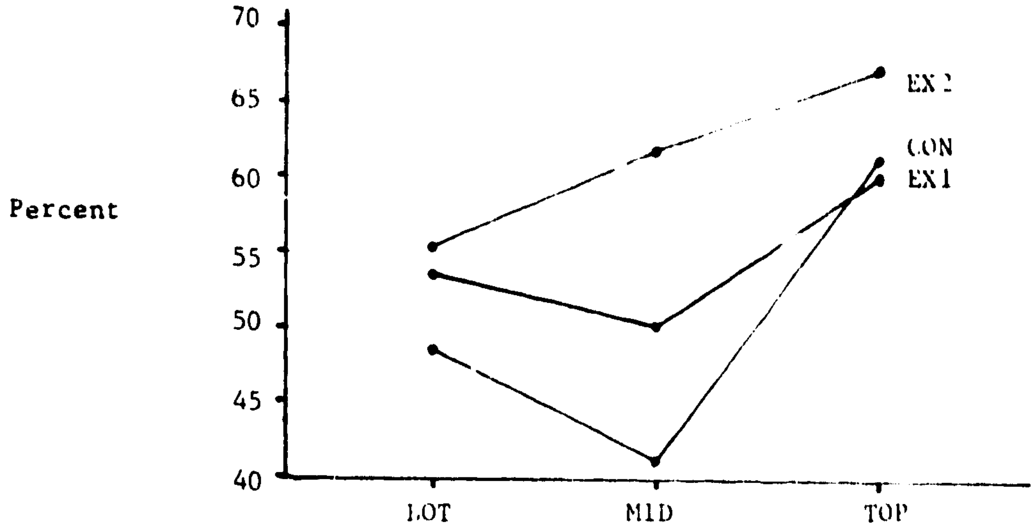

FIGURE 4. Final Exam Marks: Groups by Levels

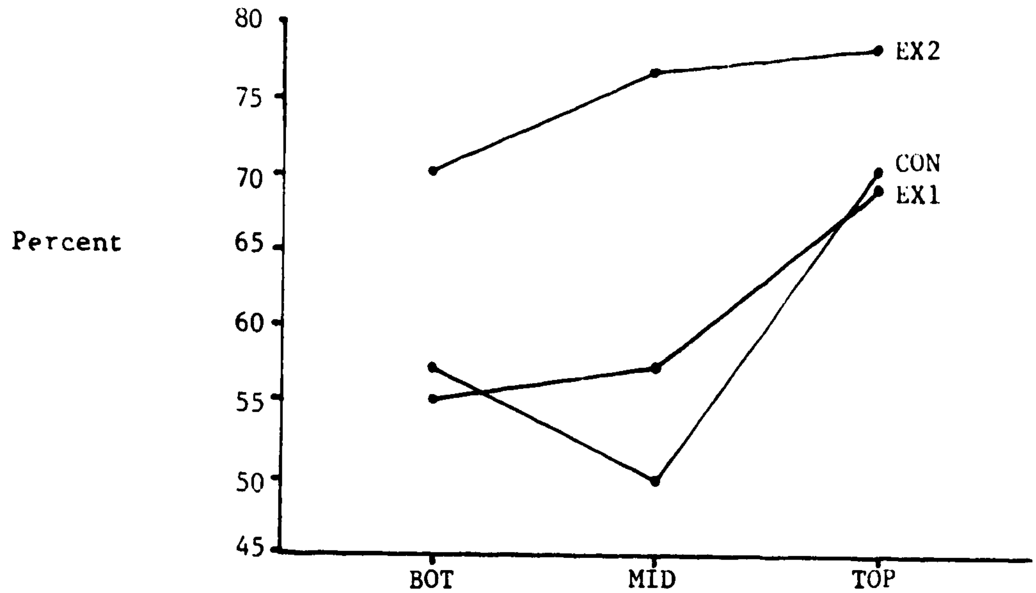

Figure 5. Final M.C. Marks: Groups by Levels

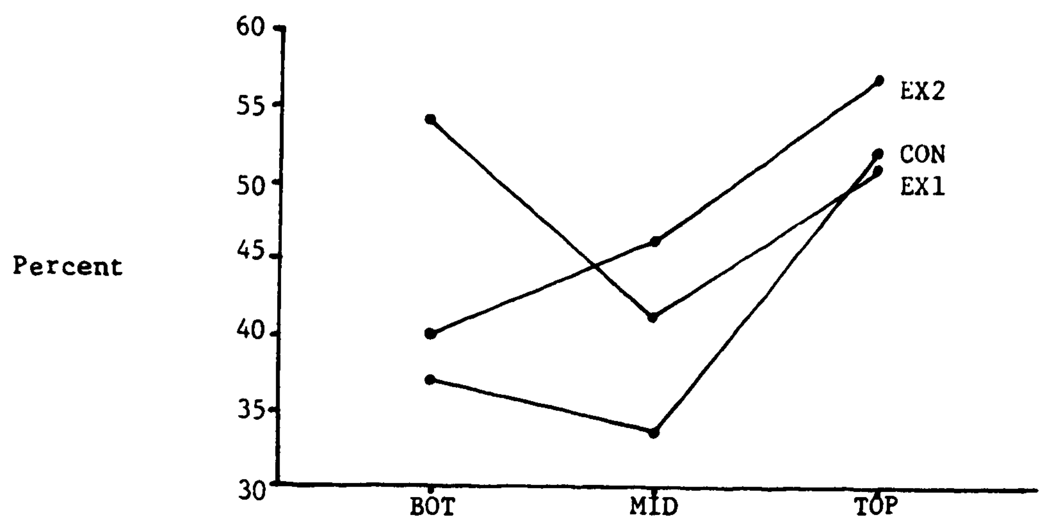

FIGURE 6. Final E.A. Marks: Groups by Levels

3. Discussion

In general, the results from Concordia College support the conclusions from the first study; there are, at least, trends indicating an improvement of the two SOLO Taxonomy groups over the control group, and further improvement of the second experimental group which received more detailed instruction in the SOLO Taxonomy. Whereas the marks on the final examination of the control group and the first experimental group differed by only 5%, the second experimental group outperformed the control group by almost 12%.

It may be considered surprising that the added emphasis on the SOLO Taxonomy instruction in the second experimental group resulted in greater increases in multiple choice performance than in essay performance. This would be expected, however, to the extent that multiple choice questions demand more in the area of higher level procedural thought than simply rote memory, and would also be predicted by cognitive theories emphasising the structuring of detailed information in terms of organizing schemas.

While these results are not conclusive due to possible group differences in age and motivation, they do suggest the employment of the SOLO Taxonomy as an instructional method is effective for the mid-range, motivated student, and certainly warrants further testing and improvement in its delivery and as a teaching method.

V. STRUCTURE OF THE SOLO TAXONOMY QUESTIONS AND ANSWERS

Biggs and Collis (1982) suggest that the adaptation of the SOLO Taxonomy to undergraduate instruction may be problematic for several reasons. First, due to the wide and shallow perspective taken by instructors in teaching introductory courses, most students do not have enough of an opportunity to develop comprehension of course material much beyond the multistructural level. Second, due to low instructor and institution cognitive demands, few students have the motivation to work hard enough at structuring and integrating their course material to achieve the higher SOLO Taxonomy levels. Their predictions have been borne out in the present studies as only a few students achieved answers at the extended abstract level. It was the case that these students appeared to enter the course with the ability to write and think at naturally higher levels, and picked up the SOLO Taxonomy's demands very quickly, while the students with lower incoming abilities according to the pre-tests did not. Again, it was the mid-range student, moving from uni- to multi-structural answers or from multistructural to relational that accounted for most of the differences. This chapter will examine the above, and other, questions in an indepth, qualitative manner in order to more fully delineate the differences between students and between the experimental groups in terms of their answers. Examples of the questions and answers at all levels will first be presented, followed by descriptive measures of where the groups scored in SOLO Taxonomy terms.

A. INSTRUCTIONAL PHASE

This section will deal descriptively with the series of exams throughout the term, in terms of the exam questions and student responses. As articulated earlier, the series of four hour-long examinations throughout the term was the main part of the experimental manipulation, designed to give the students instruction and practice in structuring essay answers in SOLO Taxonomy terms. Again, an examination given to the experimental group consisted of four hierarchically arranged questions such that the first required unistructural knowledge of the area, the second multistructural, the third relational, and the fourth extended abstract.

As presented earlier in Table I, there is much subjective disagreement among judges as to what level of knowledge is needed to answer a particular question. Agreement was highest for unistructural questions. Most of these questions simply asked for definitions or lists such as "define random selection" and "what is an action potential" that was easily judged as demanding a one-concept answer.

The lowest level of agreement concerned the demands of a multistructural question. A question such as "what are the five basic models of psychology?" was seen by two of the judges as unistructural, asking simply for a memorized list of 5 words, whereas two other judges termed this multistructural in that each of the models required separate cognitive structural representation. Other questions such as "what is neurological habituation and how does it work?" was seen as multistructural by two judges asking for two separate lines of thought, and relational by the other two as asking for a relating answer between a definition and an application. The only question with unanimous agreement as being multistructural was "Briefly describe the two main cognitive biases" which specifically required two separate answers in one.

Relational questions also received a high degree of agreement, conceivably because it is easy to ask students to identify relationships in specific terms. Unanimous judgements of relational questions included "How are scripts and schemas the same/different," "Of what use are computers in studying human intelligence," and "Give three basic assumptions of the cognitive model that make it different from other models."

Judgements concerning extended abstract questions were in agreement for the most part if the question asked "why." For example, "Why are there different perspectives in the study of psychology," "Why did consciousness evolve," and "Why are psychologists interested in perceptual illusions," were all unanimous. On the other hand, "why would a psychologist study a rat learning to press a lever" received two judgements as extended abstract, one as relational and one as unistructural. Similarly, "How is the scientific method better/worse than common sense as a tool for knowing the world" was seen as requiring evaluative extended abstract knowledge by two judges, relational knowledge by one judge, and multistructural by the fourth judge. In conclusion, the judgement of what knowledge is required to answer a question, as distinct from the judgement of what structure of answer is given, is a very subjective and problematic area.

As mentioned earlier, the training in what the SOLO Taxonomy criteria required was given to the experimental group both verbally and through examples, while practice and feedback were provided through the series of short examinations. An elaboration of the independent variable follows, with sample student responses to the partial credit questions. The first examination, given after a month of instruction, partially consisted of the four questions listed below.

1. Define "random selection." (Unistructural)
2. What's an "operational definition" and what does it define? (Multistructural)
3. What's the relationship between a dependent and an independent variable, and how is it achieved? (Relational)
4. How, and in what areas, is the scientific method

better/worse than common sense as a tool for knowing the world? (Extended abstract)

Answers to each question were marked on a four point scale (0 - 3), partially to differentiate students for evaluation purposes and partially to provide feedback to them as to what extent they were succeeding at reaching the required SOLO Taxonomy levels. There was no attempt to assign marks according to the SOLO Taxonomy structures as student stratification and feedback were the only objectives. Examples of each mark value are listed below. Student answers are presented here verbatim, without correcting spelling or grammatical errors.

1. Define random selection:

-Random selection is that a group of people are chosen randomly without other conditions in a research. (0 points)

-Random selection is a process by which research subjects are chosen from a population in such a way that the entire population is represented equally. (1 point)

-Selecting a population without any specific preference. Each individual/subject has an equal chance of being selected. (2 points)

-Each person has an equal chance of being chosen and choice of one person does not limit the fact that others may be chosen. The sample must be representative of the group being studied. (3 points)

2. What's an operational definition and what does it define:

-It is a definition given to a certain operation based on chemical and physical research. It defines the phenomenon that was studied based on the information received from the equipment used in this study. (0 points)

-Operational definitions are definitions of behaviors wanting to be studied. An experimenter must clearly define what behaviors he is observing in order for future experimentors to understand and possibly test his original experiment. (1 point)

-An operational definition states what a person specifically means by general words in an experiment. e.g. motivation (causes people to) work: an operational def of work could be pushing a lever, an operational def of motivation could be money. Everyone has his own idea of what defines a word (eg. work) so by giving it an operational definition, other will know exactly what you meant by the word in your experiment. (2 points)

-An operational definition is a statement that allows an experimenter to work with a concept. It usually defines a concept in terms of how it is measured. An example would be time. We can't talk about it unless we talk about how it is measured or unless we give it an operational definition. (3 points)

3. What's the relationship between a dependent variable and

an independent variable and how is it achieved:

-There is no relationship. The dependent variable remains the same. The independent variable is increased or decreased to check what will happen when it is. (0 points)

-The change in the dependent variable relies on the changes made to the independent variable. The dependent variable NEEDS the independent variable. (1 point)

-The relationship is directly related between a dependent and independent variable. The independent variable has control over the outcome of the dependent variable. This relationship is achieved by simply controlling the independent variable, then observing the response of the dependent variable. (2 points)

-A cause and effect relationship exists between the IV and the DV. The IV is the cause and the DV is the effect. By using controlled procedures or by treating all groups identically that are part of the experiment and by selecting all groups by random selection the effect of the IV can be isolated (ie all other possible reasons for the DV occuring are controlled or inhibited and the only difference in the treatment and experimental groups is that one experiences the IV and one does not). (3 points)

4. How, and in what areas, is the scientific method better/worse than common sense as a tool for knowing the

world?

-The scientific method is better since it doesn't "pass off" certain behaviors by explaining them with the use of stereotypes or generalizations. (0 points)

-The scientific method is better than common sense because it gives a logical sequence of steps to attain results which give us testable evidence. This helps in any aspect of science or everyday life where proof of something is desired. (1 point)

-The scientific method is better than common sense as a tool for knowing the world in that it's a universal, logical procedure that controls biases as best as possible. The scientific method would be used in scientific studies and psychological rather than common sense. The scientific method could be worse in a situation where immediate common sense and sensitivity are needed. E.g. if someone was about to jump off a bridge, you would need to use your common sense right away, not the scientific method. (2 points)

-The scientific method is better than common sense in areas where variables are testable. The scientific method allows one to be objective about making decisions, about cause and effect relationships by providing a process by which all confounding variables like experimenter bias etc can be controlled and removed. It allows one to isolate the independent variable so that a specific cause and effect relationship can be determined. The scientific method, however, does not provide the tools to obtain information about situations or variables that cannot be defined and that are not testable. Common sense is important to know and understand situations such as these, by remembering common or similar experiences people are able to understand and survive. (3 points)

There were few apparent relationships from question to question for any of the students. "Better" students had a tendency to get higher marks on all levels of questions while "poorer" students did poorly on all as well. The majority of students ranged in marks from 0 to 3 for any given question in what appeared to be a completely haphazard fashion, giving the appearance of having studied simply in a memorization fashion and trusting to luck as to whether or not what they studied would be on the test. There was no sense of answers being hierarchically arranged; marks were no lower for the higher level questions than for the lower levels. This could have been the case for several reasons: either the marker eased criteria for the higher level questions to keep overall exam marks from being too low and discouraging, or the students were simply memorizing bits of the material and the questions were not truly hierarchically arranged. It may be the case that even for the most apparent extended abstract questions such as "why are there different perspectives in the study of psychology," students could simply recite book or lecture passages that would satisfactorily answer the question. If this is the case, the partial credit questions are simply tapping into prestructural and unistructural knowledge and would not be very effective in stimulating or demanding higher level thought.

Neither was there any quantitative evidence of marks getting better over the term, again possibly due to more stringent marking over the four exams as marks were assigned in a norm-referenced rather than a criteria-referenced fashion. It was also the case that the exams were not cumulative; each examination dealt with novel class material that could not be heirarchically arranged over the term. Biggs and Collis (1982) emphasize that the levels are not maturationally determined in the Piagetian sense and that the structure of answers is sensitive to instruction and motivation. Therefore, variation in the levels of answers for any particular student, both within a single examination and across the term's examinations, would be expected.

B. FINAL EXAMINATION

The essay portion of the final examination consisted of three questions requiring extensive answers that were to be marked according to the SOLO Taxonomy. The questions are listed below.

1. Imagine the Department of Defense has come to you with a question. They have developed a computer that can tell if a plane is in the gunsights of an enemy plane, and they wish to develop a system that will inform the pilot of that fact immediately so he can take evasive action. Their question is as follows. What is the best and fastest way to let the pilot know he's about to get shot at: a visual stimulus such as a flashing light on the control panel, an auditory stimulus such as a buzzer or bell, or a tactile stimulus such as a poke in the back from a plunger embedded in the seat of the plane? Design an experiment that will answer this question for them.

2. Compare and contrast the Behaviorist and the Cognitivist paradigms. Define paradigm, relate how the two schools of thought see human behavior and its causes, how they are similar and different, their strengths and weaknesses, their appropriate areas of application and so on. On the basis of what you know of these paradigms, imagine you are about to become a research psychologist, and summarize your above answer by telling me in which paradigm you would choose to work, and why.

3. The "nature-nurture" controversy has been raging for years; that is, how much of human behavior is innate and how much is learned through experience. Current psychology no longer answers this question in terms of nature "versus" nurture but in terms of nature "and" nurture, that is, how the two influences work together, how nature provides the basic

potential upon which nurture builds. Examine this question in terms of information from any two of the following: a) neurophysiology, b) sensation, c) perception, d) consciousness, e) learning, f) cognition.

If the SOLO Taxonomy instruction had an effect, it would be expected that the experimental groups would show a lower proportion of students in the lower levels and a higher proportion in the higher levels than the control group. These results, summed over the three questions, are presented in Table VIII. A chi square test for independence showed there is a non-significant tendency in that direction ($\underline{X}$ (8) = 6.39, $\underline{p}$ < 0.05). Individually, question one showed the greatest difference in this direction ($\underline{X}$ (8) = 15.3, $\underline{p}$ = 0.05), with no differences in question two. Question three showed a much different distribution of marks with a higher proportion of the second experimental group at the multistructural level and proportionally lower in both the higher and lower SOLO Taxonomy levels. This resulted in a significant but not meaningful difference between groups ($\underline{X}$ (8) = 39.11, $\underline{p}$ < 0.005).

TABLE VIII.

PROPORTIONS OF STUDENTS IN THE SOLO LEVELS; ALL QUESTIONS

	CONTROL	EXP 1	EXP 2
PRE:	.06	.01	.02
UNI:	.31	.28	.27
MULTI:	.44	.45	.48
REL:	.13	.18	.15
ABSTR:	.06	.08	.08

In summary, the majority of student answers in all three groups scored in the multistructural level, and only a few went on to relational and extended abstract comprehension. As Biggs and Collis suspected, it may indeed be the case that undergraduate courses may be too "introductory" in that little opportunity is available for students to develop higher levels of structured knowledge, except for a few exceptional and motivated students that would do so regardless of instruction. The effect of instruction is very small.

The variablity within the samples of student answers was very large. To demonstrate this point, and to fu ther exemplify the marking scheme, examples of answers within the SOLO Taxonomy levels are presented below, drawn from the first question requiring a design of a controlled experiment. Again, student spelling and grammatical errors have not been corrected, and show a cursory negative correlation with the level of structure.

UNISTRUCTURAL:

"Hypothesis -> it can be stated that a stimulus such as a poke in the back would be the quickest way to alert a piolet that someone is shooting at them.

I believe this would work the best because if he were to be poked it would have a reflex action that would start the adrenaline pumping boom...he'd hit the steering wheel and up they would go. This would be better than a visual warning because the light has to go through the process of being seen through the brain and then through to the reflexes. A poke in the back would be an automatic reflexive action.

To test the different situations you could test different pilots. For some of them their visual reflexes may be quicker, for others it may be auditory and for others it may be tactile stimulus.

Experiment. -> Test 100 pilots of all different training, experience and age.

-> seat them in a controlled environment one similar to that of a cockpit. Try the three different stimuli to see which will work the quickest.

1) have a light on the panel in front of them, they will have their hands on the steering wheel. When the light goes off time how long it will take for the pilot to go up.

2) ring the bell and time how fast it takes for them to go up.

3) When they are poked in the back see how long it will take for them to pull upon the steering wheel.

I think the tactile one will work the best because most human beings are use to seeing lights going on/off and hearing bells and buzzers their reflexes to these stimuli may not to as quick, whereas we are not us to being poked in the back with a stick very often. And such a harsh stimulus would come without any warning whatsoever and it would scare us therefore makeing our reflexes quicker.

All the pilots will probably not be the same. Some may find the poking annoying and dicomforting but when it come to being shot out of the sky and a poke in the back, I would tend to favor the poke to tell me to get out of there.

If the pilot was nervous then any little noise may startle them same with the light because there are going to be lights on your control board. This may startle them too. The poke in the back is pretty much distinguishable from something else."

This student is only marginally out of the prestructural level, as it is difficult to pick out any central structure to her answer. She is unable to differentiate herself from the subject matter; rather than testing pilots in an objective manner, she concentrates more on her feelings and emotions and how she would respond in the testing or flying situation. Her answer is embellished with emotional phrases such as "boom...he'd hit the steering wheel and up they would go" and "being shot out of the sky." She does not understand the purpose of testing in order to answer the given question, so closes immediately in her conclusion that a tactile stimulus is best for her, but will be different for other people. She feels no need for consistency and manages to separate tactile and visual stimuli into "reflexes" and going "through the brain" respectively in order to support her conclusion. The one identifiable structure is the generally stated process of comparing the times of 100 pilots reacting to the three stimuli.

MULTISTRUCTURAL:

"Design a hypothesis. Using a random choice of three pilots, place them in the experimental setting. Tell the subjects what the experiment is about and what to expect, either a visual stimulus (flashing light), auditory stimulus (buzzer or bell), or tactile stimulus (poke in the back), signalling the shooting.

Have a controlled experiment setting. With the independent variable being the 3 different stimulus and the dependent variable being the responses. Have all the environments identical except for these stimulus.

Have all the subjects set in a situation where it seems as if he/she is going to be shot at. When they are about to be shot at, give them the stimulus. Record reaction time.

Repeat [illegible]s experiment 2 more times for each stimulus and take the av[illegible]ge of the three times and compare the data for the stimuluses. If the best reaction rate corresponds to your hypothesis, incorporate it into a theory."

This student is demonstrating the ability to hold several concepts in mind at one time; random selection, controlled procedures, variables, recording and comparing data, and so forth. Further, they are strung together in an acceptable fashion. What is lacking is a clear relating of the procedural elements so as to answer the given question; there is no application to the problem at hand. This becomes clear in the student's perseveration in the last sentence; the point about "theory" shows she is simply reciting the points she has memorized from the lecture, without interrelating these concepts as a method for answering practical questions. It is interesting to note that this answer is much shorter, yet much better, than the previous unistructural effort, as this student gets right down to the task at hand.

RELATIONAL:

If a certain stimulus was given to a pilot during war then he would react quicker in a dangerous situation.

Operational Definition: A visual stimulus, a flashing light, will be given to one experimental group, a bell will be given to another experimental group, and a tactile stimulus, a poke in the back, will be given to another experimental group. We will observe how long the reaction time is between the stimulus and the response.

First, will have to get 30 pilots, all between the ages of 30-35 years old. They will all have to be married or single. Then randomly they will be placed into three experimental groups, each individual having equal chance to get into either group. All groups are equal and must be treated equal, except for the independent variable. Then the groups are given the independent variable, either the visual, auditory or physical stimulus, and "blind" observers are recording the observations. The observers are recording the reaction time between the independent variable and the response or dependent variable. The researcher also gives false alarms to exclude any subject bias. After the experiment is complete you must analyze and find the statistics. If statistics show the results are less than 5 percent you can be sure it is the independent variable determining the results or responses. Then there must be replication.

Conclusion: I believe that the poke in the back will create a faster reaction time as long as the stimulus isn't given too much time to cause inhibition of the cells so they quit firing. This could be more a reflex action so you would react before you thought about it. I believe the visual and auditory stimulus take longer because of all the noise from actually controlling the plane, you have probably other signals and noises to contend with while flying."

This student is struggling mightily with a topic that is too much for her, but somehow manages to construct a relational answer. As in the previous multistructural answer, the necessary elements of variables, controlled procedures and data analysis are present, even if the student has no idea what statistics are all about. What distinguishes this answer from the previous is the student's insistence on random assignment, isolating the independent variable and controlling for experimenter and subject biases. She has a sense, although dimly perceived, that the experimental method is followed in order to clearly discover an answer to the given question, and it is around this point that her concepts are interrelated.

EXTENDED ABSTRACT:

"The Evasive-Action Experiment"

Before an experiment is designed, the problem must first be clearly stated. In this case, the problem is as follows: What type of stimulus will provide the most effective orienting response and the shortest reaction time for a given motor response; a visual, tactile or acoustic stimulus? Next the premises on which the experiment is based must be outlined. These premises will form the basis for the hypothesis. Because a pilot is consciously attending the visual stimuli on both his/her control board and outside his/her cockpit, additional visual stimuli would have a reduced probability of having sufficient salient value to get past the sensory filter (Broadbent's theory) and into short-term memory. Further, because a pilot's attention is primarily focussed on the airspace around him, a visual stimulus could distract him from actually flying the plane, perhaps leading to a crash. A tactile stimulus would also not be an effective way of eliciting the motor response of moving the control stick. A tactile stimulus could result in the pilot reflexively pulling away from the "poke", perhaps interferring with his/her control of the plane. Further, cutaneous sensation (i.e. pressure, or in the extreme, pain) may not be viewed as a true sensation. There are Pacinian corpuscles in the integument which register pressure, but the "sense" can be over-shadowed by distracting one's attention away from the stimulus, a technique often used in pain control. In the context of an air-battle, the pilot's attention will be very much focussed on visual stimuli; attending to the visual stimuli will hold more survival value for the pilot and thus, the tactile stimulus may be ignored. Because of this, it may be hypothesised that a tactile stimulus would not be an effective method of signalling that an enemy plane has the plane in its line of sight.

The acoustic stimulus would be the most salient stimulus in the context of a cockpit, and would result in the shortest reaction time. An impulse which had physical properties from other sounds heard via the pilot's earphones would be most effective, as described in Broadbent's dichotic listening experiments (i.e. in the unattended ear, an automatic orienting response could be elicited by presenting an acoustic stimulus which differed in pitch and/or amplitude from other acoustic stimuli, such as conversation with other pilots).

To prove that the acoustic stimulus would elicit the best orienting response and the shortest reaction time, the following experimental design could be used. In the laboratory, a stratified sample of the general population will be used; that is, only pilots will be used as subjects so that the general knowledge and procedures taught in the military would be basically constant. Out of the population of military pilots, random sampling would be used such that each pilot would have an equal chance of being chosen and each choice would be independent of all others. The pilot would be given a complex visual distraction task (such as a flight simulator) to mimic the environment in the cockpit. They would be provided with a control stick like the one in the cockpit. The independent variables would be a visual stimulus (a flashing red light), an auditory stimulus (a high-pitched tone) and a tactile stimulus (a poke in the pilot's back).

Each pilot would undergo each stimulus condition such that within-subjects variables could be compared, and also such that large deviations in results (between pilots) could perhaps be explained by such extraneous variables as personality types. Each pilot would undergo three trials with each stimulus. This is an arbitrarily chosen figure for number of trials, but it is chosen because many trials would produce a practice effect, skewing the results. The effectiveness of a given stimuli would be operationally defined as the reaction time between presentation of the stimulus and moving the control stick.

When the data were all collected, the information for the different kinds of stimuli would be separately analysed using descriptive statistics. Using inferential significance, the differences between the different reaction times would be determined. If statistical significance were reached (i.e. the probability that the difference in reaction times were due to chance is less than 5%), the data could be accepted as valid, and the stimulus showing the shortest reaction time would be deemed most effective in informing a pilot that he/she can take evasive action.

As stated in introducing the hypothesis, the auditory stimulus would probably be the most effective way of eliciting evasive action because of the preattentive processing of gross physical features which takes place when other types of stimuli are being consciously attended."

This over-achiever has presented as fine an answer to the question as could be expected from a time-constrained student who has been exposed to psychological concepts for a total of three months. She has, of course, over-done the answer, but it is this level of motivation that makes for exceptional students. What makes this an extended abstract answer is that the student has done such a thorough job of analyzing the problem in terms of what she knows of sensation and perceptual theories, and of situating her research in terms of those theories. Although she has not really stated it, one understands that she is not only correctly applying the experimental procedure but understanding its limitations, as she hints that the laboratory results must be carefully interpreted because the procedures m y not adequately simulate actual flying conditions.

C. MAINTENANCE OF LEVELS TO POST TEST

Levels of the post test answers were compared to the same students' levels at the final exam to discover if the students' level of comprehension had changed over the six months. The proportions of the SOLO Taxonomy levels that students achieved on the final and the post test are presented in Table IX. In general, it can be seen that both groups dropped levels on the post test showing that course material had been lost over the six months. However, the experimental group dropped much less than the control group did.

For the control group, seven answers were at the same level at post test, four had dropped one level, six had dropped two levels, and one had dropped three levels; an average decrease of 1.06 levels. For the experimental group, six answers had stayed the same, ten had dropped one level and two had dropped 2 levels. Interestingly, one student answer had gone up one level, and two had increased two levels, for an overall average decrease of .4 levels. In conclusion, the experimental group, influenced by the SOLO Taxonomy method of marking, showed a greater ability to retain course information over the six month period.

TABLE IX

PROPORTIONS OF ANSWERS AT EACH SOLO TAXONOMY LEVEL; FINAL AND POST TEST

EXPERIMENTAL (N = 7, 3 questions each)

	PRE	UNI	MULT	REL	E.A.
FINAL	.00	.14	.71	.14	.00
POST	.00	.57	.29	.10	.05

CONTROL (N = 6, 3 questions each)

	PRE	UNI	MULT	REL	E.A.
FINAL	.00	.33	.44	.11	.11
POST	.33	.56	.11	.00	.00

VI. EVALUATION OF THE SOLO TAXONOMY

The educational community is currently concerned with developing a valid understanding and measurement of academic learning in terms of the goals of education, methods of instruction to achieve these goals, and assessment to discover if students are reaching those goals. In the previous chapters, these three areas have been defined and expanded, the SOLO Taxonomy has been presented and examined from a theoretical cognitive perspective, and the experimental application of the SOLO Taxonomy as an instructional tool has been presented. Throughout, the emphasis has been on an evaluation of the SOLO Taxonomy as a contributor to the field of undergraduate education as a basis for a definition of academic learning. The evaluation of the SOLO Taxonomy has involved three areas: its cognitive theoretical basis, its practical and conceptual use as a tool of instructional design, and its empirical viability in the undergraduate classroom. This chapter will summarize the above evaluation and will present conclusions on the use of the SOLO Taxonomy in undergraduate instruction.

Primarily, the use of the SOLO Taxonomy as an assessment technique, and secondarily as an instructional method, results in better performance in undergraduate psychology courses. The demands of the midterm examinations encouraging students to work through the levels of the partial credit questions appear to be inducing students to generalize this approach to the final examination. The practice and feedback in structuring answers for the midterms have resulted in better structured essay answers for the experimental groups. This is especially true for the student of mid-range ability and high motivation, and especially for the retainment of course information over a longer period of time. That, in itself, is reason enough to institute it in undergraduate classes, but the SOLO Taxonomy also provides a valid and valuable perspective on conceptualizing student learning, knowledge and performance, as detailed below.

The SOLO Taxonomy, from the perspective of cognitive psychology, promises to provide a workable and measurable definition for learning in undergraduate education. Concepts of schematic representation of knowledge, as inferred from the structure of observed essay answers, provide a valid conceptualization of academic knowledge and of the extent to which students have successfully structured declarative course material. Concepts of scripted actions can be extended to define the procedural processes of dealing with that material. The SOLO Taxonomy may provide a basis for a definition of the educated or knowledgeable person: essentially one who has easy, automatic access or retrievability to structured information such that it can be applied to the novel situation of questions or problems demanding extended abstract answers. The SOLO Taxonomy's definition of learning may be seen to provide the basis for the goals of liberal arts education as defined above, in terms of the declarative and procedural knowledge that exemplifies the critical and rational thinker.

Instruction in what the SOLO Taxonomy expects on examination answers may stimulate metacognitive monitoring on the part of the student as he or she conceptualizes new course material in structural terms. Students can be taught to question their own perspectives on course material, monitoring themselves as to whether they are approaching it in uni-, multi-, relational or abstract structural terms, and further, what more needs to be done to reach the higher levels. In terms of fulfilling attitudinal goals, the reinforcing value of specifically defining what behavior is expected of the student is very effective in creating positive values for the learning process, and in motivating the student to achieve these levels.

The SOLO Taxonomy may provide the basis for structuring course content in a definitive way for both student and course designer. From the SOLO Taxonomy's perspective, "synthesizing" of course material would be conceptualized around the organizing principle of structures, multistructures, related structures and extended abstract structures, all in terms of schematic processing. Any segment of an undergraduate psychology course can be presented as a central structured principle with attendent details, related to other structures of similar material, and compared, contrasted and evaluated in relation to other concepts. "Sequencing" could be conceived as inducing students to move progressively through these stages. The SOLO Taxonomy is more amenable to a part-to-whole sequence similar to Gagne (1968) or to Bruner's (1966) spiral approach, as its postulated levels are cumulative and progressively increasing in complexity. As an instructional technique, the SOLO Taxonomy demands student involvement at the metacognitive level and encourages the activity of self-structuring of course material. A carefully constructed marking scheme, similar to the one developed above with the Partial Credit Model, provides the feedback for the students to discover to what extent they are approximating the required levels with their examination answers.

The SOLO Taxonomy may provide a workable measure for evaluation of both student performance and the efficacy of the learning institution's curriculum. Careful implementation of the SOLO Taxonomy as an instructional tool may specifically define the course goals and measure whether or not students are meeting them. Further, the level of student answers will suggest whether the instructor is effectively implementing the technique in terms of lecture structure and testing. The SOLO Taxonomy examination demands an active, constructive and generative process on the part of the student, and both measures and contributes to those processes. Essay answers evaluated in this way can serve as an operational definition of learning in the perspective developed in the previous chapters.

Student answers to SOLO Taxonomy questions can be compared to an "absolute" frame of reference for minimum or adequate performance for mastering course content rather than relying entirely on students' relative standings. This is, of course, not to say that the SOLO Taxonomy should be the entire evaluative criterion for marking. Instructors' capabilities as teachers must always be controlled for by using relative marking procedures; that is, students must not be penalized if the instructor is not providing adequate instruction to allow students to acquire extended abstract structures. They must also not be penalized if the knowledge level of incoming students is too low to permit acquisition of high levels of knowledge within the class time frame. But the SOLO Taxonomy will also provide feedback to the instructor on incoming knowledge levels and whether or not teaching procedures and course content are at levels capable of producing comprehension understanding. The SOLO Taxonomy is therefore very useful as an evaluative tool for curriculum and instructional design. It can provide feedback as to where students are regarding course content and at what level course material should be delivered, according to course pre-tests and periodic feedback throughout the term.

The expectation of having examinations marked by the SOLO Taxonomy criterion may put higher expectations on the students in terms of quality and quantity of work. It has been demonstrated that higher standards lead to higher performance (Natriello & Dornbush, 1984) in most students, but only to a point. If the standards are seen as unattainable, students tend to give up (Rosswork, 1977). The effectiveness of examination feedback lies in its strength in identifying where each student is in terms of his or her comprehension, and exactly what more is needed to move upward through the taxonomy. If each student is encouraged to take "one step at a time" through the SOLO Taxonomy stages, goals and standards of the next level can be seen as within reach, and most students will more likely continue to show improvement.

In summary, the SOLO Taxonomy, as an instructional and an assessment technique, may be very effective in undergraduate education. It is, however, not a technique that can be easily implemented by an instructor. Personal experiences in its use in the above studies show many difficulties. The concept of structured knowledge, in terms of schemas or networks, must be thoroughly understood by the instructor in order to create a structure of course material. There always seems to be a question as to whether or not enough structure is being presented in lectures, and whether the goals of the SOLO Taxonomy have been sufficiently articulated. Examinations are very difficult to construct, as there always seems to be the question of whether a given exam question is requiring or allowing a unistructural or multistructural response. Judges' estimates on the level of student comprehension needed to answer a given question is far from perfect agreement, especially between a single structure question with multiple details and a multi-structured question. And last, there is still an uncomfortable subjective judgement involved in marking examinations. Many students' answers are clearly sufficient for the demands of the partial credit questions, but there are also many that are debatable. Also, the marking of essay answers shows that criteria for judging the level of an answer is very dependent upon the class norms and the students' relative performance, and is not as objective as one might imagine.

It was, however, very gratifying to see students rising to meet the challenges that the SOLO Taxonomy offered. While there was some complaint as to the fact that these examinations demanded much more work than did other courses, students seemed appreciative of the method. Many saw the training of how to build knowledge structures as very beneficial, and it was heartening to see them drawing diagramatic hierarchies and schemas in their notebooks for studying purposes. Many comments were made from later classes concerning how they were forced to change their learning strategies from memorization of detail to comprehension and application of principles, and the expectation that this information would be of more use and would stay with them for a longer period of time. It would be interesting to more formally question students as to how this method of instruction has changed their studying and learning strategies.

One of the SOLO Taxonomy's strongest points is that its use is extremely effective as a feedback tool for the instructor. Far more effectively than impersonal multiple choice or short answer examinations, essay answers give the instructor a strong understanding of where the student is in terms of knowledge and understanding of course content. Further, an instructor gets a very firm idea of what ideas are getting across to the majority of students and which are falling flat, what works in the classroom and what needs work. The use of the SOLO Taxonomy as instructional feedback can only result in a more aware and conscientious instructor.

The use of the SOLO Taxonomy demands continuing use, experience, feedback and constant revision of instructional and asssessment techniques. Yet this in no way disqualifies it from use, as any instructional technique must be constantly evaluated in this way. Seen from this perspective, the SOLO Taxonomy offers an excellent tool around which to design, implement and assess undergraduate education, and has the potential to contribute conceptually and empirically to Reigeluth's (1984) call for a science of instruction.

REFERENCES

Adler, M.J. (1982). The Paideia proposal. New York: MacMillan.

Adler, M.J. & Van Doren, C. (1984). The conduct of seminars. In M. Adler (Ed.), The Paideia program: An educational syllabus. New York: MacMillan.

Anderson, J.R. (1990). Cognitive psychology and its implications (3rd ed.). San Francisco: W.H. Freeman.

Anderson, J.R. & Bower, G.H. (1973). Human associative memory. Washington, DC: Winston & Sons.

Andrews, D.H. & Goodson, L.A. (1980). A comparative analysis of models of instructional design. Journal of Instructional Development, 3(4), 2-16.

Angelo, T.A. (1990). Classroom assessment: Improving learning quality where it matters most. New Directions for Teaching and Learning, The Changing Face of College Teaching, 42, 71-82.

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In G.H. Bower (Ed.), Human memory: Basic processes. New York: Academic Press.

Ausubel, D.P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 51, 267-272.

Ausubel, D.P. (1968). Educational psychology: A cognitive view, New York: Holt, Rinehart & Winston.

Ausubel, D.P. (1977). The facilitation of meaningful verbal learning in the classroom. Educational Psychologist, 12, 162-178.

Baine, D. (1986). Memory and instruction. Educational Technology Publications, Englewood Cliffs, NJ.

Bartlett, F.C. (1932). Remembering. Cambridge University Press.

Bartlett, F.C. (1958). Thinking. New York: Basic Books.

Belmont, J.M. (1989). Cognitive strategies and strategic learning: The socio-instructional approach. American Psychologist, 44(2), 142-148.

Best, J.B. (1986). Cognitive psychology. West Publishing Company.

Beyer, B.K. (1987). Practical strategies for the teaching of thinking. Allyn & Bacon, Inc.

Biggs, J.B. & Collis, K.F. (1982). Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome. Academic Press.

Biggs, J.B., Holbrook, J.B., Ki, W.W., Lam, R.Y.H., Li, W.O., Pong, W.Y. & Stimpson, P.G. (1989). An objective format for evaluating the quality of learning in various secondary subjects: A symposium.

Bloom, B.S. (1956). A taxonomy of educational objectives: Handbook I, the cognitive domain. New York: Longman.

Bobrow, D.G. & Norman, D.A. (1975). Some principles of memory schemata. In D.G. Bobrow & A. Collins (Eds.) Representation and understanding: Studies in cognitive science, Academic Press.

Bower, G.H., Black, J.B. & Turner, T.J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.

Bransford, J.D. (1979). Human cognition. Belmont, CA: Wadsworth.

Bransford, J.D. & Johnson, M. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-220.

Bransford, J.D., Sherwood, R., Vye, N. & Reisser, J. (1986). Teaching thinking and problem solving. American Psychologist, 41(10), 1078-1089.

Bransford, J.D. & Stein, B.S. (1984). The IDEAL problem solver. New York: W.H. Freeman.

Brewer, W.F. & Treyens, J.C. (1981). Role of schemata in memory for places. Cognitive Psychology, 13, 207-230.

Briggs, L.J. (1977). Instructional design: Principles and applications. Englewood Cliffs, NJ: Educational Technology Publications.

Bruner, J.S. (1966). Toward a theory of instruction. New York: W.W. Norton.

Bruner, J.S. (1985). Models of the learner. Educational Researcher, 14(6), 5-8.

Case, R. & Bereiter, C. (1984). From Behaviorism to Cognitive Behaviorism to Cognitive Development: Steps in the evolution of instructional design. Instructional Science, 13, 141-158.

Chase, W.G. & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.

Collins, A.M. & Quillian, M.R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Collis, K.F. & Davey, H.A. (1986). A technique for evaluating skills in high school science. Journal of Research in Science Teaching, 23, 65-73.

Corno, L. & Snow, R.E. (1986). Adapting teaching to individual differences among learners. In M.Wittrock (Ed.), Handbook of research on teaching, New York: MacMillan.

Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Cranton, P. (1989). Planning instruction for adult learners. Toronto: Wall & Thompson.

Cronbach, L.J. & Snow, R.E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington.

Crooks, T. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58, 438-481.

Crooks, T. & Collins, E.A.G. (1986). What do first year university examinations assess? New Zealand Journal of Educational Studies, 21, 123-132.

de Groot, A. (1965). Thought and choice in chess. The Hague: Mouton.

Dempster, F.N. (1987). Time and the production of classroom learning: Discerning implications from basic research. Educational Psychologist, 22, 1-21.

Dick, W. (1981). Instructional design models: Future trends and issues. Educational Technology, 21(7), 29-32.

Domjen, M. (1987). Animal learning comes of age. American Psychologist, 42(6), 556-564.

Donald, J.G. (1983). Knowledge structures: Methods for exploring course content. Journal of Higher Education, 54, 31-41.

Dunn, R., Dunn, K. & Price, G.E. (1981). Learning Styles: Research versus opinion. Phi Delta Kappan, 645-646.

d'Ydewalle, G., Swerts, A. & De Corte, E. (1983). Study time and test performance as a function of test expectations. Contemporary Educational Psychology, 8, 55-67.

Educational Researcher, (1989), 18(9).

Elias, J.L. & Merriam, S. (1980). Philosophical foundations of adult education. Robert E. Krieger Pub Co: New York.

Elton, L.R.B. & Laurillard, D.M. (1979). Trends in research on student learning. Studies in Higher Education, 4, 87 -102.

Ennis, R.H. (1962). A concept of critical thinking. Harvard Educational Review, 32(1), 81-111.

Ennis, R.H. (1989). Critical thinking and subject specificity: Clarification and needed research. Educational Researcher, 18(3), 4-10.

Entwistle, N.J. & Ramsden, P. (1983). Understanding student learning. London: Croom Helm.

Erickson, F. (1977). Some approaches to inquiry in school community ethnography. Anthropology and Education Quarterly, 8(2), 58-69.

Fiske, S.T. & Taylor, S.E. (1984). Social cognition. Reading, MA: Addison - Wesley.

Flavell, J.H. & Wellman, H.M. (1977). Metamemory. In R.V. Kail & J.H. Hagen (Eds.), Perspectives on the development of memory and cognition. HIllsdale, New Jersey: Erlbaum.

Frederickson, J.R. & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18(9), 27-31.

Gagne, R.M. (1968). Learning hierarchies. Educational Psychologist, 6, 3-6.

Gagne, R.M. (1977). The conditions of learning. (3rd edition). New York: Holt, Rinehart & Winston.

Gagne, R.M., Briggs, L.J. & Wager W.W. (1988). Principles of instructional design. Holt, Rinehart & Winston.

Gergen, K.J. (1985). The social constructionist movement in modern psychology. American Psychologist, 40(3), 266-275.

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

Glass, A.L. & Holyoak, K.J. (1986). Cognition. Random House: New York.

Gregorc, T. (1982). An open letter to an educator; Part II. Challenge: Reaching and teaching the gifted child. 1(1), 37-40.

Gow, L. & Kember, D. (1990). Does higher education promote learning? Higher Education, 19(3), 307-322.

Graesser, A.C., Gordon, S.E. & Sawyer, J.D. (1979). Recognition memory for typical and atypical actions in scripted activities; Test of a Script Pointer Plus Tag hypothesis. Journal of Verbal Learning and Verbal Behavior, 18, 319-332.

Greeno, J.G. (1976). Indefinite goals in well-structured problems. Psychological Review, 83, 478-491.

G eeno, J.G. (1980). Some examples of cognitive task analysis with instructional applications. In R.E. Snow, P.A. Federico & W.E. Montague (Eds.), Aptitude, learning and instruction. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Greeno, J.G. (1989). A perspective on thinking. American Psychologist, 44(2), 134-141.

Gullickson, A.R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

Haertel, E. (1986). Choosing and using classroom tests: Teachers' perspectives on assessment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Harrison, C.J. (1990). Teaching abstract concepts in psychology. Journal of Instructional Psychology, 17(2), 68-70.

Hastie, R. (1981). Schematic principles in human memory. In E.T. Higgins, C.P. Herman & M.P. Zanna (Eds.), Social Cognition: The Ontario Symposium. Lawrence Erlbaum Associates.

Herrnstein, R.J., Nickerson, R.S., de Sanchez, M. & Swerts, J.A. (1986). Teaching thinking skills. American Psychologist, 41(11), 1279-1289.

Hinsley, D.A., Hayes, J.R. & Simon, H.A. (1978). From words to equations: Meaning and representation in algebra word problems. In P.A. Carpenter and M.A. Just (Eds.), Cognitive processes in comprehension. Hilsdale, N.J.: Erlbaum.

Hiscock, M. & MacKay, M. (1985). Neuropsychological approaches to the study of individual differences. In C.R. Reynolds & V.L. Willson (Eds.), Methodological and statistical advances inthe studyof individual differences. Plenum: NY.

Hopkins, K.D., Stanley, J.C. & Hopkins, B.R. (1991). Educational and Psychological Measurement and Evaluation. Prentice Hall: Englewood Cliffs, N.J.

Hunkins, F.P. (1969). Effects of analysis and evaluation questions on various levles of achievement. Journal of Experimental Education, 38, 45-58.

Hutchins, R.M. (1952). Great Books of the Western World: The Great Conversation. Encyclopedia Brittanica: Chicago.

Inhelder, J. & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic.

Jaeger, M.E. & Rosnow, R.L. (1988). Contextualism and its implications for psychological inquiry. British Journal of Psychology, 75(1), 63-75.

Jones, H.E. (1923). Experimental studies of college teaching: The effect of examination on permanence of learning. Archives of Psychology, 10, 1-70.

Kahnemann D. & Tversky, A. (1972). Subjective probability: A judgement of representativeness. Cognitive Psychology, 3, 430-454.

Kameenui, E.J. & Simmons, D.C. (1990). Designing instructional strategies. Merrill Pub. Co.

Kidd, J.R. (1988). How adults learn. Cambridge, NY.

Knowles, M.S. (1984). Andragogy in action: Applying modern principles of adult learning. San Francisco: Jossey-Bass.

Kulhavy, R.W. (1977). Feedback in written instruction. Review of Educational Research, 47, 211-232.

Kulik, J.A. & Kulik, C.C. (1988). Timing of feedback and verbal learning. Review of Educational Research, 58, 79-97.

Langenbach, M. (1988). Curriculum models in adult education. Robert E. Krieger Pub Co.: Florida.

Lehman, D.R., Lempert, R.O. & Nisbett, R.E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about every-day events. American Psychologist, 43, 431-442.

Light, L. & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. Journal of Verbal Learning and Verbal Behavior, 9, 1-11.

Mandler, J.M. (1978). Categorical and schematic organization in memory. In C.R. Puff (Ed.), Memory, organization and structure. New York: Academic Press.

Mandler, J.M. (1984). Stories, scripts and scenes: Aspects of schema theory. Hillsdale, N.J.: Erlbaum.

Mandler, J.M. & Johnson N.S. (1976). Some of the thousand words a picture is worth. Journal of Experimental Psychology: Human Learning and Memory, 3, 386-396.

Marton, F. & Saljo, R. (1976a). On qualitative differences in learning: 2. Outcome and process. British Journal of Educational Psychology, 46, 4-11.

Marton, F. & Saljo, R. (1976b). On qualitative differences in learning: 2. Outcome as a function of the learner's conception of the task. British Journal of Educational Psychology, 46, 115-127.

Masters, G.N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Masters, G.N. & Mislevy, R.J. (1991). New views of student learning: Implications for educational measurement. Educational Testing Service, Princeton, New Jersey.

McKeachie, W.J., Pintrich, P.R., Lin, Y.G. & Smith, D.A.F. (1986). Teaching and learning in the college classroom. University of Michigan.

Messick, S. (1989). Meaning and values of test validation: The science and ethics of assessment. Educational Researcher, 18(2), 5-11.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), The psychology of computer vision, New York: McGraw-Hill.

Natriello, G. & Dornbush, S.M. (1984). Teacher evaluative standards and student effort. New York: Longman.

Neisser, U. (1976). Cognition and reality. San Francisco: Freeman.

Newell, A. & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall.

Nickerson, R.S. (1989). New directions in educational assessment. Educational Researcher, 18(9), 3-7.

Nickerson, R.S., Perkins, D.N. & Smith, E.E. (1985). The teaching of thinking. Hillsdale, N.J.: Erlbaum.

Norris, S.P. (1985). The choice of standard conditions in defining critical thinking competence. Educational Theory, 35, 97-107.

Norris, S.P. (1989). Can we test validly for critical thinking? Educational Researcher, 18(9), 21-26.

Palincsar, A. & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognitive Instruction, 1, 1[illegible]-176.

Percy, K.A. & Salter, F.W. (1976). Student and staff perceptions and the pursuit of excellence in British higher education. Higher Education, 5, 447-473.

Perkins, D.N. & Salomon, G. (1989). Are cognitive skills context-bound? Educational Researcher, 18(1), 16-25.

Piaget, J. & Inhelder, B. (1969). The psychology of the child. New York: Basic Books.

Pintrich, P.R., Cross, D.R., Kozman, R.B. & McKeachie, W.J. (1986). Instructional psychology. Annual Review of Psychology, 37, 611-651.

Pintrich, P.R. & Johnson, G.R. (1990). Assessing and improving students' learning strategies. New directions for teaching and learning: The changing face of college teaching, 42, 83-92.

Reigeluth, C.M. (1979). In search of a better way to organize instruction: The elaboration theory. Journal of Instructional Development, 2, 8-15.

Reigeluth, C.M. (1983). Meaningfulness and instruction: relating what is being le ned towhat a student knows. Instructional Science, 12, 97-218.

Reigeluth, C.M. (1984). The ution of instructional science: Toward a comm ledge base. Educational Technology, 24(11), 20-2.

Rennels, M.R. (1976). Cerebral symmetry: An urgent concern for education. Phi Delta Kappan, 57, 471-472.

Rosch, E.H. (1973). On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press.

Rosenthal, R. (1966). Experimenter effects in behavioral research. New York: Appleton-Century-Crofts.

Rosnow, R.L. & Georgoudi, M. (1986). The spirit of contextualism. In R.L. Rosnow & M. Georgoudi (Eds.), Contextualism and understanding in behavioral science (pp 3-22). New York: Praeger.

Rosswork, S.G. (1977). Goal setting: The effects of an academic task with varying magnitudes of incentive. Journal of Educational Psychology, 69, 710-715.

Rumelhart, D.E. (1980). Understanding understanding. University of California, San Diego. Program in Cognitive Science Tech. Report.

Rumelhart, D.E. & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, J.J. Sprio & W.E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, New Jersey: Erlbaum.

Samples, R. (1975). Are you teaching only one side of the brain? Learning, 3, 25-28.

Schachter, S. & Singer, J.E. (1962). Cognitive, social and physiological determinants of emotional state. Psychological Review, 69, 379-399.

Schank, R.C. & Abelson, X. (1977). Scripts, plans, goals and understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Schwab, J.J. (1962). The concept of the structure of a discipline. Educational Record, 197-205.

Smith, E.E., Rips, L.J., & Shoben, E.J. (1974). Structure and processes in semantic memory: A featural model for semantic decisions. Psychological Review, 81, 214-241.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. Educational Researcher, 18(9), 8-14.

Sternberg, R.J. (1982). Who's intelligent? Psychology Today, pp 30-39.

Sternberg, R.J. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge: Cambridge University Press.

Stiggins, R.J. & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

Stiggins, R.J., Conklin, N.F. & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice, 5(2), 5-17.

Thorndike, E.L. & Woodworth, R.S. (1901). The influence of improvement of one mental function upon the efficiency of other functions. Psychological Review, 8, 247-261.

Tiedemann, J. (1989). Measures of cognitive styles: A critical review. Educational Psychologist, 24(3), 261-275.

Tobias, S. (1989). Another look at research on the adaptation of instruction to student characteristics. Educational Psychologist, 24(3), 213-227.

Tversky, A. & Kahnemann, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Van Patten, J., Chao, C.I., & Reigeluth, C.M. (1986). A review of strategies for sequencing and synthesizing instruction. Review of Educational Research, 56(4), 437 -471.

Wales, C.E. & Nardi, A. (1984). Successful decision making. Morgantown: West Virginia University Centre for Guided Design.

Wales, C.E. & Stager, R.A. (1977). Guided design. Morgantown: West Virginia University Centre for Guided Design.

Weinstein, C.E. & Mayer, R.E. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), Handbook of research on teaching, NY: MacMillan.

West, C.K., Farmer, J.A. & Wolff, P.M. (1991). Instructional design: Implications from cognitive science. Prentice Hall, Englewood Cliffs, N.J.

White, B.Y. & Horwitz, P. (1987). ThinkerTools: Enabling children to understand physical laws. Proceedings of the Ninth Annual Conference to the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Whorf, B.L. (1956). Science and linguistics. In J.B. Carroll (Ed.), Language, thought and reality: Selected writings of Benjamin Lee Whorf. Cambridge, Mass: MIT Press.

Winn, W. (1990). Some implications of cognitive theory for instructional design. Instructional Science, 19, 53-69.

Wittrock, M.C. (1990). Generative processes of comprehension. Educational Psychologist, 24(4), 345-376.

APPENDIX A

SAMPLE EXAM

An example of the SOLO marking scheme (Structure of the Observed Learning Outcome).

Question: What is a kitchen?

Answer #1: A kitchen is a place with a lot of things in it, like a stove, some chairs, and flowered curtains on the windows.

This answer is <u>prestructural</u>, there is some information presented as isolated details, no relationship between details is shown, no discrimination is made between necessary and arbitrary elements. See page 3. 1 point.

Answer #2: A kitchen is a room that has a refrigerator, stove, cupboards, a table, and often a microwave.

This answer is <u>unistructural</u>, there is one central theme, essential elements are presented as to what makes a kitchen a kitchen, and arbitrary elements are left out. 2 points.

Answer #3: A kitchen is a room that contains several appliances. There is a refrigerator for keeping food cold so it won't spoil. There is a stove or a microwave for cooking or warming food. There are cupboards for storing plates and utensils, and a table for sitting and eating.

This answer is <u>multistructural</u>, four themes (each appliance) are presented separately, and developed as to their function. 3 points.

Answer #4: A kitchen is a room in a house used for the purpose of storing, preserving, preparing and eating food. When food is purchased, it is stored in cupboards or, if perishable, in a refrigerator, in order to keep it from spoiling. Before mealtime, food is taken from the storage facilities and prepared on a stove or microwave to ready it for eating. At mealtime, the food is laid out on the table for eating, using plates and utensils from the cupboards.

This answer is <u>relational</u>, containing the same four structures from answer #3, but related in terms of the purpose and sequence of food consumption. 4 points.

Answer #5: Houses generally contain kitchens, bedrooms, living rooms, bathrooms, closets and so forth, each of which serves a specific purpose or function for the people living in the house. While bedrooms are for sleeping and living rooms are for socializing and watching television, the purpose of the kitchen is centered around the storage, preservation, preparation and consumption of food. When food is purchased...(all of answer #4)...from the cupboards.

This answer is extended abstract, containing everything from answer #4, but the concept of a kitchen is embedded in the larger context of the house, serving one purpose of daily living as do each of the other rooms. 5 points.

NOTE:

1. Striving for higher levels of answers without first developing lower levels doesn't work. For example:

Answer #6: A kitchen is a room in a house that generally consists of kitchens, bedrooms, bathrooms, living rooms and closets, each used in daily living.

While this answer contains the same higher context as answer #5, it is developed in a unistructural manner; one central theme with attendant details. 2 points.

2. It is true that better answers are generally longer than poorer answers, but the reverse is not true; longer answers are not better answers. It is possible to write pages and pages of answers, and say practically nothing. This is known as "padding" and is not appreciated. Generally, this shows the writer has not yet abstracted out the essential points from the background mass of details.

3. If you study by memorization, you'll average 2 out of 5, if you memorize very well, possibly 3 out of 5. This may be enough to pass the course. If you think about the material, structure it while you study, and incorporate and apply the information into your own perspectives and understanding, you will do very well.

Cognitive structures are often imagined as hierarchies. The following are hypothetical structures for each of the above answers.

```
#1.      Stove                Chairs            Curtains


#2.                          Kitchens
                           -refrigerator
                            -stove
                            -cupboards
                            -table
                            -(microwave)


#3.  (& #4)                     Kitchens
                                   I
                                   I
          ------------------(food)-----------------------
          I                 I               I              I
          I                 I               I              I
       refrigerator       stove         cupboards        table
       -keeps cool       -warming       -storage       -sitting
       -food spoils       -cooking       -utensils       -eating


#5.                                House
                                     I
                                     I
         ----------------------------------------------------
         I                I                I                I
         I                I                I                I
     Kitchens         Bedrooms       Living rooms      Closets
         I
      (as above)
```