UNIVERSITY OF ALBERTA

THE DEVELOPMENT OF PLANNING SKILLS IN CHILDREN

BY

RAUNO KALEVI PARRILA ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1996

ISBN 0-612-18092-1

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR: Rauno Kalevi Parrila

TITLE OF THESIS: The Development of Planning Skills in Children

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1996

Rauno K. Parrila

Karilantie

40420 Oulu

FINLAND

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 Development of Planning Skills in Children submitted by Rauno K. Parrila in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dr. J. P. Das, supervisor

Dr. R. H. Short

Dr. L. L. Stewin

Dr. F. D. Snart

Dr. J. Bisanz

Dr. R. J. Sternberg, external

Date: July 27, 1996

To: The Faculty of Graduate Studies and Research
University of Alberta

From: Dr. J. P. Das, Professor
Department of Educational Psychology
Faculty of Education
University of Alberta

Date: July 3, 1996

To whom it may concern,

This letter is to certify that my student, Rauno K. Parrila, has been given permission to reproduce the article "Development of Plannig and Its Relation to Other Cognitive Processes" authored by Rauno K. Parrila, J. P. Dash, and U. N. Dash as Chapter III of his dissertation submitted to the Faculty of Graduate Studies and Research as a partial fulfillment of his degree of Doctor of Philosophy.

Sincerely Yours,

Dr. J. P. Das, Professor
Dr. U. N. Dash, Professor

To: The Faculty of Graduate Studies and Research
University of Alberta

From: Dr. J. P. Das, Professor
Department of Educational Psychology
Faculty of Education
University of Alberta

Date: July 3, 1996

To whom it may concern,

This letter is to certify that my student, Rauno K. Parrila, has been given the rights to reproduce figures of the Crack-the-Code task, in his thesis submitted to the Faculty of Graduate Studies and Research as a partial fulfillment of his degree of Doctor of Philosophy (Appendix A, pp. 149-154).

Sincerely Yours,

Dr. J. P. Das, Professor

Dedicated

to

my late father Eero,

who taught me to question;

to

Dr. J. P. Das,

who taught me how to find answers;

and to

my wife Persephone,

who helped me to live with unanswered questions.

ABSTRACT

This thesis consists of three separate papers. The first paper discusses Vygotsky's ideas about the role of language in the development of planning skills in children. The relation between language and planning is discussed from three perspectives: (a) the regulatory function of language; (b) language as "a tool of thought"; and (c) the role of communication and social interaction in the development of planning. The article concludes with considerations of the possible educational implications of Vygotskian ideas.

The second paper presents an empirical study conducted in India. The purpose of this study was to (a) delineate the course of development of school children's planning skills, (b) examine whether different levels of planning exist, and (c) explore the role of other cognitive processes as cognitive correlates of planning. Two-hundred fifty students, 50 from each of the five targeted grades (3, 5, 7, 9, and 11), completed a battery of planning, attention, and simultaneous and successive processing tasks. MANOVA indicated that the main effect of Grade was significant, whereas Gender had no siginificant effect on planning. Correlation and regression analyses showed that the relation between planning tasks and attention, and simultaneous and successive processing scores varied as a function of the planning task and the grade level. Furthermore, two planning factors were found and cluster analyses of variables indicated that one of the tasks, Crack-the-Code, may represent a different kind of planning than the other planning tasks used.

The third paper reports an empirical study conducted in Canada. Twenty-nine students (10 from Grade 4, 10 from Grade 6, and 9 from Grade 8) completed six items of the computerized Crack-the-Code task while thinking aloud. Both computer and verbal protocols were used to examine developmental and individual differences in the four planning components: representation, anticipation, execution, and regulation. Time and accuracy measures together with error analyses indicated that (a) older participants solved more items correctly within the same time, and (b) Grade 4 participants accepted incorrect

answers more readily and were reluctant to replace a disk that had already been placed. Analyses of both verbal and computer protocols, in turn, demonstrated that (a) more planning components were present in the performances of older participants, and (b) they also used these components more efficiently than the younger participants.

ACKNOWLEDGEMENTS

Several individuals have provided helpful comments and suggestions for various parts of this disseration. First and foremost, my sincere gratitude goes to my supervisor, Dr. J. P. Das, for constantly challenging and supporting me throughout the preparation of the dissertation. I would also like to thank Dr. B. Gindis, Dr. R. H. Short, and Mr. C. P. Boodt for their comments on Chapter II, Dr. U. N. Dash, T. O. Maguire and anonymous reviewers for their comments on Chapter III, and Dr. J. Bisanz, Dr. R. H. Short, Dr. F. Snart, and Mr. T. Papadopoulos for their comments and suggestions on Chapter IV.

Dr. S. Hein, Ms. M. Kendrick, and Ms. P. Natos have greatly improved the quality of my writing for which I am grateful. I would also like to acknowledge the Social Sciences and Humanities Research Council of Canada for providing the financial assistance for my graduate studies. Finally, I am thankful to my family, both here in Canada and in Finland, for their understanding and encouragement throughout my years of graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Planning has been an important concept in cognitive psychology since the emergence of this field. Until recently, the development of planning skills has not received the attention of researchers that it rightfully deserves. In a fast changing world, planning skills may well be one of the best predictors of an individual's successful adaptation. While knowledge that is important today may become obscure tomorrow, planning skills are not likely to lose their importance. On the contrary, when an individual's environment is saturated with rapidly increasing amounts of data, good planning skills may be a necessary prerequisite for maintaining one's intellectual independence and integrity.

This dissertation consists of three separate studies, each dealing with the development of children's planning skills from a distinct but interrelated point of view. The first study (Chapter II) examines the role of language in planning development. It presents briefly some of Vygotsky's original ideas about the role of language in mental functioning and then considers how these relate to the development of planning skills in children. The second study (Chapter III) examines the relation of planning to other cognitive processes identified by the PASS model, namely attention and simultaneous and successive processing. Moreover, the second study attempts to delineate the course of development of children's planning skills during school years and examines whether different levels of cognitive planning exist. The third study (Chapter IV) focuses on individual and developmental differences in the planning process. The common content of the first two studies is the identification of the cognitive correlates of planning. The third study attempts to divide cognitive planning into four components — representation, anticipation, execution, and regulation — and examines development of planning as an interplay of these components. Thus, in their distinct ways, all three studies address the question: What is planning development the development of?

Language and Planning in Children

Chapter II presents a paper that was published in *School Psychology International*'s special issue on L. S. Vygotsky and contemporary school psychology (Parrila, 1995). The chapter discusses Vygotsky's ideas about the role of language in the development of cognitive functioning in general and in the development of planning skills in particular. Planning is broadly defined in this chapter as a self-organizing reflective activity that integrates information from external and internal sources to create meaningful behavioral responses and environments. Moreover, two additions are suggested: (1) planning is a cognitive function that can control and regulate the functioning of other cognitive processes, as well as behavioral responses, in the service of some goal or purpose; and (2) as a higher cognitive function, planning is mediated by a symbolic system.

Language is introduced as the most important symbolic system that mediates thinking and planning. In Vygotsky's terminology, language is considered to be a "psychological tool" that acts as a mediating factor between environmental stimuli and an individual's response. In this sense, the concept of language is used to refer to one's internal language, or "inner speech". The relation between planning and inner speech is discussed from two perspectives: (a) the regulatory function of language, and (b) language as "a tool of thought".

When children's behavior becomes verbally mediated, they gradually gain better control of both their internal and external functioning, that is, they learn to regulate their behavior verbally and internally, and thus break the simple stimulus-response chain. As a consequence, behavior acquires a two-phase structure, which is a necessary condition for planning. Chapter II reviews both Soviet (Luria, 1959; Levina, 1981; Vygotsky, 1978) as well as some Western (Díaz, Neal, & Amaya-Williams, 1990) ideas about the way in which the capacity to regulate one's behavior according to a plan develops. Briefly, this development could be summarized as follows: firstly, the speech of other people obtains a directive function in organizing the child's behavior; secondly, verbally mediated self-control develops, enabling the child to comply with another person's directives in the absence of that person; and, finally, the child's verbally mediated self-regulation, or the "capacity to plan, guide, and monitor his or her behavior from within and flexibly according to changing circumstances" (Díaz et al., 1990, p. 130), takes over the role of organizing the child's behavior. According to Levina (1981), children should be capable of "inner speech planning" towards the end of the first decade of their lives.

The role that language plays in thinking has been debated vigorously, first in philosophy and later in psychology and psycholinguistics. In his review, Jenkins (1969) suggested that there are three major established positions on the relations of thought and language. The first position denies the difference between these two entities, or asserts that thinking cannot be studied separately from language. The extreme, and outdated, form of this position is best illustrated by Watson's (1930, p. 238) famous argument that "what the psychologists have hitherto called thought is, in short, nothing but talking to ourselves." The position that thinking cannot be studied separately from language has its roots in the writings of von Humboldt (1836/1989) who saw speech playing a decisive role in objectifying and, consequently, in developing one's thoughts:

> Without this feature, that is, without this continuous regression of objectivity to the subject, in which language collaborates, the formation of concepts (and consequently all true thinking) is impossible .... speech is a necessary condition for reflection in solitude. As a phenomenon, however, language develops only in social intercourse, and humans understand themselves only by having tested the comprehensibility of their words on others. (p. 101).

At the heart of this argument is the methodological problem of how to grasp thought processes and their results without reference to linguistic processes.

The second position is opposite to the first and argues that language is dependent on thought. The extreme form of this argument is not popular, but many theorists, especially in the field of cognitive psychology, have argued in favor of the independence of thinking from language. Usually they assume that cognition precedes language, but language, as a special tool, then affects and aids thinking in various ways. However, language is not usually seen as indispensable to thinking and some other signal system could serve the role that language customarily plays (Jenkins, 1969).

The third position states that thought is dependent on language. To varying extents and in slightly different forms, this position has since been advocated by such influential theorists as Sapir (1921), Whorf (1956), and, more recently, McNeill (1987), Hunt and Agnoli (1991), and Lucy (1992). Vygotsky (1987) essentially took this position in emphasizing the role of inner speech in forming thoughts. Although different in form and function from external speech, Vygotsky suggested that inner speech nevertheless has its origins in external, social speech from which it develops via egocentric speech. Vygotsky (1987) saw inner speech as "an internal plane of verbal thinking which mediates the dynamic relationship between thought and word" (p. 279). He suggested that thought undergoes several changes as it is formed in speech. He saw thought to be initially an "amorphous whole" that has to find its expression in speech. During this process, thought is first formed in inner speech and then in outer, external speech or writing. Both inner speech and external speech do not merely express thought but also change it to fit the available means of expression (see Luria, 1982, for a detailed description of this process). Thus, inner speech plays an important function in mediating between thought and its external expression, the function of which is communication.

Natural language, that is, the linguistic code, also is the main tool for communication (both oral and written). Thus, Chapter II also discusses what role communication and social interaction in general play in the development of planning. The argument essentially proceeds as follows: The early transmission of culturally accumulated knowledge takes place almost exclusively in the child's interaction with adults and more competent peers. Inadequate communication skills can lead to insufficient acculturation due to a limited access to information, which, in turn, can affect the child's cognitive functioning. The detrimental effects of insufficient acculturation, in turn, should be most prominent in those areas of functioning that are profoundly socially determined. Since everyday planning is profoundly socially determined and the bulk of planning skills are learned through verbally mediated social interactions with other, more capable planners, then speech plays a significant role in learning to plan. Speech is the tool for communication and further, for reaching intersubjectivity (Rommetveit, 1974), or at least some level of shared understanding of the planning situation that is a necessary condition for teaching/learning planning.

Chapter II concludes by discussing the educational implications of the presented ideas. The main themes in this discussion are (a) the distinction between empirical and theoretical concepts and learning, and (b) the form that a "Vygotskian" instructional dialogue would take. I will return to some of these ideas in the Conclusion chapter.

Planning in Relation to Other Cognitive Processes

Chapter III of this dissertation presents a study that is accepted for publication by the *Journal of Applied Developmental Psychology* (Parrila, Das, & Dash, in press). Whereas the first study was essentially a review paper that focused on some of Vygotsky's ideas and developed them in the context of planning development, the second study is empirical and utilizes quantitative research methods to examine three questions: (1) Do planning skills develop uniformly over the school years? (2) Is cognitive planning a uniform psychological construct? and (3) What role may other cognitive processes play in planning development?

The research that is reviewed in the introduction section of Chapter III suggests that at approximately age 10, children are already capable of performing simple planning tasks at or near adult performance level, whereas with the more complex tasks, development continues well into adolescence. The first goal of Chapter III is to replicate these findings with a more representative and larger sample of school-age children. Moreover, in order to increase the comparability and replicability of our results, we mainly used standardized planning measures rather than constructed new ones. Our results essentially confirmed that different tasks have different discriminating values for different age groups. The most complex planning task, Crack-the-Code, did not differentiate between the three younger grade levels. In contrast, neither Matching Numbers nor Planned Search differentiated between the three oldest grade levels suggesting that dramatic developmental changes take place earlier in these simple planning tasks.

Our second question concerns the uniformity of planning as an explanatory construct of cognitive functioning. The idea that planning is not a 'monolithic' construct is not new (see for example, Welsh, Pennington, & Groisser, 1991). Even the definitions of planning offered by different researchers are often incompatible. As a consequence, two planning studies seldom share the same focus, often making the results from different studies difficult to compare. With little exaggeration, one could argue that planning research has failed to produce accumulative knowledge because of extensive fragmentation and a lack of a unifying framework. To overcome this fragmentation and to give structure to our research, we suggest a tripartite model of planning for conceptualizing planning and its relationship to such closely related concepts as problem-solving or strategies. Activity-planning, action-planning, and operation-planning are introduced briefly as a nested system that can be used to explain behavior on both molar and molecular levels of analysis. We believe that this model can have great intuitive value both in guiding future research and in making sense of existing findings.

The last question addressed in Chapter III has two parts: (1) What are the "cognitive correlates" of effective performance in action-planning (or complex planning) tasks on the one hand, and in operation-planning (or simple planning) tasks on the other? (2) Are these cognitive correlates essentially stable across different age-groups? The rationale for studying cognitive correlates of planning is simple. Good planning is at least partly based on the fluent functioning of such other cognitive processes as encoding, attention processes, and working memory. Yet few studies have concentrated on this question and even fewer have produced significant findings. The most consistent finding in the research seems to be that "general intelligence", as measured by various traditional IQ tests, does not correlate significantly with planning skills. The possible cognitive correlates examined in the second paper were taken from the PASS (Planning-Attention-Simultaneous processing-Successive processing) cognitive processes (see e.g., Das, Kirby, & Jarman, 1979; Das, Naglieri, & Kirby, 1994; Naglieri & Das, 1990). Since the PASS model treats planning, attention, simultaneous and successive processes as interdependent and jointly responsible for producing mental activity, these processes were regarded as particularly suitable candidates for the cognitive correlates of planning development. Not surprisingly, our results show that significant predictors change both as a function of age and as a function of the task. Development of planning is not only quantitative but also qualitative and, accordingly, different components should not be expected to correlate equally in different ages.

Developmental Changes in the Planning Process

Chapter IV presents an as yet unpublished study that focuses on developmental and individual differences in planning. Chapter IV describes in some detail how participants in three different grade levels solve a complex computerized action-planning task. The main source of information on the participants' planning processes were their concurrent think-aloud protocols that were analyzed together with computer protocols. Thus, one particular strength of this study was a rich protocol data from two independent sources.

In Chapter IV, planning is defined as a self-organizing reflective activity that integrates information from external and internal sources to create and execute meaningful behavioral responses. Moreover, we suggest that the planning process can be divided into four components: representation, anticipation, execution, and regulation. While these components are briefly introduced in the second article, the third article defines them in more detail. The four components have their roots in the neuropsychological literature on planning as a frontal lobe function (Fuster, 1989; Luria, 1973; Stuss & Benson, 1986, 1987). The definitions offered in Chapter IV attempt to "redefine" these concepts on the cognitive level of analysis. Thus, one purpose of the study was to examine whether a model of the planning process derived from neuropsychological literature could be successfully used to describe developmental changes in planning.

The result section of Chapter IV is lengthy and mostly descriptive for two reasons: Firstly, the task, Crack-the-Code, is relatively new (the computerized version was developed for this study) and has never been explored in this detail before. As a consequence, it is important to provide readers with sufficient information about the task and the results to allow them to judge which elements of our findings may be artifacts of the task and which are more generalizable across different tasks. For example, the results of this study indicate that the time a participant spent looking at the item before moving any disks (first-move latency) was not correlated with the accuracy of the solution, or even with the accuracy of the first move. It is possible that this time, which should reflect advance planning, may with this particular task reflect the visual processing demands of the colorful display.

The second reason for the lengthy treatment of results is that the development of the four planning components and how they function in unison to produce appropriate planning responses has also not been directly addressed before. For this reason, we were able to suggest only a few hypotheses in the introduction to guide the data analyses. Moreover, many of our findings seem to gain more importance when cross-referenced with data from other sources. Thus, the general discussion section of Chapter IV is devoted to the integration of information from different sources.

The dissertation concludes with a discussion of the main findings and limitations of the presented studies. Also, educational implications and some ideas for future research are presented in Chapter V.

References

Das, J. P., Kirby, J., & Jarman, R. F. (1979). *Simultaneous and successive cognitive processes.* New York: Academic.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes.* Needham Heights, MA: Allyn and Bacon.

Díaz, R. M., Neal, C. J., & Amaya-Williams, M. (1990). The social origins of self-regulation. In L. C. Moll (Ed.), *Vygotsky and education* (pp. 127-154). New York: Cambridge University Press.

Fuster, J. M. (1989). *The prefrontal cortex.* (2nd ed.) New York: Raven Press.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review, 98,* 377-389.

Jenkins, J. J. (1969). Language and thought. In J. F. Voss (Ed.), *Approaches to thought* (pp. 211-237). Columbus, OH: Merrill.

Levina, R. E. (1981). L. S. Vygotsky's ideas about the planning function of speech in children. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 279-299). New York: M. E. Sharpe.

Lucy, J. A. (1992). *Language and diversity of thought.* Cambridge: Cambridge University Press.

Luria, A. R. (1959). The directive function of speech in development and dissolution. *Word, 15,* 341-352.

Luria, A. R. (1982). *Language and Cognition.* New York: John Wiley & Sons.

Luria, A. R. (1973). *The working brain.* New York: Basic Books.

McNeill, D. (1970). *The acquisition of language: The study of developmental psycholinguistics.* New York: Harper & Row.

Naglieri, J. A., & Das, J. P. (1990). Planning, attention, simultaneous and successive (PASS) cognitive processes as a model for intelligence. *Journal of Psychoeducational Assessment, 8,* 303-337.

Parrila, R. K. (1995). Vygotskian views on language and planning in children. *School Psychology International, 16,* 167-183.

Parrila, R. K., Das, J. P., & Dash, U. N. (in press). Development of planning and its relation to other cognitive processes. *Journal of Applied Developmental Psychology.*

Rommetveit, R. (1974). *On message structure.* London: John Wiley & Sons.

Sapir, E. (1921). *Language: An introduction to the study of speech.* New York: Harcourt & Brace.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Stuss, D. T., & Benson, D. F. (1987). The frontal lobes and control of cognition and memory. In E. Perecman (Ed.), *Frontal lobes revisited* (pp. 141-158). New York: IRBN Press.

von Humboldt, W. (1989). The nature and conformation of language. In K. Mueller-Vollmer (ed.), *The hermeneutics reader* (pp. 99-105). New York: Continuum.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber, & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky* (Vol. 1, pp. 39-285). New York: Plenum Press.

Watson, J. B. (1930). *Behaviorism.* Chicago: University of Chicago Press.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function. *Developmental Neuropsychology, 7,* 131-149.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf.* Cambridge, MA: The MIT Press.

II. VYGOTSKIAN VIEWS ON LANGUAGE AND PLANNING IN CHILDREN

Introduction

The purpose of this article is to briefly present some of Vygotsky's original ideas about the role of language in mental functioning and then consider how they relate to the development of planning skills in children. Since Vygotsky wrote relatively little about planning skills and what may effect their development, his ideas are complemented with applicable contemporary research where necessary. Finally, the article concludes with some consideration of the educational implications of the ideas discussed.

Planning is defined here as a self-organizing reflective activity that integrates information from external and internal sources to create meaningful behavioral responses and environments (see also Das, Naglieri, & Kirby, 1994; Parrila, Äystö, & Das, 1994). Planning is a cognitive function that can control and regulate the functioning of other cognitive processes, as well as behavioral responses, in service of some goal or purpose. As a higher cognitive function, planning is mediated by a symbolic system, most often by language. Planning a lesson or an experiment are good examples of planning as it is defined here.

Until recently, planning skills have not received the attention of researchers and educators that they rightfully deserve. In a fast changing world, planning skills may well be one of the best predictors of an individual's later successful functioning. While knowledge that is important today may become obscure tomorrow, planning skills will never lose their importance. On the contrary, when an individual's environment is saturated with rapidly increasing amounts of data, good planning skills may be a necessary condition for maintaining one's intellectual independence and integrity.

Since planning is defined here as a semiotically mediated activity, it is necessary to first review briefly the role of language and semiotic mediation in Vygotsky's theory.

Vygotsky on Semiotic Mediation

There are three general themes that run through Vygotsky's writings and that according to Wertsch (1990) outline his theoretical vision: (1) reliance on a genetic, or developmental method; (2) the claim that higher mental functions in the individual have their origins in social activity and interaction; and (3) the claim that the defining property of human mental functioning is its mediated structure, in other words, that semiotic systems such as human language mediate social and psychological processes.

The three general themes are closely intertwined. At the center of Vygotsky's developmental theory is the claim that the most important qualitative transition during development occurs when behavioral responses to environmental stimuli acquire mediated structure. The mediated structure of behavior means that language, or some other sign system (which Vygotsky called "psychological tools"), always acts as a mediating factor between environmental stimuli and an individual's response. For example, when we are engaged in planning, we form a mental representation of the situation and our actions with the help of words (or other suitable symbols) prior to acting. As a result of mediation by signs, the whole process of behavior changes:

> The inclusion of a tool in the process of behavior (a) introduces several new functions connected with the use of the given tool and its control; (b) abolishes and makes unnecessary several natural processes, whose work is accomplished by the tool; and (c) alters the course and individual features (the intensity, duration, sequence, etc.) of all the mental processes that enter into the composition of the instrumental act, replacing some functions with others (i.e., it re-creates and reorganizes the whole structure of behavior just as a technical tool re-creates the whole structure of labor operations). (Vygotsky, 1981b, pp. 139-140)

Moreover, when children's behavior and intrapsychological cognitive functions acquire a verbally mediated form, the process of development profoundly and irreversibly changes:

> *The nature of the development itself changes,* from biological to sociohistorical. Verbal thought is not an innate, natural form of behavior, but is determined by a historical-cultural process and has specific properties and laws that cannot be found in the natural forms of thought and speech. (Vygotsky, 1986, p. 94)

The role of speech as a mediational tool is also central with regards to Vygotsky's second theme: social origins of higher mental functions. According to Vygotsky (1986), the most important function of speech is communication, i.e., facilitating social contact. Communication is necessary for the successful internalization of social activities and thus for the emergence of higher cognitive functions. In short, Vygotsky (1981a) believed that this process consists of the following steps: firstly, other people act on the child;

secondly, the child enters into interaction with those around him or her; thirdly, the child begins to act on others; and, finally, the child begins to act on himself or herself, i.e., the child has internalized the social interaction as a part of his or her cognitive functioning. An important part of this process is the internalization and subsequent transformation of external, social speech into internal, private inner speech (Vygotsky, 1936).

To summarize, language plays a central role in all the major themes of Vygotsky's writings. Language enables us to interact with more capable peers and adults and later with written material etc., thus giving us a chance to share the accumulated knowledge of our environment. Language is also the major source of signs and concepts used in semiotic mediation at the intrapsychological level, i.e., in inner speech and verbal thinking. The latter, in turn, make up most of the psychological tools that enable abstract, conceptual thinking. Furthermore, when language is used as a mediational means — either on the intra- or the interpsychological level — it anchors mental functioning to cultural, historical and institutional settings (Wertsch, 1990). When children learn to speak they acquire a system of signs that develops according to sociohistorical principles and attains meaning from the sociocultural settings in which it is learned.

Language and Planning

Since planning is a higher cognitive function mediated by a symbolic system, the relationship between language (or speech[1]) and planning is essentially the same as that between language and thinking. According to Vygotsky (1986), language and thinking have different developmental roots and are separate abilities in small children. He suggested further that the moment that language and thinking converge is the most significant moment in the course of intellectual development. This moment "gives birth to the purely human forms of practical and abstract intelligence" (Vygotsky, 1978, p. 24). Despite this convergence, language and thinking do not become one; instead, their relationship remains variable throughout development. Vygotsky (1986) expressed this relationship as follows: "Schematically, we may imagine thought and speech as two intersecting circles. In their overlapping parts, thought and speech coincide to produce what is called verbal thought. Verbal thought, however, does not by any means include all forms of thought or all forms of speech" (p. 88).

Vygotsky implied that language is an important factor in the development of

---

[1] The terms language and speech are used interchangeably in this article. See Wertsch, 1990, for the distinction between the two.

thinking in at least three different ways: First, with the help of language, children gradually gain better control of both their internal and external functioning, i.e., they learn to regulate their behavior verbally and internally, and thus break the simple stimulus-response chain. As a consequence, behavior acquires a two-phase structure, which is a necessary condition for planning. Secondly, abstract thoughts are formed with the help of language: Language is the means by which the individual becomes capable of rising above the sensorimotor level of thinking. He suggested further that thoughts are not only expressed in language but they are formed in language (Vygotsky, 1986). And, lastly, language is the means by which children get access to the accumulated sociocultural experience in their environment. We shall discuss each of these points briefly below.

The Regulatory Function of Language

The act of planning implies that the planner is capable of regulating his or her behavior according to the plan — planning devoid of this possibility is equal to day-dreaming or stating the obvious. The capacity to regulate one's behavior according to a plan is, however, not present in infants but develops gradually during childhood. According to Luria (cited in Vocate, 1987), the developmental stages of speech in organizing the child's behavior are: (1) speech having no effect; (2) external speech having an activating effect on the child's movements but no inhibitory effect; (3) external speech eliciting significative connections and organizing the child's movements; and (4) control shifting to the child's internal speech where there is an attendant formation of verbal principles that guide the child's movement.

Luria (1959) reported several experiments that demonstrated the first three of these developmental stages. According to these experiments, the behavior of children as young as one year of age can be directed by adult's speech in a simple situation that lacks conflict. This directive role of speech, however, is not maintained if the speech conflicts with the inert connections that arise at an earlier instruction or which began with the child's own activity. Furthermore, Luria demonstrated that during the second year of children's lives, visual signals are more effective than verbal signals in overcoming the inertia of motor connections. Towards the end of the third year, however, speech starts to play a strong excitatory role in children's behavior. The directive role of speech becomes central, but children still have trouble with the semantic aspects of the message. As a consequence, the inhibition of established response connection with the help of speech is still difficult. Around the age of four years, the semantic aspect of a message finally becomes dominant and the directive role of speech becomes more complete.

The directive function of speech is essential for the development of verbally mediated self-control. Verbally mediated self-control is the capacity to comply with another person's commands and directives in the absence of the person (Díaz, Neal, & Amaya-Williams, 1990). It is a necessary precursor of self-regulation that, in turn, can be defined as "the child's capacity to plan, guide, and monitor his or her behavior from within and flexibly according to changing circumstances" (Díaz et al., 1990, p. 130). These two functions have a distinctly different structure. According to Díaz et al. (1990), in self-control the behavior is produced in response to internalized commands or directives; it is organized in rigid S-R connections that environmental cues activate. In self-regulation, on the other hand, behavior is organized into hierarchical, functional systems that can be adjusted to changing goals and situations. Furthermore, behavior is now guided by self-formulated plans and goals and aspects of the environment are used as tools and mediators to attain the goals. In other words, self-control is ultimately externally directed and linear in structure, whereas self-regulation is an active and flexible form of adaptation under internalized verbal control.

The origins of verbal self-control, and consequently self-regulation, are social: the development of verbal self-control starts when the caregivers try to divert children's attention and initiate specific actions through verbal requests (Schubert, 1983). Levina (1981), in presenting Vygotsky's ideas on the regulative function of language, argued that the regulative function of children's own speech also appears first in the social, or "interpsychological" (Vygotsky, 1978), level, i.e., children will try to regulate others' behavior with the help of speech before their self-regulative capacities emerge. Levina suggests further that the child's early behavior is not goal directed in the true sense of the notion; it is more likely to be influenced by factors in the external environment than by a conscious goal-directed plan. For Levina, speech is the means by which the child gains the power to impose a goal on the environment. With the help of speech the child is able to 'rearm' more primitive concrete-visual cognition with analytical weapons — words. As a result, the child's cognitive operations gain flexibility, freedom, and independence from the concrete stimulus field. With the help of speech, children can, for example, bring to their problem solving processes elements that are not immediately present (Díaz et al., 1990). In this way speech allows children to liberate themselves from immanent but nonessential aspects of the environment and to focus on the essential: the representation of future action (see also, Vygotsky, 1978).

However, before all this is possible, a complicated developmental process lasting most of the child's first decade of life is required. The first stage of this process is characterized by speech that describes ongoing action, i.e., "constituting" speech

(Vygotsky, 1978; Levina, 1981). A positive signal of the first stage is the presence of speech that accompanies children's activity during the performance of any task. Constituting speech appears well before the elements of verbal planning begin to appear in the child's speech. At this stage speech does not yet direct action but is thoroughly intertwined with other elements of the child's behavior. However, speech plays a useful role as a device for exploring and labeling the environment and accumulating experience. Gradually, and mostly through older individual's example, children's speech takes on a more indicatory role and children start to use speech to separate the environment into objects, first as part of social communication and then for themselves. In this way speech becomes an investigative tool, a mechanism for mastering the surrounding world (Levina, 1981).

It is important to notice that children's behavior during the first stage may often appear to be planful. This planfulness is, however, prelingual and more of a property of the task or of the environment than that of the child. Speech is not used to control and regulate action but to accompany it, and problems in practical situations are solved very infrequently and only in cases in which verbal planning is observed (Levina, 1981).

At the next stage, children's speech becomes less concerned with the surrounding world and more concerned with their actions. Utterances concerned with representing the problem become more noticeable and utterances having to do with verbal planning gradually begin to appear. Speech is no longer restricted to describing the action or behavior but begins to assume an inclusive nature: it sums up the action in which the child is currently engaged. According to Levina, summarizing speech that recapitulates the basic action is the genetic precursor to planning speech. The only difference is that in planning speech the child first arrives at the connections in words and then carries out the plan in action. In this way, the child's attempts acquire a two-phased structure: they are first prepared by a verbal plan and only after that they are put into action (Vygotsky, 1978; Levina, 1981). This signals the beginning of verbal planning.

Planning for others is the first stage in the development of verbal planning. The first plans are positively present when a child turns to peers or adults and tells them about future action before carrying it out. At this stage, it becomes possible to stimulate verbal planning by asking the right questions or giving suggestions to children who are trying to solve complex problems. Also, if children's speech is prevented, they will stop planning and their behavior regresses to the earlier level (see Sokolov, 1972, for a review of this phenomenon). The following quotation from Levina (1981) summarizes clearly the significance of speech at this stage:

> First, while trying to get the goal-object in a stereotyped manner, the child is restricted to the visual field. The need to formulate his/her intentions verbally is the motivating force for examining the elements of the situation. A verbal formulation emerges in his/her actions; and, consequently, the need to examine the situation emerges. At first, only what is directly presented to the eye exists. Then the child begins to look around. Because of this, he/she increases his/her ability to overcome difficulties. The child begins to look around in the strict sense, but he/she also "looks around" at his/her past experience and thus goes beyond the boundaries of the immediately present, concrete situation. . . . The child moves from the existence in the concrete situation to an existence in time by means of the speech that precedes his/her actions. (pp. 293-294)

Vygotsky (1978) also emphasized the emergence of the planning function of speech as a landmark in the child's psychological development. Speech used for planning is not simply an appendage of behavior, rather, it has the essential function of guiding and directing all of the child's behavior (Levina, 1981). Like other cultural functions, verbal planning emerges first as a function created by others and directed towards others, that is, in the interpsychological level (Vygotsky, 1978). Then, without changing externally, it is directed toward one's own behavior (in the form of egocentric speech); and, finally, it is transformed by ceasing to appear externally and is abbreviated to form an internal device, inner speech. For Vygotsky this later development had significant consequences:

> The greatest change in children's capacity to use language as a problem-solving tool takes place somewhat later in their development, when socialized speech (which has previously been used to address an adult) *is turned inward.* Instead of appealing to the adult, children appeal to themselves; language thus takes on an *intrapersonal function* in addition to its *interpersonal use.* When children develop a method of behavior for guiding themselves that had previously been used in relation to another person, when they organize their own activities according to a social form of behavior, they succeed in applying a social attitude to themselves. The history of the process of *the internalization of social speech* is also the history of the socialization of children's practical intellect. (Vygotsky, 1978, p. 27)

According to Levina (1981), verbal planning will assume this internal form — inner speech planning — by the age of eight to ten years. After this, further development takes the form of strengthening of inner speech and disappearing of external speech. At this later stage external speech ceases to be a useful device in thinking and can even inhibit cognitive operations (Sokolov, 1969).

Language, Inner Speech, and Communication

When writing about the relationship between thought and language, Vygotsky (1987) emphasized the role of inner speech in forming thoughts. The term inner speech is usually used to signify "soundless, mental speech, arising at the instant we think about something, plan or solve problems in our mind, recall books read or conversations heard, read and write silently" (Sokolov, 1972, p. 1). In these instances, inner speech is equated with speech to oneself, or concealed verbalization. Vygotsky (1987) saw inner speech somewhat more broadly as "an internal plane of verbal thinking which mediates the dynamic relationship between thought and word" (p. 279). Vygotsky suggested that thought undergoes several changes as it is formed in speech. He saw thought to be initially an "amorphous whole" (1986, p. 219) that has to find its expression in speech. During this process, thought is first formed in inner speech and then in outer, external speech or writing. Both inner speech and external speech do not merely express it but also change it to fit the available means of expression (see Luria, 1982, for a detailed description of this process). Thus, inner speech plays an important function in mediating between thought and its external expression, the function of which is communication. Although different in form and function from external speech, inner speech nevertheless has its origins in external, social speech from which it develops via egocentric speech (Vygotsky, 1986).

If we accept the linguistic genesis of inner speech and its role in forming thoughts, we have to assume that the possession of relevant linguistic knowledge can determine an individual's success in a given planning situation. Furthermore, this leads to the idea that natural language per se plays a role as a source of inner speech. Thus, the extensiveness of our vocabulary and understanding of relevant concepts should affect our performance in planning tasks. This assumption, however, needs to be qualified: symbols used to solve a planning task that has very little linguistic content may be already so far removed from their linguistic origins that their verbal communication, and thus the verification of their existence with the help of natural language, may be difficult. Take mathematics as an example. The symbol system used in mathematics is defined with the help of natural language but later gains 'a life of its own' to the extent that a mathematician may have overwhelming difficulties in transforming his or her solution to words. Also, as Huttenlocher (1976) notes, verbal fluency involves several special skills that do not necessarily influence one's fluency in thinking, just one's competence in communicating the results of this process.

Furthermore, verbal mediation may have a detrimental effect on performance, as suggested by Kearins (1981, 1986). By using visual spatial memory tasks, Kearins showed that the Australian aboriginal children who showed very little evidence of using verbal mediation and, instead, seemed to rely on visual strategies, consistently outperformed Australian children of European origin who tried to employ strategies grounded on verbal mediation.

These findings suggest that the role of language in planning should always be considered in context with task demands. Planning tasks with more verbal content and of a structure that presupposes, or at least benefits from verbal mediation, should be affected by language abilities to a greater extent than planning tasks with less verbal content and, for example, more visual structure. Therefore, planning a driving route from one part of the town to the other may not be affected by language skills to the same effect than, for example, planning a driving lesson.

Language is also the most important medium for communication with others. Vygotsky (1978, 1986) emphasized the role of social interaction in the development of cognitive processes. In interaction with others, a child not only learns the language, which then becomes a source for inner speech, but also what is already known by the adults and more competent peers. The early transmission of culturally accumulated knowledge takes place almost exclusively in interpersonal relationships, i.e., in the child's interaction with adults and more competent peers. Thus, inadequate communication skills can lead to insufficient acculturation due to a limited access to information, which, in turn, can affect the child's cognitive functioning. Furthermore, the detrimental effects of insufficient acculturation should be most prominent in those areas of functioning that are profoundly socially determined.

The social nature of planning has been emphasized by several researchers (see, for example, Baker-Sennett, Matusov, & Rogoff, 1993; Goodnow, 1987). The decision to plan is in most cases socially determined. Not all situations warrant planning and children seem to acquire the knowledge about the appropriateness of planning in different situations during the school years (see, for example, Kreitler & Kreitler, 1987; Pea, 1982). Furthermore, as Baker-Sennett et al. (1993) have aptly expressed, "in everyday life, planning occurs in culturally organized institutions and social situations in which individuals work with others to prepare for and carry out joint action, often necessitating adjustments in planning to fit with the social distribution of both the planning and the execution of the plan" (pp. 271-272).

Thus, the plans we use are socially constructed and our interaction with others has a formative, as well as a facilitative, function in planning development. The bulk of planning skills are probably learned through verbally mediated social interactions with other, more capable planners. This implies that speech plays a significant role in learning to plan. Speech is the tool for communication, and further, for reaching intersubjectivity (Rommetveit, 1974), or at least some level of shared understanding of the planning situation, that is a necessary condition for teaching/learning planning.

Wertsch (1979) offers us a description of how this shared understanding develops during a dialogue. He was concerned with how children develop the ability to participate fully in a communicative situation involving a new type of activity, a simple problem solving task. He wanted to show how during the adult-child interaction the adult's regulation gradually becomes internalized by the child in his or her process of establishing and maintaining coherence between his or her own action and the adult's speech. According to Wertsch (1979), the child creates this coherence first "by carrying out the behaviors specified by the adult and *then* building a coherent account of the relationships among speech, definition of situation and behavior" (p. 20). This means that the child comes to share the adult's definition of the situation through following his or her instruction during a practical activity. At the same time, the child also internalizes the adult's verbal directives and they become part of the child's inner speech. Thus, the child both learns to understand the specific situation and internalizes a model of the process of building an understanding.

Wertsch's description can also be read as an example of how an adult can teach a child to use a plan of action in a simple problem solving situation. To be effective, this teaching necessitates successful communication of ideas in an increasingly complex level, in other words, the development of shared understanding through reoccurring dialogues. Wertsch (1979) states that "other-regulation by means of uninterpretable directives seems to be an important way of 'luring' the child further and further into the communication by building up his/her definition of situation" (p. 20). This communication, in turn, models the child's thinking processes towards those of the adult.

Perhaps the development of planning essentially consists of a series of analogous microgenetic processes that Wertsch described. More difficult tasks and more complex environmental constraints, especially when adolescents enter in work life that is completely governed by planning (Dreher & Oerter, 1987), demand increasingly sophisticated plans and planning procedures that individuals learn with the help of adults and more competent peers. Adults and peers act as models who engage in planning and demonstrate the usefulness and the right procedure. Through social interaction with an (ever wider) environment, individuals learn how to plan as well as what kind of plans work in different situations and why. Their planning processes become more sophisticated, and their planning repertoire both grows and gets sharper.

Implications for School Psychology

Vygotsky's ideas are empowering for educators: Education as a form of social interaction has a fundamental role in forming higher cognitive processes. We do not learn to think because our brain matures in certain ways. Instead, we learn to employ the maturing brain in certain ways due to acculturation into the sociocultural environment that we live in. A great deal of this acculturation takes place in formal schooling.

Schools are responsible for teaching the majority of literacy skills that are necessary for successful adaptation in society. "Schools should teach students to be literate in the most general sense — capable of reading, writing, speaking, computing, reasoning, and manipulating verbal (and visual) symbols and concepts" (Gallimore & Tharp, 1990, p. 192). One obvious implication of this to the relationship between planning and language is that teaching planning and problem-solving skills should not be left to the science or mathematics teachers exclusively. If language skills even partially determine our planning abilities, then teaching those skills should be a priority in any school system that concentrates on teaching thinking as well as curriculum content. But what exactly are the language or literacy skills that should be taught?

Teaching the central concepts in any subject is the way to start, as most teachers know. According to Vygotsky (1986), it is especially the "scientific" concepts that are learned through formal education. Scientific concepts are logically defined in relation to other concepts, as opposed to "spontaneous" concepts that emerge from the everyday experience of the children (Kozulin, 1986). Thus, scientific concepts focus student's attention on sign-sign relationship (Wertsch & Stone, 1985) and to language as a system of signs that can be manipulated. This, in turn, helps children to become conscious of the act of thinking itself. Hence, scientific concepts provide a system of generality that is necessary for the decontextualization of thinking processes (Lee, 1987).

But scientific concepts have to be connected to student's "spontaneous" concepts for understanding of them to be optimal. Gallimore and Tharp (1990, p. 194) suggest that this is the only way to achieve "the highest order of word meaning" and to "ensure that tools of thought will be manipulated for the solution of practical problems of the experienced world." Thus, the language that we teach should be rich both in its scientific and in its practical connections and connotations.

How scientific concepts are best taught and how the connection between scientific and spontaneous concepts is created are both crucial educational questions. An important distinction here is that between "theoretical learning" and "empirical learning" (see e.g., Davydov, 1990; Karpov & Bransford, 1995; Kozulin, 1990). In theoretical learning, a student is taught to understand a concept through its transitions and connections rather than through its verbal definition or observable examples. Thus, the student's attention is focused to the essential characteristics of the phenomenon to be learned; essential referring here to the connections between relevant phenomena and laws of their development (Davydov, 1990). After a concept is learned "theoretically", it is then used to solve concrete problems (Karpov & Bransford, 1995). In empirical learning, the realm of a concept is formed by comparing a number of objects or observable phenomena exemplifying it and identifying similar features in them. Thus, empirical learning is based on identifying common features in a group of objects or phenomena. The second step in empirical learning consists of defining the scientific concept behind the empirical phenomenon.

While both methods can lead to the same end result, i.e., obtaining a scientific understanding of a concept (however, see Davydov, 1990, for a counterargument) they do not teach the same thinking processes. Empirical learning implies that the essential characteristics of a concept can be derived from its observable characteristics through the processes of comparison, classification, and/or abstraction (Kozulin, 1990). Theoretical learning, in contrast, is based on conscious understanding of the essence of a concept or phenomenon within a system of interconnected phenomena. According to Kozulin (1990, p. 259), the goal of this understanding is "contentful abstraction", i.e., understanding of the central relationships in a symbolic form, which is then related to empirical manifestations of the concept in question. As a result, a second order abstraction of the concept is formed, which will gradually include both the scientific definition of the concept as well as its empirical applications. At first, theoretical learning sounds feasible only in upper grade levels but Davydov (1990) and Schmittau (1993) cite examples in which this method was used successfully with subjects as young as 7 to 9 years of age.

The second important implication of Vygotskian views is that communication skills and social interaction should receive more attention in educational research and practice. According to Cole (1990), studies of language use in schools have revealed a distinctive communication pattern called "instructional discourse". Instructional discourse aims at giving children information about the content of the curriculum and feedback of their success in learning it. The teacher, in turn, is provided with information regarding students' progress. Thus, during instructional discourse, the teacher first shares a part of his or her knowledge of the subject matter and then asks questions in order to find out how much students have actually learned, or memorized. The primary purpose of the questioning is to evaluate the student's performance level rather than to obtain more information about the subject matter. This type of questioning is rare outside of educational settings, and learning to respond to questions whose answer is already known to the teacher is an important task for a young learner (Mehan, 1979). Instructional discourse, however, may not be conducive for the formation of good thinking and study skills. In its simplest form, the teacher delivers the knowledge, the meaning, and the student's task is to memorize it in order to know the answer to the evaluative questions. Control is external and the thinking skills that are implicitly taught are severely limited.

Contrast the above (and maybe extensively negative) description of instructional discourse with "Socratic dialogue" in the SPELT (Strategies Program for Effective Learning/Thinking; see e.g., Mulcahy, Marfo, Peat, & Andrews, 1986; Peat, Mulcahy, & Darko-Yeboah, 1989), or with Palincsar and Brown's (1984, 1988) account of "reciprocal teaching" and how it can facilitate student's reading comprehension. During a Socratic dialogue the teacher leads the student through questioning to discover relationships between different phenomena. This type of dialogue follows more teacher focused sessions during which students are given the strategic knowledge base of a content area, and aims directly at generalization via communication and comparison of ideas. The teacher's responsibility is to guide the discussion and, consequently, students' thinking by asking leading questions, extending students' ideas, clarifying and challenging their responses etc.

In a similar manner, reciprocal teaching employs structured group discussions as a means to teach students to think while they are reading or listening a text. The discussion is conducted by the instructional group while the teacher participates both as a leader and as a respondent. The dialogue is structured to the extent that the teacher directs the discussion towards a set of predetermined strategies (in this case reading comprehension strategies of predicting, questioning, summarizing, and clarifying).

Reciprocal teaching is an excellent example of a structured dialogue that is aimed at generating the target processes within the participants, i.e., as their thinking processes. Furthermore, it utilizes both experts (teacher) and more knowledgeable peers (other students) as instruments of cognitive growth. As the authors summarize the approach, "in reciprocal teaching, the students are learning these [reading comprehension] strategies in a social context reflective of a social phenomenon with which students have much experience—discussion" (Palincsar & Brown, 1988, p. 58).

The reciprocal teaching approach could also be used successfully to teach the students both everyday and academic planning skills, such as applying for a job or composing an essay. This requires, however, that the teacher is cognizant of the planning requirements of the tasks and can structure the discussion so that these requirements will become a natural part of the dialogue. To define what is good planning for a certain age-group in any given task, we need to collect information on how those students who are highly successful perform the task. If this information is not available in educational literature, teachers can obtain it for example by asking successful students to analyze their thinking process (if they are old enough), or by asking them to think aloud while performing the task. When this procedural data from successful individuals is combined with the teacher's own task-analysis, sufficient information is available to decide on the structure of instructional discussions and the structure of the interpsychological planning processes that are being taught.

Summary

Sufficient language skills are essential for the development of higher level planning skills. Planning would not be possible without some kind of semiotic mediation that enables both the self-regulation and the restructuring of the decision making process that are necessary conditions for planning. When signs begin to mediate the decision making process, a new psychological process is created in which the direct impulse to react is inhibited, and, instead, the future action is first planned on a symbolic level before taking action. In this way, "the use of signs leads humans to a specific structure of behavior that breaks away from mere biological development and creates new forms of a culturally-based psychological process" (Vygotsky, 1978, p. 40).

The acquisition and development of planning skills depends on the social interactions that we enter into, as well as on the language skills available for planning and communication. If our tools for communicating are not up to the task, we are unlikely to obtain the complex planning skills from our environment. Similarly, if our psychological tools for thinking are not powerful enough, we are unlikely to come up with the best possible solution (Carroll, 1964). Although natural language is only one medium of communication and one source of psychological tools, it nevertheless is the most important one for both.

In Vygotsky's writings language has a profoundly empowering and "liberating" function for the individual. First, language gives us the control of our actions by breaking the stimulus-response chain. Later in ontogenesis, language helps us to distance ourselves from the experienced reality and reflect upon it on a conceptual level. And finally, literacy skills that we acquire in school open up the whole new world of dialogues, ideas, experiences, and possibilities beyond the space and time limitations of our physical existence. In this sense, language, an inherently social product, can also help us to rise above the confinements of our immediate environment.

References

Baker-Sennett, J., Matusov, E., & Rogoff, B. (1993). Planning as developmental process. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 24, pp. 253-281). San Diego: Academic Press.

Carroll, J. B. (1964). *Language and thought.* Englewood Cliffs, NJ: Prentice-Hall.

Cole, M. (1990). Cognitive development and formal schooling: The evidence from cross-cultural research. In L. Moll (Ed.), *Vygotsky and education* (pp. 89-110). Cambridge: Cambridge University Press.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes.* Needham Heights, MA: Allyn and Bacon.

Davydov, V. V. (1990). *Types of generalization in instruction: Logical and psychological problems in the structuring of school curricula.* Reston, VA: National Council of Teachers of Mathematics.

Díaz, R. M., Neal, C. J., & Amaya-Williams, M. (1990). The social origins of self-regulation. In L. C. Moll (Ed.), *Vygotsky and education* (pp. 127-154). New York: Cambridge University Press.

Dreher, M., & Oerter, R. (1987). Action planning competencies during adolescence and early adulthood. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 321-355). New York: Cambridge University Press.

Gallimore, R., & Tharp, R. (1990). Teaching mind in society: Teaching, schooling, and literate discourse. In L. C. Moll (Ed.), *Vygotsky and education* (pp. 175-205). Cambridge: Cambridge University Press.

Goodnow, J. J. (1987). Social aspects of planning. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 179-201). New York: Cambridge University Press.

Huttenlocher, J. (1976). Language and intelligence. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 261-281). New York: John Wiley & Sons.

Karpov, Y. V., & Bransford, J. D. (1995). L. S. Vygotsky and the doctrine of "empirical" and "theoretical" learning. *Educational Psychologist, 30,* 61-66.

Kearins, J. M. (1981). Visual spatial memory in Australian Aboriginal children of desert regions. *Cognitive Psychology, 13,* 434-460.

Kearins, J. M. (1986). Visual spatial memory in Aboriginal and White Australian children. *Australian Journal of Psychology, 38,* 203-214.

Kozulin, A. (1986). Introduction. In L. S. Vygotsky, *Thought and language.* Cambridge, MA: The MIT Press.

Kozulin, A. (1990). *Vygotsky's psychology: A biography of ideas.* New York: Harvester/Wheatseaf.

Kreitler, S., & Kreitler, H. (1987). Conceptions and processes of planning: the developmental perspective. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 205-272). New York: Cambridge University Press.

Lee, B. (1987). Recontextualizing Vygotsky. In M. Hickmann (Ed.), *Social and functional approaches to language and thought* (pp. 87-104). New York: Academic Press.

Levina, R. E. (1981). L. S. Vygotsky's ideas about the planning function of speech in children. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 279-299). New York: M. E. Sharpe.

Luria, A. R. (1959). The directive function of speech in development and dissolution. *Word, 15,* 341-352.

Luria, A. R. (1982). *Language and cognition.* New York: John Wiley & Sons.

Mehan, H. (1979). *Learning lessons.* Cambridge, MA: Harvard University Press.

Mulcahy, R. F., Marfo, K., Peat, D., & Andrews, J. (1986). *A strategies program for effective learning and thinking (SPELT): Teacher's manual.* Edmonton, AB: Cognitive Education Project, University of Alberta.

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

Palincsar, A. S., & Brown, A. L. (1988). Teaching and practicing thinking skills to promote comprehension in the context of group problem solving. *Remedial and Special Education, 9* (1), 53-59.

Parrila, R. K., Äystö, S., & Das, J. P. (1994). Development of planning in relation to age, attention, simultaneous, and successive processing. *Journal of Psychoeducational Assessment, 12,* 212-227.

Pea, R. D. (1982). What is planning development the development of? In D. Forbes, & M. T. Greenberg (Eds.), *New Directions for Child Development: Children's Planning Strategies* (Vol. 18, pp. 5-26). San Francisco: Jossey-Bass.

Peat, D., Mulcahy, R. F., & Darko-Yeboah, J. (1989). SPELT (Strategies Program for Effective Learning/Thinking): A description and analysis of instructional procedures. *Instructional Science, 18,* 95-118.

Rommetveit, R. (1974). *On message structure.* London: John Wiley & Sons.

Schmittau, J. (1993). Vygotskian psychology and dialectical logic: A psychological-epistemological foundation for contemporary pedagogy. *The Review of Education, 15,* 13-20.

Schubert, J. (1983). The implications of Luria's theories to cross-cultural research on language and intelligence. In B. Bain (Ed.), *The sociogenesis of language and human conduct* (pp. 59-77). New York: Plenum Press.

Sokolov, A. N. (1969). Studies of speech mechanisms of thinking. In M. Cole & I. Maltzman (Eds.), *Handbook of contemporary Soviet psychology* (pp. 531-573). New York: Basic Books.

Sokolov, A. N. (1972). *Inner speech and thought.* New York: Plenum Press.

Vocate, D. R. (1987). *The theory of A. R. Luria.* Hillsdale, NJ: Erlbaum.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1981a). The development of higher forms of attention in childhood. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 189-240). New York: M. E. Sharpe.

Vygotsky, L. S. (1981b). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 144-188). New York: M. E. Sharpe.

Vygotsky, L. S. (1986). *Thought and language.* Cambridge, MA: The MIT Press.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber, & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky* (Vol. 1, pp. 39-285). New York: Plenum Press.

Wertsch, J. V. (1979). From social interaction to higher psychological processes. *Human Development, 22,* 1-22.

Wertsch, J. V. (1990). Dialogue and dialogism in a socio-cultural approach to mind. In I. Markovà, & K. Foppa (Eds.), *The dynamics of dialogue* (pp. 62-82). New York: Harvester/Wheatsheaf.

Wertsch, J. V., & Stone, C. A. (1985). The concept of internalization in Vygotsky's account of the genesis of higher mental functions. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 162-179). Cambridge: Cambridge University Press.

III. DEVELOPMENT OF PLANNING AND ITS RELATION TO OTHER COGNITIVE PROCESSES

Introduction

This study focuses on the (a) development of planning skills over the school years, (b) the uniformity of the psychological construct of planning, and (c) the cognitive correlates of planning.

Planning has been a central concept in cognitive psychology since the publication of Miller, Galanter, and Pribram's seminal book *Plans and the Structure of Behavior* in 1960. Planning permeates all voluntary behavior and, as Kagan (1987) noted, it is not possible to propose a theory of human behavior without a set of constructs referring to plans or planning.

Studies concentrating on the development of planning skills cover a wide range of ages. Recent studies, for example, have shown that at least some components of planning such as the ability to span temporal separation between two actions, inhibit prepotent responses, and form a goal and produce activity that is directed toward it are already apparent during infancy (Diamond & Goldman-Rakic, 1983, 1985, 1986; Willatts, 1984, 1990). Accordingly, children as young as 9 to 12 months of age have exhibited rudimentary forms of planning (Rogoff, Mistry, Radziszewska, & Germond, 1992; Willatts, 1984, 1990). Most researchers, however, have expressed some doubt regarding infants' ability to form plans and strategies. For example, planning is commonly associated with the development of the capacity to manipulate mental representations and to regulate behavior verbally and hence, it is not expected to appear until later in ontogenesis (see e.g., Levina, 1981; Luria, 1973; Piaget, 1963).

Luria (1959) suggested that the self-regulatory function of speech undergoes major development during the preschool years. As children develop, they are increasingly able to both initiate and inhibit activities guided by internal speech. Studies have shown that preschool and kindergarten children are capable of devising and executing simple plans (see e.g., Casey, Bronson, Tivnan, Riley, & Spenciner, 1991; Fabricius, 1988; Gauvain, 1989; Haake, Somerville, & Wellman, 1980; Klahr & Robinson, 1981; Wellman, Fabricius, & Sophian, 1985), particularly if these plans are based on their knowledge of familiar events (Hudson & Fivush, 1991), and are aided by the utilization of overt speech (Cocking & Copple, 1987; Göncü & Kessel, 1988). Moreover, before entering formal schooling, children have at least some "metaplanning" skills. That is, they have a preliminary conceptual understanding of what planning is and when it is needed (Gauvain, 1989; Kreitler & Kreitler, 1987).

Planning skills, however, seem to develop particularly rapidly during the school years. Children learn to apply their skills in planning to a broad range of tasks. By the age of 8, for example, children learn to complete tasks such as simple forms of Tower of Hanoi (Klahr & Robinson, 1981; Welsh, Pennington, & Groisser, 1991), mazes and errand running tasks (Gardner & Rogoff, 1990; Gauvain & Rogoff, 1989; Pea & Hawkins, 1987), and Delayed Alternation (Levin et al., 1991). By age 10, children should be proficient in the Wisconsin Card Sorting Task (Kirk & Kelly, 1986; Levin et al., 1991; Welsh et al., 1991), the Rey-Osterrieth Complex Figure (Kirk & Kelly, 1986), and the Matching Familiar Figures Test (Welsh et al., 1991). Development of planning skills, particularly in more complex and demanding tasks, may be an especially prolonged process that continues well into adolescence and beyond (see, e.g., Becker, Isaac, & Hynd, 1987; Dreher & Oerter, 1987; Levin et al., 1991; Parrila, Äystö, & Das, 1994; Passler, Isaac, & Hynd, 1985; Pitt, 1983; Welsh et al., 1991).

The existing research, then, seems to validate the intuitive expectation that relatively simple planning tasks are mastered by school-age children in the early grades. Performance on more complex planning tasks, however, continues to develop beyond middle childhood and through adolescence. The first goal of the current study was to replicate these earlier findings with a more representative and larger sample of school-age children.

Our second question deals with the uniformity of planning as an explanatory construct of cognitive functioning. The inclusion of a planning construct in a model of intellectual functioning has theoretical, empirical, as well as clinical support (Telzrow, 1990). Yet, very few standardized planning measures are available and researchers commonly design their own tasks. Moreover, the term *planning* has been used to account for a multitude of behaviors that, overtly, may not seem to have much in common. As a consequence, the definition of the term has become vague and two theorists using the terms "planning" or "plan" may not share the same focus (Scholnick & Friedman, 1987). For example, Newell, Shaw, and Simon (1959) viewed planning as problem solving in a simplified, abstracted problem space, whereas Miller et al. (1960) defined plans as hierarchical control processes. For Hayes-Roth and Hayes-Roth (1979), planning consisted of anticipating a goal-directed course of action. In their integrative review of the conceptualization of planning, Scholnick and Friedman (1987) included six components (forming a representation of the problem, choosing a goal, deciding to plan, formulating a plan, executing and monitoring the plan, and learning from the plan) and three different levels of functioning (in the reality of a problem, in accordance with an imagined scheme, and in the role of mediator between the scheme and behavior). We suggest that a fertile way of conceptualizing planning and its relationship to such closely related concepts as problem-solving and strategies is to consider how planning relates to the three levels of analysis—activity, action, and operation—as introduced by Leontjev (1978, 1981).

Three Levels of Planning

At the level of *activity,* planning can be conceptualized as a method of realizing or aiming toward one's general life goals and motives, such as self-fulfillment, self-improvement, education, career development, or planning for a retired life. When our mental activity has an objective or goal, planning appears as a separate *activity* distinct from imagery. As an activity, planning is a molar unit of analysis that can be used to explain an individual's behavior in general. For example, plans for a life after retirement will provide a framework within which a person's behavior can be explained and understood. The function of activity-planning is to mediate between a person's life goals and the external, objective world. In order to achieve this, activity-planning entails components that are not necessarily present in other forms of planning. The components unique to activity-planning include selection and shaping of one's environments so that they maximally support, or minimally impede, the fulfillment of one's life goals. Problem finding, or the creation and definition of relevant problems that need solution, is also unique to activity-planning. Other possible components of activity-planning, such as forming a representation of external and internal variables that may influence goal attainment, choosing subgoals, and anticipating the course of action that is needed to realize the goals and subgoals, can also be present in planning as an action, which is discussed next.

Action-planning is equivalent to problem solving. While activity-planning is best understood as movement toward realizing one's general life goals, action-planning aims at achieving a particular goal or solving a particular problem. Everyday examples of action-planning include scheduling daily meetings, running errands efficiently, or ferrying all the children to their respective schools, clubs, visiting places and back. Problems and goals of action-plans can be components of activity-plans as well. For example, if a general life-goal is to obtain a secure and well-paying profession, then the activity-plan may involve such components as deciding upon the most suitable educational institution, financing one's studies, and finding the right type of employment. Action-planning can involve forming a mental representation of the problem, the (external and internal) constraints on planning, the goal, and the course of action to be taken (i.e., formulating a plan in advance), as well as executing the resulting plan and monitoring the whole process. Figure 3-1 presents the main components of action-planning. Note that these components are present in activity-planning as well.

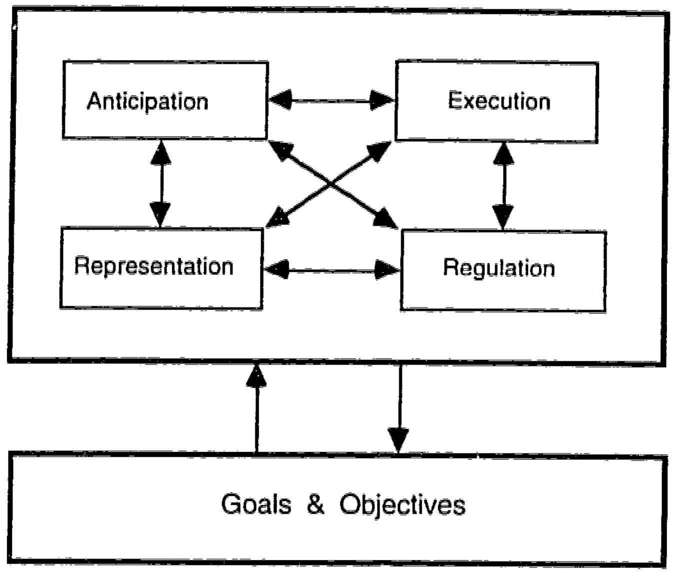


Figure 3-1. Main Components of Action-Planning.

But action-planning can also be an opportunistic process, or "planning-in-action", when task demands or the planner's skills favor this approach (Hayes-Roth & Hayes-Roth, 1979; Rogoff, Gauvain, & Gardner, 1987). Planning-in-action involves continuous evaluations and revisions of plans while they are being implemented. This recursive nature of planning-in-action is described in Figure 3-1 by the bidirectional arrows connecting the four components. The main feature of action-planning is that it emerges as a response to a given problem; therefore, it is oriented toward the present as well as the future.

At the level of *operations,* plans are equivalent to strategies and tactics, and consist of working toward the solution of a problem (or a part of it) in accordance with task-imposed constraints (i.e., meeting environmental conditions). Everyday examples of operation-plans would include locating a book in a library or using household machines. The main feature of an *operation-plan* is that it needs to satisfy the specific conditions associated with the task and, consequently, it is oriented towards the present. Because the goal, or the end-result, is often given, operation-planning involves forming a representation of the task and conditions, choosing the possible operations to be undertaken, and then executing these steps. Thus, operation-planning includes all the components of action-planning identified in Figure 3-1, but the difference is that the arrows are now unidirectional due to increased task-constraints. Another difference is that the process of choosing between possible operations or ways to proceed is not necessarily conscious if we have available "prepackaged information" (Scholnick & Friedman, 1993), that is, automated tactics or strategies that are associated with positive outcomes. For example, the Planned Search task (see below for description) can be solved efficiently by utilizing visual search strategies that are already automated due to practice with real life visual search tasks such as locating a friend in a crowd or car keys on a front lawn. When a task allows for only one possible method of proceeding, then planning entails finding that method and executing it. At this end of the continuum, planning disappears when the task imposes both the goal and the operations required, and the subject is allowed no degrees of freedom.

According to this conceptualization, 'planning' and 'plan' are generic terms that refer to any of the three levels of analysis, whereas 'problem solving' refers mainly to the action or operation levels of analysis. Strategies and tactics refer only to operation-planning. It should also be noted that our conceptualization of *activity* is more in line with that of Vygotsky (see e.g., Kozulin, 1986) and emphasizes the role of symbolic cognitive activity rather than material activity that was stressed by Leontjev.

The components in Figure 3-1 — anticipation, representation, execution, and regulation — illustrate the neuropsychological view of planning as a frontal lobe function (Stuss & Benson, 1986, 1987). The figure could be extended to include "consciousness" as an overarching, or perhaps, meta-component of planning that gives meaning to, and guides an individual's activities, actions, and operations. Sperry (1993) conceptualized consciousness as a determining force of mental activities, as opposed to merely an emergent phenomenon of the physical basis of mental activities. We think that the extension of planning to incorporate the central role of consciousness will make it easier to understand the activity-action-operation trichotomy. Activity-planning is guided by life goals. But even in childhood, activity is guided by a child's self concept. For example, the statement "I am a bookworm like my grandma" will shape and determine a child's activity-planning differently than the statement "I will be like Wayne Gretzky". The recent interest in improving children's self-esteem in order to better their academic performance is another illustration of the importance of *consciousness* in explaining, justifying, and guiding children's activity.

The majority of the planning studies that were reviewed above have focused on action-planning and operation-planning, and on the formulation of a plan in advance rather than on opportunistic planning. Research on activity-planning is difficult, mainly because realizing one's life-goals typically takes years. This creates many problems for experimental designs and their execution (however, see Nurmi, 1989, for an attempt to circumvent this problem).

The second goal of the present study was to examine whether action- and operation-planning could be differentiated with concise and abstract planning tasks. Action-planning is especially important in tasks where finding the solution requires integration of multiple steps into a coherent process and in tasks that are open to more than one way of proceeding. The Crack-the-Code task fits into this description. In this task, the participant is shown two to five information lines that contain three to five colored chips that are in a particular order. They are also shown a label indicating how many of the colored chips are in their correct places. The participant's task is to integrate information from all of the information lines and place her or his set of colored chips on the answer line in such an order that all the information lines are true. For example, Item 5 contains the following information lines (actual colored chips are used):

- Line 1: Blue, Black, Yellow, White; "0 correct"
- Line 2: Yellow, White, Black, Blue; "1 correct"
- Line 3: Black, Yellow, White, Blue; "0 correct"

The only possible order of chips that would be consistent with the information lines is: White, Blue, Black, Yellow. An effective way of reaching this solution could, for example, include (a) noticing the constraints (e.g., Blue cannot be correct in Line 2 since it is in the same place in Line 3, which has 0 correct), (b) building a hypothesis (e.g., Yellow, White, or Black is correct in Line 2 — try Yellow first), (c) testing and evaluating the hypothesis (e.g., if Yellow is correct in Line 2, where would the other chips go — there is no legitimate place left for White, so Yellow cannot be correct in Line 2), and so on, until the correct answer is found. Think aloud protocols obtained from elementary and junior high school students (Parrila & Das, 1995) clearly show that some participants proceed by planning before taking action. In other words, they go through the above cycles mentally before actually placing any of the chips on the answer line. Others, in contrast, seem to proceed by planning-in-action; for example, they first build a hypothesized answer line using the available chips and then proceed to evaluate it. Naturally, most of the participants fall somewhere in between and plan some steps beforehand and others during the action. Thus, this task clearly includes several steps and allows more than one way of proceeding. However, some procedures are more efficient than others and, thus, more likely to produce a correct answer within the time limit.

The other two tasks used in this study, Matching Numbers and Planned Search, are more constrained in nature. Planned Search requires the participant to first identify the target figure (picture, number, or letter) situated inside a box in the middle of a search field, and then find an instance of the target figure in a search field that includes distractor items belonging to the same category as the target (e.g., picture among pictures). To complete the task effectively, the participant needs to keep the target figure in active short-term memory for comparisons, develop an efficient way of scanning each field for the target figure, and control for impulsivity in order to avoid choosing wrong answers. However, because the necessary information is given directly, finding the solution does not necessarily demand conscious, multistep planning. Instead, the efficient scanning strategy may be nonconscious and automatized due to extensive experience with real-life search tasks. Nevertheless, visual search does tap into a basic mechanism of planning. This is supported by several previous studies reported in Das, Naglieri, and Kirby (1994). For example, a group of high school students were first divided into two groups, low and high performers on visual search, and then compared on reading comprehension. The results indicated that the groups were respectively less and more efficient on several comprehension measures that indexed forward and backward inferencing and organization in remembering the text. Moreover, Das and Heemsbergen (1983) observed that college students who were proficient in visual search were superior to the less proficient ones in the 'Master Mind' game, which resembles the Crack-the-Code task. Miller et al. (1960) proposed a classic distinction between *search* and *prediction* as alternative paradigms to explain thinking and problem solving. Visual search is likely a good example of the search processes that can be considered to be the foundation of many planning activities.

Matching Numbers, in turn, requires the participant to identify and underline two numbers that are the same in each of the eight rows on a page. There are six numbers per row and the digit length increases from 1 to 3 on Card 1 and from 4 to 5 on Card 2. An efficient performance in Card 1 can result from the very simple strategy of scanning each row from left to right, since the numbers are short and relatively easy to remember. In Card 2, however, such a strategy may not work because its working memory demands grow too large for most of the participants. Instead, the participant needs to utilize different strategies such as comparing the first two digits of each number, or comparing the first and the last digit of each number.

It is clear from the above descriptions that while Matching Numbers and Planned Search can both be conceptualized as planning tasks, compared to Crack-the-Code, the number of stages through which the solution must pass is smaller and relatively constrained in both of them. Moreover, successful performance in these tasks can be based on the existing repertoire of strategies and tactics that are implemented without conscious reflection. These characteristics make Matching Numbers and Planned Search congruent with our definition of operation-planning tasks. Thus, the second hypothesis was that Crack-the-Code would have a significant distance from the other planning tasks.

Cognitive Correlates of Planning

The last question addressed in this study has two parts. First, what are the "cognitive correlates" of effective performance in action-planning (or complex planning) tasks on the one hand, and in operation-planning (or simple planning) tasks on the other? Second, are these cognitive correlates essentially stable across different age-groups?

We know of only a few studies that have directly attempted to identify the cognitive correlates of planning. The rationale for studying cognitive correlates of planning is simple. Good planning is at least partly based on the fluent functioning of such other cognitive processes as encoding, attention processes, and working memory. Deciding precisely what these prerequisite processes are exactly is an inherently difficult task since most researchers agree that planning is one of the more complex cognitive functions that utilizes — as well as affects — other cognitive functions. Before introducing the cognitive correlates included in this study, we will briefly review the cognitive correlates that have been considered in previous research.

In considering the role of memory span, Pea and Hawkins (1987) found that the Digit Span task from the WAIS did not have a significant correlation with the frequency of "high level planning decisions" when measured in two groups of school children (8 to 9 year-olds and 11 to 12 year-olds). This finding was partly supported by Parrila et al. (1994), who reported that a factor defined predominantly by memory tasks had few significant correlations with planning tasks. We examine this relationship again in the present study.

The role of language (or speech) in planning has been emphasized particularly in the Vygotskian tradition (see Levina, 1981; Parrila, 1995). Some experimental studies also provide evidence that language skills influence skillful planning. For example, McGillicuddy-De Lisi, De Lisi, Flaugher, and Sigel (1987) compared noncommunication handicapped (NH) children with those who had a communication handicap (CH). Their results indicated that children with CH proposed fewer initial plans and accepted plans proposed by their siblings more often than NH children. However, children with CH did suggest alternative plans after their siblings presented initial plans. This suggests that after the initial external organization of the task, these children were capable of producing plans on their own. In the same study, the Crichton Productive Vocabulary Test, which assesses expressive language skills, had a significant correlation (.35) with the number of initial plans suggested by children in 86 planning dyads (both NH and CH target children were included). In contrast, the Peabody Picture Vocabulary Test (PPVT), which assesses receptive language skills, did not correlate significantly with the number of proposed plans. Kreitler and Kreitler (1987), in turn, found that the PPVT had no correlation with total planning score for age 5, but had a highly significant correlation (.67) for age 7, and a moderate correlation for ages 9 and 11 (.26 for both).

Planning has also emerged as a separate construct from general intelligence. This seems to be the common conclusion of researchers who have used widely different measures of planning and intelligence on both preschool and school age children (see e.g., Casey, Bronson, Tivnan, Riley, & Spenciner, 1991; Das & Dash, 1983; McGillicuddy-De Lisi et al., 1987; Nurmi, 1989; Welsh, Pennington, & Groisser, 1991).

The present study examines how the relationship between planning and the three other PASS (Planning-Attention-Simultaneous processing-Successive processing) cognitive processes may change across the school years. Research on PASS has not focused on this relationship before, with the exception of Parrila et al. (1994). Since the PASS model is described in detail in previous reports (Das, Kirby, & Jarman, 1979; Das et al., 1994; Naglieri & Das, 1990), only a brief summary is presented here.

The PASS model is based on Luria's (1966, 1973) theory of the functional units of the brain. According to Luria (1973), the human brain engages in three types of basic functions, namely arousal-attention, coding, and planning. In the PASS model, planning consists of programming, regulation, and verification of behavior (Luria, 1966), and the definition of planning is generally compatible with the model of planning presented here. Attention is a mental process that comprises at least three components as suggested by a factor analysis of attention tasks. These components are selectivity, ability for shifting, and resistance to distraction (Das et al., 1994). Selective attention is a specialized form of activation (Luria, 1973) and even if the single resource theory of attention is no longer tenable, selection-for-action must occur to control perceptual-motor activities (Allport, 1993). Such activities must be maintained until they are completed, and then attention should shift to a new activity. However, attention may have to shift earlier if there is an urgent demand from the environment. Under normal conditions, though, distractions are to be resisted.

Simultaneous and successive processing relate to reception, encoding, and storage of information arriving from the outside world through sensory receptors (Luria, 1973). The incoming information can be coded in two ways. The first, simultaneous coding (or processing), organizes information in a quasi-spatial scheme integrating separate elements into a coherent whole. This type of coding is evident in relational thinking and in understanding comparative and logico-grammatical constructions. The second, successive synthesis (or successive processing in PASS), organizes information into sequential, temporally based schemes. Successive synthesis plays an important role in serial reproduction and learning, spelling, perception of syntax, and in the initial phases of reading and writing (Das et al., 1994).

The four processes are naturally interdependent. Thus, we expect that attention and simultaneous and successive processes will emerge as significant cognitive correlates of planning. The model of planning described above implies that the planner is capable of attending to, and encoding, relevant information and utilizing it in planning. These ideas are not peculiar to the PASS model. Others such as Gauvain (1992) have suggested that allocating attention to the various demands during the planning process is an important aspect of the development of planning skills. Siegler (1991, p. 285) suggested that "how well children encode critical information in the task and how well they can use the encodings to form mental models are among the key determinants of their success on many problems." Thus, it is reasonable to assume that in many planning tasks, a poor performance may result from children's inability to perceive all of the relevant features of the task or from inadequate attention resources, rather than from poor planning skills. When children reach adequate performance levels in attention, simultaneous, and successive processing, their planning performance can improve because (a) children will have better 'raw material' to work with in planning, and (b) they can allocate more cognitive resources to planning when many of its components are automated.

By using a design similar to one presented here, Parrila et al. (1994) tested this hypothesis in order to determine if attention, simultaneous processing, and successive processing scores would predict children's planning performance. The results indicated that most planning scores were predicted significantly by other cognitive processing components and that the contribution of the other processing components varied as a function of the planning task. The present study will expand on these findings by including more participants and a more developmental focus.

Method

Participants

The participants were school children from several public schools operated by the city school board of Bhubaneswar, the state capital of Orissa, a southeastern province in

India. The participants came from families spanning a wide range of socio-economic levels, excluding the extremely poor. The total number of participants was 250; 50 from each of grades 3 (22 females), 5 (16 females), 7 (22 females), 9 (27 females), and 11 (21 females). Mean ages (*SDs* in parenthesis) for grades from 3 to 11 were 8.34 (.75), 10.24 (.62), 11.96 (.70), 13.56 (.79), and 16.92 (.94), respectively.

Since our participants came from schools in India, it is necessary to ask how or whether the cultural differences can drastically change the interrelationship of the four PASS processes. Cultural differences of this magnitude (urban India vs. North America) do not seem to play a significant part in whether a child performs well or poorly on tasks that we have used in previous investigations. Several studies on children from the same region in India have shown that these tests retained their construct validity (Dash & Mahapatra, 1989; Dash, Puhan, & Mahapatra, 1985). Kar, Dash, Das, and Carlson (1993) found that the planning tasks retained their essential characteristics as measures of planning. Also, in a study relating reading ability and performance on the PASS tasks, the results that were obtained in a Canadian sample (Das, Mensink, & Mishra, 1990) were essentially replicated with a sample of children from Orissa (Dash & Mohanty, 1992). Thus, we anticipated no problem in measuring the PASS processes with these tests in our Indian sample.

Tasks

Planning Tasks

Crack-the-Code. This task is based on the popular Master Mind game and was used by Das, Mensink and Janzen (1990) and Parrila et al. (1994) to measure planning. The task requires the participant to determine what the correct sequence of colored chips is when a limited amount of information is provided in the instruction line(s) (see example in the introductory section above). The version used in this study had eight items. In the first item, two instruction lines and three chips of different colors were used. The number of chips was then increased step-by-step so that in the last item, five instruction lines and five chips of different colors were used. The participant was given one trial to figure out the correct order of chips for each item and the time limit for each item was 3 minutes. The task was interrupted after two consecutive failures. The participant's score was the number of correctly solved items.

Matching Numbers. This task and Planned Search (see below) were adapted from the battery of PASS tasks described in Das et al. (1994). Matching Numbers was developed by Naglieri and Das (1987) and has loaded on a planning factor in previous research (Naglieri & Das, 1988). The task requires the participant to find two numbers that are the same on each of the eight rows on a card. There were six numbers of the same length in each row. In Card 1, the first row contained 1 digit numbers and the eighth row contained 3 digit numbers. In Card 2, the number length ranged from 4 to 5 digits. Maximum time per card was 180 seconds. The participant's score for Card 1 and Card 2 was the total time divided by the number of correct responses. This is an index of matching efficiency that is sensitive to speed-accuracy tradeoff.

Planned Search. This task is similar to the Visual Search task developed by Teuber, Battersby, and Bender (1949) and has been found to load on a planning factor by Ashman and Das (1980), Naglieri and Das (1988), and Naglieri, Prewett, and Bardos (1989). Planned Search requires the individual to develop an efficient scanning strategy to find a particular target stimuli situated inside a box in the middle of a search field, and then find an instance of the target figure in a search field that includes distractor items belonging to the same category as the target. The version used in this study consisted of four items that used pictures as targets and distractors. Each item included two searches, one located at the top and the other located on the bottom of an 8 1/2" x 11" page. The time taken per item was recorded and the participant's planned search score, the mean time taken for the four planned search items, was calculated.

Attention Tasks

Receptive Attention. This task is adapted from Posner and Boies (1971), who used a letter-stimuli version of it to measure receptive attention. The Receptive Attention task is a broad measure of attention since it involves discrimination, resistance to distraction, and shifting attention from one target to another for a prolonged period of time. The version used in this study consisted of two conditions: physical match and name match. In the physical match condition, the participant was given a sheet consisting of one-hundred picture pairs (trees, fruits, flowers, birds, houses, or human faces) arranged in a matrix form. He or she was then instructed to point only to those pairs of pictures that were visually alike. The time taken by the participant to complete the task was divided by the number of pairs correctly identified to provide an index of average time taken for each correctly identified picture pair. The administration and the scoring procedure was the same for the name match condition except that the participant was asked to identify the pairs of pictures that belonged to the same taxonomic category.

Simultaneous Processing Tasks

Matrices. This task, developed by Naglieri and Das (1987), involves the completion of figural analogies using a progressive matrix format. The participant is required to choose one of six options that best completes the abstract analogy. The version used in this study consisted of 35 items and the participant's score was the number of correctly completed items. The requirement that each component of the matrix must be interrelated to the others makes this task congruent with the simultaneous paradigm, and this test has previously been found to load on a simultaneous factor (Naglieri & Das, 1987; Naglieri et al., 1989).

Figure Memory. This task was used as a simultaneous marker test by Naglieri and Das (1987), and Naglieri et al. (1989). In this task, the participant is exposed to a geometric design such as a square or triangle for five seconds. The design is then removed and the participant is asked to outline the original stimulus figure within a more complex design that includes the original figure. For a response to be scored as correct, all lines of the original design had to be indicated without any additions or omissions. The requirement that the design has to be incorporated into memory as a whole, so that all the parts of the figure are interrelated, makes this task congruent with the simultaneous paradigm. There were twenty items in the test. The participant's score was the total number of items correctly reproduced.

Simultaneous Verbal. This is a verbal marker test of simultaneous processing. This task involves evaluation of logical-grammatical relationships by the participant. The version used in this study consisted of twenty-six items, with each item containing six competing illustrations of design/pictorial configurations. The participant was asked to point to one of the six figures that was commensurate with a verbal statement, such as "the ball in a basket on a table" or "the girl pointing to the ruler with the pencil". The test was scored for the number of correct responses.

Successive Processing Tasks

Digit Span. This successive marker test was abstracted directly from the WISC-R. A series of digits of increasing length were read out to the participants, who were required to recall the digits in the correct serial order. The task was discontinued after two unsuccessful attempts. The participant's score was the number of correctly recalled digits in the longest successfully completed series.

Word Series. This is a marker test of successive processing. The version used in the present study consisted of twelve lists of words that began with a four-word series and progressed to a six-word series. There were four lists for each of the four-, five-, and six-word series. All the words were highly familiar two or three letter words in the participants' first language (Oriya). The participants were required to recall each series in correct serial order. The number of words recalled in correct serial position constituted the serial recall score of the participant. The maximum possible score on this test was 60.

Procedure

All participants were individually administered the experimental tasks between November 1993 and March 1994 in their respective schools by three trained investigators who had master's degrees either in psychology or in education. The tasks were administered in a predetermined order starting with simultaneous tasks and followed by successive, attention and planning tasks. After establishing adequate rapport with the participants, they were tested in a private room during two separate sessions. In order to make the instructions comprehensible, all of the tests contained practice items at the beginning.

Results and Discussion

Development of Planning Skills

Multivariate analysis of variance with planning tasks as dependent variables and Grade (5) and Gender (2) as independent variables was performed first. The results showed a significant main effect of Grade, Wilks' $\Lambda = .490$, $F(16, 724.68) = 11.92$, $p < .001$. The main effect of Gender was not significant, Wilks' $\Lambda = .992$, $F(4, 237.00) = .456$, $p = .740$, and there was no significant interaction effect, Wilks' $\Lambda = .935$, $F(16, 724.68) = 1.00$, $p = .453$. Because no Gender or Gender X Grade effects were found, scores from male and female participants were pooled for the rest of the analyses.

The *F* values from univariate analyses together with the performance means and standard deviations within each grade level are reported in Table 3-1.

Table 3-1 shows that the *F* values corresponding to planning tasks are highly significant, suggesting that all the planning tasks were developmentally sensitive. To illustrate the differences in performance means on different grade levels, Figure 3-2 displays the standardized performance means (time scores were first multiplied by -1) of Matching Numbers Card 1, Matching Numbers Card 2, Planned Search, and Crack-the-Code as a function of grade.

Table 3-1

Means and Standard Deviations (in parenthesis) of Planning Variables within Different Grade Levels (N=250)

	Means and standard deviations					
Variables	Grade 3	Grade 5	Grade 7	Grade 9	Grade 11	F^a
Crack-the-Code	1.58	1.98	2.18	3.32	3.78	13.79***
	(1.21)	(1.56)	(1.71)	(2.10)	(2.28)	
Matching Numbers						
Card 1	14.35	9.13	6.97	6.90	6.86	22.34***
	(7.81)	(4.70)	(2.75)	(2.49)	(3.88)	
Card 2	43.41	27.77	19.00	19.20	15.39	36.60***
	(21.21)	(14.70)	(8.11)	(10.21)	(6.47)	
Planned Search	11.16	9.51	7.77	7.27	6.37	15.28***
	(4.72)	(3.75)	(2.71)	(2.56)	(2.69)	

Note. [a] F values for the main effect of grade, $df(4, 240)$.

*** $p < .001$.

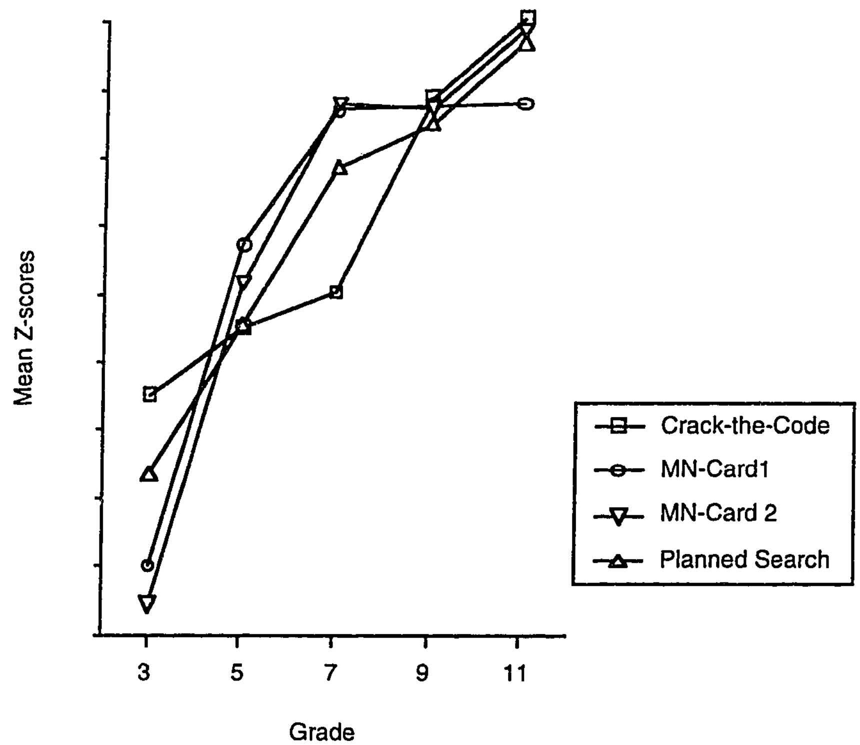


Figure 3-2. Mean z-scores for Crack-the-Code, Matching Numbers Card 1, Matching Numbers Card 2, and Planned Search in different grade levels ($N = 250$).

Figure 3-2 suggests that the developmental trends varied as a function of the planning task. Results from post hoc Scheffé tests ($p < .05$) showed that Grade 3 participants performed significantly poorer than Grade 5 participants on both Matching Numbers Card 1 and Card 2, but not on Crack-the-Code or Planned Search. Compared to Grade 7, Grade 3 participants' performance was significantly poorer on all planning tasks except Crack-the-Code. And, finally, when compared with Grade 9 and Grade 11 participants, Grade 3 participants' performance was significantly poorer on all tasks.

Grade 5 students, in turn, differed significantly from Grade 7 students only on Matching Numbers Card 2. In contrast, Grade 5 students differed significantly from Grade 9 and Grade 11 students on Matching Numbers Card 2, Planned Search, and Crack-the-Code. Grade 7 students performed significantly poorer than Grade 9 and Grade 11 students only in Crack-the-Code. None of the four planning tasks differentiated between grades 9 and 11.

It is interesting to note that Crack-the-Code did not differentiate between grades 3, 5 and 7, and then was the only task that differentiated between Grade 7 and the two older groups. This suggests that only the two older groups reasonably succeeded in performing this complex task. Frequency distribution of scores (minimum 0, maximum 8) confirmed this: The number of participants who received a score of 2 or less was 41 in Grade 3, 39 in Grade 5, and 36 in Grade 7, indicating that a vast majority of Grade 7 students were not able to proceed beyond the simplest items. According to previous research, the ability to selectively attend to relevant stimuli, to construct better hypotheses by systematically limiting the search space, and to form and remember multistep plans are among those components of planning that continue to develop during adolescence. Successful performance in Crack-the-Code demands all of these components.

In contrast, neither Matching Numbers nor Planned Search differentiated between the three oldest grade levels, suggesting that dramatic developmental changes take place earlier in these simple planning tasks. We should note, however, that both Matching Numbers Card 2 and Planned Search show some developmental differences after Grade 7, although they were not significant. Moreover, for the simplest planning task (Matching Numbers Card 1), development seemed to level off almost completely after Grade 5. This supports the earlier findings that at approximately age 10, children are already capable of performing simple planning tasks at or near adult performance level (as represented here by Grade 11 students), whereas in the more complex tasks, development continues well into adolescence. Previous research has also suggested that speed of responding continues to develop into late adolescence and early adulthood. Our results do not support this finding since both Matching Numbers and Planned Search are sensitive to speed of responding and neither of these tasks differentiated between the three older grade levels.

Different Levels of Planning

The question of different levels of planning was addressed next. Two different methods were used: factor analysis (principal axis factoring with oblique rotation) and cluster analysis for variables (agglomerative hierarchical cluster analysis; squared Euclidean distance; average linkage between groups method). Grades 3 and 5 were combined to form the elementary school group (n = 100), and grades 9 and 11 were combined to form the high school group (n = 100). Planning scores for these analyses were calculated by using regression analysis to obtain standardized residuals of the four planning scores with the effect of age partialled out within both groups.

Table 3-2 summarizes the results from the principal axis factor analysis for the elementary and high school samples.

Table 3-2

Factor Loadings and Communalities of the Planning Tasks for Elementary School Sample (N=100) and High School Sample (N=100; Principal axis factor analysis with oblique rotation)

Variables	Communality	Factor 1	Factor 2
Elementary School Sample			
Crack-the-Code	.154	-.407	.120
Matching Numbers Card 1	.436	.572	.216
Matching Numbers Card 2	.688	.636	.390
Planned Search	.237	.001	.487
High School Sample			
Crack-the-Code	.365	-.113	.552
Matching Numbers Card 1	.233	.479	-.010
Matching Numbers Card 2	.590	.771	.006
Planned Search	.203	-.051	-.467

Factor analysis produced two factors with eigenvalues > 1 for both the elementary and the high school sample. In the elementary school sample, Factor 1 had high loadings from both Matching Numbers tasks (.572 for Card 1 and .636 for Card 2), and a moderate loading from Crack-the-Code (-.407). Planned Search loaded highest on Factor 2 (.487), which also had secondary loading from Matching Numbers Card 2 (.390). Communalities were lowest for Crack-the-Code (.154) and Planned Search (.237), and highest for Matching Numbers Card 2 (.688). This two-factor solution explained 37.9% of the total variance, and the correlation between factors was .263.

In the high school sample Factor 1 had the highest loading from Matching Numbers Card 2 (.771). Matching Numbers Card 1 (.479) loaded also exclusively on Factor 1. Crack-the-Code loaded highest on Factor 2 (.552) together with Planned Search (-.467). Communalities for Planned Search and Matching Numbers Card 1 were low (.203 and .233). This solution explained 34.8% of the total variance, and the correlation between the factors was -.380.

Cluster analysis on both the elementary school sample and the high school sample combined first the two Matching Numbers scores in Step 1, then Planned Search with Matching Numbers cluster in Step 2, and, finally, all other tasks with Crack-the-Code in Step 3.

What support do these results provide for our distinction between action- and operation-planning? First, factor analyses produced two factors with eigenvalues greater than 1 in both the elementary and the high school samples. Next, our task analyses suggested that Crack-the-Code is an action-planning task, whereas Matching Numbers and Planned Search tasks would be operation-planning tasks. Both cluster analyses indicated that Crack-the-Code did have a significant distance from the other planning tasks. The results of the factor analyses, however, were not as clear. In the elementary school sample, Crack-the-Code was poorly explained by the obtained two-factor solution. In the high school sample, it loaded highest on the second factor that also had Planned Search loading on it . Thus, our expectations about Crack-the-Code were mostly confirmed.

Cognitive Correlates of Planning

The effects of attention, simultaneous processing, and successive processing on action- and operation-planning tasks were analyzed separately for each grade level by first computing raw correlations, and then using multiple linear regression analysis (forward selection). The processing scores needed for these analyses were obtained separately for each of the six grade levels as follows: the Attention score was formed by summing the two Receptive Attention z-scores, the Simultaneous score was formed by summing the z-scores of Matrices, Figure Memory, and Simultaneous Verbal, and, finally, the Successive score was formed by summing the z-scores of Digit Span and Word Series.

Simultaneous and Successive scores correlated significantly in all other grade levels except Grade 3. These correlations varied from .519 in Grade 5 to .613 in Grade 9. The Attention score correlated significantly with the Simultaneous score in Grade 5 (-.285), and with both Simultaneous (.-.495) and Successive (-.468) scores in Grade 11. Thus, there are considerably more significant raw correlations in Table 3-3 below than there are significant predictors in Table 3-4.

Table 3-3 displays the raw correlations between the planning tasks and the three processing scores in different grade levels.

Table 3-3

Correlations Between Planning Tasks and Attention, Simultaneous, and Successive Scores in Different Grade Levels (N=250)

	Grade				
Planning Task/ Processing score	3	5	7	9	11
Crack-the-Code					
Attention	-.288*	-.106	-.015	-.219	-.309*
Simultaneous	.194	.471**	.413**	.654***	.583***
Successive	.198	.472**	.341*	.518***	.410**
Matching Numbers Card 1					
Attention	.535***	.406**	.337*	.195	.282*
Simultaneous	-.309*	-.447**	-.591***	-.518***	-.142
Successive	-.269	-.552***	-.383**	-.398**	.107
Matching Numbers Card 2					
Attention	.370**	.382**	.193	.237	.448**
Simultaneous	-.227	-.378**	-.230	-.439**	-.317*
Successive	-.433**	-.390**	-.288*	-.125	-.117
Planned Search					
Attention	.157	.310*	.050	.099	.255
Simultaneous	-.066	-.345*	-.065	-.215	-.360*
Successive	-.068	-.165	-.077	-.221	-.132

Note. $*p < .05$. $** p < .01$. $*** p < .001$.

Table 3-4 displays the significant predictors of planning variables as indicated by regression analyses. In Grade 3, Attention was the only significant predictor of Crack-the-

Code, $\underline{R}^2$ = .083, $\underline{F}(1, 48)$ = 4.35, $\underline{p} > .05$. In Grade 5, Successive ($\underline{R}^2$ = .226, $\underline{F}(1, 48)$ = 14.03, $\underline{p} < .001$) and Simultaneous scores together explained about 29% of the Crack-the-Code variance, $\underline{R}^2$ = .292, $\underline{F}(2, 47)$ = 9.70, $\underline{p} < .001$. Simultaneous was the only significant predictor of Crack-the-Code in Grade 7 ($\underline{R}^2$ = .171, $\underline{F}(1, 48)$ = 9.87, $\underline{p} < .01$), Grade 9 ($\underline{R}^2$ = .427, $\underline{F}(1, 48)$ = 35.82, $\underline{p} < .001$), and in Grade 11 ($\underline{R}^2$ = .340, $\underline{F}(1, 48)$ = 24.76, $\underline{p} < .001$).

Table 3-4

Significant Predictors of Planning Variables in Different Grade Levels (Step of entry, cumulative R^2, and significance level displayed; N=250)

Planning Task/ Predictor	Grade 3	Grade 5	Grade 7	Grade 9	Grade 11
Crack-the-Code					
Attention	(1) .083*				
Simultaneous		(2) .292***	(1) .171**	(1) .427***	(1) .340***
Successive		(1) .222***			
Matching Numbers Card 1					
Attention	(1) .286***	(2) .392***	(2) .405***		(1) .079*
Simultaneous			(1) .350***	(1) .268***	
Successive		(1) .304***			(2) .153*
Matching Numbers Card 2					
Attention	(2) .284***	(2) .246**			(1) .200**
Simultaneous				(1) .193**	
Successive	(1) .187**	(1) .152**	(1) .083*		
Planned Search					
Simultaneous		(1) .119*			(1) .129*

Note. Step of entry into regression equation is shown in parenthesis first, followed by cumulative R^2 and the significance level. *$p < .05$. ** $p < .01$. *** $p < .001$.

Table 3-4 shows that Attention was the only significant predictor of Matching Numbers Card 1 in Grade 3, R^2 = .286, $F(1, 48)$ = 19.21, $p < .001$. Successive and Attention scores together predicted almost 40% (R^2 = .392, $F(2, 47)$ = 15.13, $p < .001$) of the Card 1 variance in Grade 5, and about 15% in Grade 11, R^2 = .153, $F(2, 47)$ =

4.24, $p < .001$. In Grade 7, the significant predictors were Simultaneous and Attention, explaining about 40% of the Card 1 variance, $R^2 = .405$, $F(2, 47) = 16.02$, $p < .001$). Finally, the Simultaneous score was the only significant predictor in Grade 9 ($R^2 = .273$, $F(1, 48) = 18.01$, $p < .001$).

Matching Numbers Card 2 was not predicted uniformly by any of the three processing scores (Table 3-4). The only consistent trend for Matching Numbers Card 2 seems to be the diminishing importance of the Successive score over the five grade levels. Both in Grade 3 ($R^2 = .284$, $F(2, 47) = 9.31$, $p < .001$) and Grade 5 ($R^2 = .246$, $F(2, 47) = 7.65$, $p < .01$), Successive and Attention scores together explained about 1/4 of the Card 2 variance. In Grade 7, the Successive score was the only significant predictor, $R^2 = .083$, $F(1, 48) = 4.33$, $p < .05$, whereas in Grade 9, the Simultaneous score was the only significant predictor of the Card 2 variance, $R^2 = .193$, $F(1, 48) = 11.45$, $p < .01$. Finally, in Grade 11, the Attention score was the only significant predictor, $R^2 = .200$, $F(1, 48) = 12.02$, $p < .01$, of our participants performance on Matching Numbers Card 2.

Planned Search was not predicted significantly by any of the processing scores in grades 3, 7, and 9 (Table 3-4). The simultaneous score was as a significant predictor of Planned Search in Grade 5, $R^2 = .119$, $F(1, 48) = 6.48$, $p < .05$, and in Grade 11, $R^2 = .129$, $F(1, 48) = 7.13$, $p < .05$.

To summarize, the Crack-the-Code task was predicted mostly by the Simultaneous processing score, particularly in the older grade levels, where it explained more than 1/3 of the Crack-the-Code variance. Matching Numbers Card 1 had about 1/3 of its variance explained in all other grade levels except in Grade 11; first by Attention in Grade 3, then by Successive and Attention together in Grade 5, then by Simultaneous and Attention together in Grade 7, and, finally, by Simultaneous in Grade 9. Attention was a significant predictor of Matching Numbers Card 2 in three of the five grade levels targeted; Successive was a significant predictor in grades 3, 5, and 7, after which it seemed to lose its predictive value. Planned Search scores were for the most part not predicted substantially by the three processing scores.

In general, attention and simultaneous and successive processing seem to be relevant cognitive correlates for performance on Crack-the-Code and Matching Numbers, but not necessarily for Planned Search. The coding and attention demands for Planned Search are more limited than for the other two tasks, so in retrospect this result is not surprising. What the relevant cognitive correlates for this type of visual search tasks are remains to be seen.

How can we explain the changes in significant predictors of planning scores from one grade level to another? The disappearance of the Successive processing score as a significant predictor of Matching Numbers Card 2 in the older grade levels would seem to suggest that compared to the younger participants, Grade 9 and Grade 11 students approached the task in a qualitatively, rather than quantitatively, different manner. Perhaps planning development in this task consists of developing better methods for solving the task rather than becoming more efficient in applying the same method. For example, in Matching Numbers Card 2, the younger participants may try to compare the whole 4 and 5 digit numbers, whereas the older participants may notice that comparing only parts of these numbers is more efficient and places less strain on memory. This would explain why Successive processing, which was assessed by two memory tasks, would lose its importance as a predictor.

In contrast, it is possible to suggest an alternative interpretation of planning development for Crack-the-Code and Matching Numbers Card 1 which were predicted more consistently by the same processing score (Simultaneous and Attention, respectively). Perhaps what develops in these tasks is the participant's competence in applying a particular method. The change is quantitative rather than qualitative. Although this interpretation appears reasonable for Matching Numbers Card 1, which is a simple and quickly executed task, the same cannot be said for Crack-the-Code, which is a complex multistep planning task. The amount of information that a participant has to process and integrate into a functional schema is considerably higher in Crack-the-Code than in any other planning tasks used in this study. Therefore it is understandable that it would correlate significantly with encoding factors such as simultaneous and successive processing. Why this is not so in Grade 3 is more difficult to explain.

Table 3-4 gives impetus to another question: How distinct is Matching Numbers Card 1 from the Receptive Attention task if it is predicted by it rather consistently? While part of the common variability may simply result from the fact that they are both timed tests, it is not likely to explain all or even most of the shared variance (after all, Planned Search is also a timed test and it was not predicted significantly by the Attention score). On Matching Numbers Card 1, the participant has to find two matching numbers (1 to 3 digits) in a row of 8 numbers of the same length. In Receptive Attention, the participant has to decide whether two pictures given as a pair match, either on the basis of their simple taxonomic category (trees, fruits, flowers, birds, houses, or human faces), or on the basis of their appearance (i.e., do they look the same). Both of these tasks require resistance to distraction and control of impulsivity, as well as shifting attention from one target to another. The participant may also scan the rows on Card 1 in the same manner to which the rows on Receptive Attention are scanned (one at a time from left to right). In this case, however, the working memory requirements for these tasks seem to differ. Card 1 requires the participant to compare new stimuli numbers simultaneously with all the previous numbers, whereas the only working memory requirement of Receptive Attention is the decision criteria given in the beginning of the item (either the same name or the same appearance). The other possibility is that the participant may decide to scan the row of numbers differently than from left to right, for example, starting from the middle or going back and forth. Particularly in Card 2, these alternative scanning methods should be more prominent as the length of numbers to be compared increases. While Card 1 certainly has commonalities with Receptive Attention, we believe that these differences are sufficient to justify our categorization of them as planning and attention tasks, respectively.

It should not be overlooked, however, that all planning measures include attention, successive, and simultaneous components, and vice versa. The answer is not to look for "pure" measures of planning; instead, we should look for tasks in which the planning demands outweight the demands for the other cognitive processes. Thus, while factor analyses should not be used as sole determiners of what tasks measure for example planning, together with task analyses they do provide important information regarding what tasks can be reliably labeled as planning, attention, simultaneous, and successive measures.

Our results also support the earlier findings that suggest that working memory is generally not a significant predictor of planning performance. In this study, successive processing was measured by two working memory tasks and in general, it emerged as a significant predictor of planning only in Grade 5.

We should note here that our processing scores were not optimal because we had data only from one Attention task (with two parts). This may have affected the results and in future research, more attention tasks should be included.

General Discussion

Our results have direct implications for the assessment of planning. As we mentioned before, the inclusion of a planning construct into a model of intellectual functioning has theoretical, clinical, and empirical support (Telzrow, 1990). How this planning is efficiently assessed is the natural follow-up question. We believe that researchers need to develop standardized planning measures that can be administered and interpreted. Most planning research has involved errand running, scheduling, or complex problem solving tasks that simulate real-life planning situations. While these tasks certainly have greater ecological validity than, for example, the Tower of Hanoi task or the planning tasks used in this study, they may not be practical for clinical use. Moreover, intervening and in real-life planning tasks important factors such as familiarity and practice may affect the results.

In contrast, laboratory planning tasks, such as the ones used in this study, offer an abstracted planning space where the participant has complete control over the necessary resources and the outcome of the plan (Scholnick & Friedman, 1993). Assuming that people do plan in these situations, their performance should reflect true variability in planning skills. But sometimes participants bring their own expectations and goals to the testing situation, and these have an effect on how they perform (Scholnick & Friedman, 1993). Also, because the goals are provided, one important component of everyday planning is eliminated, that is, the process of setting a goal. Thus, certain caution is necessary in interpreting the results of the present study. These results need to be replicated and fine-grained analyses of the process of planning as it develops in performing these tasks is necessary — this is the topic of our current research.

Our results suggest that different tasks have different discriminating values for different age groups. This is not surprising, given that Scholnick and Friedman (1993) suggested that developmental differences are more apparent in experimental tasks that are novel and computationally demanding than in tasks that require application of knowledge to a familiar situation. The question is: When does an experimental task lose its novelty as a measure of cognitive processing? Figure 3-1 suggests that by Grade 7, the simplest planning tasks were already beyond this stage, whereas the same was not true for more complex tasks. Thus, simpler tasks would have limited value in diagnosing differences in planning skills during the later school years. They may, however, still retain clinical usefulness as identifiers of gross planning deficits, an issue that was not addressed in this study.

The most complex planning task, Crack-the-Code, did not differentiate between the three younger grade levels. There are at least three possible explanations for younger participants' poor performance on this task: (1) They were incapable of formulating a functional plan; (2) they were capable of formulating a functional plan but failed in applying it; or (3) the task failed to elicit a planning response in the first place. Preliminary data from think aloud protocols from Grade 4 and 6 students solving the same task (Parrila & Das, 1995) suggests that in most cases, the last explanation is not accurate. Students do produce plans to solve the problem but these plans are either insufficient or their application fails.

Our second question dealt with the uniformity of the planning construct. As expected, planning tasks did not load on one factor and Crack-the-Code seemed to have a considerable distance from the other planning tasks. The two planning factors were intercorrelated; after all, they are assumed to measure two different sides of the same entity — cognitive planning. Thus, our conceptualization of action- and operation-planning was partially supported. The three level model of planning has, as suggested in this article, intuitive value in attempting to integrate within one framework diverse findings in the planning literature. Whether it has clinical or educational applications remains to be seen. Developmental data presented in this study suggests that operation-planning develops, and is entrenched, before action-planning. Similarly, it may be possible to identify subgroups of children with developmental disabilities on the basis of their performance in operation- and action-planning tasks (see e.g., Snow, 1992). If operation-planning develops earlier, then tasks assessing it should be less sensitive indicators of planning deficiencies than action-planning tasks. Results from a study comparing the performance of two groups of children (with and without mental retardation) on an action-planning task (shopping in a make-believe grocery store) suggests that this may be the case (Szepkouski, Gauvain, & Carberry, 1994). Children without mental retardation were significantly better planners than mentally challenged children (mean IQ 49.3). However, no significant differences were found in the variable (item location strategy) that can be conceptualized as an operation-planning subtask within an action-planning task. This suggests that mentally challenged children were capable of utilizing operation-planning but not necessarily action-planning.

Activity-planning, the third level in our model, is the most difficult to study. Again, this level makes intuitive sense: We all know of someone who appears to be a bright and capable problem-solver but time after time ends up making wrong choices in his or her personal life. Thus, it seems that at least in some cases, problem solving ability (operation- or action-planning) does not correlate with activity-planning. We also have some preliminary empirical support for this view. Nurmi (1989), for example, asked 11- and 15-year-old adolescents about their future hopes and how they intended to realize these. Participants' responses were then scored in terms of the level of planning that their responses displayed. Both groups were clearly interested in their future and had plans. Also, level of planning increased with age but this measure did not correlate significantly with the logical reasoning task in either age group.

If the three levels of planning can be differentiated, then the question is: Are we teaching relevant planning skills in school to develop all levels of this complex cognitive activity? Mathematical and science problem solving and composition writing have traditionally been avenues for teaching planning skills. All of these areas, while undoubtedly important, seem to concentrate on operation- and action-planning. Activity-planning may be difficult to teach directly but it seems likely that the better conceptual knowledge and understanding of social reality one has, the easier it is to come up with a plan that has chances of taking one towards the important goals rather than away from them. But again, we need to know more about activity-planning before we can say anything definite. What makes a good activity-plan? What are the cognitive, as well as noncognitive, correlates of activity-planning? These are only some of the questions that should be addressed in future studies.

The third section of our study tested the significance of attention, simultaneous, and successive processing as cognitive correlates of operation- and action-planning. In our earlier work (Parrila et al., 1994) we found that significant predictors changed as a function of task. Not surprisingly, current results showed that significant predictors changed both as a function of age and as a function of the task. Development of planning is not only quantitative but also qualitative and, accordingly, different components should not be expected to correlate equally in different ages. Also, planning is often context sensitive and not all planning tasks make the same processing demands. Both of these facts are reflected in the correlations and significant predictors of planning performances displayed in Tables 3-3 and 3-4.

Planning, attention, simultaneous, and successive processing are closely interrelated in Luria's (1973) model of brain functioning. They are also to some extent hierarchically organized, and while planning as a cognitive function is always dependent on the functioning of the attention-arousal and coding units, the reverse can also be true. For example, the same information can be coded in many different ways and how one manages the information is a planning function that influences the coding approach used. Similarly, what information is attended to, and to what extent, can be influenced by a plan controlling one's overall approach to the task at hand. Thus, when we observe higher correlations between Crack-the-Code and Simultaneous processing scores in grades 9 and 11 than in younger grades, this could be due to both the increasing influence of action-planning skills on simultaneous coding and the importance of simultaneous processing skills in this visually challenging task.

It is likely that attention, and simultaneous and successive processing would not have such an impact, at least directly, on activity-planning. The PASS *processes* are relevant for activity-, action-, and operation-planning, but the *tasks* that we have used should directly relate only to action- and operation-planning. Our conceptualization of planning in three levels emphasizes the richness and complexity of planning behaviors, which is often neglected in experimental planning studies (Benson, Haith, & Bihun, 1995; Scholnick & Friedman, 1993). Accordingly, cognitive correlates or "components" of planning should also be varied and rich. We suggest that self-concept and self-esteem, as well as personal values such as 'one should be fair' or 'one should share' should be studied as independent variables in activity-planning together with more cognitive variables. To develop appropriate measures for activity-planning, then, remains a challenge for future research.

We wish to add two further comments in concluding this paper. The first relates to the cultural context of cognitive planning, and indeed of cognitive processing represented by the tasks in this study. Although in our present study the sample consisted of students in India and in the previous study (Parrila et al., 1994) the sample consisted of students in Finland, we observed very similar patterns. Generally speaking, the cognitive processes measured by the tasks in our two studies — and also in several other studies reported by Das et al. (1994) — fit within a universal context across many cultures. Previous research, however, has also shown that different cultural groups may approach the cognitive tasks in different ways. For example, Das et al. (1994) noted that Canadian Native children do not typically use successive processing, preferring to use simultaneous processing instead. In contrast, high-caste Indian children seem to prefer successive processing strategies for solving simultaneous tasks. In terms of planning, our results suggest that operation-planning and action-planning tasks are likely to retain their common and essential processing characteristics across considerable cultural differences. Moreover, recent results from a Canadian elementary and junior high school population (Parrila & Das, in preparation) display the same two-factor structure for planning tasks evident in the present study. Cultural influences on activity-planning, however, are likely to be considerably greater due to the inherently social nature of this type of planning. Also, social and cultural influences in planning are more likely to be apparent in tasks that simulate real-life planning situations than in laboratory planning tasks such as those used in this study.

Our second comment concerns the 'applied' use of the cognitive tasks, especially, the planning tasks. The planning concept and tests are related to real-world problems. Planning has practical value not only for the understanding and assessment of cognitive abilities of developmentally normal and deviant children, but also for designing prescriptive remedial training. Some of the other contexts in which planning has been used as an explanatory concept include children's writing of compositions and managers' success in vigilant decision-making (see Das et al., 1994). Planning as it is conceptualized in our tripartite model permeates all purposeful and goal-oriented behavior. We believe that future research will expand its applications as it advances the concept of planning.

References

Allport, A. (1993). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631-682). Cambridge, MA: The MIT Press.

Ashman, A. F. & Das, J. P. (1980) Relation between planning and simultaneous-successive processing. *Perceptual and Motor Skills, 51,* 371-382.

Becker, M. G., Isaac, W., & Hynd, G. W. (1987). Neuropsychological development of nonverbal behaviors attributed to "frontal lobe" functioning. *Developmental Neuropsychology, 3,* 275-298.

Benson, J. B., Haith, M. M., & Bihun, J. T. (1995). The development of planning: Where's the future? Paper presented at the SRCD conference in Indianapolis, March 1995.

Casey, M. B., Bronson, M. B., Tivnan, T., Riley, E., & Spenciner, L. (1991). Differentiating preschoolers' sequential planning ability from their general intelligence: A study of organization, systematic responding, and efficiency in young children. *Journal of Applied Developmental Psychology, 12* (1), 19-32.

Cocking, R. R. & Copple, C. E. (1987). Social influences on representational awareness: plans for representing and plans as representation. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 428-465). New York: Cambridge University Press.

Das, J. P. & Dash, U. N. (1983). Hierarchical factor solution of coding and planning processes: Any new insights? *Intelligence, 7,* 27-37.

Das, J. P. & Heemsbergen, D. B. (1983). Planning as a factor in the assessment of cognitive processes. *Journal of Psychoeducational Assessment, 1,* 1-15.

Das, J. P., Kirby, J., & Jarman, R. F. (1979). *Simultaneous and successive cognitive processes.* New York: Academic.

Das, J. P., Mensink, D., & Janzen, H. (1990). The K-ABC, coding, and planning: an investigation of cognitive processes. *Journal of School Psychology, 28,* 1-11.

Das, J. P., Mensink, D. & Mishra, R. K. (1990). Cognitive processes separating good and poor readers when IQ is covaried. *Learning and Individual Differences, 2,* 423-436.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes.* Needham Heights, MA: Allyn and Bacon.

Dash, U. N. & Mahapatra, S. (1989). Development and differentiation of simultaneous, successive and planning processes. *Psychological Studies, 34* (2), 75-83.

Dash, U. N. & Mohanty, A. (1992). Relationship of reading comprehension with metalinguistic awareness. *Social Science International, 8,* 5-13.

Dash, U. N., Puhan, B. N., & Mahapatra, S. (1985). Information integration modes in concrete operational skills of Grade two children. *Psychological Studies, 30* (1), 20-26.

Diamond, A. & Goldman-Rakic, P. S. (1983). Comparison of performance on a Piagetian object permanence task in human infants and rhesus monkeys: Evidence for involvement of prefrontal cortex. *Neuroscience Abstracts, 9,* 641.

Diamond, A. & Goldman-Rakic, P. S. (1985). Evidence for involvement of prefrontal cortex in cognitive changes during the first year of life: Comparison of human infants and rhesus monkeys on a detour task with transparent barrier. *Neuroscience Abstracts, 11,* 832.

Diamond, A. & Goldman-Rakic, P. S. (1986). Comparative development in human infants and infant rhesus monkeys of cognitive functions that depend on prefrontal cortex. *Neuroscience Abstracts, 12,* 742.

Dreher, M. & Oerter, R. (1987). Action planning competencies during adolescence and early adulthood. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 321-355). New York: Cambridge University Press.

Fabricius, W. V. (1988). The development of forward search planning in preschoolers. *Child Development, 59,* 1473-1488.

Gardner, W. & Rogoff, B. (1990). Children's deliberateness of planning according to task circumstances. *Developmental Psychology, 26,* 480-487.

Gauvain, M. (1989). Children's planning in social contexts: An observational study of Kindergarteners' planning in the classroom. In L. T. Winegar (Ed.), *Social interaction and the development of children's understanding* (pp. 95-117). Norwood, NJ: Ablex.

Gauvain, M. (1992). Social influences on the development of planning in advance and during action. *International Journal of Behavioral Development, 15,* 377-398.

Gauvain, M. & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25,* 139-151.

Göncü, A. & Kessel, F. (1988). Preschoolers' collaborative construction in planning and maintaining imaginative play. *International Journal of Behavioral Development, 11,* 327-344.

Haake, R. J., Somerville, S. C., & Wellman, H. M. (1980). Logical ability of young children in searching a large-scale environment. *Child Development, 51,* 1299-1302.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3,* 275-310.

Hudson, J. A. & Fivush, R. (1991). Planning in the preschool years: The emergence of plans from general event knowledge. *Cognitive Development, 6,* 393-415.

Kagan, J. (1987). Foreword. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. xi-xiv). New York: Cambridge University Press.

Kar, B. C., Dash, U. N., Das, J. P., & Carlson, J. (1993). Two experiments on the dynamic assessment of planning. *Learning and Individual Differences, 5,* 13-29.

Kirk, U. & Kelly, M. S. (1986). Children's differential performance on selected dorsolateral prefrontal and posterior cortical functions: A developmental perspective. *Journal of Clinical and Experimental Neuropsychology, 7,* 604.

Klahr, D. & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology, 13,* 113-148.

Kozulin, A. (1986). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. *American Psychologist, 41,* 264-274.

Kreitler, S. & Kreitler, H. (1987). Conceptions and processes of planning: The developmental perspective. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 205-272). New York: Cambridge University Press.

Leontjev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 37-71). New York: M. E. Sharpe.

Leontjev, A. N. (1978). *Activity, consciousness, and personality.* Englewood Cliffs, NJ: Prentice-Hall.

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H., Ringholz, G., Ewing-Cobbs, L., & Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 6,* 377-395.

Levina, R. E. (1981). L. S. Vygotsky's ideas about the planning function of speech in children. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 279-299). New York: M. E. Sharpe.

Luria, A. R. (1959). The directive function of speech in development and dissolution. *Word, 15,* 341-352.

Luria, A. R. (1966). *Human brain and psychological processes.* New York: Harper & Row.

Luria, A. R. (1973). *The working brain.* New York: Basic Books.

McGillicuddy-De Lisi, A. V., De Lisi, R., Flaugher, J., & Sigel, I. E. (1987). Familial influences on planning. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 395-427). New York: Cambridge University Press.

Miller, G. A., Galanter, E. H., & Pribram, K. H. (1960). *Plans and the structure of* *behavior.* New York: Holt, Rinehert & Wilson.

Naglieri, J. A., & Das, J. P. (1990). Planning, attention, simultaneous and successive (PASS) cognitive processes as a model for intelligence. *Journal of Psychoeducational Assessment, 8,* 303-337.

Naglieri, J. A., & Das, J. P. (1988). Planning-Arousal-Simultaneous-Successive (PASS): A model for assessment. *Journal of School Psychology, 26,* 35-48.

Naglieri, J. A., & Das, J. P. (1987). Construct and criterion-related validity of planning, simultaneous, and successive cognitive processing tasks. *Journal of Psychoeducational Assessment, 4,* 353-363.

Naglieri, J. A., Prewett, P. N., & Bardos, A. N. (1989). An exploratory study of planning, attention, simultaneous, and successive cognitive processes. *Journal of School Psychology, 27,* 347-364.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on general problem-solving program. Proceedings of the International Conference on Information Processing, Paris.

Nurmi, J. (1989). Development of orientation to the future during early adolescence: A four-year longitudinal study and two cross-sectional comparisons. *International Journal of Psychology, 2,* 195-214.

Parrila, R. K. (1995). Vygotskian views on language and planning in children. *School Psychology International, 16,* 167-183.

Parrila, R. K. & Das, J. P. (1995). Developmental changes in the planning process: A think-aloud study. Manuscript in preparation. University of Alberta, Edmonton, Canada.

Parrila, R. K., Äystö, S., & Das, J. P. (1994). Development of planning in relation to age, attention, simultaneous, and successive processing. *Journal of Psychoeducational Assessment, 12,* 212-227.

Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behaviors attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 1,* 349-370.

Pea, R. D., & Hawkins, J. (1987). Planning in the chore-scheduling task. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 273-302). New York: Cambridge University Press.

Piaget, J. (1963). *The origins of intelligence in children.* New York: Norton.

Pitt, R. B. (1983). Development of a general problem-solving scheme in adolescence and early adulthood. *Journal of Experimental Psychology: General, 112,* 547-584.

Posner, M. I. & Boies, S. J. (1971). Components of attention. *Psychological Review,* *78*, 391-408.

Rogoff, B., Gauvain, M., & Gardner, W. (1987). The development of children's skills in adjusting plans to circumstances. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 303-320). New York: Cambridge University Press.

Rogoff, B., Mistry, J., Radziszewska, B., & Germond, J. (1992). Infants' instrumental social interaction with adults. In S. Feinman (Ed.), *Social referencing and the social construction of reality in infancy* (pp. 323-348). New York: Plenum Press.

Scholnick, E. K. & Friedman, S. L. (1987). The planning construct in the psychological literature. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking*, (pp. 1-38). New York: Cambridge University Press.

Scholnick, E. K. & Friedman, S. L. (1993). Planning in context: Developmental and situational considerations. *International Journal of Behavioral Development, 16*, 145-167.

Siegler, R. S. (1991). *Children's thinking* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Snow, J. H. (1992). Mental flexibility and planning skills in children and adolescents with learning disabilities. *Journal of Learning Disabilities, 25*, 265-270.

Sperry, R. W. (1993). The impact and promise of the cognitive revolution. *American Psychologist, 48*, 878-885.

Stuss, D. T. & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Stuss, D. T. & Benson, D. F. (1987). The frontal lobes and control of cognition and memory. In E. Perecman (Ed.), *Frontal lobes revisited* (pp. 141-158). New York: IRBN Press.

Szepkouski, G. M., Gauvain, M., & Carberry, M. (1994). The development of planning skills in children with and without mental retardation. *Journal of Applied Developmental Psychology, 15*, 187-206.

Telzrow, C. F. (1990). Does PASS pass the test? A critique of the Das-Naglieri Cognitive Assessment System. *Journal of Psychoeducational Assessment, 8*, 344-355.

Teuber, H.-L., Battersby, W. S., & Bender, M. B. (1949). Changes in visual searching performance following cerebral lesions. *The American Journal of Physiology, 159*, 592.

Wellman, H. M., Fabricius, W. V., & Sophian, C. (1985). The early development of planning. In H. M. Wellman (Ed.), *Children's searching* (pp. 123-149). Hillsdale, NJ: Erlbaum.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function. *Developmental Neuropsychology, 7,* 131-149.

Willatts, P. (1984). The stage-IV infant's solution of problems requiring the use of supports. *Infant Behaviour and Development, 7,* 125-134.

Willatts, P. (1990). Development of problem-solving strategies in infancy. In D. F. Bjorklund (Ed.), *Children's strategies* (pp. 23-66). Hillsdale, NJ: Erlbaum.

IV. DEVELOPMENTAL CHANGES IN THE PLANNING PROCESS: A THINK-ALOUD STUDY

Introduction

This study focuses on developmental and individual differences in the planning process as displayed by students from Grades 4, 6, and 8 in response to a novel and abstract task. That children reason differently within an age level or even between two trials of the same task has been demonstrated in several studies (see e.g., Lemaire & Siegler, 1995; McGilly & Siegler, 1989; Siegler, 1995). The current study expands on these findings and utilizes both verbal and computer protocols to examine the development of planning skills.

The term *planning* has been used to denote a multitude of behaviors that, overtly, may not seem to have much in common. As a consequence, the definition of the term has become vague and two theorists using the terms "planning" or "plan" may not share the same focus (Scholnick & Friedman, 1987). For example, Newell, Shaw, and Simon (1959) viewed planning as problem solving in a simplified, abstracted problem space, whereas for Hayes-Roth and Hayes-Roth (1979), planning consisted of anticipating a goal-directed course of action. In their integrative review of the conceptualization of planning, Scholnick and Friedman (1987) identified six components of planning: (1) forming a representation of the problem, (2) choosing a goal, (3) deciding to plan, (4) formulating a plan, (5) executing and monitoring the plan, and (6) learning from the plan. For the purpose of this article, we define planning broadly as a self-organizing reflective activity that integrates information from external and internal sources to create and execute meaningful behavioral responses. When defined this broadly, planning includes such related concepts as strategic thinking, problem solving, and some forms of executive functioning. Moreover, we assume that the planning process consists of four components, which are displayed in Figure 4-1.

The planning components in Figure 4-1 — representation, anticipation, execution, and regulation -- are derived from the neuropsychological view of planning as a frontal lobe function (Stuss & Benson, 1986, 1987). We believe that these components also adequately summarize planning as a cognitive function in knowledge-lean tasks, such as the one used in this study. The following sections will describe each of the planning components in more detail.

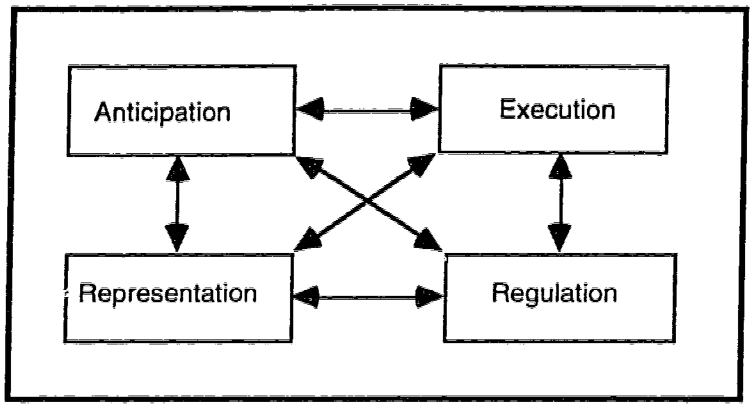


Figure 4-1. The four planning components.

Representation

Representation, or understanding, in a multi-step task can be divided into three interrelated areas: (1) understanding of the initial problem state, (2) generation of the necessary search methods or strategies, and (3) understanding of the end state or the goal (VanLehn, 1989).

The main function of the initial problem representation, called a *task definition,* is to generate the planner's internal representation of the spatial and causal structure of the task (Scholnick & Friedman, 1987). In knowledge-lean tasks, the necessary information is given in instructions and/or a task display and the participant's task is to represent this information in a format that maximally supports the solution process. This tends to be a demanding task for young participants and part of the information necessary to complete the task may be ignored. According to Scholnick and Friedman (1987), theories of planning vary according to the attributed possible inadequateness of representation. Those in favor of "veridical" models argue that representations are faithful to the problem space but they may be incomplete. In contrast, "idiosyncratic" models suggest that failures in problem solving result from distortions (rather than gaps) in representation.

A participant can also progress beyond the information provided and engage in "elaboration" (VanLehn, 1989), or "selective encoding" and "selective combination" of information (Davidson & Sternberg, 1984). A task definition consists of a set of assertions that reflect the participants' current knowledge about the task. Participants can add, delete, and modify these assertions as they generate more information. Elaboration is a new assertion that is added to the task definition "without removing any of the old assertions or decreasing their potential relevance. . . . A new assertion qualifies as an elaboration because it does not negate, remove, or obviate any of the older assertions" (VanLehn, 1989, pp. 539-541). Elaborations that selectively combine relevant existing assertions are particularly powerful in the Crack-the-Code task used in this study since they can rule out some imaginable moves and focus the planner's attention on the essential features of the task. Readers familiar with the math and science problem solving literature may find that the concept of a task definition is relatively close to such constructs as "knowledge generation" (see e.g., Lawson & Chinnappan, 1994) and "conceptual understanding" (see e.g., Schauble, 1996).

Search methods refer to the various strategies that a participant may have available or can generate, and are used for finding the solution to the problem at hand. Search methods are not provided in the instructions. Instead, participants have to either create new (unique) methods particularly suited for this task (based on their task definition) or import more general "weak-methods" from their knowledge base of what worked earlier for similar kinds of problems. The selection of different methods is a function of which search methods are available and what the planner's task definition is. If a planner has only one method available, then this method is used. In contrast, if a planner has several alternative methods available in his or her repertoire, then the interplay between specific task constraints and the planner's developing understanding of them will determine which method is chosen.

Goal representation refers to the participant's understanding of the desired goal state and when it is reached. Goal representation is often neglected in experimental studies. The reason for this is that the goal state may be given explicitly to the planner or it may be easily recognized, such as in the Tower of Hanoi task. In the Crack-the-Code task, the various conditions for a correct answer are given in the instructions. The planner has to keep all of them active and continuously evaluate performance against them. In this sense, the planner is constantly reconstructing the goal state.

Anticipation

Anticipation is closely related to representation and interaction between these two components (which are represented in Figure 4-1 by a two-headed arrow) can greatly facilitate planning. Anticipation in this context refers to the ability to predict the consequences of a plan or a partial-plan prior to acting upon them. Thus, in its most elaborate form, it includes the covert testing of plans constructed by the representation component. A more limited use of anticipation would include evaluating the appropriateness of a single move before its completion. The main function of anticipation in a knowledge-lean task such as Crack-the-Code is to minimize erroneous moves and decisions. Note that anticipation is particularly important in tasks in which later corrections are costly and advance planning is emphasized, whereas in Crack-the-Code, which allows for corrections, the use of anticipation may be deliberately minimized due to its relatively large cognitive demands.

We should note here that anticipation is not equal to "anticipatory planning." Anderson (1983, p. 167) suggested that "if it can be shown that a system reorders a preferred sequence of actions in anticipation of goal conflict, then that system is engaging in planning." Anticipation is used here in a more limited sense: If the participant rejects even one decision before its completion in action, then he or she is engaged in anticipation.

Execution

Execution refers to the application of plans in the environment. This can take place simultaneously with the plan formation or after the plan has been finalized mentally. The first type of planning is generally referred to as opportunistic planning (see e.g., Hayes-Roth & Hayes-Roth, 1979) or planning-in-action (De Lisi, 1987; Parrila, Äystö, & Das, 1994; Parrila, Das, & Dash, in press). The second type of planning implies that execution of a plan is a separate process during which the blueprint or plan is carried out in the material world. Planning models that adhere to this definition are often referred to as top-down or anticipatory models. In both cases, execution refers to overt behavior on the planner's part. Note that in top-down models, it is possible for a participant to construct a perfect plan in advance but then fail in its overt execution.

Regulation

Regulation refers to a myriad of skills and activities that are responsible for keeping the planner "on track." Regulation includes monitoring and controlling of behavior according to the plan and evaluating the plan and its appropriateness once it is executed or while it is being executed. The lack of monitoring and controlling of behavior may lead to execution failures and task inappropriate behavior. The second regulatory function, evaluation, can compensate for the lack anticipation; if there is no anticipation, then the need for evaluation is greater. In contrast, good anticipation may preempt the need for evaluation (but not the need for the regulation of behavior). Naturally, the most important function of evaluation is to detect when the goal is reached. The success in goal detection, in turn, depends directly on the representation of all the rules and conditions of the task that the goal has to meet. Hence, as also noted by Scholnick and Friedman (1987), the failures in evaluation can result from inappropriate standards as well as inaction (lack of evaluation).

We assume that the planning process in knowledge-lean multistep tasks, such as Crack-the-Code, proceeds as a continuous cycle of refining representation, anticipating outcomes, executing plans and subplans, evaluating the outcomes, and regulating the behavior accordingly. The two-headed arrows in Figure 4-1 attempt to capture this recursive nature of planning. This cycle does not, however, necessarily include all the components. Anticipation of the outcomes may be missing and the participant may "jump" into action (e.g., "trial-and-error" behavior) or in the most extreme case, both anticipation and evaluation may be missing. Representation and execution, in contrast, have to be present by definition, although the former may be severely limited and thus not very useful for the solution process.

We should also acknowledge that the planning model presented in Figure 4-1 is not adequate to account for most real-life planning tasks. At least three constructs would need to be added for that purpose: selection of goals and objectives, knowledge base, and consciousness. In a knowledge-lean task, the goal and knowledge base necessary for planning are given by definition. In contrast, many real-life planning tasks can have several different goals, none of which are more or less correct than the others, or the goal can be constructed and modified by the planner during the process. While this is also partly the case in Crack-the-Code, it does have a defined correct goal state. Success in real-life planning tasks often requires that planners have the relevant knowledge and strategies available to them. For example, older planners can be more successful because they know more about the task at hand, about planning in general, and about their own strengths and weaknesses as planners in particular. Moreover, real-life planning skills are often learned from more competent planners, as has been suggested by Vygotsky and his coworkers (e.g., Levina, 1981), and more recently by Goodnow (1987) and Rogoff and her associates (e.g., Baker-Sennett, Rogoff, & Matusov, 1993). In order to explain planning in real-life settings, Figure 4-1 should be extended to include "consciousness" as an overarching, or perhaps, meta-component of planning that gives meaning to, and guides an individual's activities, actions, and operations (Parrila et al., in press). Sperry (1993) conceptualized consciousness as a determining force of mental activities, as opposed to merely an emergent phenomenon of the physical basis of mental activities. We think that the extension of planning to incorporate consciousness will make it easier to understand those forms of planning that are guided by one's life goals. Even in childhood, planning can be guided by a child's self concept. For example, the statement "I am a bookworm like my grandma" will shape and determine a child's planning of activities differently than the statement "I will be like Wayne Gretzky". Recent interest in improving children's self-esteem in order to better their academic performance is another illustration of the importance of consciousness in explaining, justifying, and guiding children's activity.

Development of Planning

Not unlike many other areas of the developmental literature, studies on planning development have focused far more on identifying general trends or specific group differences than producing more detailed descriptions of individual differences and the mechanisms of change. Accordingly, we seem to have more knowledge about the problems that school-age children learn to solve at different age-levels than we do about *how* they solve these problems, or how they learn to plan. We know, for example, that some components of planning such as the ability to span temporal separation between two actions, inhibit prepotent responses, and form a goal and produce activity that is directed toward it are already apparent during infancy (Diamond & Goldman-Rakic, 1983, 1985, 1986; Willatts, 1984, 1990). Most researchers, however, have expressed some doubt regarding infants' ability to form plans and strategies. For example, planning is commonly associated with the development of the capacity to manipulate mental representations and to regulate behavior verbally and hence, it is not expected to appear until later in ontogenesis (see e.g., Levina, 1981; Luria, 1973; Piaget, 1963).

As children's language skills develop, they are increasingly able to both initiate and inhibit activities guided by internal speech. Preschool and kindergarten children are capable of devising and executing simple plans (see e.g., Casey, Bronson, Tivnan, Riley, & Spenciner, 1991; Fabricius, 1988; Gauvain, 1989; Haake, Somerville, & Wellman, 1980; Klahr & Robinson, 1981; Wellman, Fabricius, & Sophian, 1985), particularly if these plans are based on their knowledge of familiar events (Hudson & Fivush, 1991) and are aided by the utilization of overt speech (Cocking & Copple, 1987; Göncü & Kessel, 1988). Moreover, before entering formal schooling, children have at least some "metaplanning" skills; in other words, they have a preliminary conceptual understanding of what planning is and when it is needed (Gauvain, 1989; Kreitler & Kreitler, 1987).

Planning skills seem to develop particularly rapidly during the school years. Children learn to apply their skills in planning to a broad range of tasks. By age 8, for example, children learn to complete tasks such as simple forms of Tower of Hanoi (Klahr & Robinson, 1981; Welsh, Pennington, & Groisser, 1991), mazes and errand running tasks (Gardner & Rogoff, 1990; Gauvain & Rogoff, 1989; Pea & Hawkins, 1987), and Delayed Alternation (Levin et al., 1991). By age 10, children should be proficient in the Wisconsin Card Sorting task (Kirk & Kelly, 1986; Levin et al., 1991; Welsh et al., 1991), the Rey-Osterrieth Complex Figure task (Kirk & Kelly, 1986), and the Matching Familiar Figures Test (Welsh et al., 1991). Development of planning skills, particularly in more complex and demanding tasks, may be an especially prolonged process that continues well into adolescence and beyond (see, e.g., Becker, Isaac, & Hynd, 1987; Dreher & Oerter, 1987; Levin et al., 1991; Parrila et al., 1994; Passler, Isaac, & Hynd, 1985; Pitt, 1983; Welsh et al., 1991).

Whenever qualitative changes have been hypothesized to explain the observed quantitative changes in performance levels, these have more often than not been ad hoc additions rather than variables that were the focus of study. Also, such confounding questions as "Can children apply a successful strategy or plan if they possess one?" have received little attention. Children may, for example, be able to construct partial plans but still fail to complete the task successfully (Klahr & Robinson, 1981; Wilkinson, 1982), or they may have problems following a multistep plan to its conclusion, as was demonstrated by Hudson and Fivush (1991) with preschool age participants. In this case, studies concentrating on the end product would provide a more negative picture of children's planning skills than studies concentrating on the process of planning.

What makes the review of planning development even more complicated is the fact that few researchers have agreed on their definition of planning or what core components a developmental description should include (see also Friedman, Scholnick, & Cocking, 1987). As a consequence, the literature on planning development has multiple foci and results from different studies can seldom be reliably compared.

In what follows, we will briefly review selected studies that have provided information on the developmental changes in our four planning components. When possible, we will also explicate what predictions regarding the results of this study were made on the basis of these results. According to Ellis and Siegler (1994), the development of problem solving in knowledge-lean experimental tasks, such as Crack-the-Code, follows a different course from tasks with which children have extensive experience. In accordance with the focus of this study, we will mainly review studies that have used knowledge-lean tasks, such as the balance scale task and Tower of Hanoi, and confounding issues such as knowledge base or the effect of schooling and other kinds of social interactions are not dealt with here.

Development of representation. Development of planning in knowledge-lean tasks often results from children building a better representation of the task. A good representation of the problem and the information provided is as veridical and encompassing as possible (Ellis & Siegler, 1994). Siegler (1991, p. 285) suggested that "how well children encode critical information in the task and how well they can use the encodings to form mental models are among the key determinants of their success on many problems." Thus, the first condition for forming a functional task definition is that the participant encodes the key structural features of the task. In terms of the Crack-the-Code task, Parrila et al. (in press) demonstrated that school-age participants' simultaneous coding (see e.g., Das, Naglieri, & Kirby, 1994) skills predicted significantly the accuracy of their performance on Crack-the-Code. Better encoding of information has been shown to affect problem solving efficiency in such diverse tasks as analogical reasoning (Sternberg & Rifkin, 1979), insight problem solving (Davidson & Sternberg, 1984), Tower of Hanoi (Klahr & Robinson, 1981), and balance scale (Siegler, 1976). Better encoding, however, does not amount to faster encoding. Instead, several studies have shown that higher performing subjects may actually spend more time generating and manipulating information at the beginning of the task than lower performing subjects.

Also, the answer does not seem to be that better encoders can store more information in memory. Pea and Hawkins (1987), for example, found that a short-term memory task (Digit Span) did not have a significant correlation with the frequency of "high level planning decisions" when measured in two groups of school children (8 to 9 year-olds and 11 to 12 year-olds). This finding was supported by Parrila et al. (1994), who reported that a successive processing factor defined mostly by memory tasks had few significant correlations with planning tasks. The correlation between Crack-the-Code and the successive factor was .13 after the effect of age was controlled for. Parrila et al. (in press) found that although Crack-the-Code correlated significantly with a successive processing score (defined by 2 memory tasks) for Grades 5, 7, 9, and 11 in a sample of East-Indian students, it was a significant predictor of planning scores only for Grade 5. For other grade levels, a simultaneous processing score correlated even higher with Crack-the-Code and thus was entered first into a regression equation.

The above results suggest that rather than the "size" or span of working memory, what makes encoding superior is how information is manipulated once it has been entered into working memory. Bidell and Fischer (1994) proposed that developmental changes in the Tower of Hanoi task can be explained within the framework of the dynamic skills model. In the dynamic skills model, change in working memory capacity depends more on the nature of the structural relationships among the parts of representation than on quantitative increments in the amount of information stored. The emergence of new forms of structural relations in working memory is explained by an active process of coordination of component representational skills. The coordination of component skills leads not only to increased information capacity but also to qualitative changes in the ability to organize information on-line (see also Karmiloff-Smith, 1992, for a related argument).

Davidson and Sternberg (1984; Sternberg, 1988) suggested three processes that could enhance encoding and manipulation of information in working memory: selective encoding, selective combination, and selective comparison. The first two are relevant in the context of knowledge-lean abstract tasks. Selective encoding involves differentiating between relevant and irrelevant information and perceiving information that is not immediately obvious in a problem. Selective combination, in turn, involves synthesizing originally unrelated (or at least not obviously related) pieces of information into a functional whole that elaborates the representation. Thus, selective encoding involves deciding what information is relevant, and selective combination involves taking this relevant information and combining it into a more functional representation of the task and its constraints. Davidson and Sternberg (1984) also demonstrated that gifted children outperformed regular children in both of these representational skills. If this distinction also captures developmental differences in Crack-the-Code, we would expect to see between grade differences in our sample in both selective encoding (i.e., older children will verbalize less irrelevant information and more relevant information) and selective combination (i.e., older children will verbalize more elaborations on encoded information).

The second main component of representation is search methods or strategies. In general, older children have been found to use more advanced strategies than younger children. This not-so-unexpected result has been verified with several knowledge-lean tasks such as balance scale (e.g., Siegler, 1976), Tic-Tac-Toe (Crowley & Siegler, 1993), Mazes (Gardner & Rogoff, 1990), and Tower of Hanoi (Borys, Spitz, & Dorans, 1982). One reason for the better performance of older children is that they are generally believed to have more strategies in their repertoire, which they can also use more flexibly than the younger children (see e.g., Gardner & Rogoff, 1990).

Welsh (1991), however, found that although older participants (12-year-olds) made fewer errors on the Tower of Hanoi task, the type of errors were considerably similar across ages. She interpreted this finding to mean that when a participant's cognitive system was overloaded with depth of search requirements, he or she relied on "simpler procedures that had either worked in the past or had some likelihood of working under the current conditions" (Welsh, 1991, p. 72). When successful performance on the task demanded longer sequences of moves to be covertly hypothesized and tested prior to action, subjects who could not meet these demands characteristically returned to using "fall back rules" (Siegler, 1981). These fall back rules, or backup strategies (Siegler & Shipley, 1995), in turn, are the same strategies that younger participants use most of the time. Thus, even after children master a superior strategy they continue to sporadically use less superior ones when experiencing difficulties.

Children also seem to continue to experiment with other methods even after finding one fully functional method (Siegler, 1995). According to Siegler and Shipley (1995), prolonged use of multiple strategies has been verified in such diverse domains as addition, multiplication, spelling, time telling, serial recall, number conservation, and spatial reasoning, among others. This variability in children's thinking exists at every level: between children of different ages, between different children of the same age, within an individual solving the same problem twice, and even within an individual on a single trial. If variability is a defining feature of developmental changes in search methods, then we would expect that while older participants in general may use better search methods and solve more Crack-the-Code items correctly, many of them will at least sporadically use less effective search methods. Younger participants, in contrast, are expected to use less effective methods more often and advanced methods only sporadically. These expectations are in accordance with "the overlapping wave model" of development suggested by Siegler (1995).

Development of anticipation. Predicting the consequences of future actions and modifying the actions accordingly is a cognitive task of considerable difficulty for young participants. Barring the effects of direct learning, one could argue, in fact, that such activity is only possible after abstract symbolic thinking develops. VanLehn (1989) seems to share this point of view in suggesting that young children's plans are usually made close in time to the action and incorporate only a few actions. Klahr (1978) also noticed that his younger participants (4-year-old children) did not seem to test their strategies before implementing them. Instead, they broke the rules of Tower of Hanoi and attempted to reach the goal directly. Klahr continued that only the most advanced older children (6-year-olds) had the ability to mentally generate subgoals and to utilize strategies that temporarily moved the child further from the goal. This last point is a clear indication of anticipation, and Klahr's account of it suggests that some 6-year-olds are already capable of utilizing it. Accordingly, we expect that all our participants are capable of anticipation. This does not mean, however, that they necessarily will utilize this approach. Anticipation plays an important role particularly in planning situations in which potential errors are costly. This is not the case in Crack-the-Code and therefore we expect that the use of anticipation will reflect stylistic variations in planning rather than developmental differences in the capacity to engage in anticipation. Thus, within-group differences in anticipation can be as significant as between-group differences.

Development of execution skills. By using an easier task in which children could rely on their existing general event knowledge, Hudson and Fivush (1991) found that while 4-year-olds still experienced problems in executing two-step plans, 5-year-old children were able to construct and successfully execute plans with two separate goals that had to be coordinated. Thus, if our participants use limited "local" plans (one- or two-steps), then the execution of these plans should be relatively easy for our sample despite the fact that they are not expected to have general event knowledge of this kind of task. It is possible, however, that our participants build complex multistep plans prior to executing them (that is, they engage in complex anticipatory planning). Such planning demands a large working memory and flexible attentional strategies. Consequently, execution failures should be more prevalent.

Development of regulation. The act of planning implies that the planner is capable of regulating his or her behavior according to the plan. The first steps of this process are often equated with the development of verbally mediated regulation of behavior.

The directive function of speech is essential for the development of verbally mediated self-control. Verbally mediated self-control is the capacity to comply with another person's commands and directives in the absence of the person (Díaz, Neal, & Amaya-Williams, 1990). It is a necessary precursor of self-regulation that, in turn, can be defined as "the child's capacity to plan, guide, and monitor his or her behavior from within and flexibly according to changing circumstances" (Díaz et al., 1990, p. 130). Díaz et al. (1990) maintain that in self-control, the behavior is produced in response to internalized commands or directives; it is organized in rigid Stimulus-Response connections that environmental cues activate. In self-regulation, on the other hand, behavior is organized into hierarchical, functional systems that can be adjusted to changing goals and situations. Furthermore, behavior is now guided by self-formulated plans and goals, and aspects of the environment are used as tools and mediators to attain the goals.

Díaz et al.'s (1990) definition of self-regulation is similar to the regulation function that we have defined above. Several authors have suggested that the origins of self-regulation are social and are internalized during the first years of life (see e.g., Levina, 1981; Schubert, 1983; Vygotsky, 1978). Levina, for example, argued that the regulative function of children's own speech also appears first in the social, or "interpsychological" (Vygotsky, 1978), level. In other words, children will try to regulate other's behavior with the help of speech before their self-regulative capacities emerge. She suggested further that the child's early behavior is more likely to be influenced by factors in the external environment than by a conscious goal-directed plan. For Levina, speech is the means by which the child gains the power to impose a goal on the environment. In essence, speech allows children to liberate themselves from eminent but nonessential aspects of the environment and to focus on the essential: the representation of future action (see also, Vygotsky, 1978).

The capacity to regulate one's behavior verbally is, however, not a sufficient condition for efficient regulation of the planning process as it was defined above. Young children can still experience difficulties in evaluating and monitoring their planning process. According to Scholnick and Friedman (1987), cognitive science models of regulation generally suggest that the burden of planning comes from resource management and allocation of attention between components. As components of planning become more practiced and automatized, they require less effort and free resources from action for monitoring and evaluation. Furthermore, accurate evaluation, which is a necessary condition for efficient regulation of behavior, may be missing due to inappropriate standards. In other words, the goal representation is incorrect.

We expect that all our participants are capable of regulating their behavior according to the plan. However, since the goal state in Crack-the-Code is difficult to detect, we expect that evaluations will not necessarily be successful.

Purpose of the Study

Given the knowledge-lean task and rich data (computer and think aloud protocols) that allow us to detect both between-group as well as within-group differences, the main questions were: (1) What developmental differences can we detect in children's planning processes across the three grade levels targeted? and (2) How diverse are children's preferred planning methods and approaches across, as well as, within the grade levels?

The task, Crack-the-Code, is a complex planning task that is open to several different approaches. In contrast to tasks such as Tower of Hanoi, the participant does not have to generate complete move sequences in advance in order to solve Crack-the-Code. Rather, Crack-the-Code is open to both planning approaches (opportunistic planning and advance planning) that are frequently mentioned in the planning literature and thus allows for more stylistic variability. Moreover, since Crack-the-Code requires participants to manipulate colored chips on the computer screen, it provides ample visual and kinesthetic feedback and allows participants to make corrections at any point of their performance based on their utilization of this feedback. Thus, determining the correct solution occurs during the entire solution process rather than relying solely on the completeness of the participant's advance plan.

To separate developmental and individual differences in planning skills from the successful application of plans, we wanted to collect both procedural and product data. Since children naturally talk to themselves while engaged in difficult tasks, collecting verbal protocols should be a viable means for obtaining procedural data. The accuracy of the solution and performance time, in turn, will provide information about the final product.

We should note that neither development of the four planning components nor developmental changes in the Crack-the-Code task have been explored in this detailed manner before. Therefore the result section will be mostly exploratory in nature and we cannot know beforehand whether all reported scores or scoring procedures will turn out to be informative. Nevertheless, we believe that due to its complexity, Crack-the-Code relies especially on late developing abilities such as flexible management of one's resources and successive refinement of one's hypotheses and task representation. Thus, this task should be a good indicator of individual differences in planning skills during the school years and beyond.

Method

Participants

Participants included ten Grade 4 (5 females; mean age 9 years, 3 months; range = 8-11 to 9-9), ten Grade 6 (7 females; mean age 11 years, 4 months; range = 10-9 to 11-9), and nine Grade 8 (6 females; mean age 13 years, 6 months; range = 13-3 to 14-3) students. They attended an elementary school (Grade 4 and 6 participants) or a junior high school (Grade 8 participants) serving a middle and upper middle class suburban residential area in Edmonton, Canada. Our previous studies (see e.g., Parrila et al., in press) have indicated that the most significant changes in Crack-the-Code take place after Grade 4.

All participants volunteered to participate in the study (written permission was obtained from their parents or guardians), spoke English as their first language, and had no known organic, educational, language, or emotional problems. The participants were matched on the basis of their scores in a district-wide Language Arts Achievement Test (LAAT), Math Achievement Test (MAT), and Canadian Cognitive Abilities Test (CCAT). CCAT is a group-administered test that has three scales: verbal, nonverbal, and quantitative. The score used in matching was the mean of the three scale scores. All the tests are administered in most schools in this district at the end of Grades 3, 6, and 9. Thus, we had results from Grade 3 for all of the participants. Table 4-1 displays the means, standard deviations, and the F values for the main effect of Grade on these variables.

The main effect of Grade was significant for the Mathematics Achievement Test, $F(2, 25) = 3.44$, $p = .048$. Post hoc pairwise comparisons (Scheffé test with significance level .05) did not show significant differences. However, the mean scores from Table 4-1 indicate that on this test, Grade 8 participants performed better than participants in the two other grade levels. Thus, our initial matching was not completely satisfactory and these differences should be considered when interpreting the results.

Table 4-1

The Mean Scores in Language Arts Achievement Test (LAAT), Mathematics Achievement Test (MAT), and Canadian Cognitive Abilities Test (CCAT) for Grade 4, 6, and 8 Participants. F values Show the Main Effect of Grade

		Grade 3 District Wide Test Results			
		Grade 4 (N = 10)	Grade 6 (N = 10)	Grade 8 (N = 9)†	F
LAAT[a]	Mean	51.80	49.80	54.63	1.25
	(SD)	(5.47)	(5.83)	(8.12)	
MAAT[a]	Mean	45.20	44.60	48.00	3.44*
	(SD)	(2.90)	(2.50)	(3.25)	
CCAT[b]	Mean	114.97	111.47	111.56	.83
	(SD)	(4.94)	(7.24)	(8.18)	

Note. † N = 8 for MAAT and LAAT scores due to 1 missing value
[a] $df\,(2, 25)$; [b] $df\,(2, 26)$; * $p < .05$.

Task

The Crack-the-Code task is based on the popular Master Mind game and the paper version of it has been used to measure planning in several previous studies (see e.g., Das, Mensink, & Janzen, 1990; Parrila et al., 1994; Parrila et al., in press). In the Crack-the-Code task, participants are shown two to five information lines that contain three to five colored disks in a particular order. They are also shown a label indicating how many of the colored disks are in their correct places in each of the information lines. The participant's task is to integrate information from all of the information lines and place her or his set of colored disks on the answer line in such an order that all of the information lines are true.

The version used in this study consisted of six items presented on a computer (Macintosh IIsi with 7 X 10 inch color monitor). Appendix A contains a black-and-white reproduction of the original items. As Appendix A shows, the six items can be divided into three pairs of formally similar and progressively more difficult items: Items 1 and 2, Items 3 and 4, and Items 5 and 6. In the first two items, two information lines and three colored disks were used. Items 3 and 4 consisted of four disks and four information lines, and Items 5 and 6 had four disks and three information lines. Items 3 and 4 can be solved simply by noticing that the "2 correct" information line has two matching disks with one of the two "0 correct lines". Despite having less information lines, Items 5 and 6 are more difficult because they have less matching placements in different information lines, and thus less possible constraints for future moves.

Information lines were presented on the computer screen one on top of the other, slightly to the right of the center of the screen. Participants were asked to place the set of colored disks, which were located randomly on the left side of the screen, in their correct places in the answer line at the bottom of the screen. By using the mouse, the disks could be dragged anywhere in the screen, provided that the participant had not terminated his or her performance by clicking the "Done" button at the lower left corner of the screen. The time limit for each item was three minutes.

For example, Item 5 contained the following information lines (see Appendix A):

- Line 1: Blue, Black, Yellow, White; "0 correct"
- Line 2: Yellow, White, Black, Blue; "1 correct"
- Line 3: Black, Yellow, White, Blue; "0 correct"

Participants' were provided with four disks (one of each color) and asked to place on the answer line. The only possible order of disks that would be consistent with the information lines is: White, Blue, Black, Yellow. An effective way of reaching this solution could, for example, include (a) noting the constraints (e.g., Blue cannot be correct in Line 2 since it is in the same place in Line 3, which has 0 correct), (b) building a hypothesis (e.g., Yellow, White, or Black is correct in Line 2 — try Yellow first), (c) testing and evaluating the hypothesis (e.g., if Yellow is correct in Line 2, where would the other disks go — there is no legitimate place left for White, so Yellow cannot be correct in Line 2), and so on, until the correct answer is found.

The computer recorded each participant's performance and provided the following information:

(1) The correctness of the final answer;

(2) The sequence in which the disks were moved, as well as the starting and ending point of each move;

(3) Performance time (the time between the exposure of the item and the termination of the performance);

(4) First-move latency (the time between the exposure of the task and the initiation of the first move by clicking the mouse on one of the disks) and subsequent move latencies (the time between placing one disk by releasing the mouse and initiating the next placement by clicking the mouse on the new disk);

(5) Move duration (the time spend moving the disk on the screen); and

(6) Evaluation time at the end of the task (the time between placing the last disk and terminating the performance by clicking "Done").

Procedure

Participants attended one individual testing session lasting from 20 to 40 minutes. The testing took place during school hours in a private room in the participants' respective schools. A trained experimenter conducted all of the sessions in a one month period during the Fall of 1994.

First, participants received instruction in thinking aloud and were provided with an example of thinking aloud while multiplying 6 times 12. Participants were then asked to multiply 4 times 13 while thinking aloud. For participants who had difficulties understanding the task or producing verbal output, a second practice task was used. The idea of not asking participants to explain to the experimenter what they were thinking was emphasized and instead, participants were encouraged to "say out loud" anything that comes to their mind. If participants were silent for a period of 10 seconds, they were reminded to "keep talking".

Think aloud instruction was followed by the Crack-the-Code task instructions. Testing sessions included three sample items to ensure that all participants knew what was expected of them. Before and during the administration of the sample items, participants received instruction in thinking aloud.

After each item, participants were provided feedback regarding the correctness of their answer. The purpose of this was to ensure that all participants would understand what was required of them and to minimize the probability of them building alternative representations of the required end result.

All testing sessions were audio tape-recorded and transcribed verbatim. The transcriptions were then used in the data analyses that are described in more detail below.

Results

Results are reported and discussed below in three sections. The first section reports developmental differences in "product measures" obtained from computer protocols. These include more traditional performance measures such as accuracy and speed of solutions, as well as potentially meaningful indicators of participants' planning process such as move latencies, move durations, number of moves, and evaluation time.

The second section focuses on "error analyses". Error analyses include both incorrect moves and generated answers, as well as corrections made during the performance.

The third section analyzes participants' verbal protocols and reports on "process properties". The first part of this section describes the central within-group and between-group differences at various stages of the planning process. The second part reports on different search methods that our participants used.

Each result section is followed by a discussion that summarizes the main findings and relates them points made in the introduction. In contrast, the purpose of the general discussion is to unite information obtained in previous sections into more qualitative descriptions of participants' planning processes. In this section, several different "prototypical" planning approaches are identified and described.

Product Measures

In this section we will report first between-group differences in *performance accuracy* and *speed*. Both of these measures are expected to differentiate between the three groups in our study. The second part of this section focuses on the extent to which participants' engage in advance planning as opposed to planning-in-action. Advance planning should be reflected in the amount of time that participants spend at the beginning of the item forming the initial task representation and deciding their first move(s). This measure, termed *first-move latency,* should indicate the extent of "global" advance planning. It is also possible that a participant may engage in a more limited form, or "local", advance planning, later in the performance. This kind of local advance planning would generally be reflected in increased move latencies in general, and more specifically in increased *move latency/move duration ratios* that are sensitive to individual differences in performance speed. In other words, those participants who plan their moves carefully in advance should have higher move latency/move duration ratios than those who favor a more direct planning-in-action approach. In terms of the four planning components, longer first-move latencies and higher move latency/move duration ratios may indicate more careful representation formation and anticipation, although this information alone is not sufficient to draw such a conclusion.

Two other product measures are also reported below: *number of moves* and *evaluation time* at the end of the item (operationalized as the time difference between placing the last disk and pressing the "Done" button). Both of these measures can indicate a more controlled and reflective performance, as well as willingness and ability to detect mistakes. Thus, they can reflect both regulation and representation components.

Unless otherwise stated, product measures were subjected to Multivariate Analysis of Variance (MANOVA) with the six Crack-the-Code item scores as dependent variables and Grade (3) as an independent variable. This was followed by separate analysis of variances (ANOVA) with the product measure as a dependent variable and Grade (3) and Correct (2) as independent variables. The latter analyses were not performed for Items 1 and 2 since most participants solved these items correctly. Due to the small sample size, these analyses are not very powerful. For this reason, the effect size is also reported whenever p values were marginally significant ($< .10$)[2].

Accuracy and Performance Time.

The mean number of correctly solved items was 2.10 (SD = .88; range = 1 to 3), 3.50 (SD = 1.27; range = 2 to 6), and 4.78 (SD = 1.20; range = 3 to 6) for Grades 4, 6, and 8, respectively. ANOVA with the mean number of correctly solved items as an independent variable and Grade (3) as a dependent variable showed that the main effect of Grade was significant, $F(2,26) = 13.44$, $p < .001$. Post hoc pairwise comparisons (all post hoc comparisons reported in this section were computed using a joint univariate Bonferroni test with a significance level of .05) indicated that all differences between the three grade levels were significant. This was true despite a ceiling effect (one Grade 6 and three Grade 8 participants solved all six items correctly).

Table 4-2 displays the number of correct answers produced by participants in Grade 4, 6, and 8 on each of the six items.

Table 4-2

Distribution of Correct Answers by Grade and Item

	Grade 4 (n = 10)	Grade 6 (n = 10)	Grade 8 (n = 9)
	pass	pass	pass
Item 1	7	10	9
Item 2	7	9	9
Item 3	1	3	7
Item 4	3	5	6
Item 5	1	4	6
Item 6	2	4	6

As indicated in Table 4-2, the vast majority of participants solved Items 1 and 2 correctly. These items are formally similar to the final sample item and include only three disks and two information lines. In contrast, both Items 3 and 4, which include four disks

---

2 Effect size $d = |\text{Mean1} - \text{Mean2}|/[(SD1^2+SD2^2)/2]^{.5}$

and four information lines, and Items 5 and 6, which include four disks and three information lines, were too difficult for most of the Grade 4 participants. Less than half of the Grade 6 participants, in turn, solved Items 3 to 6 correctly, whereas the majority of the Grade 8 participants correctly answered each of these items.

Mean performance times are presented in Table 4-3. MANOVA showed no significant overall differences between the grade levels in mean performance times, $T^2 = .968$, $F(12, 40) = 1.61$, $p = .127$. The F values for the main effect of Grade from the univariate ANOVAs are reported in Table 4-3.

Table 4-3

Mean Performance Times (standard deviations in parentheses) and F value for the Main Effect of Grade in the Six Crack-the-Code Items

		Grade			
		4 (n = 10)	6 (n = 10)	8 (n = 9)	F
Item 1	Mean	50.03	53.88	50.48	.08
	(SD)	(25.22)	(24.67)	(21.33)	
Item 2	Mean	50.57	51.50	39.61	1.23
	(SD)	(12.06)	(23.57)	(16.83)	
Item 3	Mean	68.17	104.37	104.60	2.15
	(SD)	(23.02)	(51.41)	(54.47)	
Item 4	Mean	82.44	110.77	101.53	.90
	(SD)	(46.09)	(46.58)	(52.15)	
Item 5	Mean	64.80	89.16	100.91	1.69
	(SD)	(26.01)	(44.41)	(57.47)	
Item 6	Mean	74.98	85.86	83.39	.21
	(SD)	(32.26)	(39.36)	(46.46)	

Note. df (2,26). No F values were significant at .05 level.

Table 4-3 demonstrates that Grade 4 participants were faster than Grade 6 or Grade 8 participants on all of the more complex items, although the main effect of Grade was not significant for any of the items. Post hoc pairwise comparisons found no significant differences, although for Item 3 the differences between Grade 4 and Grade 6 ($d = .91$) and Grade 4 and Grade 8 ($d = .87$) approached significance, as did the difference between Grade 4 and Grade 8 ($d = .81$) for Item 5.

Separate performance time by Grade (3) and Correct (2) ANOVAs were calculated next for Items 3 to 6 in order to assess the main effect of Correct and Correct X Grade interaction. The main effect of Correct was marginally significant, $F(1,23) = 3.19$, $p = .087$ ($d = .76$), for Item 6: The correct performances involved more time (mean = 98.24 sec.) than the incorrect performances (mean = 69.41 sec.). The Grade X Correct interaction approached significance for Item 3; the only correct performance for Grade 4 took longer than the average incorrect performance, whereas the opposite was true for Grade 8.

In sum, accuracy of performance improved with grade but speed did not. As such, the speed of performance can be a somewhat misleading measure since it is averaged over correct and incorrect performances. These results indicate, however, that younger participants produced and accepted incorrect answers faster than the older participants did.

First-Move Latency and Move Latency/Move Duration Ratio

Table 4-4 displays the mean first-move latencies on the six Crack-the-Code items for the three grade levels.

Table 4-4

Mean First-Move Latencies (standard deviations in parentheses) and F value for the Main Effect of Grade in the Six Crack-the-Code Items

		Grade			
		4 (n = 10)	6 (n = 10)	8 (n = 9)	F
Item 1	Mean	8.92	12.52	13.76	.87
	(SD)	(5.13)	(8.36)	(10.89)	
Item 2	Mean	8.76	18.09	10.82	2.54
	(SD)	(7.93)	(12.26)	(8.12)	
Item 3	Mean	16.47	27.02	24.31	1.09
	(SD)	(12.63)	(16.49)	(20.05)	
Item 4	Mean	15.06	21.85	26.13	.79
	(SD)	(12.43)	(14.35)	(28.65)	
Item 5	Mean	8.74	14.15	14.51	1.33
	(SD)	(5.19)	(9.36)	(10.99)	
Item 6	Mean	12.23	22.31	10.31	2.60
	(SD)	(8.47)	(17.79)	(8.26)	

Note. df (2, 26). No F values were significant at .05 level.

Table 4-4 indicates that Grade 4 participants had shorter first-move latencies than the other two groups for all but one item. MANOVA, however, showed no significant overall group differences, $T^2 = .88$, $F(12, 40) = 1.46$, $p = .180$. Univariate ANOVAs produced marginally significant F values for Item 2 ($p = .098$) and Item 6 ($p = .093$). For both items, Grade 6 participants clearly had the largest first move latencies compared to Grade 4 and Grade 8 participants. The difference between Grades 4 and 6 approached significance on Item 2 ($d = .90$). For Item 6, Grade 6 was significantly different from both Grade 4 and Grade 8.

Separate ANOVAs for each item showed no significant differences between correct and incorrect performances in first-move latencies. For Items 3 to 5, the differences were practically nonexistent. For Item 6, incorrect performances had somewhat shorter mean first-move latency (12.6 seconds) than correct performances (18.7 seconds) but this difference was not statistically significant.

A closer examination of the mean first move latencies within each grade level indicates that the relationship between the first-move latency and the correctness of the solution was by no means stable. For Grade 6, for example, correct performances on Item 3 had a shorter average first-move latency than incorrect performances, whereas the opposite was true for Item 6. In Grade 8, correct performances had shorter average first-move latencies for Items 3 and 5; the opposite was true for Item 6, and, finally, there was no difference between the two for Item 4.

Thus, first-move latency did not predict successful performances, as one may have expected. This result suggests that the participants either did not engage in "global planning" at the beginning of the task (i.e., they did not attempt to represent the entire solution), or that their global planning was not successful. It is possible, however, that they engaged successfully in more limited "local planning", that is, they represented and evaluated their first one or two placements with respect to constraints at that point. To assess this possibility, we compared the average first-move latencies for performances that began with a correct first move with performances that began with an incorrect first move. Subsequent t tests failed to show any significant differences between the two for Items 3 to 6. Moreover, the largest difference in average first-move latencies between the two groups was 1.6 seconds (21.8 vs. 23.4), which is hardly a psychologically meaningful difference. Thus, somewhat surprisingly, first-move latency did not predict correctness of the first move.

The move latency/move duration ratio was calculated next. Move duration represents the portion of performance time during which the participants were moving the colored disks on the computer screen, whereas move latency represents the opposite (i.e., time between the moves). The ratio score was assumed to be an indicator of a participant's planning approach: High move latency/move duration ratio scores would indicate that the participant spent more time planning-in-advance his or her moves. In terms of the four planning components, high ratio scores may indicate that the participants generate more information for each step of their performance.

Table 4-5 displays the mean move latency/move duration ratio scores for Grades 4, 6, and 8 on the six items. *F* value shows the main effect of Grade.

Table 4-5 shows that the move latency/move duration ratio scores of Grade 6 and Grade 8 participants were larger than those of Grade 4 participants for all items, indicating that their planning approach may have been different.

Table 4-5

Mean Move Latency/Move Duration Ratio Scores (standard deviations in parentheses) and F value for the Main Effect of Grade in the Six Crack-the-Code Items

		Grade			
		4 (n = 10)	6 (n = 10)	8 (n = 9)	*F*
Item 1	Mean	1.45	1.86	2.34	1.38
	(SD)	(.71)	(1.08)	(1.60)	
Item 2	Mean	1.22	2.12	2.13	2.28
	(SD)	(.50)	(1.23)	(1.34)	
Item 3	Mean	1.62	2.62	2.97	1.71
	(SD)	(1.01)	(1.64)	(2.20)	
Item 4	Mean	1.25	2.89	2.16	2.09
	(SD)	(.57)	(2.46)	(1.82)	
Item 5	Mean	1.17	2.69	2.36	1.28
	(SD)	(.58)	(2.58)	(2.85)	
Item 6	Mean	1.16	2.58	2.41	1.99
	(SD)	(.63)	(1.77)	(2.39)	

Note. $df(2, 26)$. No *F* values were significant at .05 level.

Also, post hoc pairwise comparisons indicated that the differences between Grade 4 and Grade 6 approached significance for Items 2, 4, and 6, (d's = .96, .92, and 1.07, respectively), and the differences between Grade 4 and Grade 8 approached significance for Items 2 and 3 (d's = .90 and .79, respectively). MANOVA, however, showed that the effect of Grade was not significant, $T^2 = .49$, $F(12, 40) = .81$, $p = .636$. Separate

ANOVAs for each of the items also produced no significant F values for the main effects of Correct or Grade X Correct interaction. One reason for the lack of significant F values can be found in Table 4-5: For several items, Grade 6 and Grade 8 standard deviations were as large as the means, indicating significant within-group differences for this variable. Yet these results seem to indicate that both planning approaches worked equally well for our participants.

In sum, Grade 6 participants seemed to be somewhat slower starters than Grade 4 participants, in particular, but few of the differences were significant. Moreover, neither first-move latency nor move latency/move duration ratio scores seemed to predict correct performance. The first result raises obvious questions: What then are the participants actually doing at the beginning of the task? Was their "planning" activity nonconstructive? We will try to answer these questions later with information obtained from verbal protocols.

Evaluation Time

Evaluation time was operationalized as the time between placing the last disk into its respective position and terminating the performance by pressing the "Done" button. Table 4-6 displays the mean evaluation times for Grades 4, 6, and 8 on the six Crack-the-Code items.

Table 4-6 indicates that older participants evaluated their answers longer than younger participants. Also, MANOVA indicated significant overall group differences in Evaluation time, $T^2 = 1.66$, $F(12, 32) = 2.22$, $p = .036$. This result, however, is not entirely representative since four participants had one missing value (their performances on one item exceeded the three minute time limit) and MANOVA eliminated all of their performances. For this reason, F values in Table 6 are from separate ANOVAs with Grade (3) as a single independent variable for Items 1 and 2, and Grade (3) and Correct (2) as independent variables for Items 3 to 6. These ANOVAs failed to show significant group effects, although for Item 3, F value was marginally significant ($p = .056$). The main effects of Correct and the Grade X Correct interaction were also not significant for any of the items. Particularly large standard deviations (12 out of 18 SDs were larger than their respective means) in evaluation times suggest large within-group differences. Moreover, this measure captures only the evaluation that takes place at the end of the performance. The possible "on-line" evaluation of individual moves and partial answers that may be taking place during the performance is not represented in this measure. We will return to this in the process properties section.

Table 4-6

Mean Evaluation Times (standard deviations in parentheses) and F value for the Main Effect of Grade in the Six Crack-the-Code Items (in Items 3 to 6, the model included Correct (2) and Grade X Correct)

		Grade			
		4	6	8	
		(n = 10)	(n = 10)	(n = 9)	F
Item 1	Mean	6.23	7.18	13.37	1.16
	(SD)	(4.59)	(5.14)	(18.36)	
Item 2	Mean	5.38	5.14	8.72	.82
	(SD)	(4.33)	(5.39)	(9.73)	
Item 3	Mean	4.56	9.02	15.13	3.32
	(SD)	(5.58)	(12.41)	(14.95)	
Item 4	Mean	6.48	6.25	5.45	.03
	(SD)	(4.79)	(6.47)	(6.92)	
Item 5	Mean	6.75	11.30	15.66	.32
	(SD)	(10.14)	(12.56)	(18.92)	
Item 6	Mean	4.84	12.61	12.14	1.54
	(SD)	(5.10)	(10.14)	(12.33)	

Note. *df* (2, 26) in Items 1 and 2; *df* (2, 21) in Items 3 and 4; *df* (2, 22) in Item 5; and *df* (2, 23) in Item 6. No F values were significant at .05 level.

Number of Moves

Table 4-7 displays the mean number of moves participants produced for each of the items as a function of Grade. F value indicates the main effect of Grade obtained from univariate F tests in MANOVA.

MANOVA showed no significant overall grade differences in the number of moves, T^2 = .82, $F(12, 40)$ = 1.36, p = .223. For Item 5, however, the main effect of Grade was significant, $F(2, 26)$ = 3.83, p = .035. Post hoc comparisons indicated that Grade 8 participants made significantly more moves for Item 5 than Grade 4 participants did. Also, the difference between Grade 8 and Grade 6 approached significance (d = .79). For Item 2, Grade 4 participants made considerably more moves than did Grade 8 participants (d = .83), whereas the opposite was true for Item 3 (d = .85). Neither of these differences, however, were statistically significant. Individual ANOVAs with number of moves as the dependent variable and Grade (3) and Correct (2) as independent variables showed no significant main effects of Correct. Grade X Correct interaction was also not significant for any of the items.

Table 4-7

Mean Number of Moves (standard deviations in parentheses) and F value for the Main Effect of Grade in the Six Crack-the-Code Items

		Grade			
		4	6	8	
		(n = 10)	(n = 10)	(n = 9)	*F*
Item 1	Mean	4.80	3.80	3.78	.55
	(SD)	(3.74)	(1.62)	(.97)	
Item 2	Mean	4.10	3.70	3.11	1.52
	(SD)	(1.66)	(1.25)	(.33)	
Item 3	Mean	4.40	5.80	6.11	1.62
	(SD)	(.97)	(3.15)	(1.97)	
Item 4	Mean	5.40	6.60	7.11	1.01
	(SD)	(2.55)	(2.27)	(3.30)	
Item 5	Mean	5.20	6.00	9.11	3.83*
	(SD)	(1.48)	(2.21)	(5.08)	
Item 6	Mean	7.80	5.70	7.33	1.08
	(SD)	(4.71)	(2.16)	(2.45)	

Note. df (2, 26) * $p < .05$.

In sum, many older participants made more moves on most of the items. The nature of these moves will be the focus of the next section.

Discussion of Product Measures

It seems that the above product measures captured less significant performance differences than one may have expected. Similar to the findings in many previous studies, older participants solved more items correctly. Their performance, however, was not faster than that of the younger participants. In fact, the opposite seems to be true for our sample. This is perhaps mainly because Grade 4 participants in particular seemed to accept their incorrect answers more readily and were reluctant to replace a disk that had already been placed. The first point is also reflected in their somewhat briefer evaluation times, and the second in their somewhat lower number of moves for the more difficult items.

This would seem to suggest that Grade 4 participants either lacked the evaluation and regulation skills necessary for recursive planning or they simply chose not to use them.

One of the most simple decision rules that human problem solvers employ is trying to avoid decisions that undo the effects of previous decisions (see e.g., Anderson, 1995; Klahr, 1985). Anderson (1995, p. 250) calls this "repeat-state avoidance" criterion. It is possible that reliance on "repeat-state avoidance" criterion reduced the likelihood to evaluate and make corrections for our youngest participants.

Our results suggest that not only evaluation but also careful representation of the situation may have been lacking for Grade 4 participants. Grade 4 participants exhibited the lowest move latency/move duration ratio scores, as well as the lowest first-move latencies. These variables were assumed to reflect either more "global" planning or generation of more information for each planning step. Neither of these variables, however, was correlated with correct performances or correct first moves, which may be the most surprising finding above. It is possible that first-move latency that should reflect advance planning may instead reflect visual processing demands of the colorful Crack-the-Code display. Note that first-move latency does not confirm whether our participants engaged in global advance planning or explain what type of information the participants generated to guide the planning process. We will return to these questions later.

Many of the statistical analyses failed to show significant differences between the three groups. Aggregating scores across formally similar items (Items 1 and 2, Items 3 and 4, and Items 5 and 6) also did not result in more significant differences. Two obvious reasons for the lack of statistically significant differences are the small sample size and the relatively large standard deviations for many of the variables. A study with a larger sample size (but no verbal protocols) is currently being undertaken and should produce more conclusive findings about the validity of these product measures. Large standard deviations may reflect the use of various different planning approaches within each group. If this is the case, then averaging data over different planning approaches, as was done above, could have hidden meaningful between-group differences.

While only some of the above measures alone produced significant results, taken together they may indicate some important differences between the grade levels. Grade 4 participants, for instance, were faster than the other two groups on all time measures. Combined with the fact that they produced fewer correct answers, this indicates that they probably approached the task in a less reflective manner and did not generate a sufficient amount of information to guide the planning process. In other words, we could say that they may have operated in a simplified problem space. Grade 6 participants, in contrast, were slower than Grade 8 participants in most time measures but also solved fewer items correctly. This could indicate that Grade 6 participants reflected upon the task more and generated more information but failed to integrate the information into a functional task representation with the most relevant features present. Finally, Grade 8 participants seemed to have used their time constructively and experienced fewer problems with the task. Large standard deviations, however, suggest that Grade 8 participants may have used several different routes in finding the correct answer. We will need to examine more qualitative information obtained from computer and verbal protocols before these suggestions can be validated or refuted. This topic will be addressed in the remainder of the results section.

Error Analyses

In this section, we will describe in detail the number and type of errors participants *committed* and *corrected* during their performances. We expect that error analyses of this kind can provide us with valuable information about the planning process. In terms of our planning process model, participants' representation is indexed by the type of errors that are made at different grade levels. Good anticipation skills, in turn, should be reflected in lower number of committed errors. What errors are later corrected provides information about the evaluation/regulation component. It is also possible that there is a discrepancy: Anticipation of errors may be more difficult than evaluation in general, or some particular kinds of errors may be easier to anticipate than others, whereas at the evaluation stage the reverse could be true.

Since the vast majority of participants solved Items 1 and 2 correctly without making mistakes, only Items 3 to 6 were included in error analyses. The number of committed and corrected errors is reported first, followed by the analyses of what item constraints, or rules, these errors violated and what violations were later corrected. The last part of the error analyses focuses on the incorrect answer lines that the participants accepted. This information may be useful indicator of participants' goal representation.

The Number of Errors Committed and Corrected

First, the number of incorrect moves was obtained by assessing every move made by a participant either as correct or incorrect. For this purpose, disk placement was considered correct when the disk was placed in its correct position on the answer line. Incomplete moves (when a participant starts a move but does not finish the placement) were regarded as correct if the correct place for that disk was already occupied by another disk. Finally, backup moves (when a participant removes a disk that is already placed) were regarded as correct when a participant removed an incorrectly placed disk.

Since the number of incorrect moves can be directly dependent upon the total number of moves, error percentage is reported below rather than the total number of errors. Error percentage was calculated by dividing the number of incorrect moves by the total number of moves. Table 4-8 displays the mean error percentage (standard deviations in parentheses) and the *F* value for the main effect of Grade on Items 3 to 6.

MANOVA showed that the main effect of Grade was not significant, $T^2 = .59$, $F(8, 44) = 1.63$, $p = .144$. Univariate *F* tests produced marginally significant *F* values for Item 3, (p =. 069), Item 5 ($p = .098$), and Item 6 ($p = .074$). Table 4-8 indicates that on average, about 60% of the moves made by Grade 4 participants were incorrect. The respective percentage for Grade 6 was 50% and for Grade 8, only 34%. Post hoc pairwise comparisons (Bonferroni test with a significance level of .05) showed that the mean error percentage for Grade 4 was significantly greater than for Grade 8 for Items 3 and 6, and approached significance for Item 5 (d = 1.20). None of the differences between Grade 4 and Grade 6, or between Grade 6 and 8, were significant.

Table 4-8

The Mean Error Percentage (standard deviations in parenthesis) in Grades 4, 6, and 8. F value Shows the Main Effect of Grade

		Grade			
		4	6	8	*F*
Item 3	Mean	.58	.49	.28	2.98
	(SD)	(.26)	(.32)	(.21)	
Item 4	Mean	.44	.41	.32	.38
	(SD)	(.31)	(.26)	(.32)	
Item 5	Mean	.65	.57	.36	2.54
	(SD)	(.22)	(.36)	(.26)	
Item 6	Mean	.69	.53	.39	2.88
	(SD)	(.18)	(.36)	(.24)	

Note. *df* (2, 26). No *F* values were significant at .05 level.

Instead of listing all of the moves, the first row in Table 4-9 displays the total number of moves that involved placing disks, the second row displays errors in these placements, and the third row displays subsequent corrections of these errors for the three grade levels on Items 3 to 6. Incomplete and backup moves were removed from these analyses since correcting them is more likely to be based on a new decision rather than detecting errors in the old ones, whereas the opposite is probably true for the incorrect placements.

ANOVAs could not be calculated for corrections since both the "best" (no errors made) and the "worst" (no corrections made) possible performance have the same value of zero. Instead, a new variable was formed by dividing all participants into three groups (no corrections, corrects some or all errors, no errors). Likelihood-ratio chi-squares indicated that this variable was significantly associated with grade level on Item 3, $LR(4) = 21.86$, $p < .001$, and marginally significantly on Item 6, $LR(4) = 7.95$, $p = .09$. The main difference between the groups seemed to be in the "no corrections" and "no errors" categories. The "no corrections" category indicated that all but two of the Grade 8 participants (one on Item 5 and one on Item 6) corrected at least some of their incorrect moves. In contrast, four to nine Grade 4 participants and four to five Grade 6 participants accepted all their incorrectly placed disks on Items 3 to 6. The "no errors" category, in contrast, had a total of 11 entries for Grade 8 (all items combined), 6 for Grade 6, and only 4 for Grade 4.

Table 4-9

The Total Number of Incorrectly Placed Disks and Subsequent Corrections in Different Grade Levels in Items 3 to 6

		Item 3	Item 4	Item 5	Item 6	Total
Grade 4	Placed Disks	41	46	46	63	196
	Errors	24	21	33	44	122
	Corrections	0	6	6	18	30
Grade 6	Placed Disks	48	53	53	50	204
	Errors	21	21	32	27	101
	Corrections	7	10	9	6	32
Grade 8	Placed Disks	51	52	66	59	228
	Errors	16	23	33	27	98
	Corrections	14	14	22	18	68

The differences between the three grade levels are evident in Table 4-9. On average, Grade 4 participants corrected 25% of the errors they made, compared to 32% for Grade 6 and 69% for Grade 8. Thus, Grade 8 participants were considerably more likely to correct their incorrect placements than Grade 4 or Grade 6 participants. In contrast, Table 4-9 indicates that the greatest difference between Grade 4 and 6 was in the number of errors committed (122 vs. 101; or 62% vs. 50% of the placements), rather than in the number of errors corrected. These results indicate that both Grade 6 and Grade 8 participants probably build more veridical representations of the items and/or anticipated their moves prior to executing them considerably better than Grade 4 participants. Only the oldest participants, however, were also able to engage in effective on-line monitoring and correcting of already committed errors.

An interesting detail in Table 4-9 is the increase in the number of placements, errors, and corrections for Grade 4 participants on Item 6. It seems that although Grade 4 participants were still unable to successfully evaluate their moves prior to completing them (most of the placements were still incorrect), they were more likely to detect the single move errors and attempt a new move. This increased regulation, however, did not result in more correct answers (only Grade 4 participants solved Item 6 correctly), which indicates that goal representation may have been the weakest point in their planning process.

Item Constraints

For any move to be considered correct in Crack-the-Code, it has to comply with both general task constraints as well as specific item constraints. Every item has the same task constraints, which are provided in the instructions at the beginning of the task, regulating the movement of the disks and the correctness of the answer line.

Instructions for moving the disks involve following three rules: (a) all disks from the left hand side of the computer screen must be moved to the empty places in the answer line, (b) only one disk can be moved to each of the answer line places, and (c) any move that is completed can also be reversed or changed unless the participant has pressed the "Done" button. No participant attempted to place two disks in one place or leave places empty. Several participants may have ignored the last rule in favor of a more simplistic "repeat-state avoidance" principle, as is suggested by the low frequency of moves (Table 4-7) and corrections (Table 4-9), particularly for Grade 4.

General instructions about the correctness of the solution were that all information lines must be "true" at the completion of the task. This means that if the information line says "0 correct", none of disks can be in the same place in the answer line, and if the information line says "1 correct" then one disk — and only one — can be in the same place in the answer line.

These general instructions translate into changing item constraints according to the particular information lines for each item. On Item 1 (see Appendix A), for example, the specific item constraints read as follows: the answer line can have no matches with information Line 2; the answer line must have one match with information Line 1; and the answer line cannot have 2 or 3 matches with information Line 1.

Items 3 and 4 had more information lines and, accordingly, more item constraints.

The following is a list of constraints for Item 3 (respective information lines for Item 4 are in parentheses):

1. The answer line must be different from Line 1 (Line 2)
2. The answer line must be different from Line 4 (Line 4)
3. The answer line must have one match with Line 3 (Line 3)
4. The answer line cannot have two, three, or four matches with Line 3 (Line 3)
5. The answer line must have 2 matches with Line 2 (Line 1)
6. The answer line cannot have 3 or 4 matches with Line 2 (Line 1)

Constraints 1 and 2 (C1 and C2) are called negative single constraints, since they rule out some locations as legitimate, and do not require attending to the positions of the other disks. C3 can be divided into two different constraints: the first (C3a) is called a positive single constraint, since it legitimizes one placement for each of the disks (i.e., they can be the same as in the "1 correct" line) and does not require considering the position of the other disks. The second (C3b) is particularly relevant when the participant is evaluating his or her answer and involves ensuring that one match has been produced. It is called a positive relational constraint because it requires attending to the positions of the other disks.

Similarly, C5 can be divided further into two different constraints. The first (C5a) is relevant when making decisions about individual moves and, similar to C3a, it legitimizes a placement for each of the disks. Thus, when participants follow this constraint and decide that a particular place is legitimate because the disk appears in that place on the "2 correct" line, their decision is based on a single positive constraint. The second constraint (C5b) involves ensuring that enough matches (2 instead of only one or none) have been produced and thus, requires attending to the positions of the other disks. Accordingly, it is called a positive relational constraint.

C4 and C6 are called negative relational constraints because they involve inhibiting the production of too many matching placements and require attending to the positions of the already placed disks. We therefore have four classes of constraints: negative single constraints (C1 and C2), positive single constraints (C3a and C5a), negative relational constraints (C4 and C6), and positive relational constraints (C3b and C5b). Notice, however, that positive single constraints are different from the other constraints as they do not really "constrain" participants' options and thus cannot be violated.

Participants' decision making at each step can be based on any one of these constraints or a combination of two or more. Thus, valuable information about the planning process can be obtained by attempting to analyze which constraints participants attend to and which constraints they violate when completing their moves. Since this decision is difficult to make reliably for every move, we will concentrate here first on incorrect moves and what constraints they violate. This will be followed by analyses of the participants' answer lines and which constraints were accepted in these lines.

Single Move Errors

Error analyses of single moves concentrated on negative single constraints (C1 and C2) and negative relational constraints (both C4 and C6 on Items 3 and 4, and only C4 on Items 5 and 6) because only a complete answer line can violate positive relational constraints. Violation of a negative single constraint occurs when a participant places a disk in a position similar to one in a "0 correct" line. Violation of a negative relational constraint occurs when a participant places a disk in a similar position to a line with "1 correct" or "2 correct" when she or he already has one or two disk(s) matching. It is also possible for one move to violate both single and relational constraints. Only three first errors, two on Item 4 and one on Item 5 violated both type of constraints simultaneously. However, 21 consequent incorrect moves (4 on Item 3; 13 on Item 4; 1 on Item 5; and 3 on Item 6) violated both single and relational constraints simultaneously. In these cases, both violations were included and thus the total number of violated constraints is greater than the total number of incorrect moves. Moreover, only first errors are discussed in detail, since consequent errors can result from the first error and therefore do not necessarily represent independent decisions.

We should also note here that the chance-probability of producing a negative single constraint error as the first error is considerably higher than the chance-probability of producing a negative relational constraint error. For example, on Items 3 and 4, the first move has a .5 chance of violating negative single constraints, and if it does not, the second move still has an average of .45 or .55 chance (depending on the first move) of violating negative single constraints. In contrast, the probability of producing a negative relational constraint violation with two moves is .5 X .33. Keeping this in mind, we would expect a larger frequency of negative single constraint violations if the two constraint categories are attended to either equally poorly or equally well.

Table 4-10 displays the distribution of first errors and consequent errors over negative single constraints (C1 and C2) and negative relational constraints (C4 and C6 on items 3 and 4, and C6 on items 5 and 6) for each grade on Items 3 to 6.

Table 4-10

Distribution of Single Move Errors over the Error Categories in Items 3 to 6

	First Error			Consequent Errors		
Error Category	Grade 4	Grade 6	Grade 8	Grade 4	Grade 6	Grade 8
Negative Single Constraints						
Item 3	7	3	4	10	8	6
Item 4	3	2	2	9	5	5
Item 5	5	2	4	6	4	3
Item 6	7	3	4	8	5	4
Total	24	10	14	33	22	18
Negative Relational Constraints						
Item 3	2	5	1	6	3	8
Item 4	4	6	5	10	15	13
Item 5	6	6	3	8	5	4
Item 6	3	4	2	6	3	3
Total	15	21	11	30	26	28

The results in Table 4-10 suggest that Grade 4 participants seemed to "favor" violation of single constraints as their first error, whereas Grade 6 participants violated the relational constraints more frequently. Grade 8 participants had nearly equal numbers of both types of violations on their first errors. It is somewhat surprising that so many Grade 8 students violated negative single constraints by placing a disk into a position similar to one in a "0 correct" lines as this is a rather obvious violation of the item constraints. Five of these incorrect moves were the first moves for the item. In other words, they occurred in a situation where all four places on the answer line were open to the participant.

In order to assess whether violations of item constraint categories (a separate category was added for those participants who committed no errors) were independent of grade level, a likelihood-ratio chi-square was calculated for the first errors. Two first errors on Item 4 (one for Grade 6 and one for Grade 8) and one on Item 5 (for Grade 4) violated both single and relational constraints simultaneously. All were recorded as relational constraint violations.

These analyses indicated that violations of item constraint categories were not associated with grade level ($LRs(4) = 6.90$ and 2.16 on Items 3 and 4, respectively, and $LRs(4) = 5.36$ and 7.28 on Items 5 and 6, respectively; all $ps > .10$). Thus, although

Grade 4 participants made more errors and corrected fewer of them, there did not seem to be any statistically significant differences in the type of first errors participants in different grade levels committed. We should note, however, that only for Grade 4 were negative single constraint violations more common, as expected by chance.

Corrected first errors. In considering the results presented in Table 4-10, an obvious follow-up question would be: Which type of errors did participants correct? If Grade 4 participants, for example, were more likely to produce single constraint violations than relational constraint violations, was there also a difference between the two error categories in terms of the participants noticing the errors and correcting them? We will attempt to answer this question here by examining the corrected first errors.

Grade 4 participants' first errors violated negative single constraints (C1 and C2) ten times and negative relational constraints (all C4) six times on Items 3 and 4. Only two of these errors (both on Item 4) were corrected, one from each error category. On Items 5 and 6, twelve of the first errors made by Grade 4 participants violated single constraints, and five of these were corrected. Nine first errors violated relational constraints, four of which were corrected.

Grade 6 participants produced 5 single constraint violations and 11 relational constraint violations on their first errors for Items 3 and 4. Two single constraint violations and seven relational constraint violations were subsequently corrected, indicating that Grade 6 participants were more aware of the latter type of violations. On Items 5 and 6, in contrast, 3 of the 5 single constraint violations but only 2 of the 10 relational constraint violations were subsequently corrected.

Grade 8 participants corrected all of their first errors on Items 3 and 4, and all but one on Items 5 and 6. Both of the uncorrected first errors violated negative relational constraints. This suggests that although Grade 8 participants produced a considerable number of single constraint violations, they had no difficulties in detecting these later in their performance.

In summary, it would seem that Grade 4 participants were as likely to correct either type of error (if they were to correct any errors at all); Grade 6 participants were more likely to correct a relational constraint violation on Items 3 and 4, and a single constraint violation on Items 5 and 6; and finally, Grade 8 participants corrected most of their first errors.

Answer Line Errors

Violations that participants accepted in their answer lines were analyzed next. Accepting an incorrect answer line can indicate either lack of regulation or inadequate goal representation, only the latter of which is dealt with here.

Although a single move cannot violate positive relational constraints, an answer line can. Thus, the analyses of answer line errors also included this category of constraints (constraints C3b and C5b). C3b is violated when the answer line does not contain any disks that match with the "1 correct" line. Respectively, C5b is violated on Items 3 and 4 when the answer line contains zero or only one disk(s) that matches with the "2 correct" line. Also, while Items 3 and 4 include all six constraints, Items 5 and 6 do not include C5 and C6.

Table 4-11 displays the distribution of answer line errors over the three item constraint categories for Items 3 to 6.

Table 4-11

Distribution of Answer Line Errors over the Item Constraint Categories in Items 3 to 5

Constraints	Grade 4	Grade 6	Grade 8
Negative Single			
Item 3	11	6	1
Item 4	6	4	1
Item 5	8	2	1
Item 6	10	3	2
Total	35	15	5
Negative Relational			
Item 3	7	3	1
Item 4	7	5	2
Item 5	6	5	2
Item 6	3	4	1
Total	23	17	6
Positive Relational			
Item 3	7	3	1
Item 4	4	0	2
Item 5	0	0	0
Item 6	2	1	0
Total	13	4	3

Together, the results in Tables 4-10 and 4-11 suggest that Grade 4 participants seemed to commit and accept all constraint violations more easily than the other two groups. Grade 4 participants violated negative single constraints 57 times, 61% of which were still present in the accepted answer line. The respective percentages were 47% for

Grade 6 and 16% for Grade 8. In terms of negative relational constraint errors, Grade 4 participants accepted 51% of the errors they committed, Grade 6 participants accepted 36%, and Grade 8 participants accepted only 15%. Also, not correctly placing enough disks (positive relational constraints) seemed to be a problem only for Grade 4 participants.

Discussion of Error Analyses

These results indicate that in general, the representation, anticipation, and regulation skills of the Grade 4 participants were not as well developed as the respective skills of the two older grade levels. The difference between Grades 6 and 8, in turn, was more pronounced in the regulation skills. Tables 4-8 and 4-9 show that Grade 4 participants committed more errors than Grade 6 or Grade 8 participants and seldom corrected their errors; Grade 6 participants made fewer errors but frequently failed to correct many of them; and, finally, Grade 8 participants both committed fewer errors and corrected more of them. Thus, the most significant difference between Grades 4 and 6 seemed to be in the number of errors produced, suggesting that the younger participants' capacity to build a sufficient representation of the items and anticipate the consequences of the moves was more limited. In contrast, the most notable difference between Grades 6 and 8 was in the number of errors corrected, suggesting that lack of (or inappropriate) evaluation may explain why Grade 6 participants solved fewer items correctly. Tables 4-10 and 4-11 refine these conclusions by suggesting that Grade 6 participants' representation/anticipation problems may have resulted largely from not attending to negative relational constraints, whereas their evaluation failed to register a considerable number of both negative single and negative relational constraint violations.

We should note, however, that for both Grades 4 and 6, eight out of the ten participants did correct at least one of their incorrect placements, indicating that some evaluation did take place. What seemed to be lacking was the consistent and successful use of evaluation to counterbalance the lack of anticipation as expressed in the considerable number of errors that these participants committed.

In order to avoid errors altogether, participants had to engage in successful covert testing and modifying of potential moves prior to actually executing the moves. In other words, participants who did not commit errors were able to anticipate the outcome (see Welsh, 1991, for similar argument). In order to consistently avoid errors on Items 3 to 6, participants had to plan ahead at least three moves. Only one Grade 6 participant (who had no mistakes on three of the four complex items) and two Grade 8 participants (one had no errors on three out of four items and the other completed all four items without any errors) were able to consistently avoid errors. For both Grades 6 and 8, the remaining six error-free performances were produced by five different participants. This suggests that while they may have been able to sufficiently engage in covert testing and modifying of their plans, these capacities were not employed consistently across the complex Crack-the-Code items.

A common argument in the problem solving literature is that the backup strategies used by older participants are the same as the optimal strategies used by younger participants. Welsh (1991) used results from error analyses to support this argument. Our results seem to provide tentative support for this argument. A considerable number of single constraint violations committed by Grade 8 participants suggests that when Grade 8 participants faced a task that they could not immediately solve, some of them may have resorted to "primitive" trial-and-error behavior.

Analysis of Verbal Protocols

The purpose of this section is to describe how our participants approached different stages of the planning process. We will first describe the analysis of verbal protocols and the specific information about the planning process they provided. Each subsection will begin with a summary highlighting the central differences between the three grade levels before proceeding into more detailed description of the findings within each grade level. When relevant, we also compared information from verbal and computer protocols in order to detect possible discrepancies between the two. Similar approach was used also in the section that will describe the search methods.

Since presenting this kind of information requires considerable space, we will concentrate here on Items 3 to 6 only because most participants solved Items 1 and 2 correctly limiting the expected variability. Also, the findings are related to planning components only in the discussion section that follows the analyses.

Procedure. Verbal protocols were divided into statements, each of which represented an individual decision. These statements were further classified into four broad categories: task definitions, move decisions, evaluations, and miscellaneous statements. At the beginning of the items, some participants produced statements that did not seem to have relevance to the planning process (such as: "Oh, that's big", or "Item 3") and consequently, these statements were removed. We should note that the frequency of such statements was extremely low in all grade levels and we therefore decided not to include a separate category for task irrelevant verbalizations.

A statement was categorized as a *task definition* if it involved attending to at least some of the information available in the item and was not associated with an attempted or completed move. This type of statement occurred primarily (but not solely) at the beginning of an item prior to the initiation of any moves. The simplest task definitions consisted of stating existing information, for instance, by acknowledging some or all of the information lines. An example of this would be: "zero correct, two correct, one correct, zero correct". Some participants elaborated their task definitions by adding or manipulating information without limiting the search space. The three most common elaborations were to (1) to interpret an information line (e.g., "zero correct, can't be any of these"), (2) compare two information lines ("the Blue is the same and the Yellow is the same"), and (3) make an assumption that a disk (or disks) is correct in a particular line. Finally, some task definition statements attempted to actively limit the search space by combining information from different lines either to limit possible correct disks for a line (e.g., "There is two correct in the second line, it can't be Yellow"), or limit possible places for a disk (e.g., "So Black can't go there. Where could Black go? Black can't go there and Black could go there or there"), or limit possible disks for a place (e.g., "Yellow could be there, Blue .. Blue could be, and White couldn't").

Statements that were associated with an attempted or completed move(s) were classified as *move decisions.* Move decisions varied in terms of the number of disks involved and the amount of information a participant utilized, as well as whether a participant focused on finding a correct place for a disk, a correct disk for a place, or a correct disk for a particular line (analogous to task definitions). Thus, move decisions were coded according to (1) the disk involved if the focus was to find a correct place for that particular disk (e.g., "I know White doesn't belong in the first place or in the second place. So .. I think it goes in the last place"), (2) the information line involved if the focus seemed to be to produce a correct placement(s) for a particular line (e.g., "There is one correct, which might be Black there"), or (3) the place if the goal seemed to be to find a correct disk for a particular place (e.g., "And this one here, it can't be that. Okay .. uhm .. it can't be that, so it's got to be White"). The quantity of information that move decisions included varied from naming the disk or the place (or both; e.g., "and the Black one .. I guess probably be here") to considering all of the possible alternatives. Accordingly, move decision entries with no data were coded as DISK[3], PLACE, or DISK/PLACE (or TRIAL-AND-ERROR in some particular cases when this method of decision making was verbalized). When a move decision included other data, it was coded either as DATA1, DATA2, or VALID. DATA1 was used in cases in which the move was based on only one piece of information (e.g., "Black can go here because it was correct on that line"); DATA2 identified those

---

3 When coding categories are referred to in test, they are identified by smaller uppercase letters.

statements that utilized more than one piece of information but less than all relevant information (e.g., "This can't be Yellow or Black, so it has to be White" — not all the relevant information since Blue was not mentioned); and VALID was used to signify those statements that involved attending to all relevant information in that situation (e.g., "the Yellow, since this one is incorrect it shouldn't be on this line. And it shouldn't be on this line either because this isn't correct. So it should be either the first or the second. So I'll put it in the first"). The last move was not considered valid unless it was part of a decision that involved several moves. The reason for this was to avoid false positives since the last move is completed in a situation where only one place is open. Several move decisions also acknowledged the effects of other planned or already completed moves. The label LINK was used to identify these instances.

Evaluation statements were classified as either focusing on (a) an ASSUMPTION when a participant seemed to evaluate the possible outcome of a move without initiating the move; (b) an ATTEMPT when the move was started but not completed; (c) a MOVE when the participant evaluated an already completed move; (d) a PARTIAL ANSWER when he or she evaluated two or three moves that were already completed; and (e) ANSWER when he or she evaluated a complete answer line (when all four disks are placed). All evaluations were also labeled as correct or incorrect. Evaluations of assumptions and attempted moves were correct if the correct place for the disk involved was not open, and incorrect if the correct place was open. Evaluations of moves were labeled correct if the participant either decided to accept a correctly placed disk or decided to remove an incorrectly placed disk. Evaluation of a move was labeled incorrect if the participant decided to accept an incorrectly placed disk or decided to remove a correctly placed disk. A similar rationale was also used to label evaluations of both partial and complete answers. Note, however, that a participant could correctly evaluate a partial answer or a complete answer as incorrect but then proceed to remove a correctly placed disk from the answer line. In these cases, the evaluation was still labeled as correct.

Miscellaneous statements were those that did not belong to any of the three categories mentioned above but seemed to have relevance to the planning process. The two most common entries in this category were REGULATION and CONFUSION. Regulation statements involved verbal regulation of behavior with general comments such as "let me see" and "let me think" or with more specific comments such as "start with this one up here" and "first I will look at the White before I place Blue and Yellow". Confusion statements consisted almost entirely of comments such as "I don't know" or "I'm lost".

Task Definitions

The analysis of task definitions indicated that when Grade 4 participants verbalized a task definition, these were likely to be rather simple and involved little active manipulation of information. Only one Grade 4 participant attempted to limit the search space by combining information from different lines in more than one item. Also, if a task definition was not verbalized, subsequent performances provided little indication that Grade 4 participants had engaged in forming a functional representation of the task. In contrast, most Grade 6 participants verbalized task definitions. Very few of these, however, went beyond the information that was readily available. Grade 8 participants verbalized task definitions somewhat less frequently than Grade 6 participants. A higher proportion of their task definitions, however, seemed to combine information from different information lines in order to limit the search space.

Grade 4. Five Grade 4 participants did not verbalize a task definition for Item 3. One of the five, Fiona[4], started with a regulative statement that also acknowledged some of the information lines ("I think I would take the one that has most correct and look at that"). Two verbalized task definitions acknowledged some or all of the information lines but did not elaborate on this information. The remaining three task definitions included one to three elaborative statements that either compared the places of disks in different information lines, interpreted "0 correct" information lines, or assumed one of the disks was correct in Line 3/1[5] or in Line 2/2.

On Item 4, one Grade 4 participant produced four instances of limiting the search space, three of which combined information from different lines to limit possible correct disks for a line (two of the three were incorrect). All other participants either verbalized no task definitions (4 participants), stated existing information (4 participants), or made assumptions (1 participant). Only two Grade 4 participants verbalized a task definition for Item 5; one stated existing information and the other limited possible correct disks for Line 2/1. Finally, five Grade 4 participants verbalized a task definition for Item 6. Two stated existing information, two others both limited possible correct disks for Line 3/1 and possible correct places for one of disks, whereas the remaining task definition included only the latter limitation.

Grade 6. Only one Grade 6 participant verbalized no task definition at the beginning of Item 3. Three of the verbalized task definitions consisted of acknowledging

---

4 All names are fictional but gender specific. Grade 4 names all start with F, Grade 6 names with S, and Grade 8 names with E.

5 The first number in Line 3/1 locates the information line in display as being the third line from the top and second number following the slash identifies it as a information line with 1 correct disk.

some or all of the information lines, and a fourth consisted of one assumption that was consequently evaluated and correctly rejected. All of the remaining five task definitions included at least one instance of limiting the search space. Two correctly ruled out Yellow as a possible correct disk for Line 2/2, whereas another two limited possible correct places for two separate disks each. The last task definition listed possible and illegal disks for two places.

On Item 4, two Grade 6 participants verbalized no task definitions, five stated existing information, and three limited the search space once (one of these limited places for a disk, one limited disks for a place, and one limited possible disks for Line1/2). On Items 5 and 6, only two Grade 6 participants limited the search space (the same two participants on both items). On Items 5 and 6, three and two participants, respectively, verbalized no task definitions. Four task definitions on Item 5 and two on Item 6 were statements of existing information without elaborations; one task definition on Item 5 and three on Item 6 included simple elaborative statements (assumptions or interpretations). The remaining two task definitions on Item 5 and one on Item 6 included one statement that limited the search space. One task definition on Item 6 included four statements that limited the possible disks for each of the four places.

Grade 8. Three Grade 8 participants verbalized no task definitions at the beginning of Item 3. Two of these three attempted to limit the search space after first making some unsuccessful moves. Statements that involved limiting the search space were also evident at least once in four task definitions. Only one of these limited possible disks for Line 2/2, whereas the remaining seven search space limitations all limited possible places for disks. Two long task definitions, in contrast, included stating only existing information with some interpretations and/or assumptions.

Four Grade 8 participants verbalized no task definitions on Item 4, two stated existing information and assumptions, and three verbalized at least one instance of limiting search space (Emily and Eve both verbalized as many as four). On Items 5 and 6, Grade 8 participants verbalized somewhat shorter and fewer task definitions than on Items 3 and 4. On Item 5, four participants verbalized no task definitions, only one stated existing information, and four participants limited the search space (once or twice each). Finally, on Item 6, four performances included no verbalized task definitions and three participants stated only existing information. Erin limited possible correct disks for Line 3/1 and Emily limited the possible correct places for three different disks as well as possible correct disks for one place.

Move Decisions

Move decisions can be roughly categorized into those that (a) did not seem to actively use available data, (b) used at least some data, and (c) used all the relevant data and/or included a reference to other decisions. Following this outline, the three categories "No Data", "Data", and "Valid/Link" are discussed below. The first category includes all DISK/PLACE, DISK, PLACE, and TRIAL-AND-ERROR move decisions. In these instances, no data were verbalized to guide the decision-making. The second category includes those DATA1 and DATA2 move decisions that did not include a LINK to another move. These are presented in the third category together with VALID move decisions. Moreover, if elaborate verbalized move decisions reflect a more refined decision-making process, then the probability of the resulting move being correct should be higher for "Data" and "Valid/Link" decisions than for the "No Data" decisions. To check this assumption, we will also report corresponding probabilities below.

In summary, the analysis of move decisions suggested that older, and particularly Grade 8 participants, verbalized more detailed move decisions than did younger participants. While Grade 4 participants verbalized mainly "No Data" move decisions, Grade 6 and Grade 8 participants also verbalized a considerable number of more complex "Data" and "Valid/Link" decisions. In general, more detailed move decisions were also somewhat more likely to result in correct move(s) than the less elaborate move decisions, although this was true to a lesser extent for Grade 8 participants.

Grade 4. More than three quarters (35 out of 44) of Grade 4 participants' move decisions were verbalized as DISK/PLACE or DISK, and six Grade 4 participants verbalized no other move decisions on Item 3. Thus, the vast majority of Grade 4 move decisions on Item 3 belonged to the "No Data" category (about 80%). The probability of the corresponding move(s) being correct was .39 in this category. The nine "Data" move decisions, in turn, resulted in 10 moves, 5 of which were correct. We should also note that Grade 4 participants verbalized no move decisions that were either VALID or included a LINK to other decisions.

On Item 4, Grade 4 participants verbalized 58 move decisions. Again, the majority of these (about 80%) were "No Data" decisions and five participants verbalized no other move decisions. Ten move decisions (17%) included some data. The remaining two move decisions were VALID (one with a LINK). About half of the moves that followed a "No Data" decision were correct (probability .51). The respective probability for "Data" move decisions was .64, whereas both moves that followed VALID decisions were correct.

On Item 5, Grade 4 participants verbalized 49 move decisions, 82% of which belonged to the "No Data" category. Five Grade 4 participants verbalized only this type of move decisions. "No Data" decisions resulted in 38 moves, 29% of which were correct. The nine "Data" decisions resulted in 9 moves (4 correct; probability .44). Finally, on Item 6, Grade 4 participants verbalized considerably more move decisions (64). Still, "No Data" decisions were by far the most common (73%) but were not necessarily less accurate than the "Data" move decisions (22%). The probability of a move being correct was .31 for "No Data" decisions and .27 for "Data" decisions. However, the three DATA2 decisions that also included a LINK to other decisions produced three moves, only one of which was incorrect.

Grade 6. On Item 3, Grade 6 participants showed considerably more variability in their verbalized move decisions than did younger participants. Still, half (24 out of 48) of the verbalized move decisions consisted of naming the disk and its place. Only one Grade 6 participant, however, verbalized no other move decisions. Three move decisions only named the disk. These 27 "No Data" move decisions resulted in 25 moves, 14 of which were correct (probability .56). Forty percent of the move decisions belonged to the "Data" category. "Data" move decisions resulted in 21 moves, 12 of which were correct (probability .56). Thus, there was no difference between the "No Data" and the "Data" decisions in terms of the probability of producing a correct move. Finally, the two move decisions with LINK resulted in three moves, two of which were correct.

Only one Grade 6 participant verbalized no other move decisions than "No Data" decisions on Item 4. Approximately 57% of all move decisions belonged to the "No Data" category, 40% belonged to the "Data" category, and 3% (2 move decisions) belonged to the "Valid/Link" category. These numbers are almost identical to those for Item 3. The probabilities of the corresponding moves being correct were .57 for "No Data" decisions, .41 for "Data" decisions, and 1 for "Valid/Link" decisions (2 moves, both were correct).

On Item 5, Grade 6 participants verbalized 48 move decisions, 46% of which were "No Data" decisions (one participant verbalized no other move decisions) and another 46% were "Data" decisions. The "No Data" decisions resulted in 27 moves, of which only 8 were correct (probability .30). In contrast, 15 out of 25 moves that followed "Data" decisions were correct (probability .60). Two "Data" decisions were verbalized with LINK and both resulted in incorrect moves. Also, two move decisions were classified as VALID and resulted in two moves (one was correct).

Two Grade 6 participants only verbalized "No Data" move decisions on Item 6. In total, Grade 6 participants verbalized 52 move decisions on Item 6, 56% of which were "No Data" decisions. These resulted in 33 moves with a .39 probability of being correct. Twenty (38%) "Data" decisions resulted in 19 moves, out of which 12 were correct (probability .63). Finally, three "Valid/Link" decisions resulted in 4 moves, only one of which was incorrect.

Grade 8 participants verbalized 50 move decisions on Item 3, only 15 of which were DISK/PLACE move decisions. The most commonly verbalized move decision in Grade 8 was DISK/DATA, which was verbalized 21 times. Five of these were classified as DATA1, 12 as DATA2, and 4 as VALID.

No particular type of move decisions seemed to be associated with correct or incorrect moves for Grade 8. Twenty-one "No Data" decisions resulted in 13 correct (probability .72) and only 5 incorrect moves. The "Data" decisions (also 21) resulted in 23 moves, 14 of which were correct (probability .61). Eight "Valid/Link" decisions, in turn, were associated with 5 correct (.56) and 4 incorrect moves. Somewhat surprisingly, the two DISK/DATA2 & LINK decisions produced three incorrect moves. The first of these seemed to comply with the "0 correct" lines but not with the positive information lines. Eve's DISK/DATA2 & LINK started with finding the only possible place for Black, however, she failed to accept this decision and left the move incomplete.

On Item 4, about 44% of move decisions verbalized by Grade 8 participants belonged to the "No Data" category, 46% belonged to the "Data" category, and 10% belonged to the "Valid/Link" category. Thus, compared to Item 3, there was a minor drop in the frequency of the most complex move decisions. One participant verbalized only "No Data" move decisions. The probability of the corresponding moves being correct were .48 for "No Data" decisions, .54 for "Data" decisions, and 1 for "Valid/Link" decisions (5 decisions with 5 moves, all were correct).

Three Grade 8 participants verbalized only "No Data" move decisions on Item 5. Altogether, Grade 8 participants verbalized 60 move decisions; 60% were "No Data" decisions, 28% were "Data" decisions, and 12% were "Valid/Link" decisions. About half of the moves that followed "No Data" move decisions were correct (probability .49), whereas the moves that followed "Data" move decisions were slightly less accurate (probability .40). Five out of 8 moves (probability .62) that followed "Valid/Link" decisions were correct. Finally, on Item 6, one Grade 8 participant verbalized only "No Data" move decisions. In total, Grade 8 participants verbalized 48 move decisions on Item 6. "No Data" decisions were again the most commonly verbalized (44%). About one-third (31%) of all verbalized move decisions were "Data" decisions, and one-fourth were "Valid/Link" decisions. The probability of the corresponding move being correct was .50 for both the "No Data" and the "Data" categories, and .71 for the "Valid/Link" category.

Evaluations

In terms of evaluating their performances, the most pronounced difference between

Grade 4 and Grade 6 seemed to be in the frequency of evaluations, particularly in the frequency of answer line evaluations. Grade 6 participants' evaluations were also more likely to be correct, although these differences disappeared on Items 5 and 6. Compared to Grade 4, Grade 8 participants produced both a greater number of evaluations as well as more accurate evaluations, whereas the biggest difference between Grade 8 and Grade 6 was in the accuracy of the evaluations.

Grade 4. Only three Grade 4 participants verbalized evaluations on Item 3. Felicity's three evaluation statements occurred while she was still completing the item. Two of these were very brief ("no") evaluations of an assumption and of an attempt. The first was incorrect. Her first evaluation of a move accepted a correctly placed disk. Farley's evaluation of an assumption took place before he had completed any moves: He assumed that Yellow must be the one correct in Line 3/1 ("So one correct, must be the Yellow") before eliminating this possibility out ("No, it can't be the Yellow"). His second evaluation occurred at the end of his performance and consisted of accepting an answer line that violated both single and relational constraints. Faye's only evaluation consisted of accepting an answer line that violated negative single constraints but had two correct in Line 2/2 and one correct in Line 3/1. In sum, Grade 4 participants verbalized only six evaluation statements on Item 3, three of which were incorrect.

In contrast, Grade 4 participants verbalized 21 evaluation statements on Item 4. Eight involved assumptions or attempted moves; only two of these evaluations were correct. Evaluations of completed moves were considerably more successful with 9 out of 12 being correct. The only evaluation of a complete answer line was incorrect. Thus, on Item 4, 11 out of 21 evaluations verbalized by Grade 4 participants were successful.

On Item 5, Grade 4 participants verbalized only nine evaluation statements, five of which were correct. The only evaluation of assumption was incorrect, as was one of the two evaluations of moves. Contrary to the previous items, evaluations of partial answers and complete answers were most commonly verbalized (both were verbalized three times). All evaluations of partial answers were produced by different individuals and were correct, whereas only one of the three evaluations of complete answers was correct.

Finally, Grade 4 participants verbalized both more evaluations in general (17), and more accurate evaluations in particular (12 correct) on Item 6. Assumptions were evaluated twice (incorrectly) and an attempted move was evaluated once (correctly). Five out of six evaluations of moves were correct. Partial answers were again evaluated with considerable accuracy (3 out of 4 were correct), as were also the complete answers (3 out of 4 were correct).

Grade 6. Grade 6 participants verbalized 17 evaluation statements on Item 3.

Three of these statements evaluated assumptions, two of which rejected a place for one disk (both correctly so) and the third rejected two proposed disks as correct in Line 2/2 (one of these was a correct disk). Completed moves were evaluated nine times; only one of these evaluations was incorrect and one was incomplete due to time violation. Grade 6 participants also evaluated partial answer line once (incorrectly) and complete answer lines four times (2 were correct). Thus, about two-thirds of the evaluations verbalized by Grade 6 participants were correct on Item 3.

On Item 4, Grade 6 participants verbalized 18 evaluation statements, two-thirds of which were correct. Seven of these involved evaluating assumed or attempted moves (five were correct), whereas only two evaluated completed moves (both correctly). Partial answer was evaluated three times (two were correct) and a complete answer six times (three were correct).

On Item 5, Grade 6 participants were less accurate in their evaluations with 10 out of 19 evaluations being correct. In particular, evaluations of moves (2 out of 8 were correct) and evaluations of complete answers (4 out of 7 were correct) seemed to cause difficulties. In contrast, both evaluations of assumptions as well as the only evaluation of an attempted move and a partial answer were correct. On Item 6, two-thirds of the evaluations (12 out of 18) verbalized by Grade 6 participants were also correct. Six evaluations involved assumed or attempted moves (five were correct), three involved completed moves (two were correct), and another three involved partial answers (two were also correct). Evaluations of complete answers seemed to create the most difficulties for Grade 6 participants on this item, with only three out of six being correct.

Grade 8. Grade 8 participants verbalized 17 evaluations on Item 3, one of which was incorrect and one of which was incomplete due to a time violation. The only incorrect evaluation of a move resulted in accepting a single constraint violation. Four of the six correct evaluations of moves were verbalized by Earl. Three Grade 8 participants evaluated an assumption. Two participants evaluated partial answers. Both occurred in a situation in which some of the disks were incorrectly placed but then subsequently were moved to correct places. All but one evaluation of a complete answer line took place at the end of the performance and resulted in accepting correct answers. Eric evaluated a complete answer line twice. The first instance occurred in the middle of his performance and his somewhat elaborate evaluation was followed by four quick moves, which produced a correct answer.

On Item 4, Grade 8 participants verbalized 19 evaluations. Assumptions were evaluated five times (all evaluations were correct), completed moves were evaluated four times (all were correct), partial answer lines were evaluated two times (both were correct), and complete answer lines were evaluated eight times (seven were correct). Evaluations of complete answers were also verbalized eight times on Item 5 (all evaluations were correct). An assumption was correctly evaluated once. Attempted moves were evaluated three times (two were correct) and completed moves were evaluated five times (four were correct). Evaluations of partial answers were again completely accurate with all four evaluations being correct. In total, Grade 8 participants verbalized 21 evaluations on Item 5, 19 of which were correct. Finally, Grade 8 participants verbalized 18 evaluations on Item 6. Half of these were evaluations of complete answers and most of these (8 out of 9) were also correct. Five out of six evaluations of moves were correct, whereas both evaluations of partial answers and the only evaluation of an attempted move were correct.

Miscellaneous Statements

This category consisted mostly of statements that indicated either verbal regulation of behavior or confusion. To summarize, three to five Grade 4 participants verbalized regulative statements on each of the Items 3 to 6. For Grade 6, the respective figure was one to three, and for Grade 8 it was one to two. In contrast, Grade 4 participants verbalized no statements that indicated confusion on any of the items, despite the fact that they completed most of the items incorrectly. Grade 8 participants verbalized some confusion statements and Grade 6 participants verbalized several on each item.

Grade 4 participants verbalized three regulative statements on Item 3. Felicity started by stating "let's see here" and Fiona began by focusing her attention on Line 2/2: "I think I would take the one that has most correct and look at that one." Farley's regulative statement ("let me think") took place later in his performance.

Grade 4 participants produced six regulative statements in Item 4. Fiona again focused on the "2 correct" line, whereas all of the other statements were similar to those verbalized by Felicity and Farley on Item 3. On Items 5 and 6, Grade 4 participants verbalized 10 regulative statements (6 and 4, respectively). On Item 6, they also verbalized three statements that seemed to endorse their approach ("this is easy now", "I know").

Grade 6. Sandra was the only Grade 6 participant who verbalized a regulative statement on Item 3. Her statement seemed to focus her attention on Line 2/2 and Line 3/1 ("so I'll look at the two middle information lines") and occurred after she had stated that all of the colors must be different from the two "0 correct" lines. The only other two entries in this category include Simon's comment "this is hard" and Sally's two CONFUSION statements "I don't know". Sally's second statement was followed by a move that placed Blue to Place 1, which is a single constraint violation.

On Item 4, Sally verbalized three CONFUSION statements. Three participants verbalized REGULATION statements; three instances were recorded for both Simone and

Sandra, and one instance was recorded for Simon. On Item 5, three Grade 6 participants verbalized CONFUSION statements, whereas Sandra was the only one who verbalized regulative statements (three altogether). On Item 6, in turn, two participants verbalized CONFUSION statements and three verbalized regulative statements (one participant verbalized both).

Grade 8. Two Grade 8 participants verbalized REGULATION statements on Item 3, both involved the procedure for the item (Eve: "I have to check here first" and Erica: "I'll go to the White one first"). Elaine, in contrast, stated "I don't know" twice and these were coded as CONFUSION. Remember that she also used a Trial-and-Error method in deciding some of her moves, so in her case the confusion may well have been valid.

Grade 8 participants produced two REGULATION statements on both Items 4 and 5, and one on Item 6. Finally, CONFUSION statements were verbalized once on Item 5 and twice on Item 6.

Discussion of Verbal Protocol Analysis

What do the above results tell us about the different planning components identified in the introduction? The task definitions primarily provide information about the initial representation that participants construct at the beginning of the task. These can be very detailed and impose constraints on future actions, as most top-down models of planning would predict, or they can be more cursory and may include information only for the first decision, as recursive or planning-in-action models would predict.

We should start by noting that no participant verbalized "an ideal" task definition that would have resulted in maximum constraints on future actions. Items 3 and 4 have four disks and four information lines: two "0 correct", one "1 correct", and one "2 correct". On Item 3, Line 2/2 and Line 4/0 have two disks (Yellow and White) in the same locations. Thus, the simplest method of solving this item would be to compare these two lines, notice the similarities, and deduce that the other two disks in Line 2/2 (Blue and Black) are then necessarily correct. This approach requires both selective encoding and selective combination of information (Davidson & Sternberg, 1984). No participant used this approach, although several noticed that Yellow cannot be correct in Line 2/2. Since Yellow is also in the same place in Line 3/1, the three Yellows together provided the most striking visual clue in the display. Moreover, only one computer protocol was similar to the protocol most likely to result from this approach (Blue and Black in first, then Yellow and White; correct in four moves). This performance (Emma) included no other verbalizations than four DISK/PLACE move decisions.

Item 4 is formally identical to Item 3. The two disks that are the same in Line 1/2 and Line 4/0 are now Blue and Black. One Grade 8 participant noticed both of these similarities during the task definition but failed to notice the significance of this information. Also, two Grade 8 performances that included long task definitions but no comparison of Line 1/2 to Line 4/0 produced a computer protocol similar to one predicted by the "ideal" approach. Thus, it is possible that although no participant verbalized the "ideal" approach, some Grade 8 participants did utilize it.

On Items 5 and 6, only one disk is in the same position on the "1 correct" line (Line 2 on Item 5 and Line 3 on Item 6) and "0 correct" line. One Grade 4, one Grade 6, and one Grade 8 participant verbalized noticing this similarity during their task definition on Item 5, whereas two Grade 4, one Grade 6 and one Grade 8 participant verbalized noticing the respective similarity on Item 6. Thus, it seems that the most powerful method of limiting the initial search space was not frequently used by our participants.

We should note that many participants, particularly in Grade 4, did not verbalize any task definitions. Furthermore, most of the verbalized task definitions seemed to merely state the existing information or add simple interpretations and/or assumptions. This was particularly true with many of the longer task definitions that were verbalized by Grade 4 and Grade 6 participants, which may explain why we failed to find a connection between first move latency and a correct answer or a correct first move. These participants seemed to spend considerable time at the beginning of the item stating the information that was readily available rather than encoding it selectively or successfully combining relevant aspects of it. In many cases, "selective encoding" seemed to mean that the participant attended only to the positive information lines (and produced an answer line that violated "0 correct" lines) or the negative information lines (and produced an answer line that contained too many correct disks on one of the positive information lines).

Selective combination is best represented by those statements that combine information from different lines to limit the search space. Grade 4 participants verbalized ten such statements, compared to 22 for Grade 6 and 26 for Grade 8. Thus, combining information was considerably less prevalent for Grade 4. A closer examination of the most powerful of these search space limitations — limiting possible correct disks for the line with most correct — indicated that the search space limitations may also have been lacking in "selectivity" since two of the six search space limitations of this type were incorrect for Grade 4. In contrast, all six similar limitations verbalized by Grade 6 participants were correct, as were the seven verbalized by Grade 8 participants. Thus, it would seem that Grade 4 participants were not yet fluent in more elaborate encoding methods even when they chose to use such methods.

In Grades 4 and 6, performances that included a verbalized task definition were no more likely to produce a correct answer or even a correct first move than were those performances that included no verbalized task definition. The connection between verbalized task definition and correct performance was not evident in Grade 8 either and several performances with no verbalized task definition resulted in correct answers. In Grade 8, however, performances with a verbalized task definition were somewhat more likely to produce a correct first move (average probability .71 for Items 3 to 6 combined) than were performances with no verbalized task definitions (average probability .26).

These results suggest that the initial representation constructed by our younger participants was lacking either because they failed to perceive the interconnectedness of different pieces of information or because they did not find it as necessary to generate more information at the beginning of the items that would be useful in guiding the planning process. Instead, many participants seemed to choose a more direct approach and proceeded to place disks immediately. This indicates that initial task representations may have been veridical rather than idiosyncratic, although they included significant gaps.

Move decisions provide information about the "on-line" representation that the participant generates for each decision-making situation. Results in the move decision section clearly suggest that older participants generated more information and subsequently, more detailed representations. Figure 4-2 displays the average probability of a move decision belonging to one of the three categories ("No Data", "Data", and "Valid/Link") for Items 3 to 6.

Figure 4-2 indicates that Grade 4 participants verbalized mainly "No Data" move decisions. Very few of them verbalized any valid move decisions or move decisions that included a link to other moves.

While Figure 4-2 demonstrates that "No Data" move decisions were still the most common for Grade 6, the probability of a move decision including at least some data was considerably higher than for Grade 4. Similar to Grade 4, very few move decisions were valid or included a link to other moves.

For Grade 8, move decisions that used some or all relevant data were more frequent than "No Data" move decisions. Perhaps the most significant difference between Grade 8 and the other two grade levels was the substantial increase in "Valid/Link" decisions. These decisions can require a considerably more complex representation of the situation than "Data", and particularly "No Data", move decisions; either in terms of recognizing the interconnectedness of two decisions or in terms of understanding the importance of an "exhaustive search" in a relatively limited search space and accepting the uncertainty of the end result that frequently followed.

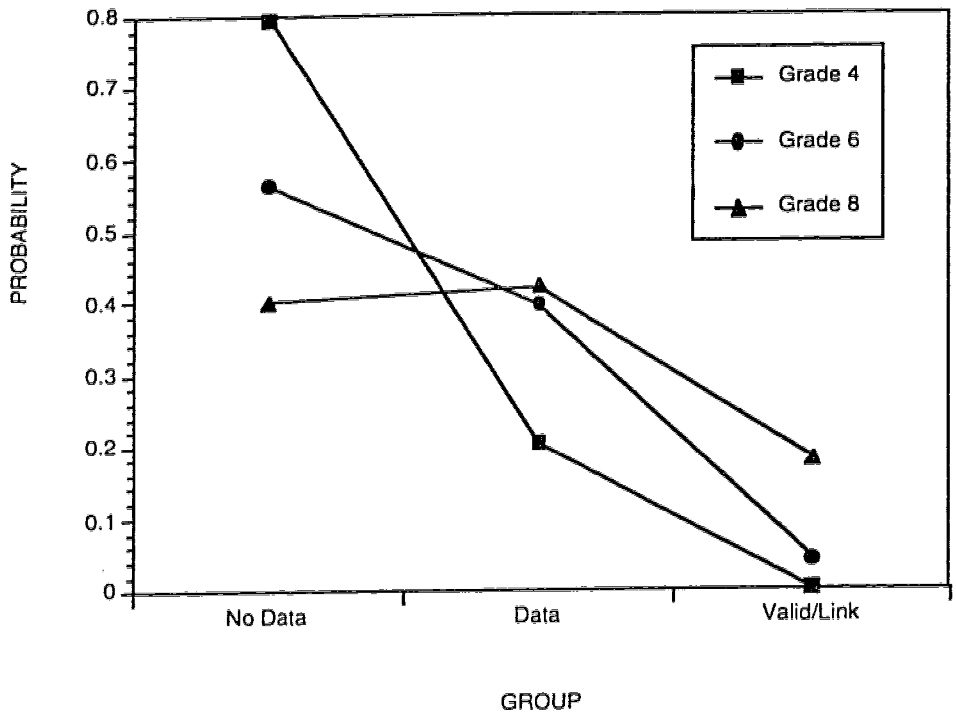


Figure 4-2. The average probabilities of "No Data", "Data", and "Valid/Link" move decisions for Items 3 to 6.

It is also possible that Grade 4 participants generated more information and formed better representations than they actually verbalized. In these cases, analysis of verbal protocols would provide an exceedingly negative picture of their planning process. As was already noted above, in Grade 4, the few move decisions that identified more information also tended to result in correct moves more often than simpler move decisions. The mean probability of the resulting move being correct was .38 for "No Data" decisions, .46 for "Data" decisions, and .88 for "Valid/Link" decisions. Thus, it seems that differences in verbalizations were likely to reflect real differences in the planning process and specifically in the representation component of it. Since most Grade 4 participants failed to produce correct answers for Items 3 to 6, we have no reason to assume that the actual move decisions would have been based on considerably more information than was verbalized. This suggests that if Grade 4 participants' initial task representation included significant gaps, these gaps were seldom filled later during the performance.

For Grade 6, the mean probability of the move being correct was .46 for "No Data" decisions, .55 for "Data" decisions, and .65 for "Valid/Link" decisions, suggesting a similar conclusion. The respective probabilities for Grade 8 were .55, .51, and .72. This result suggests that at least some older participants did not verbalize all of the information they were generating when determining the correct locations for the disks.

In our planning process model, evaluation statements are most closely linked to the regulation component. Correct evaluations, however, also require a veridical representation of the conditions that a move or answer are required to meet, and evaluation of attempted moves and assumptions indicate that the participant anticipated the consequences of his or her intended actions. Thus, evaluation statements provide information not only about regulation but also about anticipation and representation components of the planning process.

Grade 8 and Grade 6 participants verbalized considerably more evaluative statements (75 and 67, respectively) than did Grade 4 participants (53). The difference between older and younger participants was most pronounced in evaluations of complete answer lines: Grade 8 and Grade 6 participants verbalized 30 and 24 (respectively) instances of evaluating their answer lines, compared to only 10 in Grade 4. In contrast, evaluations of assumptions, attempted moves, and completed moves were verbalized equally as often in all three grade levels. This suggests that while Grade 4 participants may not have been inactive evaluators in general, they either did not perceive any reason to evaluate a complete answer line, or they were not capable of grasping this amount of information.

In terms of using inappropriate standards, our results indicate that Grade 4 participants had the most difficulties in evaluating their performances accurately, particularly on Items 3 and 4 in which their evaluations were as likely to be incorrect as correct. Evaluations of partial answers and moves were the most productive for Grade 4 participants, with all but one evaluation of partial answers being correct (all on Items 5 and 6) and 16 out of 21 evaluations of moves being correct. Evaluations of assumptions and attempted moves, in turn, seemed to be the most problematic, with only 5 out of 15 being correct. This suggests that Grade 4 participants experienced specific problems in anticipating the results of their actions and more general problems in building a sufficient criteria for evaluations, that is, the representation of a correct solution.

Grade 6 participants were somewhat more successful in their evaluations, with an average of two-thirds of all evaluations being correct. Evaluations of moves were most accurate on Items 3 and 4, whereas partial and complete answer lines were problematic. On Items 5 and 6, moves (particularly on Item 5) and answer lines were most problematic to evaluate for Grade 6 participants. This suggests that Grade 6 participants also experienced problems in building an accurate representation of the goal state and consequently used inappropriate standards.

Grade 8 participants, in contrast, evaluated their assumptions, moves, partial answers, and complete answers with almost perfect accuracy. Thus, not only did they frequently engage in on-line regulation of their behavior but they were also capable of forming a correct representation of the goal state and accurately anticipating the consequences of their actions.

It is also interesting to note that statements indicating verbal regulation of behavior were most frequent for Grade 4 participants. In contrast, Grade 4 participants verbalized no confusion statements on any of the items, although they produced the most incorrect moves and answers. Thus, it seems that our younger participants were willing to "guide" the planning process, or to "keep track" so to speak, but they were not capable of assessing the effectiveness of the process. Grade 6 participants verbalized both regulative and confusion statements, indicating more flexibility and an increasing capability to also treat the process as an object of evaluation.

The execution component was not directly represented by verbal statements; instead, we will need to compare the verbal and computer protocols and attempt to identify mismatches between the two. One such mismatch was explained above (Sandra's move decision on Item 3). Only two other mismatches were identified, both on Item 6. Sandra attempted to place Blue in a situation in which all remaining places were incorrect. She correctly noticed this but still continued to place Blue in Place 4, which she had ruled out. Earl, in turn, correctly noticed that White cannot be in Place 1 or in Place 4 and thus, it had to be placed in the "middle ones" (Place 2 or Place 3). Rather than removing a disk from Place 2 or Place 3 to accommodate White, he placed White in Place 1, which was open. The fact that we were able to identify only three such instances in 76 performances suggests that execution problems were not prevalent in our sample. One obvious reason for this is the limited nature of the plans that our participants verbalized: Rather than verbalizing complex multistep plans that are more susceptible to execution problems, most of the verbalized plans were local one- or at the most two-step plans that were followed by an immediate execution of the entire plan. Thus, possible sources of execution problems such as keeping all the components of the plan in working memory and monitoring the progress were not a concern with this sample.

Analysis of Search Methods

The final result section combines information from both verbal and computer protocols to describe the search methods that were employed by the participants. How we decide over the next move, or in Anderson's (1993) terms, how we search for the next operator, is a central question in planning and problem solving. Search methods can vary "from blind search to executing an algorithm that is guaranteed to find a minimum-step solution" (Anderson, 1993, p. 37). Anderson further argues that the two key features of problem solving are difference reduction and subgoaling. Difference reduction simply refers to the tendency of problem solvers to select moves that produce states more similar to the goal state. In other words, if the problem solver cannot reach the entire solution immediately, he or she will attempt to reach parts of it. Subgoaling is a difference reduction method linked most often with "means-ends analysis", which is frequently cited as the most common problem solving method.

Although means-ends analysis is a valid way of describing the problem solving process in many experimental tasks, the structure of the Crack-the-Code task is such that we cannot expect all aspects of a means-ends analysis to manifest in a problem solver's performance. Possible moves in the beginning of the items are not limited in a manner that would readily lead to subgoaling since the subgoal ordering in Crack-the-Code is ambiguous (any of the 3 or 4 colored chips can be placed first). In other studies, children have been found to experience difficulties in applying means-end analysis in tasks with ambiguous subgoal ordering (Klahr, 1985). Moreover, there is not one correct path through the problem space, as in the Tower of Hanoi task for example, but several different possible ones, which makes means-end analysis even more "open-ended." This is not to say that participants did not use means-end analysis. Rather, it is to say that in those instances in which means-end analysis was used, it was used more as a heuristic guide of the process ("I should try to figure which two are correct," or "get one correct"). The search method can be described better under the four categories defined below.

Thus, the problem of identifying what search methods our participants used was one of describing their difference reduction method, that is, describing the kind of reasoning on which the decisions of moves were based on. A priori task analyses and a pilot study with Grade 4 and Grade 8 students (Parrila & Papadopoulos, 1994) identified four classes of search methods:

1. *Trial-and-Error.* In this approach, the participant (a) places the colored disks in the answer line randomly, or (b) forms a hypothesis about the correct order of disks without utilizing the given information.
2. *Pattern.* Pattern search is based on one information line and a decision is made over at least two disks simultaneously. A good example of the use of Pattern is when a participant decides that the two correct disks in the "2 correct" line are, for example, Blue and Black, and then proceeds to place these colors in equivalent places in the

answer line. One specific instance of using this search method is when the participant reproduces one of the information lines, or part of it, on the answer line. This was called *False Pattern.*

3. *Climbing.* In Climbing, the participant searches for one correct placement at a time without simultaneously taking into consideration other disks. He or she can do this (a) by trying to find a correct place for one disk at a time, or (b) by trying to find a correct disk for a particular place in the answer line. While Climbing, the participant uses at least one information line as the basis for the decision, thus the placements are not random.

4. *Combination.* This is the most complex search method and includes combining information from at least two separate information lines to simultaneously decide over at least two placements. When Combination was detected, we also recorded how many disks and information lines were used to make the decision. Thus, 2 X 2 Combination included two disks and two information lines, whereas 2 X 4 Combination included two disks and four information lines (and, thus was considerably more difficult).

The hypothesized search methods vary both in their efficacy in solving different items and in their cognitive demands. Combination is the most advanced method and it is a functional way of solving all the items. However, it places high demands on working memory and therefore it may be out of reach for the younger participants. Climbing is not a fully functional method at the beginning of any of the items because it cannot produce a definite placement. In other words, when used as the first search method at the beginning of the item, Climbing leads invariably to a situation where participants must choose between two different possible placements, only one which is correct. Thus, they have a 50% chance of "guessing" correctly. In contrast, after participants have placed one or two disks in their proper places and have therefore limited the possible placement alternatives, Climbing can be a functional method, and even the preferred one since it does not load the information processing system as heavily as does the Combination method. Pattern and Trial-and-Error methods were deemed a priori to be nonfunctional approaches in all items. (This does not mean, however, that these approaches could not produce correct placements.) It is assumed that participants resort to these methods only when other, more appropriate methods, fail or when participants for some reason fail to build a functional representation of the item parameters.

Procedure. In order to determine what search methods the participants used, each verbal protocol was analyzed separately by two trained raters. As part of the training, one verbal protocol per item per grade level was analyzed together. In the transcripts of verbal protocols all moves were marked together with respective move latencies and move durations. The raters were permitted to mark down two alternatives when verbal protocols allowed for conflicting interpretations, and computer protocols were then used to determine the appropriate selection. Thus, it may be more accurate to say that the analyses of search methods were based on both computer (moves, move latencies, move durations) and verbal protocols. Since most of the protocols included more than one search method, interrater agreement was calculated based on individual decisions (whether both raters agreed on the search method), rather than on protocols as a whole. The raters agreed on 91.3% of the decisions on Item 3, 87.8% on Item 4, 85.4% on Item 5, and 86.8% on Item 6. Disagreements were solved through negotiations between the raters. Last moves were included in these analyses only if they were part of a more comprehensive False Pattern, Pattern, or Combination decision.

Perhaps the most concise way to summarize the analysis of search methods is to simply count (a) how many instances of different search methods were detected in different grade levels, and (b) how many placements were decided using each of the search methods. Note that with Climbing and Trial-and-Error, these numbers are exactly the same, whereas with Combination, Pattern, and False Pattern, the latter count is by definition higher than the former.

Grade 4. Climbing was the most frequently used search method on Item 3 with 15 instances. Combination, Pattern, and False Pattern were each used two times and involved eight, six, and eight placements, respectively. On Item 4, most placements were also based on Climbing (24 instances). Combination (involving 3 placements), Pattern (involving 4 placements), and Trial-and-Error (1 placement) were each used only once, whereas False Pattern was used three times and was responsible for 9 placements.

Item 5 performances also relied mainly on Climbing (18 instances). Pattern and False Pattern were used three times and involved eight and nine placements, respectively. Finally, Combination was used twice and involved placing five disks. Both participants who attempted Combinations first limited the search space with other methods. The only successful performance was the one that combined Climbing with Combination.

On Item 6, Grade 4 participants produced considerably more moves (Table 4-7) than on the previous items. Accordingly, they also used different search methods more frequently, 44 times altogether. Half of the placements (29 out of 58), however, were again based on Climbing. Four Pattern decisions that involved 16 placements and the one False Pattern decision (2 placements) together accounted for a third of the placements. Combination was used only once and involved two placements. The only difference in search methods compared to previous items was the increase in the use of the Trial-and-

Error search method. Nine Trial-and-Error decisions were made on Item 6, which is more than on all the other items combined.

Grade 4 participants showed little variability in their search methods. All but one used only one search method on Item 3 (mean = 1.1). On Items 4 and 5, Grade 4 participants used an average 1.4 search methods. Only on Item 6 did they attempt to use more search methods than did the Grade 6 or Grade 8 participants (mean = 1.4). Moreover, four participants attempted to solve both Items 3 and 4 in exactly the same manner despite failing Item 3. This indicates that many of the Grade 4 participants may not have had alternative methods available for them. On Item 5, Grade 4 participants' performances could be divided essentially into three subgroups: those that used only Climbing, those that used Climbing with Pattern or False Pattern, and those that used only Pattern or False Pattern. As mentioned above, Item 6 protocols were considerably longer in many cases. However, four participants approached Item 6 in essentially the same manner as Item 5. Three participants tried something different, and the remaining three had mainly "more of the same."

Grade 6. On Item 3, no Grade 6 participant used Combination. Trial-and-Error was used four times, Pattern was used three times, and False Pattern was used once. The Pattern search method accounted for 9 placements, whereas False Pattern and Trial-and-Error accounted for 4 each. In contrast, we detected as many as 25 instances of the Climbing search method in Grade 6. Thus, it seems that Grade 6 participants opted for a cautious stepwise approach on Item 3. Mean number of used search methods was 1.4.

While Climbing remained the most frequently used search method (20 instances), Combination resurfaced on Item 4 with four instances that accounted for 13 placements. Three Pattern decisions involved 6 placements and four False Pattern decisions involved altogether 11 placements. With the exception of one performance that included only one Combination decision, all other performances used Climbing at the beginning of the performance. Moreover, three performances that started with Climbing and ended with Combination produced a correct answer (one participant ran out of time while still evaluating the answer line). On average, Grade 6 participants used 2 different search methods on Item 4.

Climbing was the dominant search method on both Items 5 and 6 (32 instances on both items). On both items, approximately 70% of all the placements were based on this search method. Combination was used two times (involving five placements) on Item 5 and only once on Item 6 (two placements), whereas Pattern was used only once on both items. These decisions involved four and three placements on Items 5 and 6, respectively. The two False Pattern decisions on Item 5 involved five placements, whereas the only

False Pattern decision on Item 6 produced a complete answer line. Finally, the Trial-and-Error method accounted for one placement on Item 5 and for two placements on Item 6.

Also, Grade 6 participants used more different search methods than Grade 4 participants on Items 3 and 4. This variability was also evident in each participants performance: None of the ten Grade 6 participants attempted to solve both items in a similar manner. This was true whether or not they succeeded on Item 3. Within-group variability in search methods was more limited on Items 5 and 6 — after all, Climbing dominated most of the protocols. On average, Grade 6 participants used 1.3 search methods on both items. Moreover, there was very little intraindividual variability between the items: Only one participant changed her approach completely.

Grade 8. Climbing was also the method of choice for Grade 8 participants. On both Items 3 and 4, Climbing was used to place 25 disks. This represents over half of all placements (46 on both items). On Item 5, Climbing was used 33 times (56% of the placements) and on Item 6, Climbing was used 34 times (72% of the placements). It is interesting to note that the Trial-and-Error method was the second most common search method used by Grade 8 participants: It was used three times on Item 3, two times on Item 4, nine times on Item 5, and as much as 12 times on Item 6. Increased use of the Trial-and-Error method is less of a surprise when we note that one participant (Earl) deliberately used this method four times at the beginning of both Item 5 and Item 6 "so I don't have to keep reaching back for them." Also, the last three moves of Erin's correct performance on Item 6 were verbalized as Trial-and-Error ("I can't figure this one, that kind of looks right") but seemed to be performed more deliberately at the end of a long protocol.

Combination was used four times on Item 3, three times on Item 4, seven times on Item 5, and once on Item 6. These decisions involved 12, 7, 17, and 4 placements, respectively. Pattern, in contrast, was present only once on Item 4 and involved two placements. Finally, False Pattern search method was used two times on Item 3 and four times on Item 4 (involving 6 and 10 placements, respectively) but not at all on Items 5 and 6.

Grade 8 participants displayed significant variability on Items 3 and 4: All search methods are present in at least one performance. On average, Grade 8 participants used 1.5, 1.9, 1.8, and 1.4 search methods on Items 3, 4, 5, and 6, respectively. We should notice, however, that most performances included only Climbing and Combination. Intraindividual variability was also somewhat limited: Only two participants (Earl and Eric) produced protocols with more than two search methods present. Four participants who passed Item 3, however, seemed to approach Item 4 differently, whereas three participants mainly used the same method on both items. Two participants failed both items; one of them changed the approach whereas the other used the same search methods (Climbing and Trial-and-Error) on both items.

Within-group variability was of a more limited nature on Items 5 and 6. Climbing, Combination, and Trial-and-Error were the dominant search methods on Item 5, whereas Climbing and Trial-and-Error accounted for 46 out of 50 placements on Item 6. Intraindividual variability was maybe more prominent: Five participants who passed Item 5 changed their approach for Item 6. That Grade 8 participants used Climbing extensively in these items is maybe not so surprising, given that it seemed to work for them: six correct answers followed performances that used exclusively Climbing as the search method, and another four correct performances used both Climbing and Combination.

Discussion of Search Methods

The above results indicate that all search methods were available to at least some participants in each grade level. Figure 4-3 displays the percentages of placements in different grade levels that were based on Climbing, Combination, and nonfunctional (Pattern, False Pattern, Trial-and-Error) search methods on Items 3 and 4 combined.

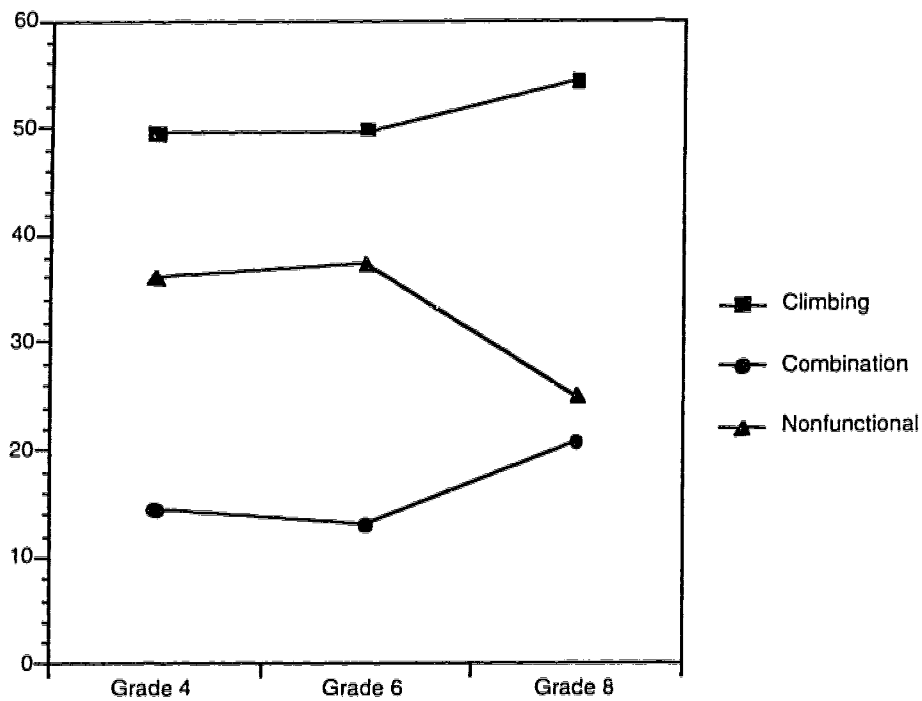


Figure 4-3. The percentages of placements based on Climbing, Combination, and nonfunctional (Pattern, False Pattern, Trial-and-Error) search methods on Items 3 and 4.

Figure 4-3 suggests Grade 4 and Grade 6 participants used different search methods to about an equal extent. Climbing was the method of choice in all grade levels, with more than half of the placements based on this search method. The only trends in Figure 4-3 that could explain why Grade 8 participants did considerably better than Grade 4 and Grade 6 participants, are the decrease in the proportion of placements based on nonfunctional strategies, and the increase in the proportion of placements based on Combination search method. In sum, this suggests that Grade 8 participants supplemented frequent use of Climbing search method with occasional Combination and nonfunctional decisions, whereas Grade 4 and 6 participants favored nonfunctional search methods over Combination as the alternative and/or supplement for Climbing.

Figure 4-4 displays the respective percentages for Items 5 and 6 combined.

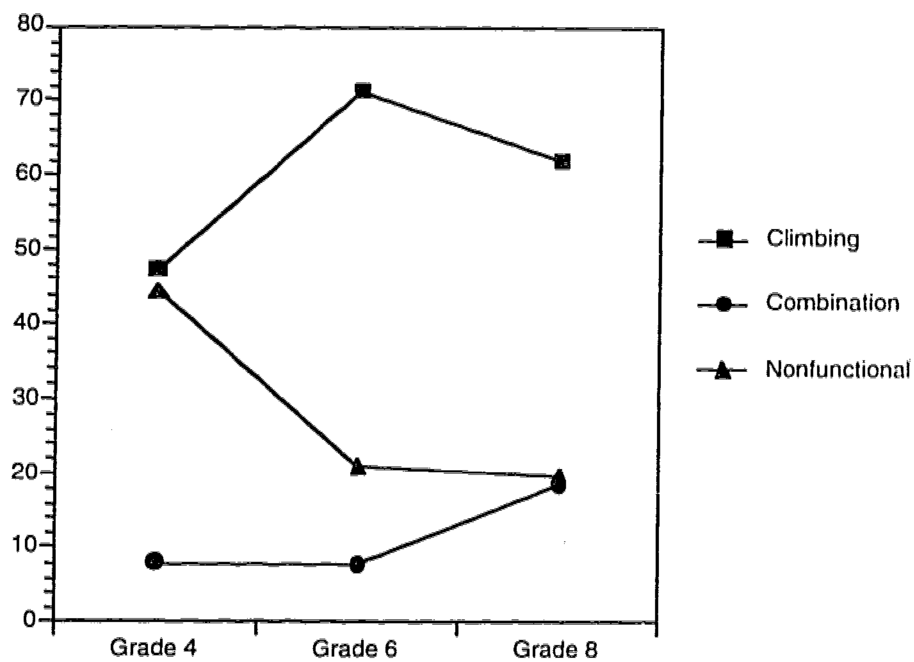


Figure 4-4. The percentages of placements based on Climbing, Combination, and nonfunctional (Pattern, False Pattern, Trial-and-Error) search methods on Items 5 and 6.

For Grades 4 and 8, the data on Figure 4-4 is rather similar to that on Figure 4-3. On Items 5 and 6, Grade 4 participants used slightly less Climbing and slightly more nonfunctional search methods, whereas the opposite was true for Grade 8. The largest difference between the figures is evident with Grade 6: On Items 5 and 6, Grade 6 participants used considerably less nonfunctional search methods and considerably more Climbing when attempting to place the disks. There were also differences within the nonfunctional methods: while Grade 4 and 6 participants used all three of them, Grade 8 participants used only Trial-and-Error. This is not an entirely inappropriate choice on Items 5 and 6, particularly when it is combined with the effective on-line evaluations that Grade 8 participants were capable of utilizing. Relatively frequent use of Trial-and-Error method does, however, explain why Grade 8 participants produced a considerable amount of single constraint violations.

As we have already noticed, most Grade 4 participants failed Items 3 to 6. Search methods can contribute to these failures in two ways: The used search methods may have been nonfunctional, or the participants may have been using functional search methods inadequately. Both explanations seem to fit part of the data for Grade 4. The only fully functional search method on Items 3 to 6 is Combination, and it was used only six times. Moreover, only in one instance of these detected Combinations, did a participant utilize all information lines — and found the correct place for all disks with the minimum number of moves. Three other more limited Combinations produced a correct answer. All of these were used at the end of the performance following several Climbing or Trial-and-Error decisions that seemed to limit the search space into a manageable size for these participant.

Climbing can also be a functional search method if combined with accurate on-line evaluation and correction of moves. As was already noted above, Grade 4 participants were not likely to correct their mistakes. Only one correct answer in Grade 4 was the result of a performance that utilized several Climbing decisions and corrections (13 moves altogether). The remaining two correct performances on Grade 4 resulted from Climbing performances in which the participants seemed to guess the first placement correctly. Thus, our results suggest that on one hand Grade 4 participants did not use the most functional search method (Combination) and over third of their placements were based on nonfunctional methods (False Pattern, Pattern, and Trial-and-Error). When they did use the best alternative search method (Climbing), it was used inadequately.

Our results suggest that Grade 6 participants were using no more functional search methods than were Grade 4 participants. Climbing search method, however, seemed to work considerably better for Grade 6 participants. As was observed above, Grade 6 participants approached the items in a more reflective manner which was evident in their better task definitions, better move decisions, and increased use of evaluation. Most Grade 6 performances that used Climbing as the preferred search method and produced a correct answer, also included successful evaluation statements or good task definitions and/or move decisions. Thus, at least some Grade 6 participants were capable of using Climbing adequately. The six Climbing performances that failed, in contrast, seemed to do so because evaluations were either incorrect or missing. In sum, Grade 6 participants succeeded when they used Climbing adequately. As in Grade 4, Grade 6 participants failed Items 3 and 4 when they used Climbing inadequately or when they used nonfunctional search methods. In contrast, inadequate use of Climbing is enough to account for Grade 6 participants failures on Items 5 and 6.

Interestingly, all but five Grade 8 performances (four by Earl) started with Climbing search method. This indicates that at the beginning of the items, Grade 8 participants deemed the search space too large to be dealt with in its entirety and opted to begin by using a difference reduction method (Climbing) to limit it. After the problem space was more limited, and probably after noticing that Climbing decisions often involved guessing between two possible locations, they switched to more comprehensive methods that involve the manipulation of several disks at once. Six protocols that started with Climbing ended with Combination and a correct solution. However, the more comprehensive method was not always functional, as is indicated by several False Pattern decisions on Items 3 and 4.

Using the False Pattern search method at the end of a protocol produces an answer line that replicates one of the information lines. This probably indicates an "idiosyncratic" goal representation rather than a veridical, but insufficient, goal representation. Eight performances in Grade 4 and five in Grade 6 ended with the False Pattern search method. In contrast, only one Grade 8 performance included False Pattern as the last search method and the participant noticed that the answer line was incorrect before running out of time. Thus, if having an idiosyncratic representation did not seem to be a problem at the beginning of the task, some Grade 4 and Grade 6 participants may have had an idiosyncratic representation of the goal state.

In sum, the most advanced search method, Combination, was used infrequently in all grade levels. Instead, most placements seemed to be based on the Climbing search method. Thus, somewhat surprisingly, the main between-group differences were not in the preferred search method but in (a) what other search methods it was combined with, and/or in (b) what information was generated for its disposal. Moreover, due to the fact that most placements were based on Climbing in all grade levels, the use of other strategies that could be classified as backup strategies was infrequent. It is interesting to note, however, that Grade 8 participants used only Trial-and-Error on Items 5 and 6, while the less functional alternatives Pattern and False Pattern were eliminated from of their repertoire completely. We should also notice that similar learning, albeit to a lesser extent, was taking place in Grade 6, as is evident in the general decreased use of nonfunctional strategies in Figure 4. What was missing in Grade 6, however, was the selection between the three nonfunctional strategies.

General Discussion

In the introduction of this article we identified two main questions: (1) What developmental differences can we detect in children's planning processes across the three targeted grade levels? and (2) How diverse are children's preferred planning methods and approaches across as well as within the grade levels? The first of these questions was covered extensively in the various results sections. In this section, we will attempt to integrate all this information into descriptions of different "prototypical" planning approaches that were identified. The emphasis will be on identifying performances that seem to share the same process structure and properties. In other words, we wish to distinguish different planning approaches in terms of differences in representation, anticipation, execution, and regulation.

Representation in our model consists of task representation, search methods, and goal representation. Initial task representation was evaluated mainly on the basis of the kind of task definition statements verbalized by the participant. Essentially, task definition statements were classified into three groups: (1) acknowledges existing information without elaborating, (2) acknowledges and elaborates (interprets, assumes, compares) without selectively combining information in order to limit the search space, and (3) limits the search space by selectively combining information. Move decisions, in turn, were used to obtain information regarding participants' later task representations. Again, three categories (explained in detail above) were used: No Data, Data, and Valid/Link.

Four categories of search methods were identified in the previous section. Most of our participants seemed to use only one, Climbing (i.e., they searched for one correct placement at a time without simultaneously taking into consideration other disks), frequently and therefore we witnessed less variability in search methods than expected. For this reason, search methods were used only as supplementary information in identifying the planning approaches.

Goal representation is evident directly in the correctness of evaluation and particularly in the correctness of answer line evaluation statements, and indirectly in using False Pattern search methods to produce an answer line. As was noticed above, only Grade 4 and Grade 6 participants used False Pattern to produce their final answer.

Anticipation was defined as the ability to predict the consequences of a plan or a partial-plan prior to acting upon them. Anticipation was represented only in evaluations of assumptions.

Execution referred to the application of plans in the environment. In general, we did not witness execution problems, probably because of the limited nature of the plans our participants actually had to execute. Thus, execution will not figure significantly in the planning approach descriptions.

Regulation was defined as the monitoring and controlling of behavior according to the plan and the evaluation of the plan and its appropriateness once executed or while being executed. The evaluations of moves and partial answers indicated the presence of on-line regulation of the planning process. Correct evaluation of the answer line, when it led to new moves, can also be regarded as part of on-line regulation.

Planning Approaches

If a performance included only four moves, the regulation component may have been lacking, particularly for those performances that produced an incorrect answer. Thus, the number of moves was used as a first criteria to divide the performances. All performances in the first four categories of planning approaches included only four moves. We should also note that six performances did not seem to fit any of the categories below and subsequently were left out. Thus, the number of performances reported below (110) is smaller than the total number of performances (116).

I. *Direct execution based on severely limited representation.* This was the "simplest" approach and, with one exception, always produced an incorrect answer with four moves. The defining feature of this approach was lack of task definition, anticipation, and on-line evaluation (regulation). Also, most of the verbalized move decisions included no data other than naming the disk and its respective place, indicating that later task representation was also limited. Sixteen performances fit this description and all but three of them were produced by Grade 4 participants.

II. *Direct execution based on insufficient representation.* Performances in this category differed from those in the first category only in that they included a task definition that consisted of stating some of the information with no or little elaboration on it. Fourteen performances fit this description, four of which were correct. Eight performances in this category were produced by Grade 4 participants whereas the remaining six were produced by Grade 6 participants. Two correct performances started with the listing of all of the information lines. The third correct performance was verbalized as several consequent guesses in between Confusion statements, whereas the fourth started with a detailed and long task definition that included two elaborative statements but no statements that limited the search space (i.e., used "selective combination"). Thus, three correct performances may have benefited from longer task definitions, although the quality of these task definitions was not high.

Verbalizations in most performances in Categories I and II were too brief and insufficient to conclusively establish the search methods that the participants utilized. Fourteen performances also included no evident pattern in move ordering or decision times; instead, they gave the impression that the participant proceeded to decide on one disk at a time without any evident data. Six performances ended with an answer line that reproduced one of the information lines (most often the one with most disks correct). All were coded as False Pattern. Eight performances seemed to be based on one information line only and were accordingly coded as Pattern. Finally, two performances were coded as Combination. One of the two was based on two information lines only.

With the exception of False Pattern performances, which may have been based on an idiosyncratic representation of the task constraints, most Category I and II performances seemed to be based on a very limited representation that acknowledged some but not enough of the task constraints. Lack of anticipation and regulation, in turn, prevented these participants from successively refining their task representation. Thus, for the most part, performances in Categories I and II seemed to be based on *inadequate search methods that operated on insufficient information.*

III. *Careful representation and execution.* The eight performances in this category ended with a correct answer in four moves based mostly on Data or Valid/Link move decisions. Four performances also included a task definition with at least one statement indicating that the participant limited the search space. For the most part, however, these performances were based on careful progress, taken one step at a time, with the participants attempting to generate as much information as possible before each individual decision, thus indicating careful "local" planning. This was also evident in lengthy decision times for most of the moves. Only one performance included an instance of evaluating the first move — no other evaluation statements were verbalized. None of the performances included anticipation. We should note that in order to produce a correct answer without any incorrect moves one has to either use the Combination search method and simultaneously consider possible placements for several disks (which was not verbalized in any of the eight performances in this category) or correctly guess between the two possible places for the first disk. Four performances in this category, all by Grade 8 participants, started with a Valid move decision that acknowledged the need to guess the first placement. All eight performances in this category were coded as using the Climbing search method. The main difference between these performances and Category I and II Climbing performances was that, in this category, the participants generated much more information to guide the decision making in each step. In other words, Category III performances used *Climbing appropriately after the initial correct guess.*

IV. *Careful starters with anticipation and representation.* All eight performances in this category included evaluations of an assumption, that is, *anticipation prior to completing the first move.* The defining feature of the five correct performances (four by Grade 8 and 1 by Grade 6 participants) was that the participants noticed that their initial assumption was incorrect and subsequently modified it. In two unsuccessful performances (one from Grade 4 and Grade 6, respectively), in contrast, the participants accepted incorrect assumptions. The third unsuccessful performance (by a Grade 6 participant) included first rejecting one incorrect assumption, then accepting another incorrect assumption. Also, all performances included a task definition, and all but one of these task definitions included elaboration. Verbalized move decisions, in turn, were of somewhat poorer quality than in Category III. Thus, it seems that *careful initial task representation combined with functional anticipation* were the defining features of correct performances in this category, whereas the incorrect performances lacked the latter. Regulation in terms of evaluating the moves and/or answer line, was not verbalized.

Three of the correct performances included the Combination search method, whereas the other two seemed to be based on Climbing. In both of the Climbing performances (both by Grade 8 participants), the participants first assumed an incorrect disk as correct on Line 2/1 (on Item 5) or Line 3/1 (on Item 6), then rejected this assumption. This indicates that the participants were simultaneously considering at least two disks at the beginning of the item without verbalizing it. Therefore, Climbing may misrepresent part of the decision making process for these participants. Out of all the performances in our data, the five correct performances in this category were the only ones that may have been using a more "global" planning approach based on veridical and sufficient representation. This is also indicated by relatively long first-move latencies among these performances.

One incorrect performance started in an exactly similar manner as the correct performances but, in this case, the participant in question accepted his incorrect assumption and ended up having two disks correct on Line 3/1 on Item 6 (relational constraint violation). This performance was coded as Climbing. The remaining two incorrect performances seemed to be based on the Pattern search method, indicating that the initial representation still included gaps. Thus, the "global" planning in these performances was probably based on insufficient representation.

Performances longer than 4 moves included at least one incomplete move, indicating anticipation, or one change, indicating on-line evaluation of a move or partial answer. This was also evident in verbal protocols: Only six performances longer than four moves included no on-line evaluation statements (evaluation during the performance). For this reason, performances longer than four moves were not grouped together with performances that included only four moves. However, Category V performances still did not include any verbalized evaluation statements and were more similar with Category IV performances than with the other performances with more than four moves.

V. *Careful starters with delayed anticipation.* In five of the six performances with no verbalized evaluation statements the extra move was an incomplete move (either second or third move) indicating that these participants may have engaged in anticipation. In one other performance anticipation was evident in the form of a verbalized evaluation of an assumption. No other on-line evaluation statements were verbalized, although three performances ended with an evaluation of an answer line. Thus, the six performances in this category differed from Category IV performances mainly in terms of having the evaluation of assumption after one or two moves rather than at the beginning of the performance. Most performances in this category included task definitions, only one of which included no elaboration, and at least some Data and/or Valid/Link move decisions. All three correct performances, as well as two of the three incorrect performances, in this category started with long task definition and successful use of the Climbing search method. Both incorrect performances seemed to subsequently switch from the Climbing to the Pattern search method, indicating that the participants were not able to modify their task representations adequately after the first changes. The third failed performance was interesting in that all move decisions were deliberate (about 20 seconds or more for each) and of high quality, yet failed to produce a correct answer. This performance started with a nonoptimal assumption that the participant accurately tested against the two zero correct lines. All the other disks were then placed based on avoiding zero correct lines (negative single constraint violations) but the last move produced a second matching placement with Line 2/1 (negative relational constraint violation). Thus, *insufficient regulation skills* combined with the *inability to modify task representation on-line* may best explain failures in this category. Success, in turn, seemed to be based on *good initial task representation and careful local planning.*

VI. *Insufficient representation, anticipation, and regulation.* Seven performances in this category (three Grade 4, three Grade 6, and one Grade 8) all failed. None of these performances included any task definition statements. Most placements were based on No Data move decisions, although three performances also included some Data or Valid/Link move decisions. Two of these performances started with good move decisions but then switched to using No Data decisions. Moreover, three moves changed a correctly placed disk into an incorrect place and another three changed a disk from one incorrect place to another incorrect place further validating the conclusion that move decisions were not of good quality in these performances. All performances in this category included at least one evaluation statement. These verbalized evaluations, however, were as likely to be incorrect as correct. Also, five moves removed correctly placed disks whereas four removed incorrectly placed disks. Search methods reflected the lack of information generated: Six performances were coded as being based on Climbing together with one of the nonfunctional search methods (False Pattern, Pattern, or Trial-and-Error), whereas the seventh was coded as being based only on Climbing. In sum, these performances were very similar to those in Categories I and II with *nonfunctional regulation and/or anticipation* components added.

VII. *Insufficient representation, anticipation, and regulation 2.* Thirteen performances, all unsuccessful, matched the description in this category; Four were produced by Grade 4 participants, six by Grade 6 participants, and three by Grade 8 participants. These performances differed from those in Category VI mainly in that Category VII performances all included a *task definition* as well. The verbalized task definitions, however, were not of high enough quality to guide the planning process to a successful solution. Seven task definitions included no elaborative statements, one task definition consisted of a single assumption statement, and the remaining five included only one statement that attempted to limit the search space. As above, most placements of disks were based on No Data move decisions and verbalized evaluations were as likely to be incorrect as correct. Thus, the defining feature of performances in this category was that all the components necessary for a well-rounded performance seemed to be present but the quality of these components was not sufficient to successfully complete the items.

Most performances in this category were based on either Climbing or a combination of Climbing and nonfunctional search methods. Three participants seemed to use Climbing together with limited Combinations that were only based on zero correct lines and on avoiding negative single constraint violations.

VIII. *Inflexible task representation.* Seven performances (four by Grade 8 participants and three by Grade 4 participants) in this category all seemed to deteriorate as they proceeded. Task definitions were either nonexistent or very brief, but all seven performances started with good move decisions (DISK/DATA2 or DISK/VALID). Moreover, verbalized evaluations of moves or first answer line attempts were also successful. The participants did not, however, seem to be capable of modifying their existing task representations on the basis of this feedback; instead, the move decisions that followed were of poorer quality (No Data) and all performances resulted in an incorrect answer. In other words, these participants seemed to get successively more confused when their high quality initial local plans did not result in correct answer. Thus, *participants in this category did not seem to be capable of successively refining their task representations.*

Two Grade 8 performances also included comments that acknowledged that the process was not going quite right. In terms of detected search methods, most of these performances seemed to start with Climbing and end with nonfunctional search methods.

IX. *Good regulation and accurate goal representation.* Thirteen performances in this category (seven by Grade 8 participants, four by Grade 6 participants, and two by Grade 4 participants) were all correct. The defining feature of these performances was that *on-line regulation was successful and the participants did not accept incorrect answer lines.* Instead, they just kept on trying until the answer was correct which also suggests that the goal representation was well defined. In contrast, task definitions were mostly nonexistent or very brief and most moves were based on No Data move decisions. Also, several performances in this category had more than ten moves with half of the moves that changed the position of a disk resulting in incorrect placement, indicating a considerable amount of "trial-and-error" behavior. This conclusion was supported by the analysis of search methods: although all of these performances were correct, eight of the twelve performances were coded as including at least one instance of using nonfunctional search methods (Trial-and-Error, Pattern, or False Pattern).

All performances in this category started with the Climbing search method. Nine performances were coded as ending with Combination, although only two performances included mentioning two disks in a one move decision. The moves that were coded as Combination followed either a peak in decision times or an incomplete move during which the participants seemed to gain insight into the problem. Accompanied verbalizations, however, were mostly very brief and only one simultaneously mentioned two disks. Thus, Combination may be an overestimation of the actual search methods used. The remaining four performances were coded mostly as Climbing, which is a functional method when used as by these participants: Combined with continuous on-line regulation and correction of detected mistakes.

X. *Good task and goal representation and excellent regulation.* This category included 11 performances, seven of which were produced by Grade 8 participants and four by Grade 6 participants. One Grade 8 performance was incorrect when the participant ran out of time (her last statement indicated that the answer line was incorrect at that point). Successful performance in this category seemed to be based on two components: *Good move decisions and flawless evaluations.* Most placements of disks were based on Data or Valid/Link decisions and all evaluation statements were correct. Moreover, all but one performance included more than one evaluation statement. This indicates that both on-line task representation and goal representation were accurate. Initial task representations, in contrast, did not seem to contribute to the successes: Four performances included no verbalized task definition, three performances included a task definition with no elaborative statements, and, finally, four performances included a verbalized task definition with one selective combination statement.

All but one performance in this category started with the Climbing search method; four ended with Combination in a similar manner as some Category IX performances. Thus, participants in this category were using *Climbing appropriately and combined either with a more functional search method and/or fully functional regulation.* In other words, if Category III performances seemed to also require luck to succeed with Climbing, performances in this category substituted luck with regulation.

XI. *Careful approach.* The last group was formed by those six performances (one Grade 4 and Grade 8 participant and three Grade 6 participants) that included long and detailed task definitions. The shortest task definition in this category included four statements, all of which limited the search space by selectively combining information. This careful approach was also evident in good move decisions (most placements were based on Data decisions) and frequent use of evaluation. The careful approach was not, however, necessarily the best possible for this task: four of the six performances in this category failed to produce a correct answer within the three minute time limit. The average first move decision time for these four performances was about 60 seconds (range 42.7 to 82.3). In other words, they had spend a third of the time allowed before making any moves. Also, none of these performances were coded as using the Combination search method; instead, the participants seemed to decide carefully over one move at a time (Climbing).

Developmental Differences

Figure 4-5 summarizes the participants performances in four categories based on computer protocols. The bars represent the percentage of performances at each grade level that fell into each of the four groups. The first group consists of those performances that failed in four moves; the second group consists of performances that failed with more than four moves; the third group consists of performances that produced a correct answer with more than four moves; and, finally, the fourth group consists of performances that produced a correct answer in four moves.

The data in Figure 4-5 clearly fits "the overlapping wave" model of development suggested by Siegler (1995), with the exception of zero entries from Grade 8 in the first group. As expected on the basis of Tables 4-2 and 4-7, Grade 4 is overrepresented in the first group with more than half of their performances ending with an incorrect answer with only four moves. In terms of the planning approaches, all but one Grade 4 performance in this group belonged either to Category I or Category II. Out of the four planning components, only representation and execution were present on these performances, and only the latter of the two was functional. In terms of search methods, these were mostly performances that either used Climbing inappropriately (i.e., not supplemented with anticipation or regulation), or used Climbing inappropriately and together with nonfunctional search methods. Thus, about half of all the performances produced by Grade 4 participants and a third of all the performances produced by Grade 6 participants could be classified as using insufficient search methods in a limited search space. One Grade 4 participant and two Grade 6 participants produced incorrect performances that were classified above in Category IV. These three performances did include anticipation that was unsuccessful. None of the Grade 8 participants produced an incorrect answer with four moves.

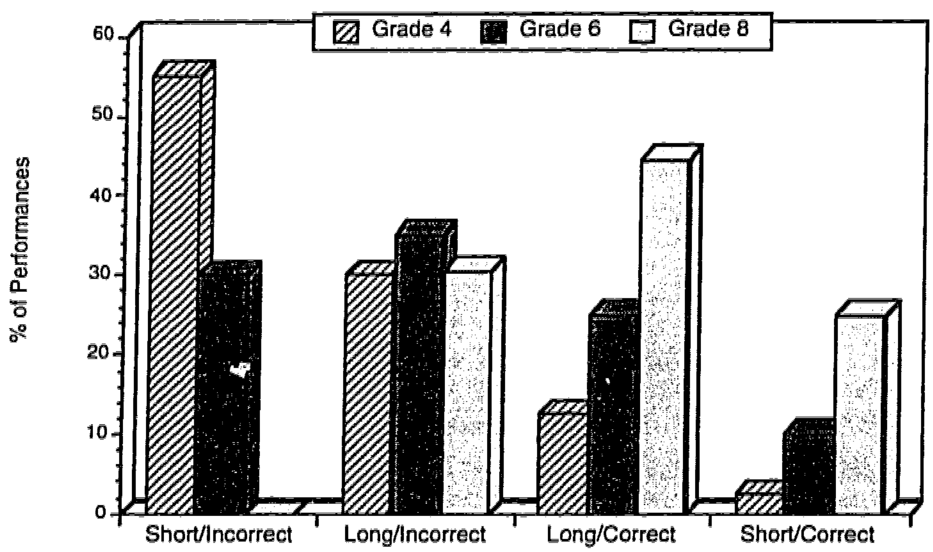


Figure 4-5. The percentages of short incorrect, long incorrect, long correct, and short correct performances in the three grade levels (based on computer protocols).

About one third of the performances from each of the three grade levels produced an incorrect answer with more than five moves. For Grade 6, this was the highest frequency in any of the groups. The "long incorrect" performances came mainly from Categories VI, VII, and VIII and were described above as including all or most of the planning components. However, either the components, or the interaction between them necessary for successful recursive planning, was insufficient. Most of these performances were coded as using either Climbing or Climbing together with nonfunctional search methods. One Grade 4, five Grade 6, and three Grade 8 performances in this group came from categories in which most performances succeeded (Categories V, IX, X, and XI). The most common reason for failures in these cases was the time violation.

The next group consisted of long correct performances. These performances belonged mostly to Categories V, IX, X, or XI and were described above as including, at the minimum, good regulation and accurate goal representation. This was the most common way of producing a correct answer in all grade levels. For Grade 8, this was also the most typical performance. Five correct Grade 4 performances were long whereas only one was successful with four moves. In Grade 6, 12 out of 16 short performances failed, whereas the longer performances were considerably more successful with 10 out of 24 being correct. This suggests that careful advance planning and anticipation needed to successfully complete the items with four moves was still out of reach for most Grade 6 participants. Generating more information and opting for a more careful stepwise approach seemed to be the most successful approach for Grade 6 and, to a lesser extent, also for Grade 4.

Eight Grade 8, four Grade 6, and 1 Grade 4 performance produced a correct answer with four moves. In terms of planning approaches, all but one of these performances came from Categories III and IV and seemed to be mostly based on careful task representation and possibly anticipation. Search method analysis suggested that either these performances utilized Climbing and generated rich information for each step (careful local planning), or they utilized Combination and a more global planning approach.

Thus, the majority of Grade 4 protocols were short and unsuccessful, lacking in representation, anticipation, and regulation. Moreover, when some or all of these components were present, they were either not sufficient to guide the planning process, or their coordination failed. The majority of Grade 6 performances, in contrast, were long. This was also an adaptive change for Grade 6: Most of the short Grade 6 performances failed whereas about 40% of the longer performances by Grade 6 participants produced a correct answer. No Grade 8 participant failed with only four moves. In other words, Grade 8 participants terminated their performance after four moves only if they had a correct answer. When Grade 8 participants failed, they seemed to do so because successive refinement of their answer lines either took too long or they were not successful in modifying their task representations on-line.

Conclusions

This was the first study to explore both the Crack-the-Code task and the planning process model in detail. Due to the exploratory nature of this study, results were presented in great length and detail in order to encourage readers to make their own judgment of the analyses methods, the planning process model, and the task. Crack-the-Code is a complex planning task that is open to several different approaches. Both of the more generally referred to planning approaches — opportunistic planning and advance planning — were witnessed, although the latter did not seem to play as prominent a role as expected. The fact that we did expect to see more advance planning may be an artifact of many studies using tasks such as Tower of Hanoi that particularly promote this approach to the exclusion of other possible approaches.

We believe that Crack-the-Code is a good indicator of planning differences, both individual and developmental, during the later school years. Grade 4 participants experienced considerable difficulties in solving all but the two easiest items and showed the least variability in their performances. In contrast, rich data were obtained from both within and between grade level differences for Grades 6 and 8.

Large standard deviations were more the norm than the exception with the time measures that were used in this study. As a consequence, many of these measures failed to show significant differences between the groups. Averaging data over different planning approaches, as was done in these analyses, probably hid meaningful between-group differences.

Error analyses indicated that Grade 4 participants seemed to be behind both of the two older grade levels in representation and/or anticipation skills as well as in regulation skills, whereas the difference between Grades 6 and 8 was more pronounced in regulation skills. The biggest difference between Grades 4 and 6 was in the number of errors produced, suggesting that the younger participants' capacity to construct an accurate representation of the items and to anticipate the consequences of the moves were more limited. In contrast, the largest difference between Grades 6 and 8 was in the number of errors corrected, suggesting that lack of (or inappropriate) evaluation may explain why Grade 6 participants solved fewer items correctly.

While the analyses of search methods produced less significant between-group differences than one would have expected, verbal and computer protocols provided rich information on both individual and developmental differences, as is evident in both the planning approaches and in Figure 5. With the exception of the execution component, which did not play a major role in planning approaches, the nature of these differences was adequately captured by the four planning components, particularly if the representation component was further divided into task representation, search methods, and goal representation.

In sum, both the task and the planning process model hold promise in advancing planning research. In future research, we need to include both older participants as well as different diagnostic groups. For example, it would be interesting to examine how and if aging affects the four planning components differently, or if they could be used to differentiate between gifted and regular students. Analysis of verbal protocols, however, is considerably labor intensive and, as such, cannot be recommended for general use. What remains to be done is to develop a more fine grained method of analyzing the computer protocols and relating them directly to the planning components.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48,* 35-44.

Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York: Freeman.

Baker-Sennett, J., Matusov, E., & Rogoff, B. (1993). Planning as developmental process. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 24, pp. 253-281). San Diego: Academic Press.

Becker, M. G., Isaac, W., & Hynd, G. W. (1987). Neuropsychological development of nonverbal behaviors attributed to "frontal lobe" functioning. *Developmental Neuropsychology, 3,* 275-298.

Bidell, T. R., & Fischer, K. W. (1994). Developmental transitions in children's on-line planning. In M. M. Haith, J. B. Benson, R. J. Roberts Jr., & B. F. Pennington (Eds.), *The developmental of future-oriented processes* (pp. 141-176). Chicago: University of Chicago Press.

Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology, 33,* 87-110.

Casey, M. B., Bronson, M. B., Tivnan, T., Riley, E., & Spenciner, L. (1991). Differentiating preschoolers' sequential planning ability from their general intelligence:

A study of organization, systematic responding, and efficiency in young children. *Journal of Applied Developmental Psychology, 12,* 19-32.

Cocking, R. R., & Copple, C. E. (1987). Social influences on representational awareness: plans for representing and plans as representation. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 428-465). New York: Cambridge University Press.

Crowley, K., & Siegler, R. S. (1993). Flexible strategy use in young children's tic-tac-toe. *Cognitive Science, 17,* 531-561.

Das, J. P., Mensink, D., & Janzen, H. (1990). The K-ABC, coding, and planning: an investigation of cognitive processes. *Journal of School Psychology, 28,* 1-11.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes.* Needham Heights, MA: Allyn and Bacon.

Davidson, J., & Sternberg, R. J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly, 28,* 58-64.

De Lisi, R. (1987). A cognitive-developmental model of planning. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 79-109). New York: Cambridge University Press.

Diamond, A., & Goldman-Rakic, P S. (1986). Comparative development in human infants and infant rhesus monkeys of cognitive functions that depend on prefrontal cortex. *Neuroscience Abstracts, 12,* 742.

Diamond, A., & Goldman-Rakic, P. S. (1983). Comparison of performance on a Piagetian object permanence task in human infants and rhesus monkeys: Evidence for involvement of prefrontal cortex. *Neuroscience Abstracts, 9,* 641.

Diamond, A., & Goldman-Rakic, P. S. (1985). Evidence for involvement of prefrontal cortex in cognitive changes during the first year of life: Comparison of human infants and rhesus monkeys on a detour task with transparent barrier. *Neuroscience Abstracts, 11,* 832.

Díaz, R. M., Neal, C. J., & Amaya-Williams, M. (1990). The social origins of self-regulation. In L. C. Moll (Ed.), *Vygotsky and education* (pp. 127-154). New York: Cambridge University Press.

Dreher, M., & Oerter, R. (1987). Action planning competencies during adolescence and early adulthood. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 321-355). New York: Cambridge University Press.

Ellis, S., & Siegler, R. S. (1994). Development of problem solving. In R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 333-367). San Diego: Academic Press.

Fabricius, W. V. (1988). The development of forward search planning in preschoolers.

Child Development, 59, 1473-1488.

Friedman, S. L., Scholnick, E. K., & Cocking, R. R. (1987). Reflections on reflections: What planning is and how it develops. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 515-534). New York: Cambridge University Press.

Gardner, W., & Rogoff, B. (1990). Children's deliberateness of planning according to task circumstances. *Developmental Psychology, 26,* 480-487.

Gauvain, M. (1989). Children's planning in social contexts: An observational study of Kindergarteners' planning in the classroom. In L. T. Winegar (Ed.), *Social interaction and the development of children's understanding* (pp. 95-117). Norwood, NJ: Ablex.

Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25,* 139-151.

Göncü, A., & Kessel, F. (1988). Preschoolers' collaborative construction in planning and maintaining imaginative play. *International Journal of Behavioral Development, 11,* 327-344.

Goodnow, J. J. (1987). Social aspects of planning. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 179-201). New York: Cambridge University Press.

Haake, R. J., Somerville, S. C., & Wellman, H. M. (1980). Logical ability of young children in searching a large-scale environment. *Child Development, 51,* 1299-1302.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3,* 275-310.

Hudson, J. A., & Fivush, R. (1991). Planning in the preschool years: The emergence of plans from general event knowledge. *Cognitive Development, 6,* 393-415.

Karmiloff-Smith, A. (1992). *Beyond modularity.* Cambridge, MA: The MIT Press.

Klahr, D. (1978). Goal formation, planning, and learning by pre-school problem solvers or: "My socks are in the dryer". In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 181-212). Hillsdale, NJ: Erlbaum.

Klahr, D. (1985). Solving problems with ambiguous subgoal ordering: Preschoolers' performance. *Child Development, 56,* 940-952.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology, 13,* 113-148.

Kirk, U., & Kelly, M. S. (1986). Children's differential performance on selected dorsolateral prefrontal and posterior cortical functions: A developmental perspective. *Journal of Clinical and Experimental Neuropsychology, 7,* 604.

Kreitler, S., & Kreitler, H. (1987). Conceptions and processes of planning: The developmental perspective. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 205-272). New York: Cambridge University Press.

Lawson, M. J., & Chinnappan, M. (1994). Generative activity during geometry problem solving: Comparison of the performance of high-achieving and low-achieving high school students. *Cognition and Instruction, 12,* 61-93.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124,* 83-97.

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H., Ringholz, G., Ewing-Cobbs, L., & Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 6,* 377-395.

Levina, R. E. (1981). L. S. Vygotsky's ideas about the planning function of speech in children. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 279-299). New York: M. E. Sharpe.

Luria, A. R. (1973). *The working brain.* New York: Basic Books.

McGilly, K., & Siegler, R. S. (1989). How children choose among serial recall strategies. *Child Development, 60,* 172-182.

Miller, G. A., Galanter, E. H., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehert & Wilson.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on general problem-solving program. Proceedings of the International Conference on Information Processing, Paris.

Parrila, R. K. (1995). Vygotskian views on language and planning in children. *School Psychology International, 16,* 167-183.

Parrila, R. K., & Papadopoulos, T. C. (1994). Developmental Aspects of Planning. Paper presented at the CSSE conference in Calgary, Canada.

Parrila, R. K., Äystö, S., & Das, J. P. (1994). Development of planning in relation to age, attention, simultaneous, and successive processing. *Journal of Psychoeducational Assessment, 12,* 212-227.

Parrila, R. K., Das, J. P., & Dash, U. N. (in press). Development of planning and its relation to other cognitive processes. *Journal of Applied Developmental Psychology.*

Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behaviors attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 1,* 349-370.

Pea, R. D., & Hawkins, J. (1987). Planning in the chore-scheduling task. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 273-302). New York: Cambridge University Press.

Piaget, J. (1963). *The origins of intelligence in children.* New York: Norton.

Pitt, R. B. (1983). Development of a general problem-solving scheme in adolescence and early adulthood. *Journal of Experimental Psychology: General, 112,* 547-584.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102-119.

Scholnick, E. K., & Friedman, S. L. (1987). The planning construct in the psychological literature. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 1-38). New York: Cambridge University Press.

Scholnick, E. K., & Friedman, S. L. (1993). Planning in context: Developmental and situational considerations. *International Journal of Behavioral Development, 16,* 145-167.

Schubert, J. (1983). The implications of Luria's theories to cross-cultural research on language and intelligence. In B. Bain (Ed.), *The sociogenesis of language and human conduct* (pp. 59-77). New York: Plenum Press.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8,* 481-520.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for the Research in Child Development, 46* (2, Serial No. 189).

Siegler, R. S. (1991). *Children's thinking* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28,* 225-273.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

Sperry, R. W. (1993). The impact and promise of the cognitive revolution. *American Psychologist, 48,* 878-885.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Viking.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology, 27,* 195-232.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Stuss, D. T., & Benson, D. F. (1987). The frontal lobes and control of cognition and memory. In E. Perecman (Ed.), *Frontal lobes revisited* (pp. 141-158). New York: IRBN Press.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527-579). Cambridge, MA: The MIT Press.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Wellman, H. M., Fabricius, W. V., & Sophian, C. (1985). The early development of planning. In H. M. Wellman (Ed.), *Children's searching* (pp. 123-149). Hillsdale, NJ: Erlbaum.

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk transfer task. *Cognitive Development, 6,* 59-76.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function. *Developmental Neuropsychology, 7,* 131-149.

Wilkinson, A. C. (1982). Partial knowledge and self-correction: Developmental studies of a quantitative concept. *Developmental Psychology, 18,* 876-893.

Willatts, P. (1990). Development of problem-solving strategies in infancy. In D. F. Bjorklund (Ed.), *Children's strategies* (pp. 23-66). Hillsdale, NJ: Erlbaum.

Willatts, P. (1984). Stages in the development of intentional search by young infants. *Developmental Psychology, 20,* 389-396.

V. CONCLUSIONS

The goal of studying the development of planning is to understand what is it that changes, or develops, over time and what are the factors that make these changes possible. These basic questions are often followed by several more specific ones: When does planning appear in ontogenesis? What factors contribute to its appearance? What factors contribute to its later development? How do changes in planning take place?

Siegler (1994) suggested recently that significant changes in thinking over the life-span are perhaps the most remarkable characteristic of human development. He continued that accounting for these changes should be the central goal of developmental research. This has not, however, been done adequately. Instead, "in most models of cognitive development, children are depicted as thinking or acting in certain way for a prolonged period of time, then undergoing a brief, rather mysterious, transition, and then thinking or acting in a different way for another prolonged period" (Siegler, 1994, p. 1). Siegler could as well have been writing about the development of planning. We have a considerable amount of knowledge about what children should be able to do at certain age levels. We also know something, although much less, about how some organismic and environmental factors may contribute to the development of planning, and about what is it that develops. To make this list complete, we could add that we know almost nothing about how the developmental changes take place. In sum, developmental studies of planning have been biased towards general trends and structural descriptions whereas individual differences and descriptions of processes have been customarily lacking.

Studying change is more difficult than studying general trends or identifying average performance levels in different groups. Study of changes includes all the conceptual and methodological demands of the latter and "imposes the added demands of determining what is changing and how the change is being accomplished" (Siegler, 1994, p. 1). Current methods of inquiry tend to conceal rather than to capture possible mechanisms of change, such as, for example, intraindividual variability and the effect of practice. Nor do they often allow concomitant manipulation of both organismic and environmental variables. According to Fischer and Silvern (1985, p. 643), only research that accounts for both sources of variation "can reveal the full range of plasticity and constraint arising from cognitive levels and environmental contexts." They continue that we need assessment techniques that "allow the detection of individual differences in developmental sequences" and are "sensitive to the possibility that people reason differently, not just more or less maturely."

At this point we should add one more item to the list of unanswered questions: How can we guarantee the best possible outcome of the developmental process? This question is particularly important for educators. While identifying general trends may be essential for designing age-appropriate teaching methods and curriculum, a more detailed study of the mechanisms and sources of variation in development may be necessary for designing successful interventions in special education, especially in the burgeoning field of cognitive remediation.

Thus, the methodological questions that we should ask are: (1) How can we identify both general trends and individual differences in the development of planning? and, (2) How can we manipulate both organismic and environmental variables in the same design? I would like to suggest that both large scale correlational studies and smaller scale experimental and microanalytic studies are needed and can contribute to understanding planning development. Studies presented in this dissertation fill only some of these shortcomings. In what follows, I will first summarize the main findings and discuss what possible educational implications the results may have and what limitations the reported studies had. Each section concludes with some suggestions for future research.

Language and Planning in Children

Chapter II suggested that sufficient language skills are essential for the development of higher level planning skills. Planning would not be possible without some kind of semiotic mediation that enables both the self-regulation and the restructuring of the decision making process that are necessary conditions for planning. Moreover, the acquisition and development of planning skills depends on the social interactions that we enter into and if our tools for communicating are not up to the task, we are unlikely to obtain the complex planning skills from our environment. Similarly, if our psychological tools for thinking are not powerful enough, we are unlikely to come up with the best possible solutions. Although natural language is only one medium of communication and only one source of psychological tools, it nevertheless is the most important medium and source.

Chapter II included a separate section dealing with the educational implications of the presented ideas. The fact that both *School Psychology International* and *Educational Psychologist* have recently celebrated the centennial of Vygotsky's birth by publishing special issues on his ideas is encouraging for educational psychologists. One can only hope that the issues of instructional dialogue and theoretical versus empirical concepts and learning will warrant serious research in the future.

While Vygotsky's ideas are appealing to psychologists, psycholinguists, and educators alike, both the empirical testing of his ideas and evaluation of their practical implications are still in their infancy. Proponents of his approach often select a decontextualized idea from his work that best fit their agenda and then extrapolate on that idea. Seldom are these ideas subjected to rigorous empirical testing, and, to be fair, we should note that many of Vygotsky's original ideas seem to escape simple experimentation. For example, one of the basic premises of Vygotsky's theory is that higher cognitive processes have their origins in the interpersonal relationships that the child enters in his or her environment. In these relationships, the accumulated knowledge of the environment is transmitted to the child by the adults (Luria & Yudovich, 1959). According to this premise, having inadequate language skills would led to insufficient acculturation due to limited access to information, which, in turn, would affect the child's thinking skills. This hypothesis is difficult to address since the pivotal role of language in organizing higher mental processes is another central premise in Vygotsky's theory, and the two possible ways that language affects a child's intellectual functioning are, to say the least, difficult to separate. This is evident in the examples given by Luria and Yudovich (1959). They cite various Soviet studies on deaf-mute subjects and conclude that deaf-mute subjects' complex perceptual processes are clearly underdeveloped because they are excluded from speech communication. Their own study with identical twins, whose language development was retarded, showed that the acquisition of speech lead to new forms of communication with other children and induced significant changes to the structure of the twins' conscious activity which, as a result, changed from the strictly practical to the verbally mediated form. In all of the studies mentioned by Luria and Yudovich, neither the amount nor the level of communication was controlled for and therefore, it is impossible to determine what role new communication skills really played in the twins intellectual development.

Designing studies that focus on the role of language in thinking is inherently tricky and such studies seldom provide conclusive evidence one way or the other. It seems, however, that the "linguistic relativity hypothesis" is far from forgotten and several recent publications have attempted to rework it into a more quantifiable and testable form (see e.g., Hunt & Agnoli, 1991; Lucy, 1992). The premise behind these approaches is that new models of cognition should indicate ways in which thought can be influenced by different aspects of language and that new research methods should be able to capture the differences reliably.

The main limitation of Chapter II is that it also does not provide any new data or suggest any new approach for studying the relationship between planning and language. Some such suggestions are offered below. In Chapter III, we introduced three levels of planning — activity, action, and operation. Activity-planning aims at realizing one's general life-goals and motives, such as career development and planning for a retired life. It probably is more socially determined than the other two levels. Accordingly, this is the level where language skills should have the most profound impact on planning. Firstly, good activity-planning requires a rich knowledge-base of the environment and the forces in it that can affect goal-attainment. Thus, both access to information as well as later processing of that information should affect the feasibility of one's activity-plans. Secondly, good activity-planning requires both understanding of other agents in the environment as well as abilities to affect their functioning. Activity-planning does not take place in a vacuum; instead, other people are frequently employed as coplanners, resource persons, or agents to be manipulated. All these functions require that the planner is able to fluently communicate his or her ideas with others. While activity-planning itself may be outside the realm of experimental methods, we should be able to create complex experimental action-planning situations in which effective communication of one's ideas is essential for reaching the goal. We already have studies available that utilize complex problem solving situations (see e.g., Funke, 1991), studies that examine the effect of social interaction on planning (see e.g., Gauvain & Rogoff, 1989; Radziszewska & Rogoff, 1988), and studies that also examine the effect of the quality of this interaction (see e.g., Radziszewska & Rogoff, 1991; Dimant & Bearison, 1991; Tudge, 1989). What we seem to be lacking is a study that combines all of these factors and obtains information of the participants' language skills as predictors of planning outcomes.

A second possible approach for studying the effect of language skills on planning would be to identify simple action-planning or operation-planning tasks that have similar formal structure but vary in their linguistic content. For example, it may be possible to identify two math problems that require the same factual, procedural, and context specific conceptual knowledge (see e.g., Bisanz & LeFevre, 1990), but differ in terms of how these problems are embedded in different linguistic contexts. If a significant portion of the observed variance in correct answers can be predicted by the participants' language skills, then the "linguistic relativity hypothesis" would be supported.

The assumption that relevant linguistic knowledge predicts success in planning needs to be qualified. Strictly verbal mediation may also have a detrimental effect on performance, as Kearins (1981, 1986) has shown. By using visual spatial memory tasks, Kearins showed that the Australian aboriginal children who showed very little evidence of using verbal mediation and, instead, seemed to rely on visual strategies, consistently outperformed European Australian children who tried to employ strategies grounded on verbal mediation.

Kearins' findings suggest that the role of language in planning should always be considered in relation to the task demands. Planning tasks with more verbal content and of a structure that presupposes, or at least benefits, from verbal mediation, should be affected by language abilities to a greater extent than planning tasks with less verbal content and, for example, more visual structure. Accordingly, Crack-the-Code, Planned Connections, and Matching Numbers (Chapter III) should be affected more by language skills than, for example, Planned Search. Whether these assumptions can be experimentally verified is yet to be seen.

Planning in Relation to Other Cognitive Processes

Larger scale correlational studies can be useful in identifying different components and precursors of planning and planning development. For example, what is the role of such other possible components of successful planning performance as memory, speed, attention, simultaneous, and successive processing in determining the outcome in different ages? We have suggested elsewhere (Parrila, Äystö, & Das, 1994) that planning may not be possible before the child has reached sufficient levels of functioning in the other components of the PASS model. The results presented in Chapter III showed that the significant predictors of planning performance changed both as a function of the task as well as a function of age. Planning is context sensitive and not all planning tasks are expected to make the same encoding and attentional demands. Moreover, development of planning is not only quantitative in terms of older children solving more problems correctly but also qualitative in terms of children in different age groups or within the same age group approaching the tasks in distinctly different manners. While the age related changes in cognitive correlates of Crack-the-Code were perhaps smaller than the changes in other planning tasks (Table 3-4), more detailed analysis of the planning process in Chapter IV identified several distinctly different planning approaches displayed by the participants while solving the items. Similar microanalytic studies with other planning tasks would certainly help us to better understand how planning skills develop qualitatively.

Chapter III also indicated that different laboratory planning tasks had different discriminating values for different age groups. This result suggests that our simpler planning tasks would have limited value in diagnosing differences in planning skills during the later school years. They may, however, still retain their clinical usefulness as identifiers of gross planning deficits, an issue that was not addressed in Chapter III. Vast literature on the effects of frontal lobe lesions on cognitive functioning supports this point of view: many neuropsychological tasks used to assess frontal lobe functioning would be best described as simple operation-planning tasks and yet they have proved useful in diagnosing frontal lobe damage. They are less useful, however, for the assessment of planning in normal populations, particularly when used with adolescents and older participants. As a consequence, we do not currently seem to have available standardized measures of planning skills that would capture important developmental and individual differences in adolescence and beyond. Perhaps this is the reason why researchers interested in planning development have mostly chosen not to include adolescents or adults in their studies.

Studying the later development of planning skills in general and the learning of planning skills in particular is, however, important. More difficult tasks, more complex environmental constraints, especially when adolescents enter work life that is governed by planning (Dreher & Oerter, 1987), require increasingly sophisticated plans that they have to learn with the help of adults and more competent peers. Peers and adults act as models who engage in planning and demonstrate both the usefulness of planning as well as the correct planning procedures for different situations. Through social interaction with (an ever broadening) environment, an adolescent learns what kind of plans work in different situations and why. Her or his planning repertoire both grows and gets sharper. Whether this development results only from building up a large repertoire of discrete skills that are not transferable to new situations and contexts (e.g., Radziszewska & Rogoff, 1988), or also affects planning as a general higher order cognitive skill or ability that can be used across various problem situations, is an interesting question. In order to answer this question, we need to utilize complex planning tasks that do not build on prior knowledge. Figure 3-1 indicates that the Crack-the-Code task may have real potential in capturing this later development.

The third major finding in Chapter III related to the uniformity of the planning construct. As expected, planning tasks did not load on one factor and the only task that was defined as an action-planning task, Crack-the-Code, seemed to have a considerable distance from the other planning tasks. Thus, our conceptualization of action- and operation-planning was partially supported. The limitations of this conclusion are obvious: the study included only one action-planning task and showing that it did not load on the same factor with our other planning tasks does not provide conclusive evidence of the multifaceted structure of cognitive planning. This finding needs to be replicated preferably with other experimental tasks that can be defined a priori as action-planning and operation-planning tasks. If the differentiation of planning levels is successful, then several other related questions follow: Can activity-planning, action-planning, and operation-planning be distinguished in all age and ability groups? If so, do they develop concurrently? More importantly, what is their relationship to such real-life measuring sticks as educational success or career development? We may hypothesize, for example, that while action-planning and operation-planning may be good indicators of early educational success, activity-planning, due to its nonconstraint nature, may be a better indicator of, for example, political or managerial skills and success.

At the same time, the three levels of planning as well as their relation to each other has to be defined in more detail than what was done in Chapter III. For example, questions such as what are the general factors that make a particular cognitive activity planning in the first place, and what are the specific factors that then differentiate between the three levels of planning, have to addressed. While some of the answers are provided in Chapter III, we should regard these only as guidelines for future theoretical and empirical work. That the three levels of planning make intuitive sense is not enough to substantiate their value in understanding existing planning research or in guiding future research. Before we can say anything more conclusive, a thorough review of both psychological and educational literature on such issues as development of strategies, problem solving skills, future orientation, life-planning skills etc. is required.

Developmental Changes in the Planning Process

Identifying the mechanisms of change — both at the microgenetic and ontogenetic levels — requires methods that focus on processes and qualitative changes in performances and manipulate both organismic and environmental factors. Chapter IV presented a study that attempted to detect both between-groups and within-groups differences in planning processes by analyzing verbal and computer protocols of participants from three different grade levels. Although the more traditional performance measures mainly failed to show significant differences due to large standard deviations and small number of participants, rich qualitative data enabled us to identify several different planning approaches that captured some of the within-group differences in how the participants solved the Crack-the-Code task. Moreover, a clear developmental picture emerged suggesting a progression from short unsuccessful performances to short successful performances via a period of practice with longer and more variable performances. With the exception of the execution component, which did not play a major role in planning approaches, the nature of these differences was adequately captured by the four planning components proposed in Chapter IV.

A major limitation of this study was that it only used 29 participants. While analyzing verbal protocols from a larger sample would probably be too time consuming, error analyses and analyses of "product measures" (accuracy, solution speed, move latencies, move durations, number of moves, and evaluation time) could be replicated with a larger sample resulting in more powerful statistical tests. What has to be done first is to develop a more refined method of analyzing the computer protocols and relating them directly to representation, anticipation, execution, and regulation. In future research, we also need to include both older participants as well as different diagnostic groups. For example, it would be interesting to examine if aging affects the four planning components differently, or if they could be used to differentiate between gifted and regular students. Moreover, it would be interesting to examine whether the more "generic" planning process model could be used to reinterpret process analyses of science and mathematics problem solving. One striking feature of science and mathematics problem solving literature is the lack of common explanatory models. Instead, even the concepts used to explain, for example, scientific reasoning (see e.g., Schauble, 1996) seem to be different from those used to explain geometry problem solving (Lawson & Chinnappan, 1994). Consequently, the educational implications of the findings tend to be equally specific. We do not seem to have a curriculum for teaching problem solving and planning in general as much as we have a curriculum for teaching specifics of solving one kind of math problem or one kind of science problem. This relates to the issue of empirical and theoretical learning raised in Chapter II. The current methods of teaching problem solving are essentially parallel to the empirical learning approach and, at least according to Detterman (1993), are notoriously poor in producing transfer. However, rather than teaching more content, as Detterman seems to suggest, maybe what we need is a more theoretical approach concentrating specifically on teaching the principles of problem solving and planning per se.

Chapter IV suggested that the Crack-the-Code task is a good indicator of developmental differences in planning. I believe that Crack-the-Code and protocol analysis could also be used to examine the role of variability in learning and development. Variability exists in children's thinking at every level: between different age and ability groups, between two children in the same group, within a child solving similar problems, or even within a child solving the same problem twice (Siegler, 1994). Developmental psychologists have generally concentrated on explaining the first type of variability, that which exists between two groups of individuals. The other forms of variability have been largely neglected since they pose problems for stage theories of human development that have dominated developmental theorizing. How do we define a stage if variability within that stage is more of a norm than an exception? While some theorist (see e.g., Fischer & Silvern, 1985) have tried to address this question by proposing less stringent definitions of stages, most have chosen to downplay the importance of these forms of variability and to treat them as a nuisance rather than as an important part of cognitive development (Siegler, 1994).

Nevertheless, "detailed analyses of tasks on which one-to-one correspondences between age and way of thinking have been postulated indicate that children's thinking is generally much more variable than past depictions have suggested" (Siegler, 1994, p. 2). When individual performances are tracked carefully we can find great variability in means to the same end. Yet conventional study designs have tended to compare groups of children at different ages and to ignore the individual pathways children follow to get to the goal (Thelen, 1992). This negligence is even more serious in light of the recent findings suggesting that variability is not just an incidental feature of thinking but contributes directly to cognitive development and learning (Siegler, 1994; Thelen, 1992).

If, for example, within-subject variability is the largest immediately before and during learning a new strategy, as suggested by Siegler (1994), then it should be a useful index of a student's zone of proximal development (Vygotsky, 1978) within which an intervention should benefit the student most. But variability also implies that what is good for one child is not necessarily good for another. Thus, identifying between-subjects variability in their preferred mode of processing, for example, is as essential for successful interventions as within-subject variability.

The Crack-the-Code task and protocol analysis methods developed for Chapter IV can identify both general trends and individual differences in planning. Moreover, by using, for example, four different items that are formally similar we should also be able identify different learning curves and "pathways," as well as verify the role of variability as a contributor to learning. In the current study, two items were used in each of the difficulty levels and we did not witness learning in terms of more participants solving the later item correctly. However, it is likely that if we add more items in each level and provide feedback on the accuracy of the solution, such learning would take place. Results in Chapter IV also allow us to make specific hypotheses about what changes we may expect to witness. With most Grade 4 participants, for example, we could expect such learning to mean that they first add more components into their short and insufficient performances. Later, probably with practice, these components would also become functional and result in correct answers.

In a different context (computational models of cognitive development), Klahr (1995, p. 356) recently remarked that "the challenge is not to construct performance systems that can adapt, but rather to construct adaptive systems that can perform." Expanding on this idea, the challenge for future psychological studies on planning development is to understand how the continuous changes and self-modifications create adaptive responses to the environment, as well as modify the environment to fit our plans. The challenge for future educational studies, then, is to develop methods on how to effectively teach the student to adapt as well as act on the environment.

References

Bisanz, J., & LeFevre, J. (1990). Strategic and nonstrategic processing in the development of mathematical cognition. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 213-244). Hillsdale, NJ: Erlbaum.

Dimant, R. J., & Bearison, D. J. (1991). Development of formal reasoning during successive peer interactions. *Development Psychology, 27,* 277-284.

Dreher, M., & Oerter, R. (1987). Action planning competencies during adolescence and early adulthood. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for thinking* (pp. 321-355). New York: Cambridge University Press.

Fischer, K. W., & Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology, 36,* 613-648.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg, & P. A. French (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Erlbaum.

Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25,* 139-151.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review, 98,* 377-389.

Kearins, J. M. (1981). Visual spatial memory in Australian Aboriginal children of desert regions. *Cognitive Psychology, 13,* 434-460.

Kearins, J. M. (1986). Visual spatial memory in Aboriginal and White Australian children. *Australian Journal of Psychology, 38,* 203-214.

Klahr, D. (1995). Computational models of cognitive change: The state of the art. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence* (pp. 355-375). Hillsdale, NJ: Erlbaum.

Lawson, M. J., & Chinnappan, M. (1994). Generative activity during geometry problem solving: Comparison of the performance of high-achieving and low-achieving high school students. *Cognition and Instruction, 12,* 61-93.

Lucy, J. A. (1992). *Language and diversity of thought.* Cambridge: Cambridge University Press.

Luria, A. R., & Yudovich, F. I. (1959). *Speech and the development of mental processes in the child.* London: Staples Press.

Parrila, R. K., Äystö, S., & Das, J. P. (1994). Development of planning in relation to age, attention, simultaneous, and successive processing. *Journal of Psychoeducational*

Assessment, 12, 212-227.

Radziszewska, B., & Rogoff, B. (1991). Children's guided participation in planning imaginary errands with skilled adult or peer partners. *Developmental Psychology, 27,* 381-389.

Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaborators on children's planning skills. *Developmental Psychology, 24,* 840-848.

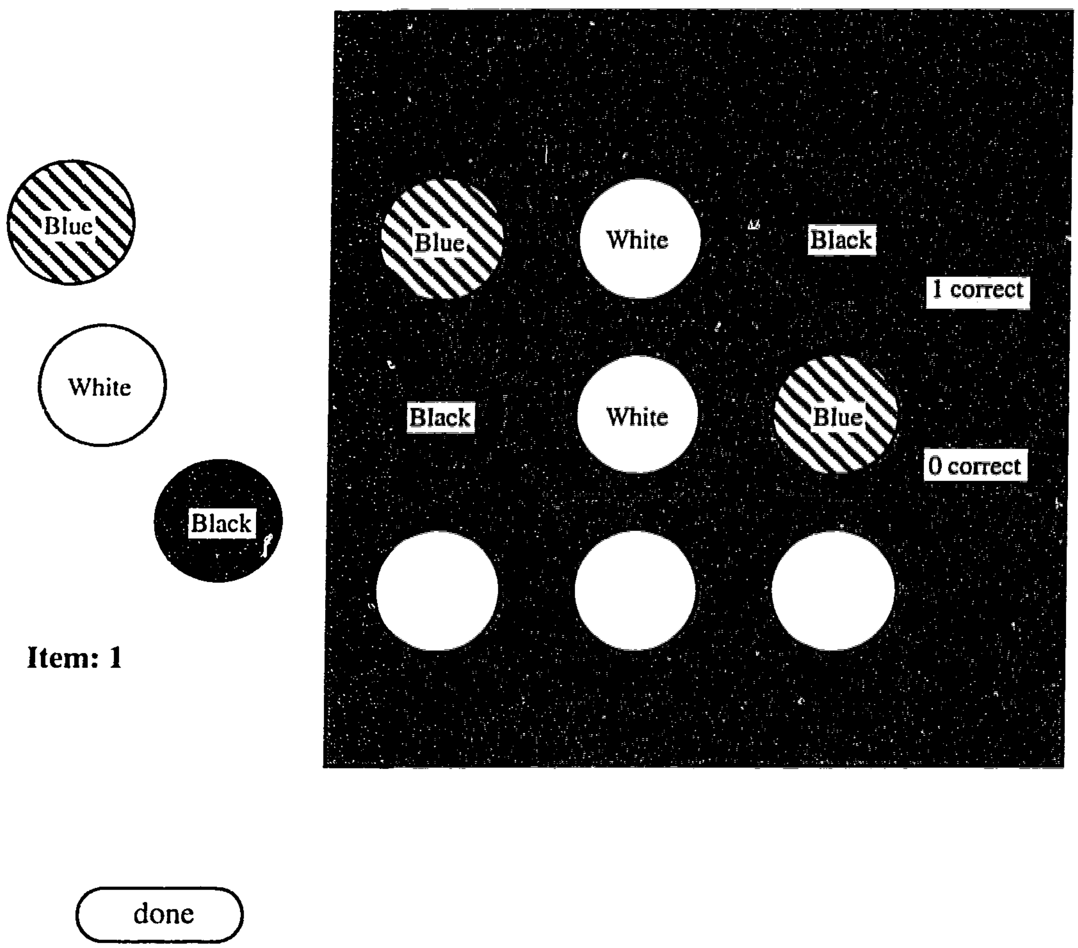
Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102-119.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science, 3,* 1-5.

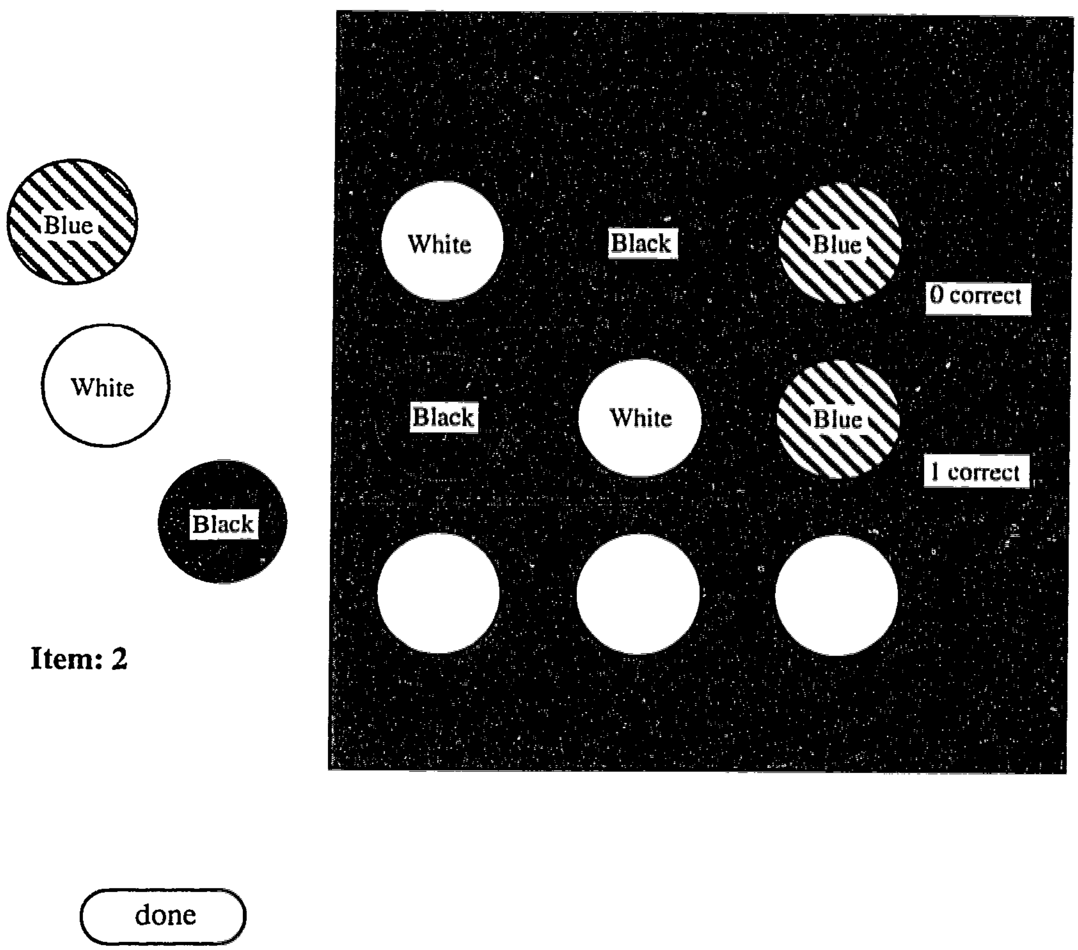
Thelen, E. (1992). Development as a dynamic system. *Current Directions in Psychological Science, 1,* 189-193.

Tudge, J. (1989). When collaboration leads to regression: Some negative consequences of socio-cognitive conflict. *European Journal of Social Psychology, 19,* 123-138.

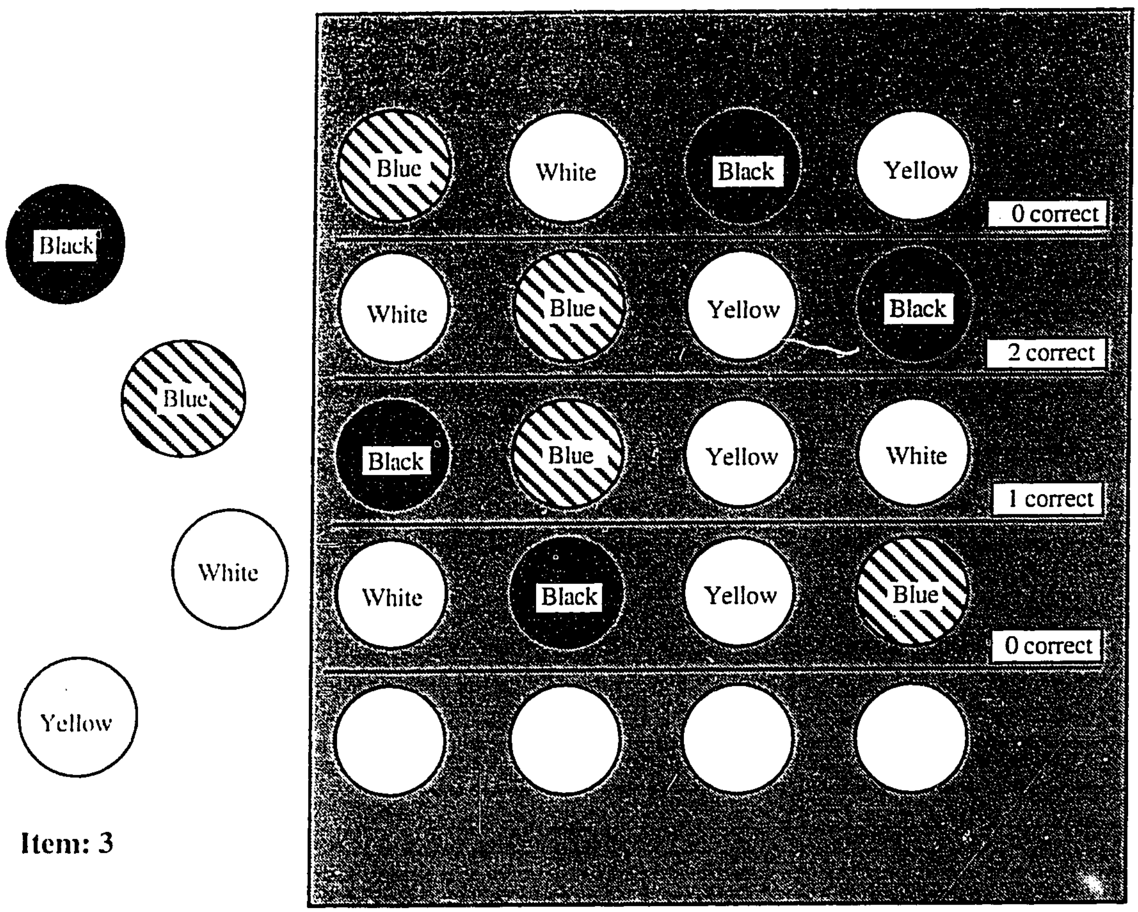
APPENDIX A
The Crack-the-Code Items



ITEM 1 of Crack-the-Code



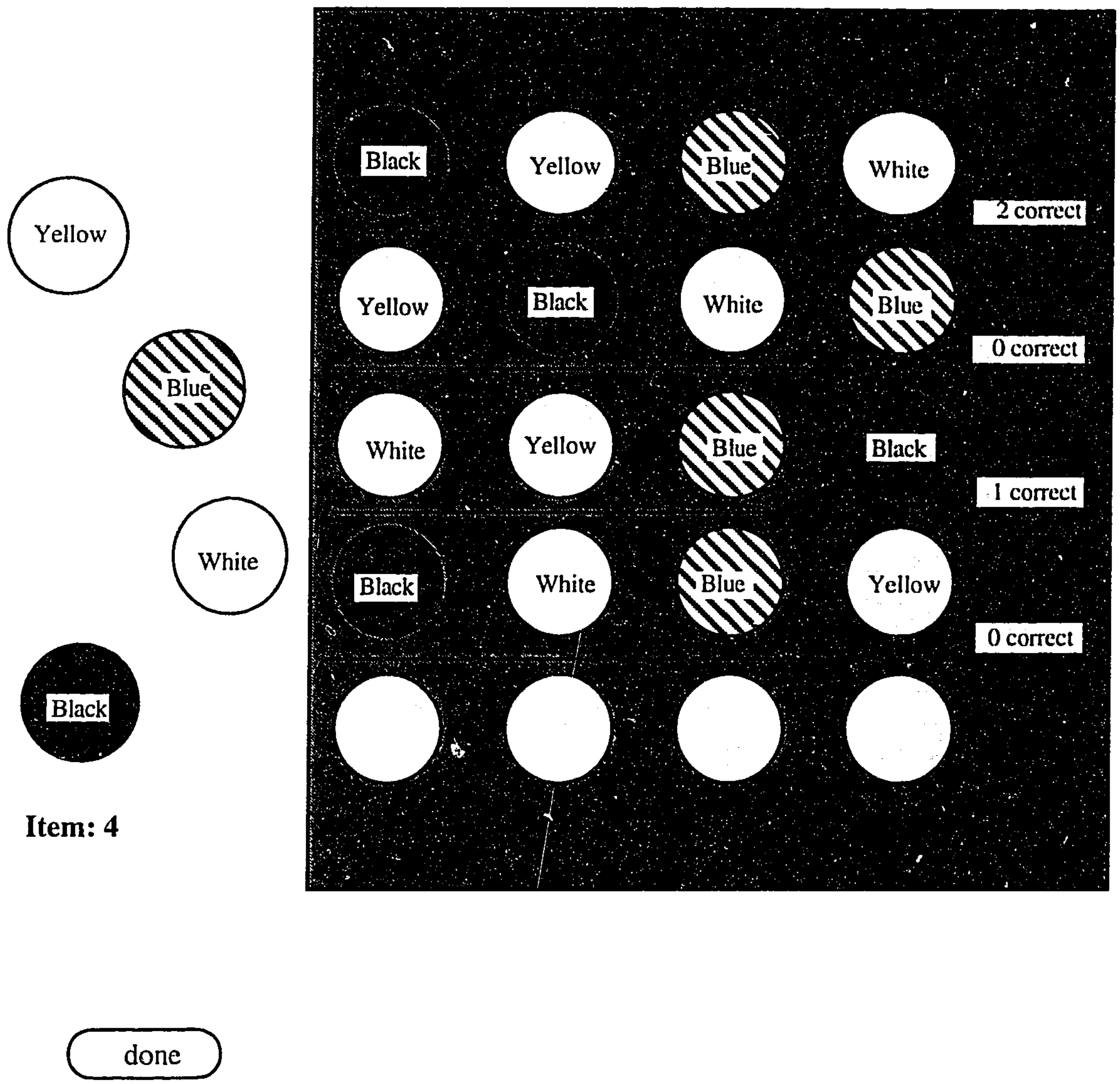
ITEM 2 of Crack-the-Code



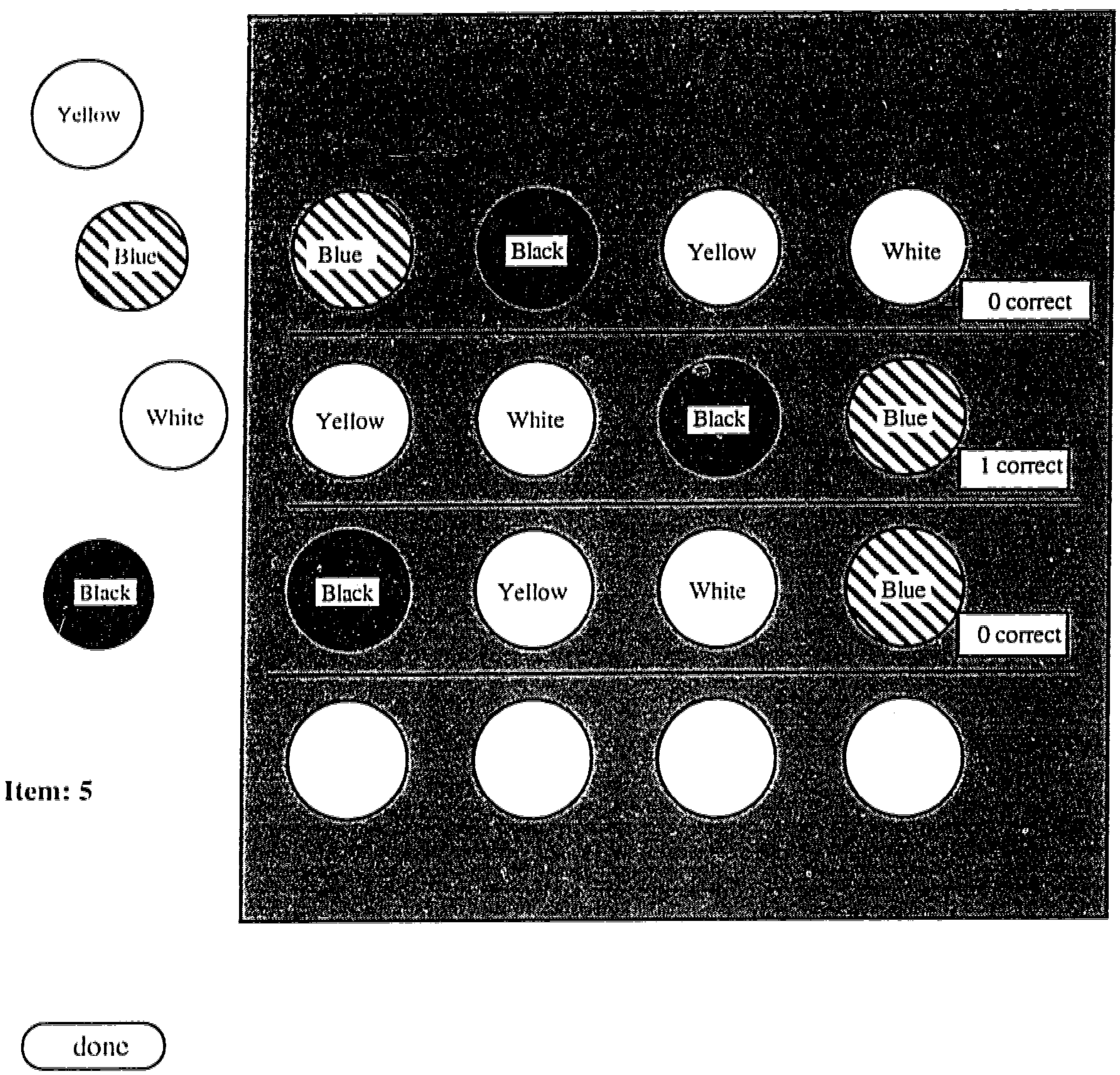
Item: 3

done

ITEM 3 of
Crack-the-Code Copyright:

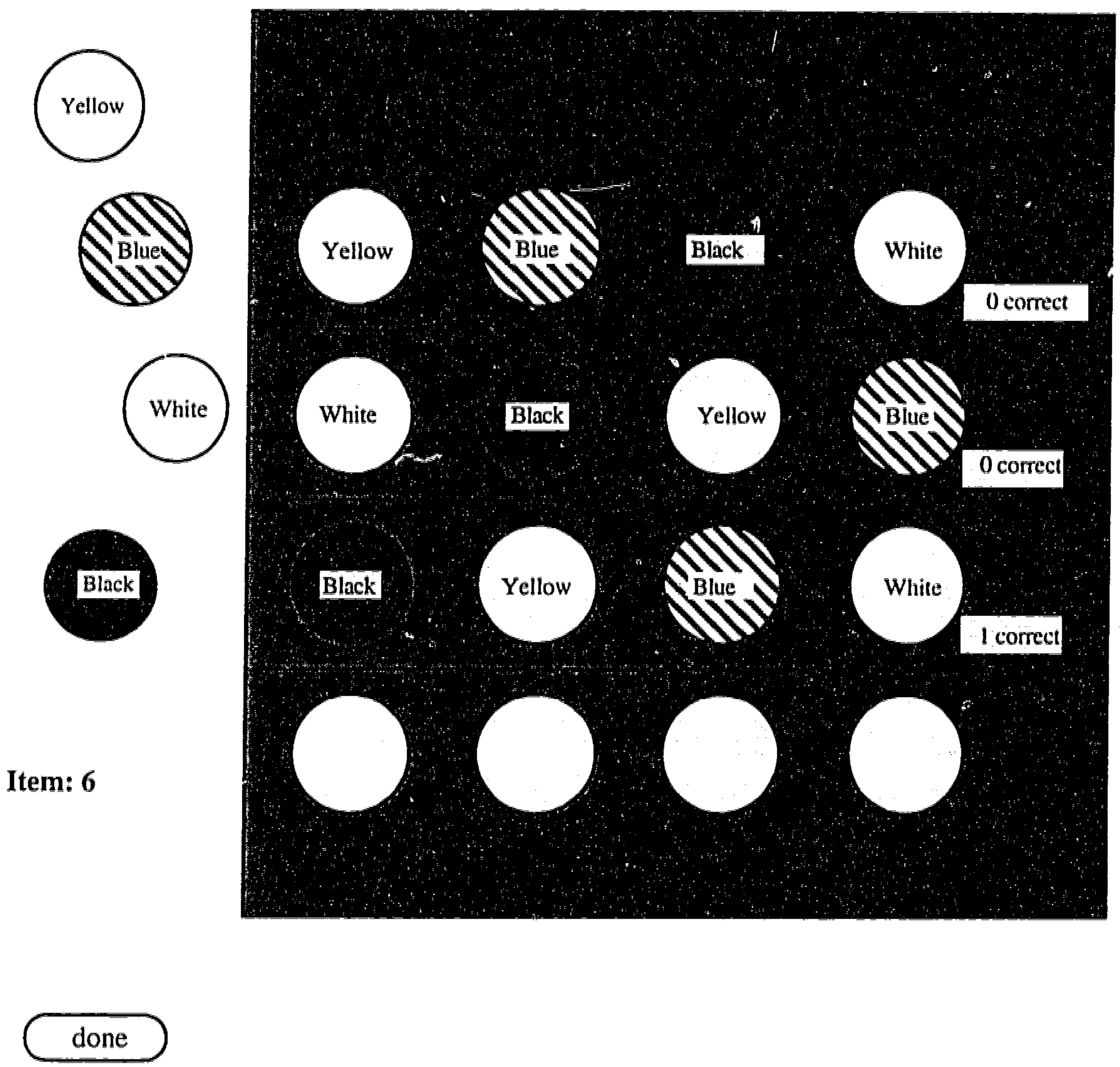


ITEM 4 of
Crack-the-Code Copyright:



ITEM 5 of

Crack-the-Code Copyright: Dr. J P. Das



ITEM 6 of
Crack-the-Code Copyright: Dr. J P. Das