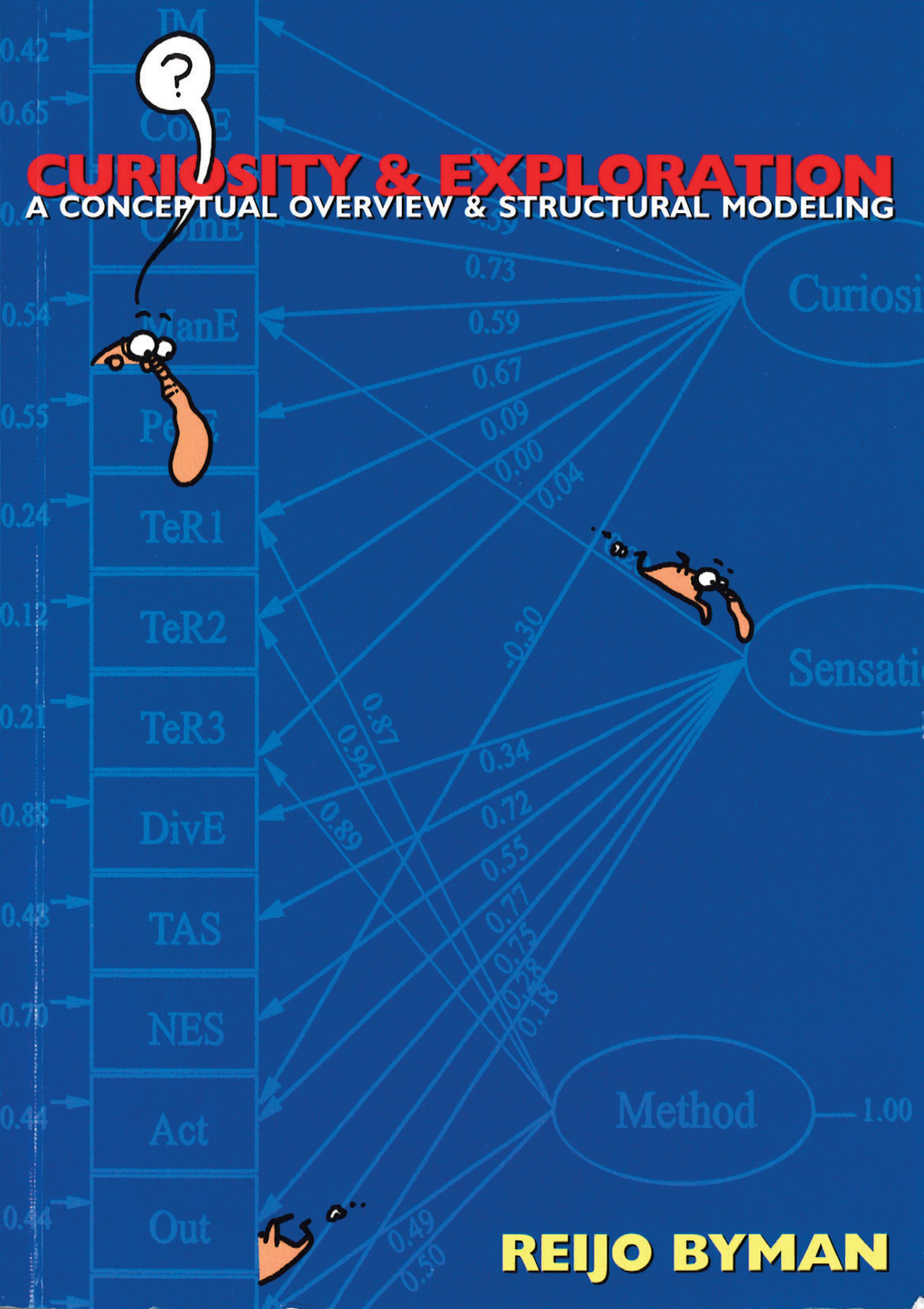
# CURIOSITY & EXPLORATION

## A CONCEPTUAL OVERVIEW & STRUCTURAL MODELING




**REIJO BYMAN**

Curiosity is an educationally interesting concept. One of the ideal motivational teaching strategies has been to bring students to explore, discover and learn in a self-directed manner. However, there is no consensus on what curiosity is. Moreover, some researchers have referred to this concept by the terms intrinsic motivation and sensation seeking. The main purpose of the present study was to clarify the concept of curiosity, starting at theoretical level. Using conceptual analysis and the results of previous studies, it was possible to construct nine alternative conceptual models of curiosity. Confirmatory factor analysis (CFA) was used to investigate the statistical fit of the nine models. Theoretically oriented readers will benefit from a comprehensive review of the scientific study of curiosity and exploration. Methodologically, the present study provides a useful application that demonstrates the flexibility of CFA for the purpose of construct validation.

FINNISH EDUCATIONAL RESEARCH ASSOCIATION

Kasvatusalan tutkimuksia • 5 • Research in Educational Sciences

ISBN 952-5401-04-9
ISSN 1458-1094
Painosalama Oy - Turku 2001

KASVATUSALAN TUTKIMUKSIA – RESEARCH IN EDUCATIONAL SCIENCES 5
FINNISH EDUCATIONAL RESEARCH ASSOCIATION

# CURIOSITY AND EXPLORATION: A CONCEPTUAL OVERVIEW AND STRUCTURAL MODELING

ENGLISH SUMMARY

REIJO BYMAN

ACADEMIC DISSERTATION

TO BE PUBLICLY EXAMINED AND DEBATED WITH DUE PERMISSION OF THE FACULTY OF EDUCATION OF THE UNIVERSITY OF HELSINKI, THE FESTIVITY HALL AT THE DEPARTMENT OF EDUCATION

SEPTEMBER 3, 2001, AT 12 NOON.

THE UNIVERSITY OF HELSINKI
HELSINKI 2001

# Curiosity and Exploration: A Conceptual Overview and Structural Modeling

Suomen Kasvatustieteellinen Seura
Samfundet för Pedagogisk Forskning i Finland
Finnish Educational Research Association

Kasvatusalan tutkimuksia – Research in Educational Sciences 5

# Curiosity and Exploration: a Conceptual Overview and Structural Modeling

Reijo Byman

FINNISH EDUCATIONAL RESEARCH ASSOCIATION
RESEARCH IN EDUCATIONAL SCIENCES

Publisher: Finnish Educational Research Association



The publications published in the series have gone through a referee-system.

Orders: Visit the Finnish Educational Research Association web-site at WWW.kasvatus.net (>publications)

Institute for Educational Research
University of Jyväskylä
Box 35, (Freda)
FIN-40351 Jyväskylä, Finland

tel. +358 14 260 3220
fax. +358 14 260 3241
e-mail: teairama@jyu.fi
www.jyu.fi/ktl/sktseura.htm

Department of Teacher Education
University of Helsinki
Box 38
FIN-00014 University of Helsinki, Finland
tel. +358 09 191 28112
fax +358 9 191 2114
e-mail: marja-liisa.lonardi@helsinki.fi



Cover: Rasmus Bengts and Tero Kylä-Junnila
Layout: Jouni Vilhonen
Printing: Painosalama Oy, Turku 2001
ISBN: 952-5401-04-9
ISSN: 1458-1094

Electronic version produced by University of Jyväskylä, Open Science Centre 2021
ISBN 978-952-7411-02-5
ISSN 2489-768X

# ABSTRACT

## CURIOSITY AND EXPLORATION: A CONCEPTUAL OVERVIEW AND STRUCTURAL MODELING 222 PAGES

*REIJO BYMAN*

Many attempts have been made to measure curiosity or intrinsic motivation. However, there is no consensus on what curiosity is. In the present study, five previously constructed curiosity inventories were translated to Finnish and modified to be convenient for fifth-graders. The main purpose of the study was to clarify the concept of curiosity. Using conceptual analysis and the results of previous studies, it was possible to construct nine alternative conceptual models of curiosity. The corresponding statistical models were expected to account for the covariances among 15 subscales measuring curiosity-related exploratory behavior. However, before proceeding to the subscale level, the structures of the inventories were tested at the item level. Confirmatory factor analysis was used to investigate the fit of the alternative models. The four-step logic suggested by Stanley A. Mulaik was used in testing the models. A sample of 529 Finnish fifth-graders from southern Finland was used. This sample was divided according to sex (263 girls and 277 boys). The best-fitting model was a three-factor model with two trait factors and one method factor. The trait factors were termed Curiosity and Sensation Seeking. Several mild moderation effects were found. The theme "What to do when an unrestricted model fails?" was also discussed.

# CONTENTS

Acknowledgements 9

**1. Introduction** 11

1.1 Basic Considerations 12

1.2 Aims of the Study 14

**2. Curiosity as a Concept in Different Psychological Theories** 17

2.1 Psychodynamic Theory 18

2.1.1 Attachment Theory 19

2.1.2 Competence Motivation 21

2.2 Humanistic Theory 22

2.3 Trait Theory 24

2.4 Neobehaviorist Theory 27

2.4.1 Daniel Berlyne's Theory 28

2.4.2 Diversive Curiosity 30

2.5 Cognitive Theory 32

2.6 Eclectic Mini-Theories 35

2.6.1 Dual-Process Theory of Curiosity and Anxiety 35

2.6.2 Gab Theory 36

2.7 Conclusions 37

**3. Concepts That are Closely Related to Curiosity** 41
3.1 Sensation Seeking 41
3.1.1 Curiosity and Sensation Seeking 44
3.2 Interest 45
3.2.1 Interest and Curiosity 48
3.3 Intrinsic Motivation 50
3.3.1 Curiosity and Intrinsic Motivation 51

**4. Measures of Curiosity** 53
4.1 Stimulus Preference Techniques 53
4.2 Performance Measures 54
4.3 Self-Report Techniques 54
4.4 Teacher-Peer Ratings 57

**5. Dimensionality of Curiosity** 59
5.1 Breadth-Depth Model of Curiosity 59
5.2 Global State-Trait Model 63
5.3 Multidimensional Models of Curiosity 65

**6. Issues and Problems in the Use of Factor Analysis** 69
6.1 The Philosophical Background of Factor Analysis 69
6.2 Confirmatory Factor Analysis 70
6.2.1 Advantages of the CFA Approach 72
6.2.2 Tests of Invariance 73
6.2.3 Higher Order Factor Analysis 73
6.2.4 Nested Factor Model 74
6.2.5 Unrestricted Factor Model 76
6.2.6 Fit Indices 77
6.2.7 Possible Reasons for the Failure of a Factor Model 78

**7. Design and Procedures of the Empirical Study** 81
7.1 Objectives of the Study 81
7.2 Subjects 83
7.3 The Measurement Instruments and the Conceptual Models behind Them 84
7.3.1 Test of Intrinsic Motivation 85
7.3.2 Diversive Curiosity 86
7.3.3 The Broad C-Trait Scale 87
7.3.4 Sensation Seeking Scale 89
7.3.5 Teacher Rating of Curiosity 91

**8. Item-level Analyses** 93
8.1 Model-testing Strategy 93
8.2 Evaluating Reasons for the Failure of a Model 95
8.2.1 Statistical Issues 95
8.2.2 Alternative Theoretical Models 96
8.2.3 Measurement 96
8.2.4 Trimmed Inventory 98
8.3 Invariance Across Gender 99
8.4 Assessment of Model Fit 100
8.5 Results of Item-level Analyses 102
8.5.1 Preliminary Analysis 102
8.5.2 Consistency of the Scales 104
8.5.3 Test of Intrinsic Motivation 106
8.5.3.1 Alternative Theoretical Model 107
8.5.3.2 Systematic Measurement Errors 108
8.5.3.3 Tests of Invariance 112
8.5.3.4 Discussion and Conclusions 113
8.5.4 Diversive Curiosity Scale 114
8.5.4.1 Systematic Measurement Errors 115
8.5.4.2 Tests of Invariance 118
8.5.4.3 Conclusion 118

8.5.5 The Broad C-Trait Scale 119
8.5.5.1 Assessment of the Statistical Issues 120
8.5.5.2 Test of Alternative Theoretical Models 121
8.5.5.3 Systematic Measurement Effects 122
8.5.5.4 Revised Scale 125
8.5.5.5 Testing for the Invariance of the 20-item Scale 128
8.5.5.6 Discussion and Conclusions 129
8.5.6 Sensation Seeking Scale 131
8.5.6.1 The Model With Four First-Order Factors 131
8.5.6.2 The NF Model with Five Orthogonal Factors 133
8.5.6.3 The Model with One Second-order Factor 135
8.5.6.4 Testing for the Invariance of the Sensation Seeking Scale 136
8.5.6.5 Discussion and Conclusions 137
8.5.7 Teacher Rating Scale 138
8.5.7.1 Tests of Invariance 141
8.5.7.2 Conclusions 141
8.6 Discussion 142

**9. Subscale Level Analysis** 147
9.1 Methodological Starting Points 147
9.2 Conceptual Models for the Subscale Data 149
9.2.1 One-Dimensional Model 150
9.2.2 Two-Dimensional Model 152
9.2.3 Three-Dimensional Models 154
9.2.4 Four-Dimensional Model 159
9.3 Preliminary Analysis 160
9.4 The One-Factor Model 161

9.5 Tests of the Unrestricted Models 162
9.6 Tests of the Measurement Models 163
9.7 Post Hoc Analysis 167
9.8 Discussion 170

**10. General Discussion** 175
10.1 The Nature of the Two Trait Factors 175
10.2 The Two Trait Factors as Parts of Personality 179
10.3 Curiosity, Sensation Seeking and Interest 180
10.4 Psychometric and Methodological Issues 182
10.5 Limitations of the Present Study 186
10.6 Further Research on Curiosity 187
10.7 Conclusions 188

References 190

Appendixes 211

# ACKNOWLEDGEMENTS

I hereby wish to thank all those who in different ways have supported me in the present study. Among them I should like to mention the following in particular.

First of all, warm thanks are due to Docent Erkki Komulainen for his continuous encouragement from start to finish. Without his support the project would not have been undertaken, and it certainly would not have been completed. He is also the person who aroused my interest in applied statistics. During this project I have learned to probe the depth of his excellent understanding of statistics, and it goes almost without saying that I am deeply grateful to him.

My gratitude also goes to Professor Pertti Kansanen, who has greatly contributed to my scientific development. His constructive criticism has done much to improve my work. He also knows how to nudge people ahead without pushing them. I am deeply grateful to him.

Associate Professor Paul Hellgren was my philosophical counsellor and critic at the beginning of my work. By asking crucial questions he compelled me to clarify my thoughts and helped me to think about conceptual issues. I am still deeply grateful to him.

Professor Stanely A. Mulaik introduced me to the logic of the unrestricted factor model. He also made valuable comments on previous versions of my manuscript. At the beginning of my work I had splendid discussions with Professor Kai Karma that deepened my understanding of reliability and validity issues in the behavioral sciences.

I wish also to extend thanks to my two reviewers Professors Raimo Konttinen and Esko Leskinen for their detailed comments on the initial manuscript of this research. After their through review, any errors or omissions are uniquely mine. I thank also Professor Henry Fullenwider, who read the manuscript prior to publication.

Important for me has been also the support of my esteemed colleagues and friends Tor Kronlund, Matti Meri, Jukka Husu and Riitta Jyrhämä, especially in periods of extreme stress and difficulty. I feel sure they know how grateful I am.

I thank The Finnish Educational Research Association for publishing this monograph. Three foundations must also be acknowledged for their support to my work. These foundations are the Helsingin Sanomain 100-vuotis säätiö, the Alfred Kordelin säätiö and the Leo ja Regina Weinstein säätiö. Without this financial support my work might not have been completed.

Last but not least, I wish to thank my wife Maria for her tolerance and understanding. Our precious boys Roni and Roi and their endless curiosity have been the inspiration of my work.

Lahti, June 20, 2001 *Reijo Byman*

# 1. INTRODUCTION

Curiosity is an essential concept in modern motivation terminology (see Murphy & Alexander, 2000). Curiosity, sensation seeking, interest, and intrinsic motivation are closely related concepts and are said to have very similar positive effects on learning, specially on its quality (see Björk-Åkesson, 1990; Bruner, 1966; Berlyne, 1965; Deci & Ryan, 1994; Dewey, 1913; Rigby, Deci, Patrick, & Ryan, 1992; Ryan & Deci, 2000; Schiefele, Krapp, & Wintler, 1992; Zuckerman, 1994). Iran-Nejad, McKeachie, and Berliner (1990) have mentioned curiosity and interest as two "crucial mentalistic concepts" when trying to develop a unified learning theory. Other such concepts are attention, motivation, and metacognition. An understanding of these concepts may also solve the old paradox which Iran-Nejad et al. formulated into the question how it is possible that children learn so much before school and sometimes so little at school.

The present study is the first part of a broader research project whose aim is to investigate the state/trait nature of curiosity and its relationships to other concepts like students interests, school achievement and attractiveness of learning tasks. The purpose of this first part is to clarify issues related to the conceptualization and measurement of curiosity, especially trait-like curiosity.

## 1.1 Basic Considerations

According to Wittgenstein (1981, §43), the meaning of a term is its use. Term *curiosity* is used both in common language and science. However, the meaning of the term curiosity is not same in these two contexts (see Byman, 1995). In everyday language, the word curiosity has both in Finnish and in English (cf. Berlyne, 1978; Voss & Keller 1983, pp. 1-5) a slightly negative connotation. It is something that is not very desirable: it is not a necessary component of good education. Usually the word curiosity creates mental images of a person who is "a eavesdropper," or "a nosy parker," and thus has an excessive and indelicate desire to know something which is in some way secret and none of his or her business. In this particular context Berlyne (1978, p. 99) used the word *inquisitive* as a synonym for *curious*.

When the word curiosity is used as a scientific concept, it does not have a negative connotation. However, the scientific use of the term curiosity has not been uniform, either. Actually, curiosity has been a very difficult concept to define. Fowler (1965, p. 23), for instance, gave up intentions and determined curiosity as "a behavior without a definition." Thus, different theoretical and operational definitions of curiosity exist. On the other hand, the terms *curiosity* (Berlyne, 1960), *intrinsic intellectual motivation* (Lloyd & Barenblatt, 1984), *intrinsic orientation* (Harter, 1981), *intrinsic motivation* (Day, 1971; Deci, 1975; Rubenstein, 1986), *reactive curiosity* (Penny & McCann, 1964), *diversive curiosity* (Day, 1968), *stimulus variation seeking* (Penny & Reinehr, 1966), *seeking curiosity* (Livson, 1967), *novelty seeking* (Pearson, 1970) and *sensation seeking* (Zuckerman, 1971) have at times been used interchangeably and at other times as an indicator of quite distinct concepts. Moreover, even the use of the term *curiosity* has been ambiguous. It has been used to refer both to a hypothetical construct (e.g., Berlyne 1960, 1978) and to an observable behavior. To eliminate this terminological ambiguity, Voss and Keller (1983, p. 150) have suggested that the term *exploration* should be used to refer to the observable behavior, and the term *curiosity* should refer to the corresponding psychological construct. Thus, for instance, what Kreitler, Zigler and Kreitler (1974, 1975) first labeled as *curiosity*, they have recently termed *exploration* (Kreitler & Kreitler, 1994). According to Kreitler and Kreitler (1994, p. 260), they now want to "distinguish strictly between the overt behavior and its motivation." On the other hand, the motivation behind exploration is not always curiosity. For instance, a person may "explore" a room because she or he has lost something Typical of this kind of exploration is that it has an external goal.

Despite varying definitions of curiosity, a common view is that curiosity is on the other hand a disposition and on the other hand an aversive state of subjective uncertainty. This internal state arises when a person is confronted by a specific object or concept which differs from his or her mental schemas. Such characteristics as novelty and complexity give rise to uncertainty. Subjective uncertainty generates a tendency to want to obtain more information about the specific object. Depending on cognitive appraisal and affective determinants, some kind of exploratory behavior can take place. The aim of this behavior is to resolve or mitigate the uncertainty by gathering more information. When the subjective uncertainty is reduced successfully, a person experiences feelings of competence and mastery (Berlyne, 1960; Keller, Schneider, & Henderson, 1994; Trudewind & Schneider, 1994; White, 1959; Wohlwill, 1981).

Curiosity has also been seen as an educationally interesting concept (e.g., Berlyne, 1965; Bruner, 1966; Day, 1982; Deci and Ryan, 1985; Hidi and McLaren, 1990; Voss and Schauble, 1992). Since Aristotele, one of the ideal motivational teaching strategies has been to get students to explore, discover and learn actively. In recent years several psychologists and educators have preferred learning motivation which is free from the pressure of grades, extrinsic rewards and control. That is, they advocate motivation where learning activities are undertaken for their own sake. But what, then, makes a person gather information for its own sake - for the pleasure and satisfaction derived from the activity? One answer has been curiosity, another interest, and a third intrinsic motivation. On the other hand, schools are often blamed for fostering the wrong kind of motivation, for example, killing the natural curiosity of children. Carl Rogers (1969, p. 157) expressed the problem as follows "*Human beings have a natural potentiality for learning.* They are curious about their world, until and unless this curiosity is blunted by their experience in our educational system." Is the situation as Rogers describes it because the word *curiosity* evokes both good and bad feelings in our minds or because teachers do not know what curiosity really is and for that reason cannot teach "the Curious Way", as Day (1982, p. 20) expressed it. Is it so that the natural curiosity of children must be fostered, not suppressed? Day and Maynes (1972) stated as follows:

> Each of us is born with a disposition to be curious, a disposition to know. Whether this curiosity survives and serves us effectively depends on the strategies used by our parents and teachers. These strategies are independent of the place of learning. (p. 69)

## 1.2 Aims of the Study

The concept of a *language game* is essential for the present investigation. According to Wittgenstein (1981), language games are rule-guided ways of verbal action or interaction between members of a community. Based on this view, the meaning of a word is determined by the rules governing its operation: we learn the meaning of a word by learning how it is used. A second important concept for the present approach is *family resemblance.* Curiosity is seen as a family-resemblance word, which means that the aim is not to find common definitions for the concepts curiosity and exploration but to investigate and understand how the common conceptions of curiosity and exploration overlap, how they are similar and how they differ. In other words, the aim is to investigate the family resemblances within the concept of curiosity (cf. Wittgenstein, 1981). Both theoretical conceptions and operationalizations of curiosity are studied.

From a scientific perspective, evidence for the construct validity of curiosity is at the preliminary stage, as it is characterized by considerable ambiguity. Several different measurement instruments exist to measure curiosity and exploration. As Ainley (1987, p. 54) noted, "construction of these measures has often been pragmatic rather than adhering closely to a particular theoretical system." Thus, in previous studies different curiosity inventories have had only low to moderate positive correlations (Ainley, 1987; Langevin, 1971, 1976; Rubenstein, 1986). Rubenstein (1986) pointed out that these kinds of results raise doubts whether curiosity is a unidimensional construct. Knowledge of the relationships between different curiosity scales helps us better understand the concept behind the measurements as well as the inconsistent results of some curiosity studies. Thus, the dimensionality problem is one aspect of the construct validation problem of curiosity.

All previous studies (e.g., Ainley, 1987; Langevin, 1971, 1976; Olson & Camp, 1984; Rubenstein, 1986) have used exploratory techniques and inductive logic to clarify the dimensionality problem of curiosity. However, the starting point of the present study is hypothetico-deductive. As Popper (1963) emphasized, a sign of a scientific theory is its testability. According to Popper, science grows through unjustifiable conjectures and refutation of conjectures that fail. Based on this logic, the purpose of the present study is to test different conceptions of the structure of curiosity construct. In practice, this means that conceptual models are first constructed and then converted to confirmatory factor analysis (CFA) models. These models are then statistically tested and compared. The discussion takes place at two levels: both item-level and subscale-level models are presented.

The following presentation consists of two parts. The aim of the first part is to provide a theoretical framework for the research problems and for the presented conceptual models. It tries to elaborate, clarify and integrate different theoretical and operational definitions of curiosity and exploration. The aim of the second part is to find out what is measured by curiosity scales and what dimensions are needed to describe the curiosity construct.

In addition, this study has two special themes. The first is related to the implicit and explicit beliefs that gender is related to curiosity. It is, for instance, a common belief that girls are more curious than boys. However, the results of earlier studies are partly contradictory (see e.g., Voss and Keller, 1983). One reason for these results may be the measurement instruments, which have not been invariant across gender and thus have different validity. Before it makes sense to compare the means between girls and boys it is necessary to get reliable information about the invariance of the measurement instrument used. In the present study the invariance of the measuring instruments are systematically tested using multigroup confirmatory factor analysis (MGCFA). The second special theme of the present study is to discuss what to do when a model fails. In this connection special attention is paid to the failure of an unrestricted confirmatory factor analysis model. This model tests the hypothesis that the presented number of factors is enough to account for the covariances among test variables (Jöreskog, 1979a; Mulaik & Millsap, 2000).

# 2. Curiosity as a Concept in Different Psychological Theories

Wittgenstein (1981, §§383-4) suggested that mental concepts should be analyzed by describing the use of words used to describe the concept. Danziger (1997) expressed the same idea by arguing that the discourse of which a word is a part gives meaning to the term. These notions are the starting point of this chapter. Danziger (1997) buts it this way:

> All psychological categories have changed their meaning through history, and so has the discourse of which they were a part. To gain an understanding of the categories in common use at the moment, we need to see them in historical perspective. When we go back to the origin of these categories we usually find that what later became hidden and taken for granted is still out in the open and questionable. (pp. 8-9)

Schools of thought in psychology have different, often hidden assumptions and implications (see e.g., Slife & Williams, 1995). Thus, the meaning of the term curiosity also differs in these communities. For Wittgenstein, the curiosity has a different role in different "language games." However, distinguishing between these theories is problematic because different views have been directly or indirectly influenced by each other. Trait theory, for instance, has been influenced by psychodynamic theory. However, the aim of following section is to describe the use of term curiosity in five psychological theories, namely psychodynamic theory, trait theory, humanistic theory, neobehaviorism, and cognitive theory.

## 2.1 PSYCHODYNAMIC THEORY

Psychodynamic theory is a very broad and heterogeneous movement that also includes psychoanalysis. Characteristic of this theory is a strong element of determinism. It argues that forces we are unaware of influence our behavior. These forces include unconscious ideas and instincts, especially sexual instinct. According to psychodynamic theory, a person can not directly control his/her motivation. Moreover, Freud argued that all motivation is sexual motivation. One starting point of the psychodynamic theory is also that the past experiences of a person cause his/her present behavior.

These assumptions have also had an effect on conceptualizations of curiosity. Freud (1971) used the term curiosity in a very limited and narrow sense. However, as he was not very consistent in his use of the term, at least two interpretations have been presented. First, Freud related curiosity closely to sexual instinct. Freud (1971) argued that curiosity is first directed to sexual parts of the body. A drive which Freud called "scoptophilia" causes the child to develop an intense interest in sexual matters. A child finds it pleasurable to explore his/her sexual organs. This early source of pleasure is the reason for the child's intense interest in sex. This interest is directed also to the exploration of the sexual organs of his/her parents and playmates. Sexual curiosity begins very early, sometimes before the third year. However, through *sublimation* the direction of this curiosity can be diverted to other objects, for instance art.

Aronoff (1962) has suggested another interpretation of Freud's concept of curiosity. According to this view, curiosity is basically a coping mechanism the child develops in response to the threat which it feels when a rival, for instance a sister or brother, emerges in its life. This newcomer challenges the child's privileged position in the family because the child must now share the love of the parents with this newcomer. In this kind of situation a child begins to think about the origins and mechanisms of birth in order to avoid this kind catastrophe in the future. According to Aronoff, it is this sort of experience which "develops in the child the capacity of curiosity."

### 2.1.1 ATTACHMENT THEORY

Bowlby (1969) has extended Freud's conception of curiosity and exploration. Freud strongly emphasized the child's attachment to the mother, which is based on the intense dependence of an infant on its mother. Object relations, that is, relationships to other people, stem from the early relationships to the mother. A child's early personality development occurs in the setting of the family (anaclitic identification). According to Freud, "early childhood" (the first five or six years) is extremely important for the growth of a child's object relations. During this time separation from the mother, for instance, can be traumatic. Freud argued that neuroses are acquired only in early childhood. Moreover, Freud also claimed that anxiety in children is originally nothing other than an expression of the fact that they are feeling the loss of the person they love (Freud, 1905b in Bowlby, 1973, p. 54).

Later Bowlby (1969) developed these ideas into a theory that has been called "attachment theory." Bowlby emphasized the meaning of the early childhood experiences in the same deterministic way. Instead of the term "object relations," Bowlby used the terms *attachment* and *attachment figure*. According to Bowlby, attachment behavior is an inborn behavioral system that provides for the survival of the species. The goal of attachment behavior is to maintain proximity. Within the first three years a child has generated typical patterns of attachment behavior towards the caregiver. Strong feelings of love and joy are connected to the attachment figure. After three years, attachment behavior diminishes both in intensity and in the frequency with which it is elicited. After that age, a subordinate attachment figure, for instance a teacher, can substitute for the mother as a "secure base." However, on the whole, the attachment behavior forms a continuum from childhood to adulthood. In adulthood the attachment experiences direct the expectations regarding significant others and self as a person of worthy of love. Thus, attachment behavior is a pivotal feature of personality across the life span.

Curiosity and exploration have a central role in attachment theory. One starting point of attachment theory is that the attachment to a caregiver provides a secure base for exploration that is essential to the development of subsequent competence. A dynamic balance exists between attachment and curiosity-exploration behavior. When the mother is not present, the child's attachment behavior is likely to be elicited (cf. separation anxiety) and its exploratory behavior inhibited. Conversely, when the mother is present, the attachment behavior is latent and exploratory behavior is likely to be elicited. Thus, exploratory behavior can also be defined negatively as the absence of attachment behavior. Moreover, waning attachment behavior or changes in its form may be caused by increasing curiosity and exploration (Bowlby, 1969). More recently, Mikulincer (1997, p. 1226) has argued that security in attachment also has a cognitive facet which includes "active information search," "openness to new information," and "flexibility of cognitive structures." Together, these things improve a child's coping mechanisms.

According to Ainsworth, Blehr, Waters and Wall (1978, p. 279) an unfamiliar or strange situation activates five behavioral systems in varying degrees of strength: exploratory behavior, wary/fearful behavior, attachment behavior, sociable behavior, and angry/resistant behavior. Wary/fearful behavior is antithetical to curiosity and exploratory behavior, as is also attachment behavior, which overrides exploratory behavior when activated. Thus, security is necessary for a child to be able to play and explore (attachment-exploration balance). A child uses its mother as a secure base from which to explore. Moreover, Bowlby (1973) claimed that the child-mother interaction also determines later curiosity-exploration activity:

> ... infants whose mothers are sensitive and responsive to them are those who later turn cheerfully to exploration and play. Their willingness to cooperate, their capacity to concentrate, and their good scores on developmental tests at twenty-one months bode well for their futures. (p. 406)

The results of several subsequent studies have given support to the claim that children explored less during their mother's absence than when she was present (Ainsworth et al., 1978).

## 2.1.2 COMPETENCE MOTIVATION

White (1959) subsumed curiosity under the heading of *competence motivation.* According to White, the motivational aspect of competence is *effectance.* Accordingly, Harter (1981) referred to this concept using the term *effectance motivation.* White was inspired by Freud's ego psychology and Woodworth's (1958) behavior-primacy theory, which stated that "all behavior is directed primarily toward dealing with the environment." According to Freud (1969), the *ego* develops from the *id* progressively and contains our conscious identity and awareness of ourselves. The influence of the surrounding external world makes a part of the id to grow out as the ego. The ego gets its energy from ego instincts, while the id receives its energy from sexual instincts. The goal of ego instincts is self-preservation.

According to White, a human being has an urgent, drive-like "intrinsic need" to deal with and master the environment. Modifying Freud's and Woodworth's ideas White (1959, p. 297) used the term *competence* to refer to a person's capacity to interact effectively and competitively with the environment. White used term competence in a broader sense than in everyday language. The subjective side of competence motivation is satisfaction, which shows up as a "feeling of efficacy." The competence motivation (competence motive) of a child is undifferentiated, but later it may lead to a life-long specific exploratory interest. Competence motivation is a broad motivational concept that covers what has been called, for instance, curiosity, children's playful exploration, mastery, or a need for excitement. Drawing on theories of the "optimal level of stimulation" (Hebb, 1955; Leuba, 1955), White argued that novel and unfamiliar objects as well as boredom arouse competence motivation. Competence motivation is a persistent, ongoing process which in young children is undifferentiated but which may later be differentiated into specific motives such as mastery and achievement. Through acquisition of competence a person gains in independence. Later, Deci (1975, p. 56) suggested that self-determination is very similar to competence or effectance motivation, since "someone who is self-determining will feel efficacy." What Deci called *intrinsic motivation* as opposed to *extrinsic motivation* is based in the tendency to be competent and self-determining in relation to the environment. According to Deci, the need to feel competent and self-determining lies behind two kinds of exploratory behavior: behavior which "seeks" optimal challenges and behavior which "conquers" challenges. Thus, people seem to actively seek and conquer challenges which are optimal for them. According to White (1959), satisfaction of the competence motivation is the primary goal of exploratory behavior, while learning is secondary.

## 2.2 HUMANISTIC THEORY

Humanistic psychology has been influenced by existential philosophy. It states that every human being is unique and has free will. According to humanistic psychology, one is motivated to actualize one's unique potential. Maslow described this concept with the word *self-actualizing*. According to Maslow (1970, p. 46), "What a man can be, he must be." In principle the concept of the self-actualizing tendency is the only motivating force, for instance, in both Rogers' and Maslow's theories. This tendency suggests, and holds out as a goal for us, what we can become if we fully develop our potential (self-fulfillment). Humanistic theorists emphasize the role of personal experience, claiming that the real meaning of behavior lies in the person's phenomenology.

According to Rogers (1969), every individual intuitively knows his or her own nature and what is required for one's own growth and actualization. Thus, in Rogers' view, the term "teaching" has a bad connotation because it means the same as "to instruct," or "to show, guide, direct." Rogers believed that instead of teaching facts, the goal of education should be the "facilitation of learning." The goal of teaching should be active, self-directed and creative learning (a learning man). According to Rogers, a good learning situation should stimulate and enhance a person's innate "unquenchable curiosity" (Rogers, 1969, p. 190). Thus, instead of emphasizing teaching static knowledge (facts), Rogers emphasized the role of the knowledge-seeking process. Rogers advised persons to free up their curiosity and open everything to questioning and exploration. A fully functioning person is curious, creative and able to live "the good life." Recently, Csikszentmihalyi (1997, p. 346) has emphasized the link between curiosity and creativity by stating that "cultivation of curiosity" is the first prerequisite of creativity. According to Csikszentmihalyi, creative individuals are always curious and open to new things. For Csikszentmihalyi, curiosity is at first diffuse and generic but later becomes specific, thus focusing on a specific domain of interest.

One premise of humanistic psychology is that a human being is an integrated, organized whole. According to Maslow (1970, p. 19), this means that "the whole person is motivated rather than just a part of him." Maslow rejected the existence of single somatically specified needs and drives. According to Maslow, a human being has five hierarchically organized basic needs, namely physiological needs, safety needs, the need to belong and to be loved, the need for social approval and self-approval, and the need for self-actualization.

Certain conditions are prerequisites for the satisfaction of basic needs. Among these conditions is "freedom to investigate and seek for information" (Maslow, 1970, p. 47). Actually Maslow (1970, p. 50) postulates a second hierarchy which includes what he calls "cognitive needs" or a "desire to know and to understand" and which is interrelated to the hierarchy of the five basic needs (part of personality). The cognitative needs are the need to know, to understand, to systematize, to organize, to analyze, to look for relations and meanings, and to construct a system of values. The intention is to search for meaning. Common to all cognitive needs is that they seem to happen for their own sake, for the sheer delight of knowing and understanding per se. Maslow (1970, p. 49) argued that mysterious, unknown, chaotic, unorganized, and unexplained things per se are attractive and interesting (the opposite of boredom).

Maslow uses the phrase "need to know" as a synonym for curiosity. For Maslow, curiosity is an innate, conative or striving need which leads an individual to gather information ("restless curiosity"). If the information is separate or atomistic, other conative needs are investigated, above all the need to understand. Thus, the need to know is prepotent to the need to understand. Maslow also linked anxiety and curiosity together. According to Maslow, a person can seek information (explore) in order to reduce anxiety. Actually Maslow (1968, p. 67) suggested that "all cognitive needs are instigated by anxiety and are only efforts to reduce anxiety." On the other hand, Maslow noted that a person can also avoid knowing in order to reduce anxiety. Thus, as did Freud, Maslow (1968, p. 62) also emphasized that "incuriosity" can be a defense ("curiosity is dangerous") and sign of anxiety and fear.

Rogers, Maslow and Csikszentmihalyi are among many others who have noticed that curiosity can be killed. Maslow (1970, p. 50) put this idea this way: "Children do not have to be taught to be curious. But they may be taught, as by institutionalization, not to be curious." However, the view of humanistic psychologists seems to be more optimistic than that of many others. As Rogers noted, curiosity is an innate tendency which can be dulled or blunted but never totally destroyed. Recently Csikszentmihalyi (1997) has even published a number of self-help suggestions for stimulating curiosity:

> Try to be surprised by something every day ... Stop thinking what all things are about. Be open to new things...Try to surprise at least one person every day ... Stop being your predictable self. Break your routine ... Write down each day what surprised you and how you surprised others ... When something strikes a spark of interest, follow it. (pp. 347-348)

## 2.3 TRAIT THEORY

Most trait theorists share three basic assumptions in common. First, they argue that people possess broad predispositions to respond in certain ways on certain occasions. These predispositions are called *traits*. Second, they develop a hierarchical view of the human personality. This means that parts of our personality have interrelationships and that some parts are more influential than others. The third assumption is that traits can be inferred from behavioral signs (Pervin, 1984; McAdams, 1997).

Trait theorists have traditionally distinguished two forms of traits: outer or behavioral traits that can be directly observed, and inner traits which explain the outer traits (Johnson, 1997). Cattell, for instance, used the terms *surface traits* and *source traits*. Surface traits are descriptions of attributes which are overt, manifest or superficial, while source traits are underlying sources of observed behavior. Moreover, a source trait can either be general or specific. General source traits affect behavior in many different situations, whereas specific source traits are individual and may operate in one situation only. Thus, source traits explain a variety of responses across many situations.

Cattell's personality theory is also characterized by a hierarchical organization which divides source traits into *dynamic*, *ability*, and *temperament* traits (Cattell, 1950). According to Cattell, the basic source of human motivation lies in dynamic traits which Cattell divided to *ergs* and *metanergs* (Cattell, 1950). Ergs are innate psycho-physical dispositions which triggers a reactive tendency to certain classes of objects more readily than to others. Cattell also connected specific emotion and goal activity to each erg. As Cattell and Kline (1977) noted, ergs are roughly equivalent to the drives and instincts identified in former theories. Using factor analysis Cattell (1957) differentiated between 16 ergs, one of which was curiosity. However, on the theoretical level Cattell never made explicit what he meant by curiosity. At the empirical level, the following items measure the factor which Cattell (1957, p. 516) labeled "curiosity":

| | |
|---|---|
| I like to read books, newspapers, and magazines. | .5 |
| I want to listen to music. | .5 |
| I want to know more about science. | .4 |
| I like to satisfy my curiosity about everything going on in my neighborhood. | .3 |
| I want to see more paintings and sculpture. | .3 |
| I want to learn more about mechanical and electrical gadgets. | .3 |
| I like to see a good movie or play. | .3 |
| I am not interested in being smartly dressed. | .3 |

Cattell stated that the emotion related to curiosity-erg is curiosity and that the corresponding goal is *exploration*. Other two motivational variables which Cattell called *metanergs* are *sentiments* and *attitudes*. How a person behaves at a specific time depends on the traits and motivational variables relevant to the situation.

In the 1970s trait theories encountered criticism from researchers with a behaviorist orientation. According to this "situationist criticism," human behavior depends on situational cues more than trait theorists supposed. For instance, Mischel (1968) claimed that personality test scores (trait) and behavioral criteria seldom correlate higher than .30. However, as Johnson (1997, pp. 75-76) has noted, situationist criticism failed to disprove the existence of traits for five reasons. First, if situations control behavior, then people must have a capacity (trait) to respond to situational cues. Second, people respond differently to the same situation. Responses differ as a function of personality. Third, having a special trait means reacting the same way to the same situation, not reacting the same way to different situations. Fourth, having a trait does not mean that one reacts the same way in the same situation every time. Instead, people may be described in terms of the likelihood of behaving in a particular way. Fifth, the inconsistency of behavior over time does not rule out the existence of emotional or cognitive traits.

In common language and in the professional terminology of psychology, people have been classified with such trait terms as "introverted," "aggressive" or "curious." However, as Wiggins (1997) has noticed, traits as attributes of persons are only one aspect of the definition. The other two are traits as attributes of behavior and traits as predictors of behavior. Thus, the modern conception of traits is that they are "consistent patterns of thoughts, feelings, or actions that distinguish people from one other" (Johnson, 1997, p. 74).

As MacAdams (1997) has noted, the most influential trait taxonomy is currently the "Big Five." This taxonomy is based on the idea that all traits are organized hierarchically from narrow and specific to broad and general and that all existing trait dimensions can be reduced to five basic categories. These five categories are (1) extraversion-introversion, (2) neuroticism, (3) openness to experience, (4) agreeableness-antagonism, and (5) conscientiouness-undirectedness. For the conceptualization of curiosity, openness to experience is the most important of these five dimensions. MacCrae and Costa (1997) conceptualized openness both as a psychic structure and a need for experience. They emphasized that curiosity and sensation seeking are relevant to an understanding of the motivational aspects of openness to experience. Active curiosity or motivation to seek out the unfamiliar is typical of open people.

Day (1968) may be the first to have distinguished explicitly between state-like and trait-like curiosity. Later, also Berlyne (1971b, p. 191) admitted that "distinct sets of traits" exist which predispose persons to specific or diversive exploration. According to the development of trait theories, this difference has been ascribed to a view of curiosity both as a motive-like trait (C-trait) and state (C-state) (see Boyle, 1983, 1989; Naylor, 1981). For example, Boyle (1983) proposed a schematic conceptual model of state-trait curiosity in which the cognitive appraisal of external or internal stimuli plays a central role. According to the state-trait distinction, individual differences in exploratory activities are thought to vary along two dimensions: (a) the trait of the individual, referring to the predisposition to manifest a state across a wide range of contexts and conditions; and (b) the state of the individual, referring to affective reactions that vary in intensity, fluctuate over time, and result from specific environmental conditions and the level of the trait that an individual possesses. It is also presumed that those possessing more C-trait experience greater intensity of C-state. Trudewind and Schneider (1994) described the relationship between C-trait and C-state as follows:

> ... we postulate an original motive to explore one's physical and social environment. The ultimate function of this behavioral disposition is the acquisition of knowledge or the assimilation of objective structures, whereas the immediate cause of exploration is assumed to be a state of subjective uncertainty created by certain aspects of the environment. (p. 152)

## 2.4 Neobehaviorist Theory

One neobehavioristic (see Berlyne, 1975; Madsen, 1974, 1981) theory has had an enormous effect on the conceptions of curiosity as a scientific concept. This is the theory of Daniel Berlyne. Several theoretical conceptions of curiosity are based directly or indirectly on the investigations of Berlyne. This also means that several of his ideas were later adopted for use in other psychological theories, for instance in cognitive psychology (cf. Beswick, 1971, 1974). Moreover, most of the modern definitions of curiosity depend on his ideas (e.g., Keller, Schneider, & Henderson, 1994).

### 2.4.1 Daniel Berlyne's Theory

Berlyne was influenced by neobehavioristic drive theory and cognitive psychology. Both Jean Piaget and Clarke Hull have left their marks on his thinking (see Flavell, 1963). Berlyne argued that curiosity is an externally stimulated, drive-like, mental state. Specially, he defined curiosity as an internal aversive state occasioned when a person is confronted by object or event which is, for instance, novel or complex. Then the person undergoes an aversive state of subjective uncertainty which generates a tendency to engage in exploratory or investigatory behavior aimed at resolving or mitigating the uncertainty. This motivation is what Berlyne (1960, 1978) meant by *curiosity*, and the behavior is *specific* exploration. Berlyne's definition was based on the theory of optimal arousal first presented by Hebb (1955) and Leuba (1955), according to which a person has an optimal level of arousal or activation which he tries to maintain. This level of arousal is controlled by extrinsic and intrinsic variables.

As did the behavioristic theorists, Berlyne (1960) stated that curiosity is awakened by an external stimulus with certain special quality. However, other than the earlier behaviorists Berlyne argued that curiosity is a mental state which then may produce exploratory behavior. Berlyne used the terms "collative properties" or "collative variables" when referring to stimulus characters which cause an interaction between the observer and the stimuli. In this process the perceiver collates or compares his or her mental schemata to the stimulus and the comparison results in a relative assessment of novelty, surprisingness, complexity, ambiguity, incongruity and other properties which contain a certain measure of unexpectedness and uncertainty. Thus, collative variables involve conflict, and therefore competing and mutually interfering response tendencies that heighten arousal and lead to an aversive internal state which Berlyne called curiosity. The tendency to engage in exploratory behavior is a result of curiosity. The aim of this behavior is to reduce this uncertainty and in this way to recover the state of optimal arousal.

Berlyne (1960) made a distinction between *perceptual* and *epistemic* curiosity. Uncertainty-relieving perceptions activate perceptual curiosity and exploratory behavior. Berlyne (1963) stated that the degree of response conflict depend upon the nature of previous experiences with similar stimulation. Moreover, it depends on the number of opposing response tendencies, the degree of their opposition, their relative strengths, and their absolute strength. According to Day and Berlyne (1971), the last factor is "somehow related to how meaningful the whole situation is to the observer, for the more meaningful, the greater is the intensity of conflict" (p. 313). If perceptual curiosity is aroused, a person tries to resolve the conflict by specific exploration, which can take the form of *receptor adjusting*, *locomotory exploration*, or *investigatory behavior*. The first happens, for instance, when a person comes into a room and fixes his or her attention on a specific object. Locomotor exploration happens when the person moves toward the source of stimulation. Investigator behavior is mainly manipulative, handling a strange object, taking parts away from it, or similar actions. Often an exploratory response includes more than one type of behavior.

Epistemic curiosity results from *conceptual conflict,* by which Berlyne (1963, 1978) meant conflict due to discrepant thoughts or beliefs or attitudes. The types of conceptual conflict are, for instance, doubt, perplexity, contradiction, conceptual incongruity, confusion and irrelevance. By the term "epistemic behavior" Berlyne referred to behavior whose function is to get information that can relieve or mitigate the conceptual conflict. Berlyne noticed that in exploratory behavior a person deals mainly with the perception of objects or events, whereas in epistemic behavior one deals with concepts and symbolic representations. Moreover, the function of exploratory behavior is to provide stimuli that will be immediately useful, whereas the function of epistemic behavior is to "equip the organism with knowledge," by which Berlyne meant "structures of symbolic responses."

Berlyne (1965) divided epistemic behavior into three categories, namely, epistemic *observation*, which includes different kinds of experimental and observational techniques, *consultation*, which includes asking other people questions or consulting books, and *directed thinking*. In everyday life, exploratory and epistemic behavior can frequently be intertwined, for instance, when a child leafs through an animal book and sees for the first time a picture of a kiwi. The child stops scanning and fixes his or her attention on the picture, after which the child directs questions like "What is this?" "Is this a bird or some other creature?" first to him- or herself and then maybe to a parent. If the parent cannot answer, the child may pose the same questions to a teacher the following day.

## 2.4.2 DIVERSIVE CURIOSITY

Day (1968) later extended Berlyne's definition of curiosity. To clarify the curiosity construct, Berlyne (1960) had divided exploration into *specific* and *diversive* exploration. Berlyne had stated that what he meant by curiosity is always specific, which means that the exploration that follows curiosity is always specific exploration. The aim of this specific exploration is to release the curious person from the subjective uncertainty caused by collative variables. By diversive exploration Berlyne meant behavior whose aim is to seek entertainment, new experiences or relief from boredom. Relevant for this behavior is that the collative properties of the sought-for stimuli are just right.

Following the thinking of Berlyne, Day (1968) suggested that curiosity can also be dichotomized into *specific* and *diversive curiosity*. According to Day, diversive curiosity is the condition of heightened arousal induced by a situation of changelessness, repetition, or monotony. Day and Berlyne (1971) specified that diversive curiosity also results from uncertainty, but leads to diversive exploration such as seeking entertainment or new experiences. Moreover, Hutt (1981) noticed that diversive exploration also includes playful behavior. Thus, the aim of exploration is not to reduce uncertainty but to increase the level of activation or to provide stimulation. That is, as Day and Berlyne (1971, p. 312) expressed it, "looking for collative variability." Day (1968) noted that what he means by diversive curiosity may be analogous to what Maw and Maw (1965) considered as "the need to seek new experiences" or to extend one's knowledge into the unknown.

In the definition of diversive curiosity presented by Day (1968, 1971), the level of optimal arousal is approached from its opposite, namely from specific curiosity. If the state of arousal, resulting perhaps from boredom, has fallen below the optimal level of arousal, then the result may be diversive curiosity. As Boyle (1983, p. 380) indicated, the resultant behavior depends on a person's cognitive appraisal of a low stimulus situation, which may induce a psychological state of either diversive curiosity or anxiety. The diversively curious person tries consciously look for new, amusing, or exciting stimuli in order to raise the level of arousal to the optimal plane. Exploration cannot be specific and diversive at the same time, but specific and diversive curiosity may temporarily form a sequence (cf. Hutt, 1970).

Later, Wohlwill (1981) reformulated the Berlyne-Day "specific-diversive" distinction with the terms "inspective" and "affective," since both concern responses to specific stimuli such as inspection for the sake of uncertainty or conflict reduction, or contemplation for the sake of enjoyment or pleasure. The inspective-affective distinction is no longer a mutually exclusive dichotomy as is the specific-diversive distinction (cf. Naylor, 1981). Instead of perceiving it as a dichotomy, Wohlwill perceived the inspective-affective distinction as a "continuum" (cf. Nunnally, 1981). Inspective and affective exploration are closely interdependent, and in practice exploration usually contains varying mixtures of these two. Wohlwill (1987, p. 64) emphasizes that "exploration of a stimulus may (and generally does) serve both an information-extraction and affect-production function." Wohlwill (1987) has since expanded the inspective-affective differentiation by adding the "genuine diversive exploration" concept to his theoretical model. Wohlwill described diversive exploration as an activity in search of stimulation, designed to relieve boredom, raise arousal, or the like. Voss (1987, p. 47) summarized Wohlwill's position as follows: "Whereas the functions of inspective and diversive exploration are uncertainty reduction and stimulus/sensation seeking, respectively, affective exploration is directed to the maintenance of an optimal hedonic tone." On the other hand the affective-emotional content relinks curiosity research with the psychology of aesthetic experiences, as Görlitz points out (Görlitz, 1987, p. 361).

## 2.5 COGNITIVE THEORY

The distinction between neobehaviorism and cognitive psychology is not sharp (see e.g., Leahey, 1991). Thus, for instance, Berlyne's neobehavioristic thinking has also had a strong influence on the cognitivist view of curiosity. On the other hand, motivational-emotional concepts have not been central to cognitive psychology. Most of the criticism received by cognitive psychology has concentrated on this point (see e.g., Eysenck & Keane, 1990). According to cognitive psychology, vague "activity" moves people. Thus, the concept of curiosity is implicit in the thinking of several cognitive theorists.

It is typical of the cognitive view that the subject is not seen as a passive perceiver but an active participant. Moreover, concept formation is also seen as an active intellectual process. Thus, cognitive psychology sees people as rational, logical, autonomous and intentional beings who interact with the external world. The interaction with the world is mediated by the mind, which is a symbol-processing, cognitive system. Cognitive theory reduces all questions and explanations of human behavior to the cognitive system (Slife & Williams, 1995). Schema and its variations is one of the central mental concepts in cognitive psychology (see Eysenck & Keane, 1990). The idea of schema first emerged in Kant's philosophical writings, but Bartlett (1932) was the first to use the term *schema*. According to Eysenck and Keane (1990, p. 275), a schema is a "structured cluster of concepts" which usually involves generic knowledge and may be used to represents events, sequences of events, precepts, situations, relations and even objects. A schema guides the perception process by offering expectations what is going to probably happen next and by directing perceptual exploration towards relevant environmental stimuli. The conceptual core of the cognitive view on curiosity is the idea that a person compares perceived objects or events to his or her schemes, and if the difference is within a moderate range, uncertainty and curiosity arise. A person attempts to reduce the cognitive uncertainty by exploration or by seeking more information. The outcome of exploratory behavior is the enrichment of the individual's cognitive structures.

According to Bruner (1966, p. 43), exploration has three aspects, namely activation, maintenance and direction. The major factor that activates exploration is an optimal level of uncertainty. Referring to Berlyne (1960), Bruner emphasized that curiosity is a response to uncertainty and ambiguity. By exploration a person attempts to restore the clarity of his or her cognitive system. In Bruner's theory curiosity has also adaptational value: it is a biologically relevant, intrinsic motive which is essential to the survival of the individual. Moreover, the curiosity to which Bruner (1966, p. 115) referred was specific in nature: "It is clear that unbridled curiosity is little more than unlimited distractibility. To be interested in everything is to be interested in nothing for long."

Jean Piaget's thoughts have strongly influenced the conceptions of curiosity, although he himself very rarely used the word *curiosity* in his writings. The concept of curiosity is mainly implicit in Piaget's theory (see e.g., Flavell, 1963; Voss & Keller, 1983). Explicitly, Piaget (1981, p. 18) defined curiosity or the "need to know" as an innate instinct-like cognitive function. However, later, McReynolds (1971) and Hunt (1963, 1971a,b), among others, made explicit much of that which Piaget left implicit concerning curiosity. Piaget related curiosity to the development of thought. Piaget's view was that the function of intelligence is to help the individual to adapt to the environment (Piaget, 1971). Adaptation occurs in stages where states of cognitive equilibrium and disequilibrium vary. Piaget refers to this process with the term *equilibration*. Equilibration is a process which tries to restore cognitive equilibrium in situations where, for instance, perceptions and the schema do not match. According to Piaget (1977), the essential driving force or motivational factor of cognitive development is "nonbalance" and the re-equilibrations which it involves. However, equilibration is not a static state but a dynamic process whose aim is "increasing equilibration." Thus, the subject's search for coherence is a central factor in the development of her or his cognitive structures. Assimilation and accommodation are fundamental processes in striving to achieve cognitive equilibrium.

McReynolds (1971) referred to curiosity with the term *cognitive motivation*. According to McReynolds, a person's experiences form cognitive structures and ultimately an overall "category system." Sensory and memory inputs are compared to this category system, and if the input "fits" the existing cognitive structure, the input is assimilated. However, if the input is incongruent with the existing cognitive structures, there are two alternatives. First, a person may alter the existing cognitive structure or invent a new one which is congruent with the input. McReynolds referred to this process with term *cognitive innovation*. Second, the input may remain unassimilated. McReynolds argued that people commit themselves to set up cognitive structures. Once one commitment is made, another tends to spring up. According to McReynolds (1971, p. 42), "man has a need to have needs, a motive to have motives." Thus, man is a goal-setting creature.

Hunt (1963, 1971a,b) included curiosity under the term *intrinsic motivation*. Hunt's thinking was based on the computer metaphor and the idea of the test-operate-test-exit (TOTE) unit. According to Miller, Galanter and Pribram (1960), an organism always tests input against some standard (test), and when *incongruity* is noticed the organism is activated (operates). Action continues until the incongruity is removed (test-exit). Following this idea, Hunt (1963) argued that there must be "motivation inherent in information processing and action." Hunt called this motivation *intrinsic motivation*. However, he noticed that intrinsic motivation is identical with what Berlyne (1960) labelled *epistemic curiosity*. According to Hunt, the notion of incongruity provides a basis for the instigation of intrinsic motivation, and the notion of *congruity* provides a basis for stopping the action. By incongruity Hunt (1963, 1981) meant the discrepancy between input stimuli and cognitive conceptions like exceptions, beliefs or previous understandings. On the other hand, an optimal amount of psychological incongruity is important for the awakening of intrinsic motivation.

Kagan (1972) based his thinking on Hunt and Berlyne. He viewed motives as cognitive representations of goals. This view implies that motives are only "a special class of idea." Kagan identified four kinds of motives: motives of sensory pleasure, hostility, resolution of uncertainty, and mastery. According to Kagan, uncertainty means the same as *cognitive conflict*, *cognitive dissonance*, and *cognitive disequilibrium*. Thus, the motive to resolve uncertainty might also have been called the motive to know or epistemic curiosity (cf. Berlyne, 1960). On the other hand, Kagan argued that one seeks uncertainty when one believes that one can deal with it.

Berlyne and Piaget have contributed a lot to the thinking of Beswick (1974). He described curiosity as a cognitive strategy which consists of an acquired predisposition to "create, maintain, and resolve conceptual conflicts" (p. 16). What Beswick meant by "conceptual conflict" seems to be very much the same as what Berlyne meant with the same term. Thus, conceptual conflict arises when the perceived situation is optimally strange, unusual, novel or unexpected if compared to pre-existing expectations. According to Beswick (1971), individual differences in curiosity are due to differences in the category system, which he called the "cognitive map" and the coding operation. Highly curious people have very differentiated cognitive maps, and thus the probability of conceptual conflicts is high. According to Beswick, assimilation changes the stimulus, whereas accommodation changes the category system. Thus, Beswick related curiosity to both openness and orderliness.

## 2.6 ECLECTIC MINI-THEORIES

There are many researchers who have borrowed ideas or constructs of curiosity from other theories. Two modern mini-theories of curiosity exist, namely the Dual Process Theory of Curiosity and Anxiety developed by Spielberger and Starr (1994) and an Information-Gap Theory advanced by Loewenstein (1994).

### 2.6.1 DUAL PROCESS THEORY OF CURIOSITY AND ANXIETY

Several researchers (e.g., Berlyne, 1960; Boyle, 1983, Keller, 1987; Schneider & Unzner, 1994) have argued that curiosity-instigating collative variables also arouse fear, neophobia and anxiety at the same time. Thus, curiosity and anxiety are antagonistic, and exploration is a compromise between both behavioral tendencies — curiosity, which is reduced only by inspecting, and anxiety, which leads to withdrawal or avoidance behavior.

Spielberger and Starr (1994) called theories which explicate incompatible exploratory and avoidance reaction to collative variables *dual-process* theories of curiosity and anxiety. These theories point out that when diversive curiosity (or sensation-seeking) is strong and anxiety is relative weak, diversive exploration is motivated. On the other hand, when anxiety is much stronger than diversive curiosity, avoidance behavior (flight reactions) will occur. However, the simultaneous experience of curiosity and anxiety seems to be symbiotic in motivating specific exploratory behavior when the physiological arousal associated with these emotional states is at or near an optimal level. According to Spielberger and Starr (1994, p. 233), at this level the "reduction in collative stimulus intensity will increase the individual's overall experience of pleasantness by reducing the unpleasantness associated with moderately high levels of anxiety, while the pleasantness associated with high curiosity remains unchanged." Thus, diversive and specific curiosity can be explained in terms of the intensity of curiosity and anxiety as emotional states. According to this view, which Spielberger and Starr called the "Optimal Stimulation/Dual Process Theory of Exploratory Behavior," the concept of specific curiosity drive appears to be redundant although it still is important to distinguish between diversive and specific exploration.

## 2.6.2 GAP THEORY

Loewenstein (1994) has proposed an "integrative interpretation of curiosity," an *information gap* theory. He combined ideas from drive theories, the incongruity perspective, and the competence approach with ideas from Gestalt psychology, social psychology and behavior decision theory. However, Loewenstein used term curiosity in a limited sense referring to state-like "specific epistemic curiosity," which he described as "an intrinsically motivated desire for specific information" (p. 87). The information gap perspective assumes that the person must be aware of the information gap before she or he can experience curiosity. By information gap Loewenstein meant a "discrepancy between what one knows and what one wishes to know." On the other hand, arousal of curiosity requires a knowledge base, as it is unlikely that curiosity about a certain topic can be created if there is no a prior knowledge about that topic. Moreover, the information gap perspective predicts a sudden increase in curiosity when the gap in information is realized. The increase of curiosity continues until the person "approach the goal of closing the information gap" (p. 89).

Loewenstein also explained the distinction between *voluntary* and *involuntary* curiosity. Although curiosity in principle is aversive, the explanation for curiosity-seeking behavior or voluntary curiosity is that "the process of satisfying curiosity is itself pleasurable" (p. 90). Loewenstein compared voluntary exposure to curiosity to a "gamble," that is, before a person participates a curiosity-inducing situation she or he estimates the likelihood that curiosity will be successfully satisfied. Loewenstein mentioned five situational factors that arouse involuntary curiosity:

1. Puzzling questions or riddles which include the hint that information is missing.
2. Presentation of a sequence of events with an anticipated but unknown outcome. This category includes, for instance, the desire to find out the murderer in a mystery novel.
3. Violation of anticipated expectations.
4. Confronting a situation where one knows that another person possesses information which is relevant also to oneself. Here Loewenstein gave the example of a situation where the parents also want to know the sex of a fetus when the doctor knows it.
5. Situations where information is "on the tip of one's tongue" are strong curiosity inducers.

## 2.7 CONCLUSIONS

The aim of this chapter was to describe the role of the term curiosity in five psychological theories. The starting point of this discussion was that the discourse in which curiosity plays an important role gives meaning to the term. The aim was not to give a synthetic definition of curiosity but to provide a theoretical background to the measurement of curiosity in the present study. Wittgenstein's concepts of language game and family resemblance were especially helpful in this process. As Wittgenstein (1980) noted in his *rope metaphor*, what holds the concept together and gives it unity is not a "single thread" running through all cases, but an overlapping of different fibres. Curiosity was analyzed using this metaphor.

The meaning of the term curiosity varies both within and between different schools of thought in psychology. Of the five schools of thought presented in the present study the most heterogeneous conception of curiosity is presented in psychodynamic theory. Nevertheless, psychodynamic theory is the only theory which tries to describe the developmental aspects of curiosity and exploration. Typical to the psychodynamic view of curiosity is that it is related strongly to the sexual instinct (Freud, 1971) or to the coping mechanisms of an individual (Bowlby, 1969). Bowlby's view that curiosity-exploration behavior is a coping mechanism and a choice behavior for attachment behavior is also characteristic only of the psychodynamic approach. Of the three psychodynamic views presented, the most effective for modern school motivation study has been White's theory of competence motivation. For White, competence motivation was a broad motivational concept that covers such behaviors as, for instance, curiosity, children's playful exploration and a need for excitement. Several modern theories of motivation (e.g., Harter, 1981; Deci & Ryan, 1985, 2000) that operate with term intrinsic motivation base their thinking to White's competence motivation concept. These theories have emphasized the relevance of competence feedback to intrinsic motivation.

Cognitive theory is a broad and heterogeneous approach to the human mind. However, Berlyne's neobehavioristic curiosity theory has influenced many cognitive psychologists. On the other hand, Piaget's theory also influenced Berlyne. Common to both approaches is that curiosity is seen as a state aroused by inner or outer perceptions that involve some kind of conflict and thus do not fit the person's existing cognitive structures (category systems). Berlyne referred to this nonbalance by term conflict and cognitive psychologist like Hunt with term incongruity. Otherwise than cognitive psychologists, Berlyne also specified that this specific state of subjective uncertainty is felt to be aversive. The aim of the resulting behavior, which Berlyne called exploratory behavior, is to resolve the conflict or discrepancy between input stimuli and cognitive structures.

If the conceptions of Berlyne and cognitive psychologists are compared on the active-passive dimension, it must be noted that Berlyne's conception is more passive than that of the cognitive psychologists. According to cognitive psychologists (e.g., Beswick, 1971), an individual is not a passive recipient of conceptual conflicts but also actively creates and maintains them. On the other hand, if

Berlyne's curiosity concept is extended by Day's diversive curiosity concept, then the neobehavioristic concept of curiosity is much broader than it is in cognitive theory. The Berlyne-Day approach has also had a strong influence on the measurement of curiosity.

Both humanistic theory and trait theory consider curiosity as an innate disposition which leads an individual to gather information. Both approaches also place curiosity as a part of a hierarchical system where it has connections to other concepts of the system. From the point of view of humanistic theory curiosity is closely related to creativity. Both Rogers (1969) and Csikszentmihalyi (1997) emphasized that curiosity is like a driving force of creativity. In humanistic theory, concept creativity thus also affects the meaning of curiosity. In modern trait theory curiosity as a part of the openness to experience dimension is also related not only to creativity but also to intelligence and cognitive abilities, for example divergent thinking. McCare and Costa (1997, p. 834) formulated this connection by arguing that "curious and imaginative people may become more involved in tasks that require flexible and fluent thought."

# 3. Concepts That are Closely Related to Curiosity

In order to better understand a concept, it is important to study the relationships it has to other concepts (see e.g., Danziger, 1997). Thus, the next objective of the present investigation is to study the use of four educationally interesting concepts and describe, if possible, how they overlap, what similarities they show to each other and how they differ. The four concepts are *curiosity*, *sensation seeking*, *interest*, and *intrinsic motivation*. The usage of these terms varies, and sometimes they are used as synonyms with almost equivalent meanings.

## 1.1 Sensation Seeking

Zuckerman's theory of sensation seeking utilizes the modern idea of optimal level(s) of arousal construct (see Zuckerman, 1984, 1987, 1994). He used term "sensation seeking" to describe a personality trait which shows itself in various forms of sensation seeking. Zuckerman used the term "sensation" instead of "stimulation" because he wished to emphasize the role of the subjective element of the stimulus; the same stimulus may produce different sensory experiences and emotions in different people. According to

Zuckerman (1979, p. 10), "it is the sensory effects of external stimulation that are most important." Recently Zuckerman (1994) has suggested that sensation seeking is a part of a broader personality trait called impulsive-sensation seeking (ImppSS). Zuckerman (1984, 1994; see also Geen, 1997) also assumed a biological basis for sensation seeking. At an early stage in his research Zuckerman (1979, p. 10) defined sensation seeking in the following terms:

> Sensation seeking is a trait defined by the need for varied, novel, and complex sensations and experiences and the willingness to take physical and social risks for the sake of such experience.

Later Zuckerman (1994, p. 27) modified this definition slightly:

> Sensation seeking is a trait defined by the *seeking* of varied, novel, complex, and intense sensations and experiences, and the willingness to take physical, social, *legal*, and *financial* risks for the sake of such experience.

In a factor analytical study (Zuckerman, 1971) 72 items related to sensation seeking were sharpened to four factors: Thrill and Adventure Seeking (TAS), Experience Seeking (ES), Disinhibition (Dis), and Boredom Susceptibility (BS). Zuckerman (1994, pp. 31-32) describes these factors as follows:

> TAS – Items on this factor reflected a liking for sports or other physical activities which include risk but provide unusual sensations of speed or defiance of gravity, such as parachuting, scuba diving, or skiing. An example item is "I sometimes like to do things that are a little frightening."
>
> ES – Items which loaded on this factor reflected a preference for seeking novel sensations and experiences through the mind and senses (music, art, and traveling). Nonconforming and unconventional life-style was also common to items which loaded on this factor. The essence of this factor is condensed to the item "I like to have new and exciting experiences and sensations even if they are a little frightening, unconventional or illegal."

Dis – Items on this factor expressed desire to seek sensation through social activites like gambling, "wild" parties, social drinking, and sex. The key item is "I like to have new and exciting experiences even if they are a little unconventional or illegal".

BS – This factor expressed a dislike of repetitive experience of any kind, including routine work or boring people. A person feels restless if things are unchanging. A typical item describing this factor is "The worst social sin is to be bore" (versus the forced-choice alternative: "The worst social sin is to be rude.")

In addition to the four subscales TAS, ES, Dis, and BS, Zuckerman (1971) also defined a general scale of sensation seeking (SSS IV). Gender moderated the results of the study and the fourth factor, boredom susceptibility, was identified only for women. The internal reliabilities for TAS, ES, and Dis scales ranged from .68 to .84 but the reliabilities for BS scale were lower, especially for women. The retest reliabilities for the general scale and the subscales TAS, ES, and Dis were good, ranging from .75 to .89. Later, three of the four factors identified showed good cross-gender and cross-cultural replicability. The BS scale is the only scale which has not been as reproducible as the other scales across populations. The ES scale has also shown some problems in factor replications. According to Zuckerman (1994), the reason for this is probably the culture-specific nature of some of the ES items. Thus, the most recent version of the SSS (form VI) uses only TAS and Dis scales and divides each type of scale into two sets: experienced activites (E) and intended or desired activites (I). The response format has also been changed to a three-point Likert-type response scale.

### 3.1.1 CURIOSITY AND SENSATION SEEKING

Several researchers (e.g., Ainley, 1987; Byman, 1993; Krapp, 1994; Trudewind & Schneider, 1994; Voss & Keller, 1983) have included the concept of sensation seeking in the concept of curiosity. Specifically, they have argued that sensation seeking comes very close to diversive curiosity in the epistemic-diversive curiosity distinction first suggested by Day (1968). Day and Berlyne (1971) described the diversive curious child as follows:

> The *Diversive Curious* child is one who seeks stimulation, creates excitement, and challenges the world around him. He seems willing to face adversity, take risks, and extend himself into new and daring situations. Of course, such a child must enjoy being in these situations and must react positively and with good feeling when he is in a situation high in collative variability. (p. 319)

In defining diversive curiosity and sensation seeking both Day and Zuckerman base their thinking on the theory of optimal arousal. According to this view an individual tries to maintain an optimal level of arousal. According to Day (1968, 1971), if the level of stimulation, resulting perhaps from boredom, sensory isolation or repetitiveness, has fallen below the optimal level, then the result may be diversive curiosity. Like Day, Zuckerman (1979) postulated that every individual has an optimal level of stimulation and arousal. Thus, Zuckerman defined sensation seeking as a need for varied, novel and complex sensations and experiences. However, Zuckerman made several things explicit which Day left implicit. He, for instance, defined the aspects of sensation seeking.

The hypothesis of a close relationship between sensation seeking and diversive curiosity is supported also by empirical studies. Olson and Camp (1984) found that the Diversive Curiosity scale correlated .45 with SSS Total. Moreover, Ainley (1987) showed that, Diversive Curiosity scale and Sensation Seeking scale loaded on the same factor. As Table 1 shows, Zuckerman (1994) has related sensation seeking and its dimensions also to several other constructs.

## 3.2 INTEREST

In everyday language, interest has many more positive connotations than curiosity: it is even desirable. However, the scientific concept of interest is hardly the same as its common-language everyday counterpart (see Byman, 1995). As a technical term, interest has been treated as an affective variable, a general arousal experience, or an emotion (Iran-Nejad, 1987). Piaget's cognitive-affective distinction has been the starting point of many studies of interest. According to Piaget (1971), every "action involves an energetic or affective aspect and a structural or cognitive aspect" (p. 5). Piaget (1971, 1981) argued that interest is the affective side of assimilation and accommodation. On the other hand, Izard (1991) emphasized that interest is the most frequently experienced positive emotion which also motivates the development of skills, competencies, and intelligence. At the experimental level Izard described interest-excitement as the feeling of being engaged, caught up, fascinated, or curious.

Krapp, Hidi, and Renninger (1992) distinguished two major points of view from which interest has been approached. One is interest as a characteristic of person and the other is interest as a psychological state aroused by specific characteristics of learning environment. Traditionally, the former approach has been termed with term "individual interest" or "topic interest" and the latter has been called "situational interest." According to Hidi (1990), individual interest develops slowly and tends to have long-lasting effects on a person's knowledge and values, whereas situational interest is an emotional state that is evoked suddenly by something in the immediate environment and that may have only a short term effect on an individual's knowledge and values. However, individual and situational interest are not dichotomous phenomena. Both types of interest concern person/environment interaction, and they are supposed to interact and influence each other's development.

Schiefele (1991) drew a conceptual distinction between a *latent* (disposition) and an *actualized* individual interest. A latent individual interest is a relatively enduring preference for certain topic, subject areas, tasks, contexts, or activities. Moreover, Schiefele suggested that interest is a content-specific concept as well as a directive force,

TABLE 1 *Summary of Relationships Between Sensation Seeking Scale (SSS) and Other Tests* (Zuckerman, 1994, p. 93)

| Near equivalent | Strong relationship | Minor relationship |
|---|---|---|
| Change Seeker Index–SSS Gen | Need Change–SSS Total, ES | Need Change–SSS TAS, Dis, BS |
| Novelty Seeking ES–SSS TAS | Novelty Seeking ES–SSS Gen | Novelty Seeking ES–SSS ES, |
| | Internal Sensations–SSS ES | Dis, BS |
| Stimulus Variation Seeking–SSS–Gen | Openness (to Experience)–SSS ES | |
| | Diversive Curiosity–SSS Total | |
| | Need Cognition–SSS ES | |
| | | Need Cognition–SSS Total, TAS |
| Venturesomeness–SSS TAS | Vent.–SSS Gen, Dis, BS | Monotony Avoidance–SSS TAS, ES |
| | Monotony Avoidance–SSS Gen, Dis, BS | |
| | RA–Musical Auditory–SSS Dis | Sensitivity–SSS (–) Total, TAS |
| RA–SSS Gen | Ra–Gen, Life-Style–SSS ES, Dis | Stimulus Screening–SSS(–)Gen |
| Arousal Seeking–SSS Gen, Total | Reactivity–SSS (–) Gen | |
| | Impulsivity–SSS Gen, Dis, BS | Impulsivity–SSS TAS, ES |
| | Cognitive Structure (–)SSS Gen | Cognitive Structure (–)SSS Dis, BS |
| | Risk Taking–SSS Total, TAS | Risk Taking–SSS Dis, BS, ES |
| | | Hypomanic–SSS Gen |
| | Need Play–SSS Gen, Dis | Need Play–SSS TAS, ES |
| | Anhedonia (–) SS Gen (patients) | Anhedonia (–) SS Gen(normals) |
| | Conservatism (–) SSS Gen | |
| | Aggression–SSS Dis (males) | Aggression–SSS Dis (females) |
| | Positive Emotional Exp.–SSS Gen | Fear of physical harm–SSS TAS |
| | Autonomy, Uniqueness–SSS Gen | Narcissism–SSS Dis |

TABLE 1 (CONT.)

| Multitrait systems | Strong relationship | Minor relationship |
|---|---|---|
| Pavlovian (Strelau) | | Strength of Excitation–SSS (+) Gen, TAS |
| | | Strength of Inhibition (–) Gen, Dis, BS |
| | | Mobility (+) SSS Gen, TAS |
| | | Balance (+) SSS Gen, TAS, BS, ES |
| Cattell | Independence–SSS Gen, Dis, BS | Superego (–) SSS Dis |
| Costa & McCrae | Openness–SSS ES | Openness–SSS TAS, Dis, BS |
| Eysenck | Agreeableness (–) SSS Dis, BS | Agreeableness (–) SSS Total |
| | Conscientiousness (–) SSS Total, BS | Conscientiousness (–) SSS TAS, ES, BS |
| | | Extraversion–SSS Total, TAS |
| | Psychoticism (P)–SSS Total, ES, Dis, TAS, BS | |

and that it consists of feeling-related and value-related valences. Feeling-related valences are feelings that are associated with a topic or an object, for instance feelings of enjoyment and involvement. Value-related valences refer to the attribution of personal significance to an object.

Unlike individual interests, which are always specific to individuals, situational interests are assumed to be spontaneous, fleeting, and shared among individuals. Situational interest may be evoked suddenly by such aspects of a situation as novelty, intensity, or complexity. If these aspects are seen as characteristics of an environment, then they contribute to the *interestingness* of the situation. Moreover, Deci (1992) emphasized that a situation or an object must also offer an "optimal challenge" in order to be interesting. However, situational interest can also be seen as a psychological state within the person.

Hidi and Anderson (1992) tried explicate the conceptual distinction between two psychological states of interest: actualized individual interest and situational interest. They hypothesized that the difference is in the area of affect. Feelings such as liking, enjoyment, and involvement are typically connected to actualized individual interest. However, the connection of situational interest and liking (or pleasingness) seems not to be so simple. Berlyne (1971a) first suggested that "interestingness may continue to rise, while pleasingness sharply declines, when moderate degrees of complexity are exceeded"(p. 217). Moreover, intense intellectual activity can also cause intense interest (Iran-Nejad 1987).

## 3.2.1 INTEREST AND CURIOSITY

The things that arouse curiosity also arouse interest, and, as such, they are what Berlyne (1960) called "collative variables". Persistence is a fundamental characteristic of both curiosity and interest. Exploration follows both curiosity and interest. Moreover, anxiety has been reported to have a negative connection to both curiosity and interest. Thus, it seems difficult to differentiate between curiosity and interest.

Tobias (1994) concluded that the curiosity construct has two advantages over interest. The first is that curiosity can be related to three apparently different states, which according to Tobias, are "an eagerness to approach some activities and situations motivated by curiosity and interest, neutral reactions, and disinterest leading to flight induced by anxiety from other material" (p. 47). The second advantage of the curiosity construct is that a number of measures are available for research. In this connection, Tobias referred to curiosity in its broad meaning, including sensation-seeking, novelty experiencing, academic curiosity, curiosity as an state, and curiosity as a trait (cf. Ainley, 1987; Byman, 1993; Langevin, 1971).

Hidi and Anderson (1992) later tried to make explicit the differences between situational interest and curiosity. According to Hidi and Anderson, the most important differences are: (a) situational interest can be elicited not only by collative variables, but also by content-specific text characteristics such as power, death, and sex; (b) situational interest may develop into relatively enduring individual interests; and (c) the inverted-U function does not necessarily characterize the relation between situational interest and the stimulus characteristics that elicit it.

Kirkland (1976) presented a sequential model that interlocks four concepts: attention, curiosity, skill and interest. The sequence begins when curiosity is triggered by "attention to an environmental anomaly." Sustained and persistent effort applied to resolve curiosity leads to skill development. Kirkland defines interest as the voluntary application of a skill. It develops from the successful resolution of puzzles. Recently Krapp (1994) advanced a theory which links the two concepts diversive curiosity and individual interest. Krapp emphasized that "diversive curiosity and exploration are not directed randomly at whatever objects or action possibilities happen to be available, but instead often exhibit a goal-oriented character. According to this view, interests are an important component of this phase of stimulus-search behavior" (p. 96). In addition, interests also play a "decisive role" in the content orientation of specific curiosity. Loewenstein (1994, p. 93) emphasized that interest "primes the pump" of specific epistemic state curiosity, which means that a person's pre-existing interests focus attention and in this way also effect their curiosity.

## 3.3 INTRINSIC MOTIVATION

Unlike the terms "curiosity" and "interest," the term "intrinsic motivation" is not used in everyday language. The concept is derived from Woodworth's (1918, 1958) "behavior primacy theory," according to which motivation consists fundamentally of dealing actively with the environment. The main idea was that even without extrinsic sources of motivation the organism would be active. According to Deci's (1992) interpretation, the concept intrinsic motivation emerged from the critique of Skinnerian operant theory and Hullian drive theory. These theories were not adequate to explain such activities as exploring novel spaces and manipulating objects, actions that seemed to have neither a direct nor an indirect relation to reinforcements.

Originally Deci (1975) split motivation into intrinsic and extrinsic motivation. Extrinsically motivated behavior is instrumental in nature. Such action are performed for the sake of some expected outcome or extrinsic reward or in order to comply with a demand. Intrinsically motivated behaviors, on the other hand, are engaged in, as Deci expressed it, "for their own sake and not because they lead to an extrinsic reward." Later Deci and Ryan (1985) limited the idea of the antagonistic nature of intrinsic and extrinsic motivation. Borrowing the concept of internalization from Schafer (1968), Deci and Ryan (1985) identified four types of extrinsic motivation: external, introjected, identified and integrated forms of regulation. Moreover, Harter and Jackson (1992) demonstrated that intrinsic-extrinsic motivation must be conceptualized both as a trait and a nontrait in order to fully understand children's motivational orientation to school subjects.

According to Deci and Ryan (1985, pp. 32-35), intrinsic motivation is based in the innate or psychological, organismic needs for competence and self-determination. Deci and Ryan used the concept competence in very much the same way as White (1959). They utilized the concept autotelic when they explained the teleology of intrinsic motivation. Csikszentmihalyi (1975) first used this concept when emphasizing the role of enjoyment or the inherent experiential aspects of intrinsically motivated behaviour: the reward is the ongoing subjective experience of enjoying the activity. Deci and Ryan (1985, p. 34) used this thought when they declared that "the emotions of enjoyment and excitement accompanying the experiences of competence and autonomy represent the rewards for intrinsically motivated behavior."

### 3.3.1 CURIOSITY AND INTRINSIC MOTIVATION

The relationship between curiosity and intrinsic motivation will depend on which conceptualization one chooses. According to Beswick (1974), curiosity is "the prototypical example" of intrinsic motivation (see also Bruner, 1966). Intrinsic motivation has also been used as a synonym for curiosity (e.g., Beswick, 1974; Rubenstein, 1986).

Usually the concept intrinsic motivation has a broader meaning than curiosity, and curiosity is only one component of intrinsic motivation (e.g., Berlyne, 1971b; Harter, 1981; Gottfried, 1985). On the other hand, if the concept curiosity is used in its broad meaning containing both specific and diversive curiosity, then the concepts intrinsic motivation and curiosity seem to be almost identical. This conclusion is implicit in Deci's theory. According to Deci (1975), there are two general kinds of intrinsically motivated behaviors. First, "when there is no stimulation people will *seek* it" (p. 61) or more detailed, people seek out challenges which are optimal for them. The other type of intrinsically motivated behavior involves, according to Deci, "*conquering* challenge or reducing incongruity."

Gottfried (1985, 1990) used the term *academic intrinsic motivation* in a broad sense to depict a special kind of intrinsic motivation for school learning. Academic intrinsic motivation involves enjoyment of school learning characterized by a mastery orientation; curiosity, persistence, taskendogeny, and the learning of challenging, difficult and novel tasks. Gottfried never exactly explains what she means by the term curiosity in her intrinsic motivation construct but implies that it is very similar to Berlyne's specific curiosity.

# 4. Measures of Curiosity

Different techniques for measuring curiosity have been developed. Following Langevin (1971) and Maw and Maw (1977), they can be classified as stimulus preference techniques, performance measures, self-description techniques and teacher-peer ratings. However this classification does not cover all existing measuring techniques of curiosity. In several studies (e.g., Moch, 1987; Trudewind & Schneider, 1994; Schneider & Unzer, 1994), for instance, question asking has been used as an indicator of curiosity, especially as an indicator of what Berlyne (1960) called "epistemic curiosity."

## 4.1 Stimulus Preference Techniques

So-called "Berlyne figures" present a classical example of stimuli used in several studies to measure state-like curiosity. Berlyne (1957, 1958) constructed a group of pictures which should awaken what he called "perceptual curiosity." Novelty, incongruity, complexity and surprisingness are characteristics of these pictures. Berlyne measured perceptual curiosity by presenting two or more pictures at the same time and measuring either the choice of pictures or the differences in looking time for each. Later, Kreitler, Zigler, and Kreitler (1974), among others, have used "Berlyne figures" to measure one aspect of curiosity. Black and white pictures, cartoons, colored pictures, and verbal absurdities have also been used to measure curiosity (see Maw and Maw, 1977).

## 4.2 Performance Measures

Those researchers (e.g., Hutt, 1981; Keller, 1994) who have measured curiosity as a state have often used performance measures to investigate curiosity. This technique has been used especially to assess the curiosity of young children. Performance measures are based on observations in a laboratory or in a free environment. In this context a standardized curiosity-instigating situation or toy has been presented to a child. Several researchers (e.g., Hutt, 1981; Henderson & Moore, 1979; Keller, 1994; Keller, Schölmerich, Miranda & Gauda, 1987) have used a special kind of toys, called "curiosity boxes," for measuring curiosity through visual, tactile, locomotary and manipulatory exploration.

Henderson (1984a,b) emphasized the role of context in curiosity studies. Parents, for instance, effect their children's exploration in several ways. They attract and direct their attention, but also ask and answer questions. On the other hand, parents often also forbid such kinds of exploration as touching and grasping. For this reason, some studies (e.g., Ainsworth et al., 1978) have been conducted in natural settings. Schneider and Unzner (1994), for instance, investigated the significance of mother-child interaction for exploration (cf. Bowlby, 1969) in natural settings. The observation environments were the home and the market.

## 4.3 Self-Report Techniques

Questionnaires have been based on various conceptualizations of curiosity. Only the most frequently used questionnaires are discussed in here.

Following the *State-Trait Anxiety Inventory* (STAI) format of Spielberger, Gorsuch and Lushene (1970), several researchers have used self-report questionnaires to investigate curiosity both as a state and as a trait. These inventories contain two almost parallel sections. In the state section the respondent answers how she or he feels "at this moment" or "right now," and in the trait section the respondent monitors how she or he "generally" feels. Naylor (1981), for instance, referred to Spielberger, Gorsuch and Lushene when he constructed the *Melbourne Curiosity Inventory* (MCIT), which includes both a trait and a state scale. The MCIT contains 40 items (20+20). Later, Olson (1986) used a state-trait distinction very similar to Naylor's in her curiosity inventory construction.

Day (1971) developed a 110-item inventory which he called the *Ontario Test of Intrinsic Motivation* (OTIM). By intrinsic motivation Day meant both specific and diversive curiosity. Thus, OTIM includes 90 items measuring specific curiosity (OTIMSC), 10 items measuring diversive curiosity (OTIMDC) and 10 items measuring socially desirable answering. The operational definition of the specially curious person was as follows:

1. he would show approach behavior in the presence of novelty, complexity, and/or ambiguity;
2. he would show some form of exploration in the presence of novelty, complexity and/or ambiguity, by attending to it, manipulating it, handling it, etc.;
3. he would investigate novel, complex sources of information which would tell him more about them;
4. he would explore novel, complex and/or ambiguous stimuli longer than familiar simple and/or clear stimuli. (Day, 1971, p. 109).

Several researchers (e.g., Ainley, 1985; Beswick, 1974; Langevin, 1976) later examined the validity of the OTIM.

In order to study the role of curiosity in motivation and learning, Leherissye (1972) developed the State Epistemic Curiosity Scale (SECS). Leherissye defined epistemic curiosity following Berlyne (1960). The content of the SECS items related to a student's desire to

1. know more about a learning task,
2. approach a novel or unfamiliar learning task,
3. approach a complex or ambiguous learning task, and
4. persist in information-seeking behavior in a learning task (Leherissey, 1972, p. 523).

Later, Boykin and Harackiewicz (1981) used "word frequency problems" to measure epistemic curiosity. The subject's task on every item was first to determine which of the four words occurred most frequently in written English. Thereafter, the subject was asked to use a 13-point scale to rate how interested they were in hearing the correct answer. This interest rating served as a measure of expressed epistemic curiosity.

Pearson (1970) has described curiosity by means of a four-dimensional concept which he named *novelty seeking*. Pearson defines novelty seeking as a "tendency to approach versus a tendency to avoid novel experiences." Later, Kohn and Annis (1975) gathered validity data in order to explore the validity of a modified version of Pearson's novelty experience scale. The results of a confirmatory item factor analysis supported the hypothesized factor structure of External Sensation, Internal Cognitive, Internal Sensations and External Cognitive; however, the factor structure fits women less well than men. As a result, Kohn and Annis conducted a separate exploratory factor analysis for each sex. The same basic factors occur for both sexes, but the composition of the factors differs in detail according to sex.

Ainley (1985) used items from 13 previously constructed inventories when she constructed the Two-Factor Curiosity Scale. Exploratory factor analysis and item/factor score correlations were used to determine the best items for the new scale. Based on the distinction first made by Langevin (1971), Ainley called the two sub-scales breadth of interest curiosity and depth of interest curiosity. Beswick (1974) and Rubenstein (1986) also used items from previously constructed scales in order to create a new curiosity scale.

## 4.4 TEACHER–PEER RATINGS

In order to validate paper and pencil measures of curiosity, Maw and Maw (1961) used peer and teacher ratings of curiosity. They defined a curious person as follows:

1. reacts positively to new, strange, incongruous or mysterious elements in his environment by moving toward them, by exploring them or by manipulating them;
2. exhibits a need or a desire to know more about himself and/or his environment;
3. scans his surroundings seeking new experiences; and
4. persists in examining and exploring stimuli in order to know more about them. (p. 299)

Later Berlyne (1963) and Day (1971) criticized the definition of Maw and Maw as confusing to the raters because it contains two forms of exploration. Parts 1 and 4 of the definition describe what Berlyne called "specific exploration" and part 3 and aspects of part 2 reflect exploration which Day labeled "diversive exploration."

# 5. Dimensionality of Curiosity

Dozens of instruments have been devised to measure curiosity (see e.g., Maw & Maw, 1977). This is a confusing situation. It is, for instance, difficult to compare results of studies which have used different curiosity measures. This has lead several researchers (e.g., Ainley, 1987; Boyle, 1989; Langevin, 1971, 1976; Olson & Camp, 1984; Rubenstein, 1986; Spielberger & Starr, 1994) to try to clarify the similarities and differences between existing scales. That is, they have tried to find out whether different measures of curiosity measure different kinds of curiosity. The answer to this question has been clearly no, but otherwise the findings of these studies have not been unanimous.

## 5.1 Breadth-Depth Model of Curiosity

Langevin (1971) was among the first who tested the hypothesis of the multifaceted nature of curiosity. Langevin (1971) investigated how seven various types of curiosity measures correlated. The results of the study revealed that two factors accounted for the correlations among the seven curiosity measures. Langevin named these curiosity factors Breadth of Interest and Depth of Interest Curiosity. Later, Ainley (1985, 1987) gave support to Langevin's two-dimensional curiosity model by factoring 12 frequently used curiosity scales or subscales (see Table 2). Ainley deviated from Langevin in that she named curiosity factors as two separate styles of behavior; that is, Breadth-of-Interest and Depth-of-Interest Curiosity. The breadth factor consists of a positive orientation towards varied and changing experiences. The depth factor consists of positive orientation towards complex ideas and puzzling phenomena and includes the attempt to understand them. According to Ainley (1985, p. 340), "both dimensions involve approach to novelty but the forms of novelty which prompt approach are distinctively different."

TABLE 2 *AINLEY'S TWO-FACTOR MODEL OF CURIOSITY*

| | Factor Loading |
|---|---|
| Factor One: Depth-of-Interest Curiosity | |
| NESIC (Pearson, 1970) | .68 |
| NESIS (Pearson, 1970) | .37 |
| NESEC (Pearson, 1970) | .75 |
| OTIMSC (Day, 1969) | .71 |
| OTIMSD (Day, 1969) | .38 |
| Test of Intrinsic Motivation (Beswick, 1974) | .78 |
| Melbourne Curiosity Inventory (Naylor, 1981) | .56 |
| Factor Two: Breadth-of-Interest Curiosity | |
| SSS TAS (Zuckerman, 1979) | .55 |
| SSS ES (Zuckerman, 1979) | .58 |
| SSS Dis (Zuckerman, 1979) | .55 |
| SSS BS (Zuckerman, 1979) | .41 |
| NESES (Pearson, 1970) | .57 |
| NESIS (Pearson, 1970) | .47 |
| OTIMDC (Day, 1969) | .38 |

The benefit of the concept pair represented by Langevin and Ainley has been questioned. Langevin (1971) himself claimed that the distinction of curiosity into "depth" and "breadth" may also reflect the distinction between specific and diversive curiosity as well as the distinction between state (C-State) and trait (C-Trait) curiosity. Giambra, Camp, and Grodsky (1992) called Ainley's two factors *Stimulation Seeking* and *Information Seeking*. On the other hand, there may also be a difference between paper-and-pencil tests and other tests in the background of these factors (cf. Langevin, 1971; Looft & Baranowski, 1971). Boyle (1989) also criticized Ainley for underfactoring and for using factor analysis in the manner of a self-fulfilling prophecy. However, Spielberger and Starr (1994) have noticed that the results of Ainley and Boyle are actually complimentary rather than contradictory. As seen in Table 3, the two-factor solution of Spielberger and Starr is very similar to Ainley's two-factor solution.

TABLE 3 *SPIELBERGER AND STARR'S TWO-FACTOR MODEL OF CURIOSITY*

| | F1 Information Seeking | | F2 Experience Seeking | |
|---|---|---|---|---|
| Subscale | Males | Females | Males | Females |
| MCIT | .76 | .77 | .10 | .05 |
| STCI[a] | .73 | .74 | .04 | .11 |
| OTIMSC | .72 | .68 | .04 | .10 |
| NESIC | .67 | .66 | -.09 | -.17 |
| NESEC | .61 | .65 | -.10 | -.11 |
| NESES | -.05 | -.04 | .75 | .82 |
| SSS-TAS | -.06 | .01 | .80 | .86 |
| SSS-ES | .24 | .05 | .42 | .51 |
| OTIMDC | -.01 | -.06 | .21 | .37 |

[a]The State-Trait Curiosity Inventory, Trait Curiosity (Spielberger & Butler, 1971)

Later, Byman (1993) reanalyzed Ainley's data by using CFA instead of EFA. The results of the reanalysis suggested that the Ainley's (1985; 1987) original two-factor model was statistically insufficient. The three-factor model was more successful in accounting for correlations among variables than the two dimensional "breadth-of-interest curiosity styles" and "depth-of-interest curiosity styles" model. It was also possible to find a meaningful name for the third dimension rejected by Ainley, which consists of the subscales SSTA and NESES and which may be the same dimension that Olson and Camp (1984) called "venturesomeness" and Kohn, Hunt and Hoffman (1982) "physical thrill-seeking."

Byman's (1993) reanalysis showed that two of the three dimensions, Experience Seeking and Venturesomeness, were not totally independent, both reflecting experience or sensation seeking. Byman suggested that it may be the emphasized subjective affective-emotional component of curiosity described by Wohlwill (1981, 1987) which links these two dimensions. On the other hand, the difference between these two factors was that the items measuring venturesomeness described actions in which it is possible to experience mild feelings of danger and fear. Byman argued that Zuckerman's (1979, p. 11) "risk" component is central to this dimension. In addition, it also seemed that a physical aspect was emphasized in this venturesomeness, and on this basis the name Physical Thrill-Seeking also suits this dimension. There was no desire to sense danger or fear in the items of the factor measuring experience seeking, the subjects seeking instead continuously new and many-sided aesthetic and other experiences from stimulants, social contacts, dressing, changing situations etc. Emphatic avoidance of boredom and routine was also characteristic of this dimension.

Byman (1993) emphasized that the factor called General Curiosity or Information Seeking seemed to be quite independent of the two other factors. On the factorial level this difference resembled the conceptual distinction which Wohlwill (1981) made between inspective curiosity and affective curiosity and Berlyne between specific curiosity and diversive curiosity. According to the narrow definition of curiosity represented by Berlyne (1971a, p. 100; see also Berlyne 1978, p. 98), diversive curiosity "has nothing to do with curiosity."

The subscale NESIS seemed to be the most problematic of the scales used by Ainley (1987), as the fit of this measure was very poor. According to Byman this could be caused by the conceptual ambiguity also referred to by Ainley. It was much more difficult to identify the reason for the poor fit of the NESEC subscale. Byman (1993) suggested that the high standardized residuals refer to a new dimension which is obscured by the number of variables (cf. Loehlin, 1987, p. 61). This factor could be very much sex-related, since almost one half of the items concerned technology, which is usually considered a masculine interest (cf. Kohn and Annis, 1975).

Taken overall, Byman's (1993) results supported Boyle's (1989) contention that the third factor rejected by Ainley has "highly significant loadings." The reanalysis also supported the finding that curiosity is not a unitary construct, and it clarified the differences and similarities among some widely-used scales. However, it must be remembered that the three-factor model suggested by Byman was only tentative, and another independent sample should be taken for confirmation.

## 5.2 GLOBAL STATE–TRAIT MODEL

Boyle (1983, 1989) criticized previous studies which tried the solve the dimensionality problem of curiosity as ad hoc interpretations of the curiosity construct. Boyle also argued that the previous studies contained several methodological weaknesses, as for instance use of excessively narrow scales and an inadequate factor-analytic methodology. Boyle agreed with Langevin's (1976) conclusion that the multifaceted nature of curiosity is an artifact of psychometrical problems in curiosity scales. Boyle (1983) viewed curiosity as a psychological system which is best described with an interactional and global state-trait model. According to this model, the resultant behavior depends on a person's cognitive appraisal of the stimulus situation (Figure 1).

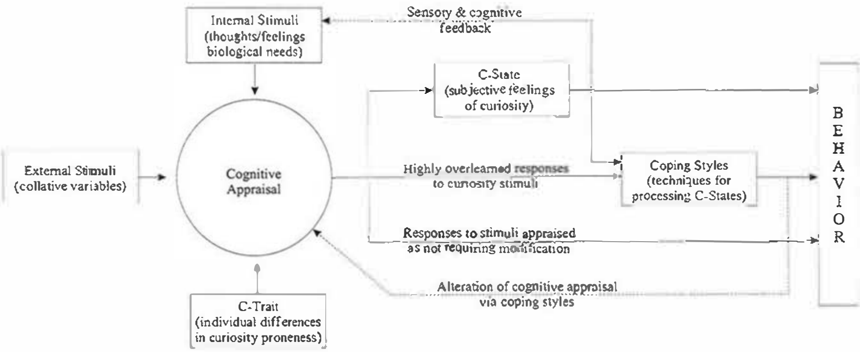



FIGURE 1. *CONCEPTUAL MODEL OF STATE-TRAIT CURIOSITY (BOYLE, 1983).*

Boyle's (1989) sample contained 300 senior secondary school students. Naylor's (1981) C-State and C-Trait scales were administered to the students together with Leherissey's (1972) State Epistemic Curiosity scale (SECS), and Spielberger, Edwards, Lushene, Monturi and Platzek's (1972) State-Trait Anxiety inventory (STAI). A 16 x 16 intercorrelation matrix was the starting point for the factor analysis of the subscale data. The interpretation of the emerging factors was based on a six-factor solution. This solution supported the state-trait distinction suggested by Boyle (1983, 1989). However, Boyle noted that much of the variance was also due to the different nature of the reversed and nonreversed item types used. Thus, both Reversed Curiosity and Nonreversed Curiosity factors came out, as well as Reversed A-State and Nonreversed A-State factors.

## 5.3 MULTIDIMENSIONAL MODELS OF CURIOSITY

Olson and Camp (1984) factor analyzed eight paper-and-pencil measures of curiosity. One of these measures was Spielberger's (1979) State and Trait Personality Inventory, which includes state-trait scales for curiosity, anxiety and anger. The results of the total score factor analysis and subscale factor analysis were very similar. In both analyses General Curiosity and Experience Seeking emerged as factors. In addition to this, Specific Curiosity and Venturesomeness factors appeared in subscale analysis (see Table 4)

Kohn, Hunt and Hoffman (1982) investigated aspects of experience seeking by factor analyzing Zuckerman's Sensation Seeking scale (Form IV), Mehrabian and Russell's (1973) measure of Arousal-Seeking Tendency (AST), Pearson's (1970) Desire-for-Novelty scale (DFN), Vando's (1970, 1974) Reducer-Augmenter Scale, and a slightly modified version (Kohn & Annis, 1975) of Pearson's (1970) Novelty Experiencing Scale (NES). The results of scale-factor analysis revealed that four factors can explain the correlations among the 12 subscales. The first factor was called the General Stimulation- and Arousal-Seeking factor, second Curiosity, the third Physical Thrill-Seeking, and the fourth Boredom and Desire for Change. Internal Cognitive Experience Seeking (NESIC), External Cognitive Experience Seeking (NESEC), and Internal Sensation Seeking (NESIS) subscales had a loading over .40 on the Curiosity factor.

Rubenstein (1986) used what she called item-level analysis to clarify the differences and similarities among six existing curiosity scales. Rubenstein used the terms "intrinsic motivation" and "curiosity" interchangeably. The research was exploratory in nature, which means that she used exploratory factor analysis techniques in order to clarify the dimensionality problem behind the curiosity construct. However, instead of analysing the total or subscale scores (cf. Ainley, 1985, 1987; Langevin, 1971), Rubenstein used individual scale items in her main analysis. However, as a preliminary analysis Rubenstein also analysed the total scale scores. The scales used were follow:

1. The Children's Reactive Curiosity Scale (RCRCS) (Penny and McCann, 1964)
2. The Choice-Motivator Scale – multiple-choice format (CMS) (Haywood, 1971)

TABLE 4 *FACTOR ANALYSIS OF SUBSCALES (OLSON & CAMP, 1984)*

| | Factor Loading |
|---|---|
| Factor One: General Curiosity (29.6%) | |
| Spielberger Trait Curiosity (Spielberger, 1979) | .93 |
| Spielberger State Curiosity (Spielberger, 1979) | .80 |
| Melbourne State Curiosity (Naylor, 1981) | .71 |
| Academic Curiosity (Vidler & Rawan, 1974) | .55 |
| Social Desirability | .39 |
| OTIMSC-Ambiguity (Day, 1971) | .35 |
| SSS ES (Zuckerman, 1979) | .35 |
| Factor Two: Specific Curiosity (12.8%) | |
| OTIMSC-Complexity (Day, 1971) | .88 |
| OTIMSC-Novelty (Day, 1971) | .87 |
| OTIMSC-Ambiguity (Day, 1971) | .71 |
| Academic Curiosity (Vidler & Rawan, 1974) | .63 |
| SSS ES (Zuckerman, 1979) | .35 |
| Melbourne Trait Curiosity (Naylor, 1981) | .31 |
| Factor Three: Experience Seeking (8.4%) | |
| SSS BS (Zuckerman, 1979) | .74 |
| SSS Dis (Zuckerman, 1979) | .74 |
| OTIMDC (Day, 1971) | .52 |
| SSS ES (Zuckerman, 1979) | .38 |
| Social Desirability | -.33 |
| Factor Five: Venturesomeness (6.0%) | |
| SSS TAS (Zuckerman, 1979) | .63 |
| Proverbs Test (Maw & Maw, 1975) | .48 |
| ACT | .37 |
| SSS ES (Zuckerman, 1979) | .31 |

3. The Ontario Test of Intrinsic Motivation (OTIM) (Day, 1971)
4. The Scale of Intrinsic Versus Extrinsic Orientation in the Classroom (SIEOC) (Harter, 1981)
5. The Intrinsic Intellectual Motivation Scale (IIMS) (Lloyd and Barenblatt, 1984)

Rubenstein did not interpret the results of the total scale score data, however the two-factor solution of the data resembled the factor structures which Ainley and Langevin got in their studies. The IIMS, OTIM-Specific, SIEOC, and CMS scales loaded on the first factor, and the second factor had only one high loading, namely OTIM-Diversive, which loaded very highly ($r$ = .97) on this factor. The two factors were almost orthogonal ($r$ = -.08).

From different item-level data solutions Rubenstein interpreted a seven-factor solution. The first factor was called Academic Curiosity, the second Enjoyment of Learning, the third Interest in Science (interest in machine/science/nature), the fourth Divergent Thinking, the fifth Openness to New Experience, and the sixth Belief in the Inherent Value of Learning. Rubenstein did not find a clear interpretation for the seventh factor. The intercorrelations among the factors ranged from .01 to .32 with a median of .16. The first two factors were most closely related to the other factors.

# 6. Issues and Problems in the Use of Factor Analysis

## 6.1 The Philosophical Background of Factor Analysis

Factor analysis has been traditionally divided into exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) (see e.g., Gorsuch, 1983; Leskinen, 1987). EFA is an inductive method which Mulaik (1987, 1988) placed in the category of exploratory descriptive statistics. Mulaik viewed exploratory statistics as a modern form of Baconianism. The aim of EFA is to find an optimal set of latent variables (factors) that account for the covariation among the observed variables. Mathematical factor analytic methods define the factors, the number of latent variables is determined in the analysis, and the interpretation of the latent variables is given afterwards (see e.g., Gorsuch, 1983, 1988). The principle of eliminative induction underlies the interpretation process: not only the salient loadings on a factor are examined, but also what variables do not load on that factor.

CFA is a hypothetico-deductive method. It based on a Kantian notion that knowledge is not based only on experience, but also on a priori mental factors which organize and structure experience (Mulaik, 1988). Scientific knowledge is obtained by continually testing our conceptions and ideas against experience. According to this view, a researcher should formulate explanatory hypothesis about, for instance, structures and relationships in nature and from this hypothesis logically deduce observable consequences. If these consequences are consistent with experience, the hypothesis is supported. According to this thinking, EFA can only suggest hypotheses which can be useful when these hypotheses later are tested with independent data (Mulaik, 1988).

CFA is a procedure in which a researcher a priori presents a hypothetical model of latent variables which influence the observed variables. The number and the interpretation of latent variables are given in advance. Whether a certain latent variable influences a certain observed variable is also specified in advance. However, the use of term "confirming" is confusing in the sense that a hypothetical model can never be confirmed (Popper, 1963). As Maruyama (1998) noted, a model can be disconfirmed or it can fail to be disconfirmed. In factor analytical thinking, this means that the model either fits the observed data or it does not.

## 6.2 CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis is a special type of structural equation modeling (SEM) (see e.g., Bollen, 1989). SEM is a technique where hypothesized relationships among a set of variables are defined by a series of statistical statements. The method evaluates the fit of the hypothetical model to the data. If the model is good, it produces a population covariance matrix that is very close to the sample covariance matrix. The closeness is evaluated with chi-square test and various kinds of fit indices. Structural equation modeling derives from path analysis. The difference between structural equation models and path analysis models is that structural equation models use both latent and observed variables, whereas path analysis models use only observed variables.

Structural equation models typically consist of two parts, the *measurement model* and the *structural model*. The measurement model relates the measured variables to the factors (or latent variables, as they are often called). The measurement model is normally a confirmatory factor analysis model (see Figure 2). The part of the model that presents the hypothesized relationships among the latent variables is called the structural model. According to Bollen (1989, p. 180), latent variables are the representations of concepts in structural

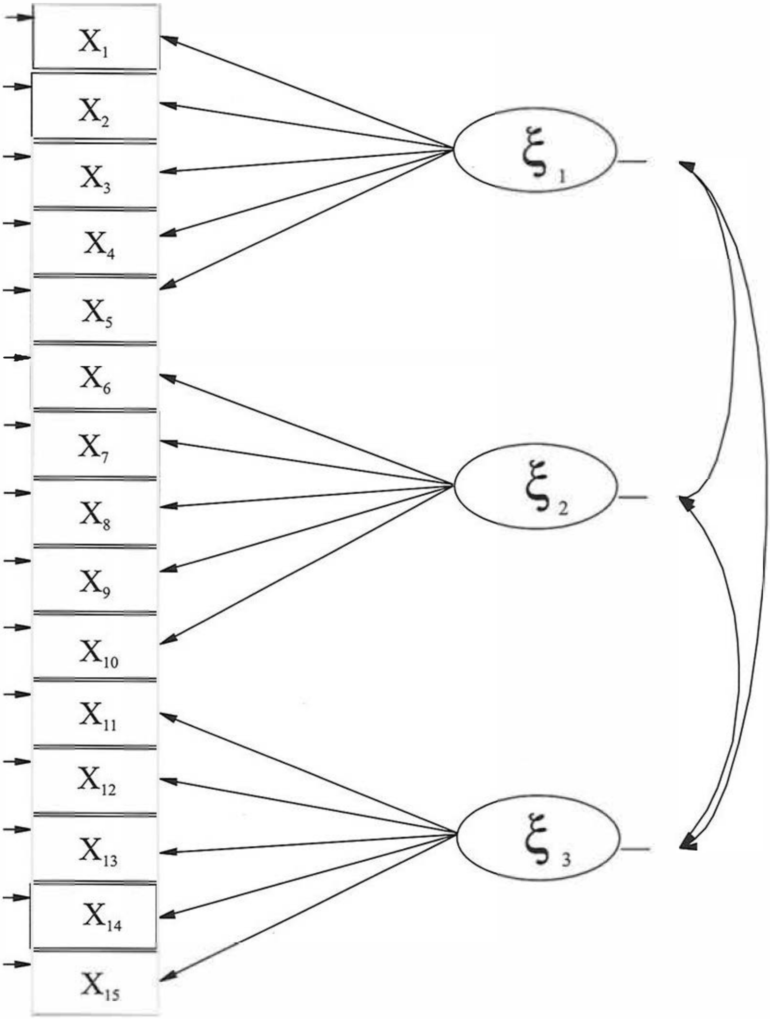



*FIGURE 2.* EXAMPLE OF A CFA MODEL.

equation models, and dimensions are the distinct aspects of a concept. Each dimension is presented by a latent variable in structural equation models.

Jöreskog (1993) distinguished between three kinds of model testing situations. In strictly confirmatory (SC) research, an a priori constructed model is tested and either accepted or rejected. In the second model several alternative or competitive models (AM) are tested, and one of these models is selected as best fitting (see e.g., Hull, Tedlie & Lehn, 1995). The third model is a situation where a tentative initial model fails and the researcher then tries to find a better-fitting model. This can mean that several post hoc models are tested (see e.g., Byman, 1993; Byrne, 1991). The respecification of each model can be theory- or data-driven. Jöreskog emphasized that this kind of approach is model-generating (MG) rather than model-testing. According to Jörskog, the MG situation is the most common. The problem with the AM situation is that it is often difficult to specify several alternative models a priori. In practice, the term confirmatory factor analysis nowadays usually means an analysis made by such programs as LISREL (Jöreskog & Sörbom, 1993a) or EQS (Bentler & Wu, 1993).

### 6.2.1 Advantages of the CFA Approach

As Bollen (1989) noted, the most distinctive difference between EFA and CFA is that in EFA the number of latent variables is not determined before the analysis. Moreover, in EFA all latent variables influence all observed variables, and measurement errors are not allowed to correlate. In contrast, in CFA a model is constructed a priori and the number of latent variables are also set in advance. In the CFA approach measurement errors may also correlate.

The CFA approach has many advantages over an EFA approach. Specially, it (a) produces a uniquely determined factorial solution; (b) makes it possible to statistically test and estimate the magnitude of fit of the whole model and the details of the model; (c) gives guidance to the improvement of the model; (d) makes it possible also to model the mismatches of the model; (e) allows the incorporation of both orthogonal and oblique factors within the same solution; and, (f) forces the researcher in advance to think the relationships of his or her concepts. In addition to this, the multi-group confirmatory factor analysis (MGCFA) approach also allows for the simultaneous analysis of several groups and the testing of factorial invariance across groups (Byrne, 1991; Hayduk, 1987, 1996; Jöreskog, 1979b). However, as Jöreskog and Sörbom (1989, p. 96; see also Bollen, 1989, pp. 226-232) noticed in practice, most studies are to some extent both exploratory and confirmatory.

### 6.2.2 TESTS OF INVARIANCE

Using the CFA approach it is also possible to test the invariance of the factor structure, that is, to test if the CFA model for one group has the same parameter values as that in another group. As Bollen (1989, p. 355) noted, there is a high risk of making serious errors when parameter values differ across groups. Several studies have shown, for instance, that gender can moderate the factorial validity of an inventory in various ways. First, the two gender groups can have different numbers of factors (e.g., Rappoport, Peters, Downey, McCann, & Huff-Corzine, 1993; Zuckerman, 1994). Second, the two groups can have similar factors, but different factor loadings (e.g., Byrne, Baron, & Campbell, 1993). Third, the two gender groups can have the same number of factors and the same loading patterns, but different variances and covariances among factors (e.g., Byrne, Shavelson & Muthén 1989). Even if the two groups are invariant in all three aspects, it is still possible that the factor means differ by gender (e.g., Bollen & Hoyle, 1990). To find moderation effects is normally an inductive process (e.g., Byrne, 1991; Byrne et al., 1989).

### 6.2.3 HIGHER ORDER FACTOR ANALYSIS

Both EFA and CFA can be divided to first-order and higher-order solutions. According to Bollen (1989), the first-order CFA is a data reduction technique where the primary goal is to explain the covariance matrix between many observed variables by means of relatively few underlying latent variables (see Figure 3). The first-order factor solution may have correlated factors. If so, it is possible to postulate second-order factors which are intended to account for the correlations among first-order factors (see Figure 3). This means

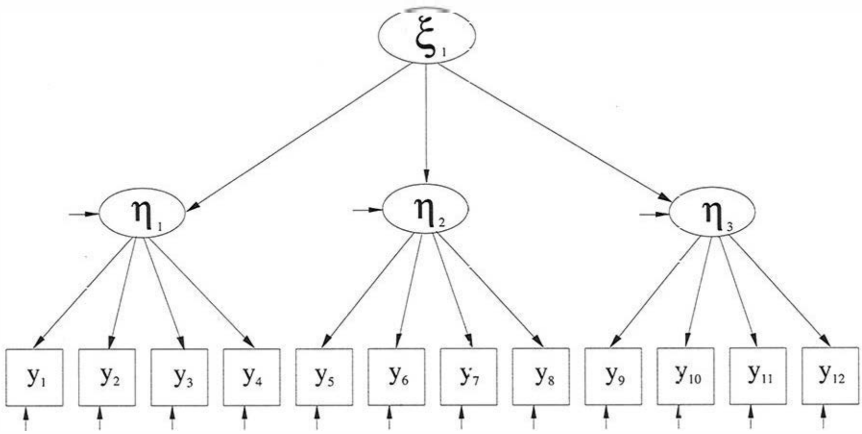



*FIGURE 3.* EXAMPLE OF A SECOND-ORDER MODEL.

that more general and abstract "second-order" latent variables determine the "first-order" latent variables (Bollen, 1989).

In principle, higher order analyses can continue still only one factor exist or there occur only uncorrelated factors exist. The main difference between first-order and higher-order factors is that the first-order factors are narrow whereas higher-order factors are broad in scope. Bollen (1989) noted that second-order factors can also explain the correlated measurement errors of the first-order factor model.

## 6.2.4 NESTED FACTOR MODEL

Alternatives to a second-order factor analysis have been presented (see Rindskopf & Rose, 1988). Recently Gustafsson and Balke (1993) have proposed a model which they called the *nested factor* (NF) model. The NF model has almost the same interpretation as the second-order factor analysis model, but it is easier to construct. A NF model contains a set of orthogonal first-order factors with different degrees of generality. The idea is that "the less general factors are nested within the more general factors" (Gustafsson & Balke 1993, p. 414). Compared to a higher-order model (HO), the NF model imposes direct effects from general factors on observed variables. In the HO model the higher-order factors have only indirect effects through lower-order factors on the observed variables. Figure 4 presents an example of an NF model with one general factor and three narrow factors.

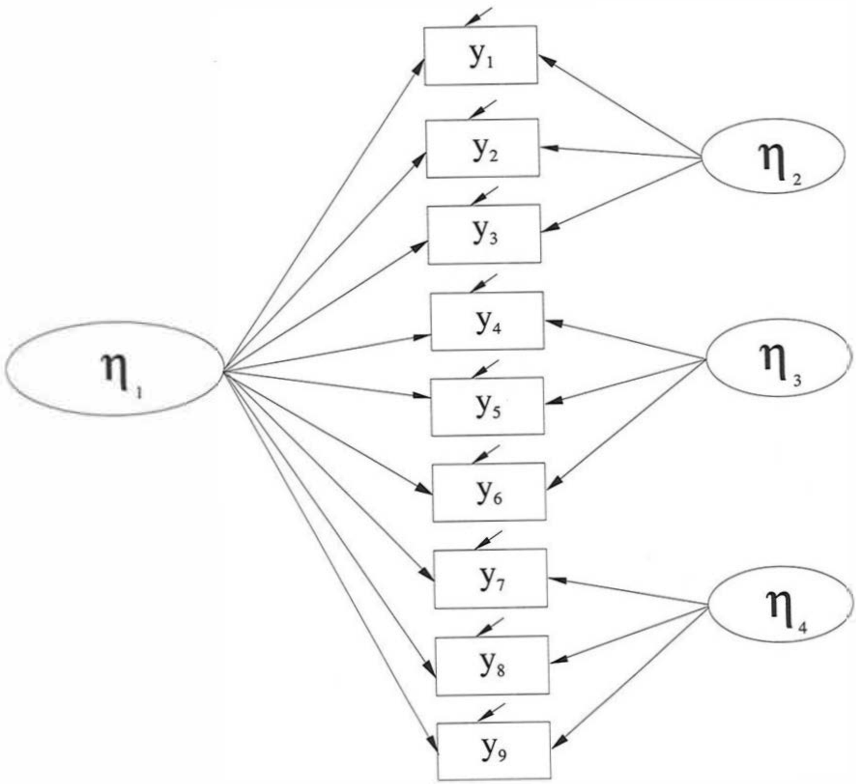



*FIGURE 4.* EXAMPLE OF A NF MODEL.

## 6.2.5 UNRESTRICTED FACTOR MODEL

James, Mulaik and Brett (1982) presented a three-step procedure for evaluating Structural Equation Models (SEM). In this procedure, a full SEM is divided into logically independent submodels, and at each step of the procedure a more restrictive model is compared to the less restrictive model of the previous step. Since the models are nested, it is possible to use the chi-square difference to test the fit of each successive model. The sequence of tests begins with the test of a *measurement model*, which usually is a confirmatory factor analysis model (CFA); the tests then continue until the null hypothesis is rejected. Based on this logic, it is possible to isolate the sources of lack-of-fit that arise in the model. Recently, Mulaik (SEMNET,[1] March 23, 1997; see also Mulaik & Millsap, 2000) has supplemented the three-step approach and created a four-step procedure by adding one new step before the test of the measurement model. This new step involves the test of an *unrestricted model*, which thus is the least restricted model tested in the nested sequence of models (cf. Jöreskog, 1979a, 1979b). The idea of this new first step is to test, without specifying a particular factor structure, whether the a priori hypothesized number of factors is correct.

Depending on the number, values, and positions of the fixed elements in $\Lambda$,$\Phi$, and $\Psi$ matrices, Jöreskog (1979a) divided the Confirmatory Factor Analysis (CFA) into unrestricted and restricted solutions. According to Jöreskog, in an unrestricted factor analysis model no restrictions are set on the common factor space, that is, $\Lambda\Phi\Lambda^T$ is left unrestricted. Jöreskog demonstrated that only $m^2$ fixed parameters, distributed appropriately on the $\Lambda$,$\Phi$, and $\Psi$ matrices, are needed to obtain an identified unrestricted model. According to Tepper and Hoyle (1996), the degrees of freedom of a correctly specified model can be expressed as $[p(p + 1)/2]-[(pm - m^2) + m(m + 1)/2+p]$, where $p$ is the number of indicators of latent variables. Conceptually, an unrestricted model is equivalent to an exploratory factor analysis model for the same number of common factors (Mulaik & Millsap, 2000).

In practice, an unrestricted factor model is specified by first fixing $m$ variances of common factors to unity and then freeing the covariances among the common factors. After this, a reference variable is chosen for each factor. The loading of this variable is freed on that factor and fixed to 0 on other factors. All remaining loadings are then freed. An alternative method is to free all elements of the Φ-matrix while fixing the loading of the reference variables to unity and 0.

### 6.2.6 FIT INDICES

Structural equation models should be evaluated from two aspects, namely the goodness-of-fit of the model and the parsimony of the model. According to parsimony principle, models should be as simple as possible. A common way to indicate the simplicity or the parsimony of SEM is to count the number of free parameters that must be estimated (Mulaik, James, Van Alstine, Bennett, Lind & Stillwell, 1989). However, McArdle and Cattell (1994, p. 64) noted that the parsimony of the multi-group model should also be based on "comparisons of the benefits of invariance over groups against the costs of complexity of a factor pattern." Moreover, the goodness-of-fit criteria for a multi-group model should be the model's capability to account for the observed relationships in several groups. Goodness-of-fit and parsimony are interdependent concepts because complex models fit data better than simple ones. On the other hand, when two models fit to the data almost equally well a common suggestion has been that the more simpler is chosen (see e.g., Bollen, 1989).

The most common index of overall fit of a model is the chi-square test. However, there are many hidden assumptions and problems behind this test (see e.g., Bollen, 1989; Jöreskog, 1993). Thus, as Chou and Bentler (1995) noted, the test statistic follows the $\chi^2$ distribution only if the assumptions of large sample, model specification, and distributions of test variables are correct. If, for instance, the distribution of variables does not fulfill the multivariate normality assumption the scaled $\chi^2$ statistic should be used. A second well-known problem with the $\chi^2$ test is that its power is strongly linked to the sample size. If the sample size is sufficiently large, the $\chi^2$ test will even reject models that approximate the covariance or correlation matrix very closely. Thus, Bollen (1989), for instance, recommended that chi-square estimates along with several other fit indices should be used when the overall fit of the model is estimated. However, there is no consensus about which alternative overall fit measure is superior. One common way to divide fit indexes is to classify them to absolute versus incremental fit indexes (e.g., Hoyle & Panter, 1995; Hu & Bentler, 1995; Tanaka, 1993). An absolute fit index directly assesses how well an a priori model matches the observed covariances. A $\chi^2$ test, for instance, has been constructed so that it actually measures the "badness of fit." Thus, a value of zero indicates optimal fit (cf. a saturated model).

The other type, the incremental fit indexes, are based on the logic whereby the model in question is compared to a more restricted baseline model, usually the null or independent model, which usually suggests that no factors underlie the observed variables (see Bollen, 1989, pp. 269-276; Tanaka, 1993). Hu and Bentler (1995) defined three groups of incremental fit indexes: types 1, 2, and 3. Bentler and Bonett's (1980) normed fit index (NFI) is a classic example of a type-1 incremental fit index. According to Hu and Bentler (1995, p. 83), "NFI represents the proportion of total covariance among observed variables explained by a target model when using the null model as a baseline model." However, Hoyle and Panter (1995) recommended that only type-2 and type-3 indexes should be used when the fit of a structural equation model (SEM) is reported.

### 6.2.7 POSSIBLE REASONS FOR THE FAILURE OF A FACTOR MODEL

Factorial validity is an instance of construct validity. According to Messick (1993, p. 19), it "corresponds to the test's external structure." Thus, a situation where an a priori suggested factor model fails means also that the construct validation or the interpretation of the data fails. That is, the test items do not behave in a manner consistent with theory. On the other hand, in disciplines such as the behavioral sciences it is implausible to assume that conceptual variables are measured perfectly. It is plain, for example, that not a single item in an inventory or scale is pure in its meaning (Cronbach, 1990). Moreover, the response format used, the method, the particular order of items etc., may influence the scores.

The possible variance components of instruments using written language are multiple. Voss and Keller (1983, p. 67) noted that "varying language levels" should be kept in mind when designing verbal self-reports for investigating curiosity and exploration. Johnson (1997) argued that when personality is assessed with questionnaires, both pragmatic and semantic misunderstandings occur. According to Johnson, "pragmatic rules are implicit social conventions about meaning that can vary across subcultures who share the same language" (p. 81). Groves (1989, p. 450) differentiated three types of measurement errors associated with words. First, because different groups use different vocabularies, it is possible that the respondent can give no meaning to a used word. Second, a word can have different meanings to the same respondent. Third, a word can have different meanings for different respondents. In several studies (e.g., Boyle, 1989; Hoyle & Lennox, 1991) a typical method effect has also been that reverse-worded items have introduced common method variance to the measurement. Hoyle and Lennox (1991) called this kind of extra factor the *response bias factor*. Moreover, Byrne (1994) has argued that highly overlapping item content can lead to systematic error variance and correlating measurement errors.

In contrast to exploratory factor analysis, in a confirmatory factor analysis it is possible that the $\varepsilon$ terms are correlated or, in other words, that measurement errors (disturbances) are correlated (Bentler and Chou, 1987; Bollen, 1989). However, when the errors are correlated and the correlations are substantially meaningful, it is doubtful that these are "errors." According to Hayduk (1987, pp. 191-193; 1996, p. 31), in such cases a better way to proceed is to replace the measurement errors with concepts and bring them into the model as model segments. Thus, using CFA methodology it is possible to separate out different kinds of variance components of a variable concerning components of meaning as well (see e.g., Carlson & Mulaik, 1993).

Unexpected extra systematic variance can be the reason why an a priori postulated model does not fit the data. When a tentative initial model fails it is usually useful to continue the analysis in exploratory mode in order to find a modified model that fits the data better and thereby explains the misfit of the initial model. Jöreskog

(1993) called this situation the model generation (MG) phase, thus emphasizing that the aim of the analysis is model generation rather than model testing. According to Jöreskog, model generation can be data- or theory-driven. However, the risk of data-driven model modification is a high likelihood of capitalization on chance factors (MacCallum, 1995; MacCallum, Roznowski & Necowitz,1992).

[1]It is possible see the SEMNET Discussion List archives in web-address: http://bama.ua.edu/archives/semnet.html

# 7. DESIGN AND PROCEDURES OF THE EMPIRICAL STUDY

## 7.1 OBJECTIVES OF THE STUDY

Factorial validity is an important component of construct validity for a measurement instrument with several subscales. Construct validity is supported if the factorial composition of the scale is consistent with theoretical expectations (Anastasi, 1988; Floyd & Widaman, 1995). Five previously constructed curiosity scales are used in the present investigation. All five of the instruments have a hypothetical factor structure. Thus, the first aim of the present study is the following:

1. To test the factorial validity of the used inventories.

Cronbach (1990, p. 159) stated that every test is to some degree impure, which means that there are always several unavoidable sources of variance in all test scores. One part of the construct validation process is to identify impurities in the measurement. One form of impurity might be that the test scores are corrupted by systematic sources of error. Groves (1989) has also suggested several measurement errors associated with words in written inventories. Following this logic, the first sub-problem of the present study is:

1.1 To investigate what "impurities" the selected measurements contain.

There can be no valid comparison of the means for girls and boys if the measurement instrument is not invariant. As Hoyle and Smith (1994) have noted, comparison of means when the measurement is non-invariant is like "comparing apples and oranges." A great deal of evidence from several studies (Moch, 1987; Olson, 1986; Keller et al., 1987; Voss & Keller, 1983) indicates that boys and girls show both qualitative and quantitative differences in exploration. However, before firm conclusions can be drawn about these gender differences, it is necessary to first confirm that the differences observed are not due to differently valid measuring instruments. For example, when a written instrument, such as a self-report inventory, is used, it is even likely that gender differences exist in understanding the meaning of the words and sentences used (Groves, 1989). Thus, the second research sub-problem of the present study is the following:

1.2 Does gender moderate the factorial validity of the inventories?

Classical test theory claims that at least eight items per factor are needed when a scale is constructed. In an SEM context this claim has frequently led to badly fitting models. A very popular strategy has been to undertake a post hoc analysis in order to eliminate badly fitting items from the model. However, the risk involved in data-driven elimination of items is that the meaning of the concepts may change. In a situation where a multiple indicator model fails, Hayduk (1996, pp. 29-30) has recommended a different strategy which guarantees that the conceptual model will not change, or at least that the changes will be conscious. Thus, the second research problem of the present study is mainly methodological. The investigation tries to answer the question:

2. What should be done when an unrestricted factor model fails?

There are both implicit and explicit beliefs that gender is related to curiosity and exploration. It is, for instance, a common belief that girls are more curious than boys. Thus, the third research problem of the present study is:

3. Are there gender differences in exploratory behavior as a trait?

de Vaus (1991, p. 49) has noted that conceptual clarification is not a "once-and-for-all process" which precedes empirical research. According to de Vaus, conceptual clarification and data analysis are interactive processes. The results of data analysis can help to clarify concepts. However, de Vaus emphasized that the clarifying process must begin at the theoretical level.

Several theories of curiosity exist. Because of this, a large number of scales measuring curiosity and related constructs have been developed. Thus, the question arises whether several different kinds of curiosity exist or whether the problem is terminological and whether these different scales actually measure the same kind of curiosity. Using conceptual analysis techniques and the results of some previous studies, it is possible to construct some conceptual models of certain trait curiosity conceptions. Moreover, using the corresponding measurement instruments, it is also possible to statistically test these models. The fourth research problem of the present study is:

4. Which of the suggested models best accounts for the covariance matrices of the selected subscales?

## 7.2 SUBJECTS

A sample of 529 Finnish fifth-graders from the southern part of Finland was used. The sample consisted of 24 classes from 14 schools and four towns, viz. Helsinki, Lahti, Orimattila and Nastola. This sample was divided according to sex (263 girls and 277 boys). The curiosity inventories were administered to all the pupils during class time. The pupils completed the inventories during two successive hours. At the beginning of both hours two sample items were presented first. The test instructions, sample items, and the questionnaires were administered by the author. All told, the administered questionnaires contained 120 items.

## 7.3 THE MEASUREMENT INSTRUMENTS AND THE CONCEPTUAL MODELS BEHIND THEM

Campbell and Fiske (1959) first demonstrated that the method used to measure human traits can introduce additional common variance to the measurement. This has been especially problematic in attempts to measure curiosity and exploration (Langevin, 1971). To make it possible to extract both method and trait variance, two kinds of methods were used in the present study in order to measure curiosity (see e.g., Maruyama, 1998): four pencil-and-paper self-report scales and one teacher rating scale. The self-report scales were chosen with three major constraints. First, the scales chosen had to be suitable for children, or it had to be possible to modify them for children. Second, the scales had to present the same type of items and contents which have been used to measure curiosity in several other studies (e.g., Ainley, 1987; Langevin, 1971, 1976; Spielberger & Starr, 1994). Third, the development of each scale should have been quite independent of that of other scales so that together they potentially present diverse conceptions of the curiosity construct. According to these criteria, the following four self-report scales were selected for the present study:

1) Curiosity Inventory for Junior High School Students (the trait scale) (Olson, 1986), 40 items
2) Test of Intrinsic Motivation (Beswick, 1974), 15 items
3) Ontario Test of Intrinsic Motivation (the Diversive Curiosity scale) (Day, 1969), 10 items
4) Sensation Seeking Scale (Björck-Åkesson, 1990), 55 items

Teacher ratings were used as an alternative way to evaluate curiosity. This method has also been used in several other studies (e.g., Langevin, 1971; Coie, 1974; Henderson, 1994). The present teacher rating scale is based on the scale developed by Maw and Maw (1964). Together, the chosen five scales included 12 subscales and 120 items. For the present investigation, all items were translated to correspond to the vocabulary of a fifth-grader. To check the translation, techniques recommended by Brislin (1986) were used. The Finnish translations of the items were translated back into English by an outside expert and then compared to the original items.

### 7.3.1 TEST OF INTRINSIC MOTIVATION

The first curiosity inventory of the present study based on the "cognitive process" theory of curiosity (Beswick, 1971, 1974). According to this theory, curiosity is a predisposition to create, maintain and resolve conceptual conflicts. By means of investigator acts (exploration), assimilation and accommodation, a person tries to resolve conceptual conflicts and, in this way, to restore the balance of his or her category system.

Beswick used intrinsic motivation as a synonym for curiosity. The test contained 16 items which Beswick gathered from other tests, for instance from OTIM (Day, 1968). Beswick labeled the 16-item scale the Test of Intrinsic Motivation (IM). Internal consistency of the items and correlations with other measures of curiosity were the criteria used in the item selection process. According to Beswick, alpha coefficients of the IM scale varied from 0.68 to 0.72. In the present study, the original five-point response format was changed to an 11-point format to obtain "continuously measured data variables" (Comrey & Lee 1992, p. 225). The measurement scale was redesigned to form a continuum from 0 (Never) to 10 (Always). Moreover, Item 11, "Some truths can only be expressed in paradoxical statements," was eliminated from the original scale because it was too difficult for fifth grade pupils. Moreover, Items 12 and 14 were simplified so that a fifth grader can understand them.

Beswick (1974) never tested the dimensionality of IM with factor analysis, but because acceptable level of internal consistency was one criterion of the item selection, it was suggested that only one factor, intrinsic motivation, underlies the responses to the 15 test items. Later, Ainley (1987) also used IM as a unidimensional measure of curiosity. In Ainley's study, IM correlated highly (r=.52) with Naylor's (1981) Melbourne Curiosity Inventory (Trait form) and Day's OTIM-specific subscale (r=.60). In Beswick's (1971) study the IM scale correlated .70 with OTIM total score. The simple conceptual model of IM is displayed in path diagram form in Figure 5.

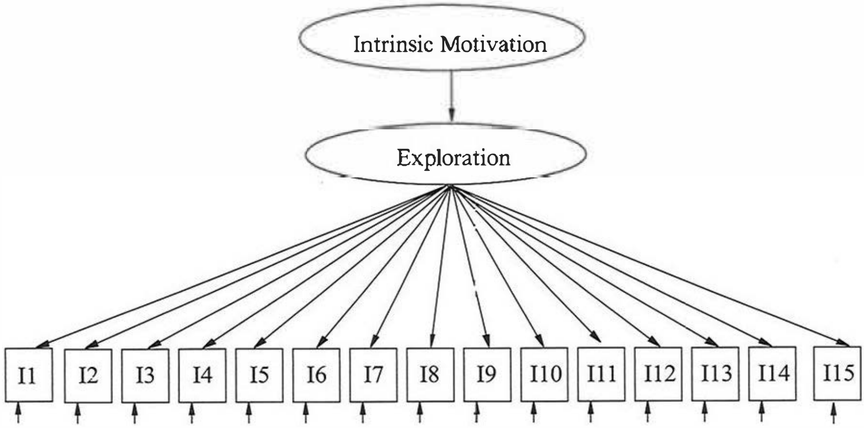



*FIGURE 5*. CONCEPTUAL MODEL FOR THE TEST OF INTRINSIC MOTIVATION.

## 7.3.2 DIVERSIVE CURIOSITY

The second inventory of the present study was a subscale of OTIM (Day, 1968, 1971), namely the diversive curiosity scale (OTIMDC). In the definition of diversive curiosity, the level of optimal arousal is approached from the opposite side to specific curiosity (cf. Berlyne, 1960). If the state of arousal, resulting perhaps from boredom, has fallen below the optimal level of arousal, then the result may be diversive curiosity. The diversively curious person tries consciously look for new, amusing or exciting stimuli in order to raise the level of arousal to the optimal plane. Thus the aim of exploration is not to resolve or mitigate the uncertainty but to increase the level of activation or to provide stimulation. The diversive curiosity scale contained 10 items which described different forms of diversive exploration. As in the study of Ainley (1987) and Rubenstein (1986), OTIMDC was used as an independent unidimensional scale in the present investigation. Thus, the path diagram presented in Figure 6 illustrates the corresponding conceptual model. To make it easy for the children to use the scale, the original five-point response format was changed to the same kind of 11-point format as in the IM scale.

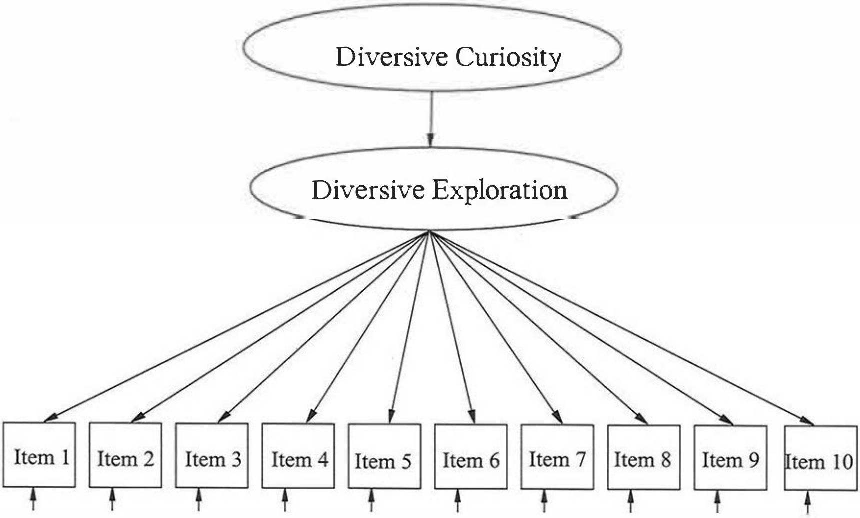



*FIGURE 6*. CONCEPTUAL MODEL FOR THE DIVERSIVE CURIOSITY SCALE.

## 7.3.3 THE BROAD C-TRAIT SCALE

The third Curiosity-Exploration inventory of the present study was based on the Curiosity Inventory for Junior High School Students created by Olson (1986). Olson's Curiosity Inventory measures four of the five exploration types first identified by Kreitler, Zigler, and Kreitler (1975; see also Kreitler & Kreitler, 1986, 1990, 1994), namely, *Manipulatory Exploration, Perceptual Exploration, Conceptual Exploration,* and *Exploration of the Complex*.

Manipulatory Exploration is elicited by objects which are new in some respect and which can be explored in order to find out how they operate and what their tactile and thermal qualities are. The Manipulatory Exploration factor may be gender-related: on all those variables which loaded highly on this factor boys scored significantly higher than girls (see also Keller, Schölmerich, Miranda, & Gauda, 1987; Voss & Keller, 1986). Perceptual Exploration refers to activity whose aim is to reveal the referent's sensory qualities and sensations, mainly what is seen and heard, often to find out what something is made of. Conceptual Exploration represents active information seeking through asking questions and analytically checking commonly accepted concepts, focusing on what is implied. Recently, Kreitler and Kreitler (1994) have noted that Conceptual Exploration corresponds to what Keller et al. (1987) have called "verbal exploration." Exploration of the Complex or Ambiguous as a factor refers to a preference for a complex stimulus when both a simple and a complex stimulus are present. In describing Exploration of the Complex, Kreitler and Kreitler (1986) emphasized the subjective or internal aspect of perceptions, namely sensations, feelings and emotions. Recently, Kreitler and Kreitler (1994) have emphasized that the four exploratory modes are personality traits and the underlying motivation may be curiosity (C-Trait). The distinction made by the Kreitlers resembles the traditional distinction made in Trait Theory between outer and inner traits (see Johnson, 1997).

Olson used a deductive approach (see e.g., Burisch, 1984) in constructing her inventory. In the content validation process, four experts were used to evaluate the content of the inventory items to ensure that they represented and covered all four exploratory concepts. Numerous revisions were made after their evaluation until a consensus about the item contents was obtained. Teachers' ratings of their students' curiosity were used for concurrent validation of the inventory. The final inventory included a trait scale and a state scale, a total of 80 items. In the present study, only the trait scale (40 items) was used. The alpha coefficient of the trait scale (N=509) as reported by Olson was 0.92, and the alpha coefficient for the four subtests ranged from .75 to .79.

As Olson noted, in addition to testing reliability, content validity and concurrent validity, it is also necessary to test the factorial validity of the scale to determine whether the 40 items really reflect four different dimensions or types of exploration. Olson's inventory consisted of 40 generally worded and specifically worded items, and responses were obtained using a five-point Likert scale (almost never, sometimes, half the time, oftentimes and almost always). The students were asked to circle the number which indicated how they *generally* felt. For the present study, the original five-point response format was changed to an 11-point format that corresponded to the format of the IM and OTIMDC scales. Finally, the places of items 1 and 27 were reversed to provide an easier and more suitable start for the fifth-graders. Otherwise, items appeared in Olson's original order. One example item was provided on the instruction sheet, and this example was reviewed with the students before administering the actual inventory. The conceptual model for the C-Trait scale is similar to Figure 3.

### 7.3.4 SENSATION SEEKING SCALE

As Zuckerman (1994) has noted, it has been a problem that only a few sensation seeking scales (SSS) for children exist. Eva Björck-Åkesson (1990) has developed one of the existing scales (SESE). It was designed to measure sensation seeking in Swedish preadolescent children (6th – 9th graders). Zuckerman's original SSS included four subscales, namely Thrill and Adventure Seeking (TAS), Experience Seeking (ES), Disinhibition (Dis), and Boredom Susceptibility (BS). However, many of these subscales contained items which were not suitable for children, for instance ES and Dis subscales contained many items concerning sex, alcohol and drugs. To make these items appropriate for children, Björck-Åkesson changed the content of many items. The final version of the SESE contained 56 items, of which 22 were original Zuckerman items (Form IV). Most of these items measured thrill and adventure seeking. The same forced-choice item format was used as in Zuckerman. To minimize the effects of culture, the sensation seeking scale used in the present investigation was based on the version of Björck-Åkesson. Items of this scale were translated into Finnish. The forced-choice response format was used also in the Finnish version of the SESE. Item 3 of Björck-Åkesson's scale, "It is fun to go to a discotheque," was eliminated because it was unsuitable for fifth graders.

Björck-Åkesson (1982, 1990) used both EFA and CFA techniques to invent and test several factor models. The best-fitting model was a nested factor (NF) model with five orthogonal first-order factors that had different degrees of generality. In that model the four less general factors were nested within the more general factor. The NF model was based on the claim made by Gustafsson and Blake (1993) that it is better to formulate hierarchical models with latent variables at one level only. According to Björck-Åkesson (1990, p. 120), the broad dimension measured "preference for arousing experiences as opposed to nonarousing experiences." This interpretation is in line with Zuckerman's (1993) later thinking, according to which people differ in the amount of stimulation that is felt as optimal. The other four factors were narrow and expected to influence only some of the items. Björck-Åkesson noted the possiblility that the narrow dimensions of the NF model measure other factors besides sensation seeking. Thus, alternative interpretations also exist for narrow dimensions. However, Björck-Åkesson called the four narrow dimensions of the NF model Thrill and Adventure Seeking (TAS), New Experience Seeking (NES), Activity (Act), and Outgoingness (Out). Interpretations of these dimensions were somewhat different from those found when SSS tests were given to adults. According to Zuckerman (1994), this was because of the exclusion of some items from the ES and Dis scales. Björck-Åkesson (1990) interpreted the factors as follows:

> TAS: Reflects the preference for extreme risk and challenge.
>
> NES: Refers to novelty, variation and a positive attitude towards being out of control.
>
> Act: Involves the preference for being an active part of the youth culture and the preference for social interaction.
>
> Out: Pertains to non-conformity with generally accepted norms, being the center of attention, preference for extreme appearance and emphasis on social feedback. (p. 125)

The Act dimension resembled the Disinhibition scale of SSS IV, and the Out scale contained items very similar to those of the ES and Dis subscales. However, if compared to inventory of Zuckerman, no BS factor emerged. The detailed conceptual description of the five-factor NF model is presented by Björck-Åkesson (1990).

In sum, based on the studies of Björck-Åkesson (1982, 1990) and Zuckerman (1979, 1984, 1994), three alternative models for the Finnish version of the Sensation Seeking Scale for children were presented. First, it was hypothesized that the responses to the SESE could be explained by four first-order factors TAS, NES, Act and Out. Second, it was suggested that the Finnish version of the Sensation Seeking Scale also measures five orthogonal dimensions with different level of generality (NF model). Third, as an alternative to the NF model, a simple second-order factor model was suggested. This model resembled the NF model, but in this higher-order model the second-order factor, namely General Sensation Seeking, was expected to explain not only the covariances among the observed variables but also the correlations among the four first-order factors (see Bollen, 1989; Rindskopf & Rose, 1988). The theoretical rationale for the second-order models based on the thinking of Zuckerman (1979, 1984, 1994). In several studies the median intercorrelations among the four subscales of Zuckerman's Sensation Seeking Scale have ranged from .27 to .58. According to Zuckerman, a broad second-order factor, namely General Sensation Seeking, accounts for the correlations among these four subscales. Thus, factors TAS, ES, Act, and Out were present in all three models. However, the meaning and the scope of these four dimensions were not the same in the three suggested models (cf. Björck-Åkesson, 1990). In principle, Figures 2, 3 and 4 present the three alternative models in path diagram form.

## 7.3.5 TEACHER RATING OF CURIOSITY

The rating instrument presented to teachers was based on the instrument first suggested by Maw and Maw (1961). However, following objections raised by Berlyne (1963) and Day (1968, 1971), five rating scale items were presented to the teachers instead of just one. Items one to three described what Berlyne called "specific exploration." Item four was expected to measure what Day (1968) labeled "diversive curiosity" and Berlyne (1971b) "diversive exploration." The content of item four was based on the definitions of Maw and Maw (1961) and Day (1968). Item five was expected to describe behavior which Zuckerman (1971) called General Sensation Seeking. Teachers were asked to rate each pupil on these five items using a 11-point scale format (see Appendix A).

Based on the modifications made to the original scale and the objections voiced by Berlyne (1963) and Day (1968, 1971), it was hypothesized that two first-order factors account for the variances among the five teacher rating scale (TeR) items. The factors were named according to Day and Berlyne (1971) as specific and diversive exploration. On the basis of previous studies (Ainley, 1987; Lanegvin, 1971, 1976; Rubenstein, 1986), it was expected that the correlation between the two exploration factors would not be very high. The corresponding conceptual model is presented in Figure 7.

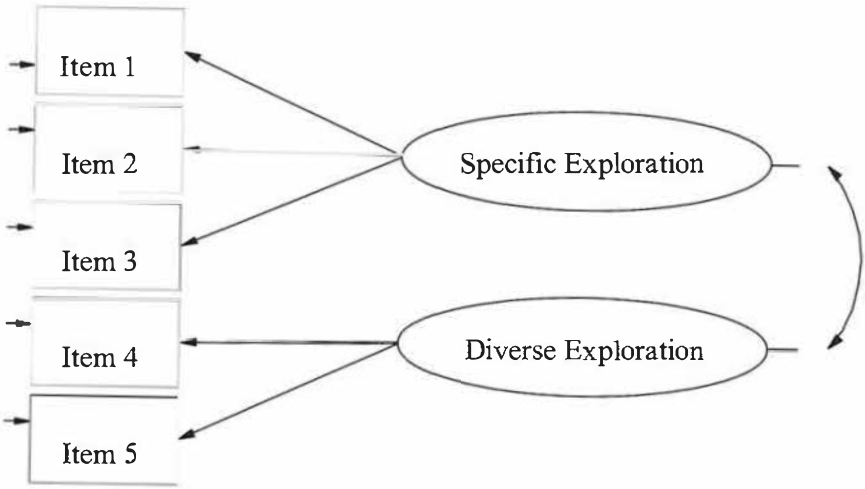



*FIGURE 7.* FACTORIAL MODEL FOR THE TEACHER RATING SCALE.

# 8. ITEM-LEVEL ANALYSIS

The empirical analyses of the present study were divided into two parts: item-level analyses and subscale-level analyses. The item-level procedures and analyses are presented first.

## 8.1 MODEL-TESTING STRATEGY

Outliers and non-normality of observed variables can have dramatic effects on the results of a SEM analysis (see e.g., Bollen, 1989; West, Finch, & Curran, 1995). Thus, the analyses of the present investigation started with preliminary analyses which tried to detect both non-normality and outliers. Preliminary analyses were conducted in two phases. First, IM, OTIMDC and C-Trait scales were analyzed because they all had the same scoring system. In the second phase, the SESE with dichotomously scored variables was analyzed. As a first preliminary analysis, tests of univariate and multivariate normality were computed using the PRELIS 2 program. Item distributions which showed skewness or kurtosis values greater than two in the test for zero skewness and kurtosis were regarded as showing skew or kurtosis. After that, each item was standardized and the z-scores were calculated for each individual in the sample in order to identify any univariate outliers, that is, cases with an extreme value on one variable. Z-score values greater than 3.0 were interpreted as potential outliers.

After the preliminary analysis, the factorial validity of the constructed scales was tested using different kinds of CFA techniques.

A special case of CFA called congeneric factor analysis was used to test the unidimensionality hypothesis of IM and OTIMDC scales. A typical distinction in modern test theory has been to distinguish between *parallel*, *tau-equivalent*, and *congeneric* measures (e.g., Bollen, 1989; Jöreskog, 1979a). Measures are said to be congeneric if they share a common factor, that is, they measure the same thing. Unlike parallel and tau-equivalent models, congeneric models do not imply that the observed scores measure the underlying factor in the same degree (Loehlin, 1987). Of the three models, the congeneric model is the most plausible in psychology (Jöreskog, 1979a). In the present study, congeneric factor analysis was also used to test the reliability of the IM and C-Trait scales. Several studies (e.g., Fleishman & Benson, 1987; Bollen, 1989) have shown that internal consistency methods like Cronbach's alpha have given biased estimates of reliability in several cases. Squared multiple correlations (SMC) were used as item-specific reliability estimates (Bollen, 1989).

The expected multidimensionality of C-Trait, SESE and TeR scales was tested using CFA. The testing sequence of C-Trait and TeR scales started with a test of an unrestricted model. This model tested whether the a priori hypothesized number of common factors is enough to account for the covariances among the observed variables (Mulaik & Millsap, 2000). Hypothetical unrestricted models were constructed also for SESE, but because of the limited sample size they did not converge. Thus, the testing process of SESE started with full measurement models.

All models were estimated separately for girls and boys. Multigroup confirmatory factor analysis techniques (MGCFA) were used only to test the invariance of the models across gender. The estimation method of the models was Maximum Likelihood (ML). The WLS method was not used because Jöreskog and Sörbom (1999) have suggested that ML should be used instead of WLS when the sample size is small or moderate (see also Hu, Bentler, & Kano, 1992).

## 8.2 EVALUATING REASONS FOR THE FAILURE OF A MODEL

Given the findings of ill-fitting initially hypothesized unrestricted factor models (UFA), analyses of IM, OTIMDC, and C-Trait scales proceeded in an exploratory mode. Following the thinking of Cronbach and Meehl (1955), *statistical*, *theoretical*, or *measurement* shortcomings can explain unfavorable results of a model test. In the context of an unrestricted model these three factors have special significance.

### 8.2.1 STATISTICAL ISSUES

To find out why the initial model produced unacceptable results, the impact of statistical issues was evaluated first. If a model fails, it is necessary at the outset to find out if statistical problems made it impossible to test the hypothetical model properly. Otherwise inappropriate and nonreplicable modifications of a theoretically adequate model may result (West, Finch & Curran, 1995). Sample size, estimation method effects, outliers, non-normality of the variables, and incorrect specification of the estimators have been reported to influence the fit indices used in SEM (Bollen, 1989; Chou & Bentler, 1995; Hu & Bentler, 1995). Partly in response to these problems, several researchers have proposed alternative approaches to estimation of fit statistics (e.g., Bentler, 1992; Jöreskog & Sörbom, 1993a, 1993b; Steiger, 1995) in order to obtain estimates that have good statistical properties. On the other hand, so-called specification errors in the CFA context are unlikely to cause the misfit of an unrestricted model because the factor-loading matrix is free except for the reference variables. Moreover, all identifiable unrestricted models of the same data produce the same fit indices (Jöreskog, 1979a). Thus, omitting secondary loadings cannot explain the misfit of an unrestricted model.

## 8.2.2 Alternative Theoretical Models

One possible reason for the misfit of a UFA model is that the factor number is not large enough to account for the data. Thus, the theory or the conceptualization which generated the initial factor model was incorrect and significant variance remained after the hypothesized dimensions were extracted. Messick (1993, 1994) referred to this kind of variance as "construct-irrelevant variance." One source of this kind of variance includes variance related to other distinct constructs. Thus, a new conceptualization is required.

Following the approach that Cronbach (1990) called "weak construct validation," the new conceptualization can be sought inductively using exploratory common factor analysis (see e.g., Byrne, 1991). After the right number of factors have been extracted, the model can be specified using post hoc CFA techniques (see e.g., Gerbing & Hamilton, 1996; Jöreskog, 1978). On the other hand, alternative interpretations of the data, or what Cronbach (1990, p. 184) called "plausible rival hypotheses," can be tested by constructing new factor analysis models. Cronbach referred to this deductive approach as "strong construct validation."

## 8.2.3 Measurement

Measurement error can be random or systematic. Of these two error types, random error has the desired properties of error variance: it is unrelated to any variables or other error terms. In contrast, nonrandom error variance is variance that is related in some systematic way to a variable or other error term. In other words, some items measure *something else* or *something in addition* to the construct they are supposed to measure (cf. Jöreskog, 1993). According to Maruyama (1998), the most common reason for this kind of systematic error variance can be an extra dimension that underlies two measures. The extra dimension can be substantive or methodological.

In several studies (e.g., Byrne, 1994; Tomás & Oliver, 1999), measurement has been shown to introduce systematic method variance to the data. Thus, in principle, different kinds of method effects can also account for an ill-fitting model. However, the problem in removing some of these kinds of unwanted properties is that multiple measures are needed (Maruyama, 1998). Thus, in the present analysis it was not possible to separate method variance from trait variance on item-level data in a reliable fashion. However, it was possible to suggest some substantively meaningful post hoc hypotheses about the existence of some systematic measurement errors.

Several studies (e.g., Byrne, 1994; Byrne & Shavelson, 1996; Fleishman & Benson, 1987) have shown that if two items have a very similar specific item content, it seems probable that this specificity introduces some systematic variance to the responses to these items. Gorsuch (1988, p. 255) has stated that the specificity can even be a common unique word. Moreover, Gorsuch (1983, p. 240) has also argued that when redundant variables are factored, several factors may occur which are narrow in scope. Comery and Lee (1992) referred to the same thing by arguing that when very similar items are used in factor analysis, they will correlate considerably higher among themselves than with other items. Because of this, these items probably define their own separate factor in solution. Taken as a whole, these findings suggest some semantic post hoc hypotheses for the present investigations. Using standardized residuals, MIs and conceptual analysis techniques, it was possible to present some possible alternative models for the basic models of each inventory.

In CFA contexts, systematic error variance can mean that correlated errors emerge. A common solution to this kind of problem has been to free up the corresponding error terms (uniquenesses) from the model. However, freeing up error terms in an unrestricted factor model is problematic due to identification problems (Jöreskog, 1979a; Mulaik & Millsap, 2000). On the other hand, when the errors are correlated and the correlations are substantially meaningful, it is doubtful that these are "errors." Both Hayduk (1987, 1996) and Maruyama (1998) have suggested that a better way to proceed in such cases is to replace the measurement errors with concepts, and to introduce them into the model as model segments. From the perspective of factor analysis, this can mean that we postulate extra common or specific factors to explain the residuals and meaningful correlations between error terms. Rummel (1970, pp. 326-327) defined a *specific factor* as a factor with only one high loading. Leskinen (1989) has broadened this concept by stating that correlating specific factor components of unique factors forms specific factors with two or three variables. Specific factors of this type are assumed to be orthogonal to each other and to the common factors. Special kind of specific factors with only two high loadings has been termed *doublets* or *doublet factors* (e.g., Harman, 1976; Rummel, 1970, p. 326). Thus, in the present study specific factors with two or three items were used to account for some of the observed systematic error variance.

### 8.2.4 Trimmed Inventory

Several researchers (e.g., Anderson & Gerbing, 1988; Benson & Bandalos, 1992; Byrne, 1991) have suggested that multiple indicator CFA model failures can be remedied by eliminating the poorly fitting items and then continuing the analysis with a subset of the original items. The risk involved in data-driven elimination of items is that the meaning of the concepts may change. In a situation where a multiple indicator model fails, Hayduk (1996, pp. 29-30) has recommended a different strategy which guarantees that the conceptual model will not change, or at least that the changes will be conscious. Hayduk suggested that in such situations one should begin again with a model which in the first phase has only one indicator per factor. This single indicator should be the best indictor of each concept. The factor loadings ($\lambda$) of these indicators are fixed to a value of 1 for each concept, and the corresponding error variances ($\theta$) are fixed to the specific values appropriate for each indicator. Fixing the error variance of a variable at a specified value means that the researcher has direct control over the meaning of the concept involved. That is, the researcher himself/herself adjusts the gap between the reference indicator and the concept. In the second phase, the second-best and third-best indicators of each concept are added to the model in separate LISREL runs.

Hayduk (1996) has strongly emphasized that two or three items per factor are suffice enough to measure a concept, otherwise confusion results instead of clarification. However, the number of items per factor depends on the broadness and fuzziness of the concept. If the concept is narrow (specific) and well-defined, even one indicator is enough. If the concept is broad (general) and complex, more items are needed. The correspondence of reported sex and actual sex, for instance, is almost perfect. Thus, only one variable is needed to measure this concept, and the measurement error is expected to be at the 1% level (see Hayduk 1987, p. 120). On the other hand, it is common knowledge that most psychological concepts are fuzzy and that it is difficult to find reliable indicators for these concepts. Thus, following the logic of Hayduk, two post hoc models were constructed in the present investigation to account for the covariance among the reduced C-Trait scale items. One was a narrow, eight-item model and the other a broad, twenty-item model.

## 8.3 INVARIANCE ACROSS GENDER

In the present investigation the invariance of the factor structures across gender was tested through a sequence of nested multigroup models. Thus, the tested models represented a continuum from the least constrained model to the most constrained model. This meant that it was possible to compare two successive models $M_1$ and $M_2$ with a chi-square difference test ($\Delta\chi^2$). That is, if $\Delta\chi^2$ was non-significant when Model $M_1$ was compared against more restricted Model $M_2$, this meant that the more restrictive model $M_2$ did not fit significantly worse than Model $M_1$. Model $M_2$ is nested in model $M_1$ if $M_2$ can be obtained from $M_1$ by constraining one or more of the free parameters in $M_1$ to be fixed or equal to other parameters (Long, 1983). Bollen (1989) divided invariance tests into two overlapping dimensions. The first dimension was the model structure, and the second dimension was the similarity in parameter values. Applying the suggestions of Leskinen (1987) and Byrne (1998), the hypothesis-testing strategy of the present study was as follows:

Hypothesis I: Testing for the validity of factor structure ($H_{form}$).

Hypothesis II: Testing for the invariance of factor loadings ($H_{\Lambda}$).

Hypothesis III: Testing for invariant factor variances and covariances ($H_{\Lambda\Phi}$)

Byrne et al. (1989; see also Byrne, 1998) have noted that most measuring instruments are actually only partially invariant across groups. Byrne et al. also demonstrated that a meaningful comparison of means is possible in situations where only partial measurement invariance is present. More recently, Steenkamp and Baumgartner (1998) have argued that comparisons of factor means are meaningful if at least one item (other than the reference item) is metrically invariant. By "metrical invariance" Steenkamp and Baumgartner meant invariance of factor loadings ($H_{\Lambda}$). Thus, the last invariance test of the present study was as follows:

Hypothesis IV: Testing for invariant factor mean structure ($H_{\Lambda\vee\kappa}$).

## 8.4 ASSESSMENT OF MODEL FIT

In the present study, three types of fit index were used (see Hu & Bentler, 1995; Hoyle & Panter, 1995). First, both Satorra-Bentler scaled chi-square (S-B$\chi^2$) and non-scaled chi-square ($\chi^2$) were used as absolute fit indices. Second, Bollen's (1989, pp. 270-272) Incremental Fit Index (IFI) was used as a type-2 fit index. IFI provides a modification of NFI which lessens the effects of sample size and takes the degrees of freedom (i.e., parsimony) into account. Third, Bentler's (1992) Comparative Fit Index (CFI) and its parsimony-adjusted variant (PCFI) was utilized as a type-3 fit index. The rationale for CFI is that it indicates the relative reduction in lack of fit when estimated by the noncentral $\chi^2$ of a target model versus a baseline model. All three types of incremental fit index are based on different rationales, and each describes somewhat different aspects of fit. The PCFI is obtained by multiplying the CFI by parsimony ratio ($df_k/df_0$, where $df_k$ is the degrees of freedom of the tested model, and $df_0$, degrees of freedom for the independent model). According to Mulaik, James, Alstine, Bennett, Lind, and Stilwell (1989), the PCFI penalizes a model for losses in degrees of freedom in the post hoc process of obtaining a well-fitting model. Akaike's Information Criterion (AIC) was used as another approach for model comparison. The advantage of AIC is that it permits model comparison also between non-nested models. The model with the lowest AIC should be preferred. On the other hand, the problem with AIC is its strong relationship to sample size. AIC favors saturated models in very large samples, and models with the greatest degrees of freedom and reasonable fit in small samples (Mulaik et al., 1989).

In the present investigation, a conventional .90 cut-off or "rule of thumb" criterion for the IFI and CFI indices was used for restricted factor models (see e.g., Hoyle, 1995). On the other hand, a value of at least .93 was expected in order for a model to be considered well-fitting (Byrne, 1994). Due to the fact that the unrestricted model is very liberal, it is important to have strong statistical support for it. Thus, following Mulaik's (personal communication to Les Hayduk on SEMNET, April 9, 1997) suggestions a .95 cut-off criterion for CFI and IFI was used for unrestricted factor models. As Byrne (1994) has noted, as PCFI values are typically lower than CFI and IFI values, a value of .80 served as rule-of-thumb lower limit cutpoint of acceptable fit for PCFI.

Steiger (1990, 1995) has provided a different view of the evaluation of model fit. He argued that fit coefficients should be based on population rationale, rather than sample rationale. According to Steiger, instead of trying to achieve a perfect fit, we should ask questions like "How bad is the fit of our model to our statistical population?" and "How accurately have we determined population badness-of-fit from our sample data?" Fit coefficients should also be relative unbiased and uninfluenced by sample size. Moreover, according to Steiger, confidence intervals should be reported, and the fit index must also compensate for model parsimony. Thus, in the present investigation, the Steiger-Lind Root Mean Square Error of Approximation (RMSEA) index was used to investigate how well individual models fit the statistical population. According to Steiger (1995), a RMSEA index value below .05 indicates a good fit. However, Browne and Cudeck (1993) have noted that in practice RMSEA values of about .08 or less indicate a reasonable error of approximation.

Residuals, modification indices (MI), and indices of expected parameter change (EPC) were used in the detailed assessment of model fit. LISREL gives two kinds of residuals: fitted and standardized. A fitted residual is an observed minus a fitted covariance (variance). A standardized residual is a fitted residual divided by its estimated standard error. LISREL also calculates a value of MI for each fixed parameter in the model. MI is an exploratory tool which provides information about the misspecifications of the model. MI indicates the drop in overall $\chi^2$-test if the fixed parameter is freed. Associated with each MI, there is EPC which gives information about the expected change and direction of the parameter (Saris, Satorra & Sörbom, 1987; Jöreskog and Sörbom, 1993a). In situations where a large MI is associated with a large EPC and when strong theoretical reasons for doing so are present it is reasonable to free a parameter. A scale-free variant of EPC is SEPC, as suggested by Kaplan (1993).

## 8.5 RESULTS OF ITEM-LEVEL ANALYSES

This section begins with the results of the preliminary analysis. After that, the test results of the item-level models are reported scale by scale.

### 8.5.1 PRELIMINARY ANALYSIS

First, IM, OTIMDC, C-Trait and TeR data were analyzed because they have similar scoring systems. Variables were treated as continuous because the number of categories in the observed variables was 11 (cf. Comrey, 1978). All preliminary analyses were carried out separately for girls and boys. Tests of zero skewness and kurtosis indicated that no variable deviated unacceptably from normality, that is, no items showed a skewness or kurtosis value greater than 2 (Cuttance, 1987). However, Mardia's multivariate tests of skewness and kurtosis showed that both samples deviated significantly ($p < .001$) from multivariate normality in all cases. Thus, to get a Satorra-Bentler scaled chi-square statistic, both the covariance matrix and the asymptotic covariance matrix were calculated for both gender groups. However, the calculation of an asymptotic covariance matrix was not possible for the C-Trait scale and SESE because of the small sample size. Since, for instance, the number of variables at the beginning of the CFA phase was 40, at least 780 observations ($k(k-1)/2$, where $k$ is the number of variables), would be required to estimate the asymptotic covariances for the C-Trait scale. However, the number of categories in the observed variables was 11, and none of the variables had skewness or kurtosis values over 2. Therefore, there was good reason to believe that the measurement was approximately on an interval scale and thus that covariances and the ML method could be used (cf. Comrey, 1978; Cuttance, 1987;

Jöreskog & Sörbom, 1989, p. 191; West et al., 1995). However, because some variables showed slight kurtosis and the data was not multivariate normal, it was expected that these features would increase the chi-square values and lead to underestimation of the fit indexes and the standard errors in the parameter estimates of first C-Trait models (Hu & Bentler, 1995; West et al., 1995).

Outliers can have dramatic effects on the results of a SEM analysis (see Bollen, 1989; West et al., 1995). Thus, each item was standardized and the z-scores were calculated for each individual in the sample in order to identify any univariate outliers, that is, cases with an extreme value on one variable. A z-score value greater than 3.0 was regarded as an potential outlier. The summary results of this analysis is presented in Table 5. To investigate the effects of the outliers on the results of analysis, corrective actions suggested by Gorsuch (1991) and West et al. (1995) were taken. These alternative analyses showed that the effects of outliers seemed to be minimal. However, on the basis of these analyses four cases were deleted from the analyses.

Next, the SESE data were investigated. Dichotomous variables with very uneven splits between the categories have been demonstrated to be problematic for SEM analyses (e.g., Bollen, 1989; West et al., 1995). Rummel (1970) suggested that dichotomous variables with 90-10 splits between categories should be deleted. Rummel gave two reasons for this operation. First, the correlation between two dichotomous variables is low if most responses to the

TABLE 5 *RESULTS OF THE PRELIMINARY ANALYSES*

| | Potential Outliers ($|z| > 3.0$) | |
|---|---|---|
| | Girls | Boys |
| IM | Item 3(10) | Item 4 (4) |
| OTIMDC | Item 6(8) | |
| C-Trait | Item 1(4), Item 10(10), Item 14(4), | Item 16(6) Item 12(10), Item 15(2), Item 16(6), Item 29(4) |

*Note.* Values enclosed in the parentheses represent the number of potential outliers.

variable fall into one category (truncated correlation coefficients). Second, the scores in the category with 10% of the cases can be regarded as outliers and are thus more influential than those in the category with 90% of the cases. Thus, in order to find variables with extreme splits, the frequency distributions of the sensation seeking scale variables were computed by the PRELIS 2 program.

In the data, dichotomous variables with 90-10 splits between categories were found for only one variable in both girls' (Item 50) and boys' (Item 35) data. On the other hand, 11 variables in the girls' data and 14 variables in the boys' data had 80-20 splits between categories. This result suggested difficulties in the item-level analysis of SEM (cf. Björck-Åkesson, 1990). Gorsuch (1983, pp.292-295) has demonstrated that factoring data where the variables have different splits usually lead to false factors. That is, extra factors emerge which are called "difficulty factors."

When the observed variables in LISREL analysis are dichotomous, Jöreskog and Sörbom (1993a, pp. 230-231) recommend using estimates of tetrachoric correlations analyzed using the WLS method. The weight matrix should be the asymptotic covariance matrix. However, because the number of variables at the beginning of the CFA phase was 55, it was not possible to estimate the asymptotic covariance matrix. Consequently it was not possible to us the WLS method or the Satorra-Bentler scaled chi-square statistic.

## 8.5.2 Consistency of the Scales

Estimates of internal consistency were calculated for all scales used. The resulting alpha coefficients are presented in Table 6.

Cronbach's alpha coefficient is designed for tau-equivalent and parallel measures. The well-known character of the alpha coefficient is that it underestimates the reliability of congeneric measures (see e.g., Bollen, 1989; Tarkkonen, 1987). All the scales and subscales used in the present study were expected to be only congeneric measures. However, although this fact was taken account, some of the alpha coefficients in Table 6 were unacceptably low, especially the alpha coefficients (KR-20) of subscales NES and Out. According to Anastasi (1988, p. 122), the alpha coefficient is influenced by two sources of error variance: 1) content sampling and 2) the heterogeneity

TABLE 6 *ALPHA COEFFICIENTS FOR THE SCALES USED*

| Scale[a] | Sample | |
|---|---|---|
| | Girls (N=263) | Boys (N=273) |
| IM (15) | .78 | .83 |
| OTIMDC (10) | .58[b] | .40[c] |
| C-Trait (40) | .88 | .88 |
| ConE (10) | .57[d] | .65 |
| ManE (10) | .75 | .71 |
| PerE (10) | .71 | .72 |
| ComE (10) | .72 | .70 |
| SESE (55)[e] | .87 | .82 |
| TAS (15) | .81 | .79 |
| NES (15) | .54 | .39[f] |
| Act (13) | .71 | .63 |
| Out (12) | .53 | .52 |

*Note.* [a]Number of items per subscale in parentheses. [b]The alpha coefficient was .69 when Items 3 and 8 were eliminated from the scale (weak negative Item-Total correlation). [c]The alpha coefficient was .59 when Items 3 and 8 were eliminated from the scale (weak negative Item-Total correlation). [d]The alpha coefficient was .61 when Items 3 was eliminated from the scale (weak negative Item-Total correlation). [e]The coefficients of the SESE subscales are Kuder-Richardson 20 (KR-20) coefficients. [f]The coefficient was .47 when Items 26 and 37 were eliminated from the scale (weak negative Item-Total correlation).

of the concept the test is trying to measure. On the other hand, the calculation of alpha coefficient is based on the average inter-item correlation. Because the preliminary analyses of the present study revealed that several SESE items had 90-10 or 80-20 splits, it was suggested that these extreme splits may also have had an effect on the Cronbach's alpha coefficients of the SESE subscales. However, because the alpha coefficient is not for single indicators, none of the items was excluded during this phase of the study.

### 8.5.3 TEST OF INTRINSIC MOTIVATION

The starting point for testing the IM scale was the conceptual model in Figure 5. According to this path diagram, it was hypothesized that Intrinsic Motivation (Curiosity) is the common cause of the 15 test items, but it was not supposed to be the common factor for those indicators. The common factor that directly accounts for the covariances among the 15 test items was expected to a behavioral trait which was named Exploration. Thus, the testing process of the IM scale began with a one-factor congeneric measurement model whose common factor was named Exploration instead of Intrinsic Motivation or Curiosity (cf. Hayduk, 1996, pp. 20-23). The 15 items were free to have different loadings on the Exploration factor and different error variances. Because also the asymptotic covariance matrix was used, the estimated chi-square statistics were Satorra-Bentler scaled. However, as indicated in Table 7, the statistical fit of Model IM1 was less than adequate for both girls and boys.

To find the reason for the misfit of the initial model, the impact of statistical, theoretical and measurement issues was studied. Because the estimated statistics were Satorra-Bentler scaled and no outliers

TABLE 7 *SUMMARY OF MODEL FIT STATISTICS FOR GIRLS AND BOYS (IM SCALE)*

| Model | df | S-B$\chi^2$ | CFI | PCFI | IFI | RMSEA |
|---|---|---|---|---|---|---|
| Girls | | | | | | |
| 1 Model IM1 | 90 | 208 | .76 | .65 | .77 | .07 |
| 2 Model IM2u | 76 | 150 | .85 | .62 | .85 | .06 |
| 3 Model IM2 | 89 | 208 | .76 | .64 | .77 | .07 |
| 4 Model IM3 | 85 | 122 | .91 | .74 | .91 | .04 |
| Boys | | | | | | |
| 1 Model IM1 | 90 | 175 | .87 | .75 | .87 | .06 |
| 2 Model IM2u | 76 | 103 | .94 | .68 | .94 | .04 |
| 3 Model IM2 | 89 | 174 | .87 | .74 | .87 | .06 |
| 4 Model IM3 | 87 | 110 | .94 | .78 | .94 | .03 |

existed, theoretical and measurement issues seemed to be the most plausible explanations for the misfit of the initial model.

### 8.5.3.1 Alternative Theoretical Model

Beswick selected the items of IM scale on the basis of internal consistency and correlations with other measures of curiosity. Moreover, the items of IM were derived from previously constructed inventories, for instance from the OTIM and Cattell's curiosity erg factor. In the study of Rubenstein (1986) items 4, 10, 11, and 13, all of which are OTIM items, loaded on a factor which she called "Interest in Science." This finding gave reason to postulate an alternative model for the initially postulated one-factor model. Figure 8 presents this Model IM2 in path diagram form.

The testing procedure of Model IM2 began with an unrestricted Model IM2u. Items 1 and 10 served as reference variables in this model. Item 10 was chosen as a reference variable because in the study of Rubenstein this item had a loading of .69 on the Interest in Science factor. Thus, reference Items 1 and 10 were fixed to value 1 on the reference factor and to value 0 on the other factor. Model IM2u was estimated separately for boys and girls.

As indicated in Table 7, the statistical fit of Model IM2u was inconsistent. The model fitted the boys data quite well but not to the girls data. Moreover, a detailed inspection of the results of Model IM2u revealed also some discriminant validity problems. The correlation between the two factors was .89 in the girls' group, which suggests that these two latent variables may not be distinct constructs. To investigate this possibility, a restricted Model IM2 was constructed first. The loadings were fixed according the conceptual Model IM2 (Figure 8).

As shown in Table 7, the goodness of fit for Model IM2 was less than adequate for both girls and boys. Moreover, a detailed inspection of the results showed that the two factors are highly correlated ($\varphi_{Girls}$=.94 and $\varphi_{Boys}$=.95). To test the post hoc hypothesis that the two latent variables are not distinct, a model was constructed where the correlation between the two factors was fixed to value 1 (Maruyama, 1998). When this model was compared to Model IM2 by using chi-square difference test, the results supported the presented post hoc

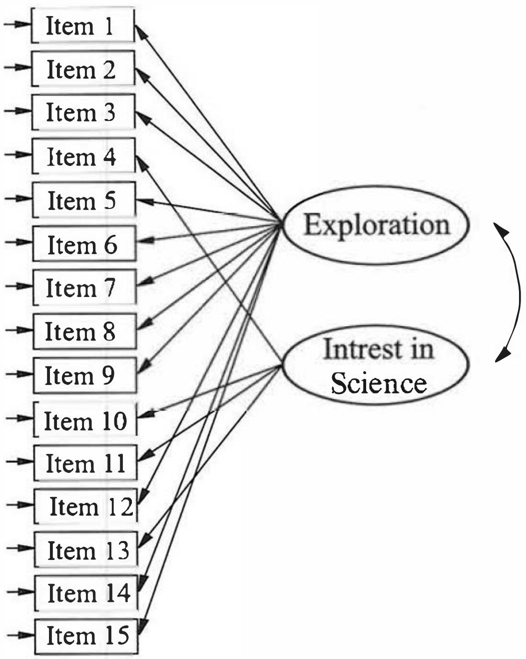



*FIGURE 8.* PATH DIAGRAM OF MODEL IM2.

hypothesis. The resulting $\Delta\chi^2$ was statistically non-significant for both gender groups, S-B$\Delta\chi^2(1)_{Girls}$=0.81 and S-B$\Delta\chi^2(1)_{Boys}$=1.22. This result meant that the two-factor model reverted to a one-factor model.

### 8.5.3.2 Systematic Measurement Errors

Next, issues related to measurement were investigated in order to find an explanation for the misfit of the initial model. Above all, an effort was made to determine whether the wording of items introduced some systematic error variance to the measurement. MIs, EPCs and standardized residuals supported this hypothesis. Thus, the pattern of correlated measurement errors and the way the IM was constructed gave reason to postulate an alternative post hoc model to Model IM1 (Model IM3). It was suggested that the development history of the IM is implicit in the wording and themes of the IM items. By carefully reading the contents of the items, it was possible to postulate one meaningful specific factor S1 and four possible correlated errors. Specific factor S1 was expected to account for the correlated errors among Items 4, 6 and 10. As already noted, in the study of Rubenstein (1986) Item 4, "Complicated machinery is fascinating to look at," and Item 10, "It is interesting to try to figure out how an unusual piece of machinery works," loaded highly ($\lambda$>.60) on a factor which she called "Interest in Science" or "Interest in machines/science/ nature." In the present study, specific factor S1, which was named Interest in Machines, was expected to account for the systematic variance between the two OTIM items having to do with machines. These items were interpreted as being of a special significance for technically oriented pupils. It also seemed theoretically reasonable that Item 6, which concerned reading, loaded negatively on this factor.

The four pairs of items with possible correlated errors were (a) Item 11 "I like to look at rocks which are made of many kinds of minerals" and Item 13 "I think about how strange plants grow"; (b) Item 14 "At times I have focused on something so hard that I went into a kind of benumbed state of consciousness, and at other times into a state of extraordinary calm and serenity" and Item 15 "If I come across something interesting I drop everything and study it. It is never a waste of time"; (c) Item 3 "If I read something which puzzles me, I keep reading until I understand it" and Item 5 "When I don't know the answer to a question on a test I look up the answer when the test is completed"; (d) Item 5 "When I don't know the answer to a question on a test I look up the answer when the test is completed" and Item 7 "I am interested in mathematical procedures that are possible with new calculating machines." Correlated errors between Items 3 and 5 and Items 5 and 7 were expected to be gender-specific, and thus they were postulated only for girls.

The relaxation of the four error covariances was based on different theoretical rationales. First, the correlation between Items 11 and 13 seemed reasonable because both items are OTIM items involving the study of nature. In Rubenstein's (1986) study both items 11 and 13 loaded on the "Interest in Science" or the "Interest in machine/ science/nature" factor. However, as the results of Model IM2 revealed, such a factor was not supported in the present investigation. Thus, it was hypothesized that, similar to Items 4 and 10, Items 11 and 13 also had a special meaning for a small subset of nature-oriented students. Second, the rationale for relaxing the error term between Items 14 and 15 was based on the claim that these two items may not have been totally locally independent (cf. Groves, 1989), and thus the relaxed parameters reflect a response bias. On the other hand, both Items 14 and 15 describe deep concentration on thinking and investigating. Thus, the relaxed parameter can also reflect specific item characteristics (cf. Aish & Jöreskog, 1990; Byrne, 1994). Theoretical arguments for the relaxation of the error term between items 3 and 5 was based on the similarity in the item content. Both items concern a need to understand things and to get correct answers in a test. On the other hand, according to the results of Model IM1 Item 5 had very little common variance (SMC=0.03) with other items in the girls' sample. This finding might mean that girls interpreted item 5 in different ways. Thus, the error covariance between Items 3 and 5 may represent an interpretation which is related to achievement behavior (see Heckhausen, 1987). The possible error covariance between items 5 and 7 was somehow more difficult to interpret. The most plausible explanation for this finding seemed to be the similarity in wording. Both items can be interpreted to have something to do with direct feedback. Thus, these items were suggested to have a special meaning for a group of girls.

As shown in Table 7, the overall fit of Model IM3 was better than for Model IM1, and it was regarded as acceptable for both gender groups. The 90 percent confidence interval for RMSEA was from 0.023 to 0.056 for girls and from 0.004 to 0.048 for boys, thus suggesting that Model IM3 is a good population model for both gender groups. However, a detailed analysis of the results revealed one interesting finding and a possible moderation effect. With two exceptions, all the loadings on the Intrinsic Motivation factor were statistically significant ($p < .05$) in both gender groups. However, the examination of the girls' results showed that Items 5 and 12 did not have a statistically significant loading on the Exploration factor. Neither were these items part of any big residuals. Thus, Items 5 and 12 contained very little common variance with other items, and the reliability of these items seemed to be questionable in the girls' sample. Figure 9 displays factor loadings for Model IM3.

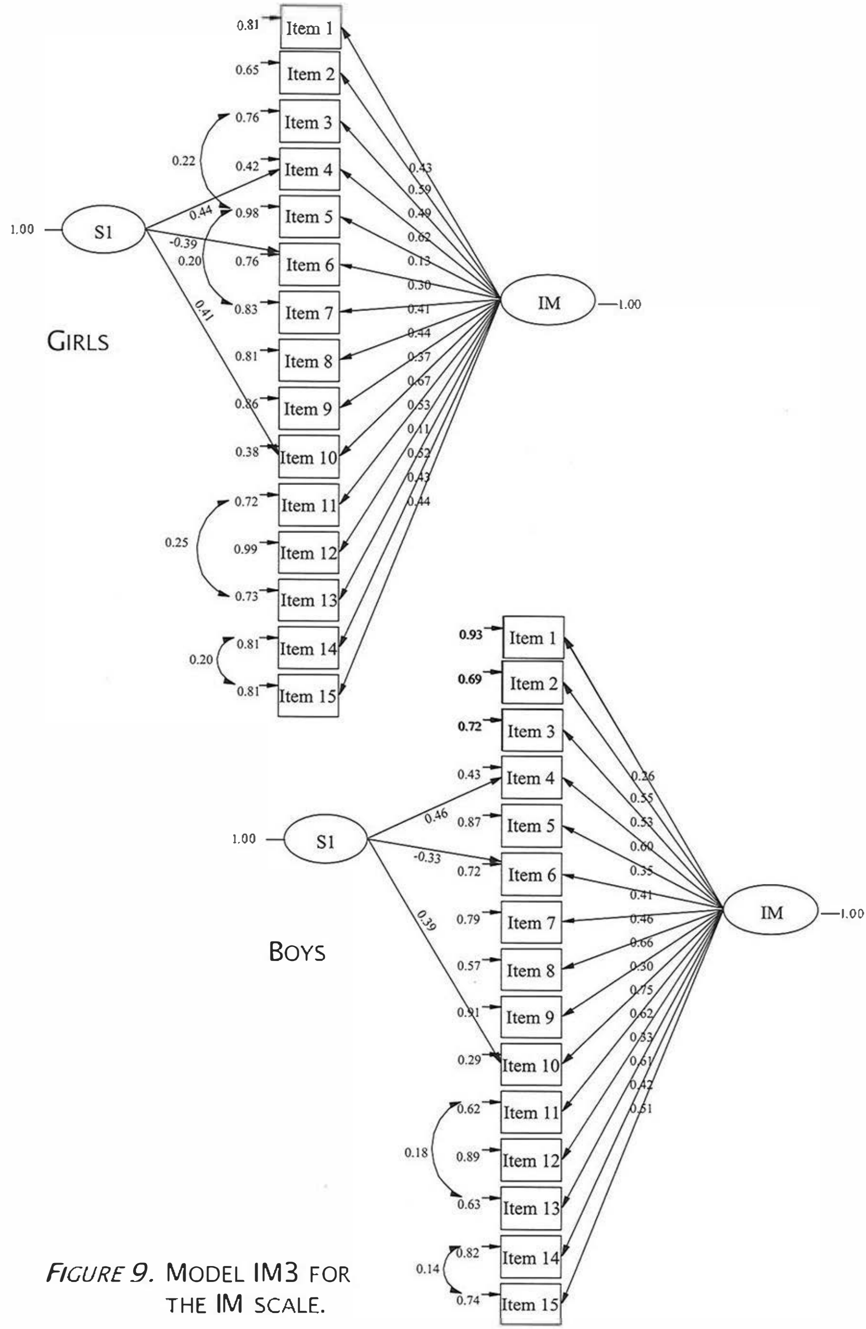



*FIGURE 9.* MODEL IM3 FOR THE IM SCALE.

### 8.5.3.3 Tests of Invariance

First, a multigroup baseline model was estimated. Because an acceptable reason for the misfit of the initial model had been found, Model IM3 was used as a baseline model for both gender groups. Thus, a corresponding multigroup confirmatory factor analysis Model $H_{form}$ was constructed first. Because chi-square is summative, the fit of this model was acceptable, S-B$\chi^2(172)$ = 226.04, CFI = .93, IFI = .94. Having received support for the preliminary test of invariance (Model $H_{form}$), the testing of gender effects proceeded in a hierarchical fashion. The model with equality constraints was compared to a less restrictive model in which the same parameters were free. Since these two models were nested, a $\Delta\chi^2$- test could be used to test the difference in $\chi^2$ (see, e.g., Byrne et al., 1989; Jöreskog & Sörbom, 1993a). Following this logic, the invariance of the measurement parameters was systemically tested next.

The invariance hypothesis, $H_{\Lambda}$, proposed that the two gender groups have equal factor loadings. All factor loadings were constrained to be equal across gender. This model was then compared to Model $H_{form}$, in which no equality constraints existed. The result of the $\Delta\chi^2$-test was statistically significant, thereby supporting rejection of the hypothesis of invariant factor loadings (S-B$\Delta\chi^2(16)$ = 32.0, $p < .01$). However, the goodness-of-fit measures remained quite high (e.g., CFI = .92 and IFI = .92). Moreover, RMSEA was .04 and its 90 percent confidence interval from .029 to .051, thus supporting the conclusion that $H_{\Lambda}$ was a good population model. Examination of MIs suggested that the factor loading of Item 8 was not invariant across gender (MI=5.4). However, because of the relatively small MIs and to avoid capitalization on chance, the hypothesis of invariant factor loadings was accepted with caution (cf. Steenkamp & Baumgartner, 1998).

The next step was to test the invariance of the factor variances ($H_{\Lambda\Phi}$). Both the goodness of fit indices and the result of the omnibus test (S-B$\Delta\chi^2(2)$ = 3.4) supported the hypothesis of invariant factor variances. Having obtained evidence for the invariant factor loadings and variances, it made sense to compare the means of the Exploration and specific factors (Model $H_{\Lambda\nu\kappa}$). This was done using the method suggested by Leskinen (1987) and Byrne (1998). Vectors TX and KA, which contained intercepts for the observed measures ($\nu_1$-$\nu_{15}$) and for the Exploration and specific factors ($\kappa_1$ and $\kappa_2$), were added to Model $H_\Lambda$. TX was set free for girls and invariant for boys. The elements of the KA matrix were fixed to value 0 for girls and free for boys. In terms of both chi-square and other goodness-of-fit indices, the fit of Model $H_{\Lambda\nu\kappa}$ was poor, S-B$\chi^2(195)$ = 338.13, CFI = .88, IFI = .88. However, when the MIs of the intercept terms were evaluated, none of them was over five. Thus, it seemed that LISREL 8.30 could not reliably estimate the MIs of Model $H_{\Lambda\nu\kappa}$. Thus, to find the reason for the misfit of Model $H_{\Lambda\nu\kappa}$ the intercept terms were freed one by one. This process revealed that at least the intercepts of Items 7 and 10 were not invariant across gender. Relaxing these measurement intercepts also made theoretical sense because items 7 and 10 both concerned machines, and it is well-known fact that machines interest boys more than girls. Successively relaxing the constraints of Items 7 and 10 yielded a substantial and statistically highly significant improvement in fit as compared to Model $H_{\Lambda\nu\kappa}$, where all intercept terms were constrained as invariant (S-B$\Delta\chi^2(2)$ = 55.31). Both the type-2 and type-3 fit indices of Model $H_{\Lambda\nu\kappa^*}$ were on an acceptable level, IFI = .91 and CFI = .91. In addition, the RMSEA estimate was .042 and its 90 percent confidence interval from .032 to .052 supported the partial scalar invariance. Thus, no further modifications were made to the model. No sex differences were found in the means on the Exploration and specific factors.

### 8.5.3.4 Discussion and Conclusions

The analysis of the IM scale revealed that the initially postulated congeneric model could not account for the covariances among the 15 test items. Neither could the alternative two-factor model. The reason for the misfit of the initially postulated one-factor model seemed to be systematic measurement errors in item responses. On the other hand, if the history of the IM scale is taken account, the presence of such correlated errors would seem to be psychometrically reasonable. In particular, their presence was not surprising, because items of IM stem from previously constructed inventories. Thus, because the construction of one three-item specific factor and the relaxation of four correlated errors are based on substantive arguments, I suggest that these elements are part of the factor model in this population.

The only moderation effect found in the invariance analyses was the non-invariant intercepts of Items 7 and 10. Both items concerned machines and thus it was no surprise that boys scored higher than girls on these items. On the other hand, the analysis of the baseline model showed that Items 5 and 12 had very little common variance with other IM items in girls' sample and their loading on Exploration factor was statistically non-significant. This moderation effect is difficult to explain and may be a sample-specific finding. Thus, because Items 5 and 12 had statistically highly significant loading on the Exploration factor in the boys' sample, these items were not eliminated from the scale.

### 8.5.4 DIVERSIVE CURIOSITY SCALE

Using the same testing logic as in the IM scale testing, it was hypothesized that one factor, Diversive Exploration, accounts for the covariances among the 10 OTIMDC-scale items (Model DC1). Thus, a congeneric factor analysis was again used to test this assumption. However, as Table 8 shows, the fit of Model DC1 was unsatisfactory for both gender groups.

TABLE 8 *GOODNESS-OF-FIT INDEXES FOR MODELS OF THE OTIMDC SCALE*

| Model | df | S-B$\chi^2$ | CFI | PCFI | IFI | RMSEA |
|---|---|---|---|---|---|---|
| Girls | | | | | | |
| 1 Model DC1 | 35 | 56 | .87 | .68 | .87 | .05 |
| 2 Model DC2 | 33 | 31 | .98 | .72 | .98 | .00 |
| 2 Model DC3 | 36 | 39 | .95 | .76 | .95 | .02 |
| 2 Model DC4 | 13 | 17 | .96 | .59 | .96 | .04 |
| Boys | | | | | | |
| 1 Model DC1 | 35 | 82 | .74 | .58 | .75 | .07 |
| 2 Model DC2 | 32 | 36 | .95 | .68 | .95 | .02 |
| 2 Model DC3 | 35 | 56 | .89 | .69 | .89 | .04 |
| 2 Model DC4 | 12 | 18 | .95 | .54 | .95 | .04 |

### 8.5.4.1 Systematic Measurement Errors

Given the rejection of the initially postulated model, the next logical step was to find a reason for the misfit of the model. To find the reason for the misfit of the initial model, the impact of statistical, theoretical and measurement issues was again inspected. This showed that issues related to measurement seemed to be the most plausible explanation for the misfit of the initial model. A detailed examination of the results of Model DC1 revealed that correlated errors may account for the misfit of Model DC1. Again, it appeared that the issues related to item wording introduced some extra systematic variance to the measurement. By carefully reading the items of the OTIMDC scale, it was possible to postulate two theoretically meaningful correlated errors which were both consistent across gender. First, one correlated error was expected to be between Item 1 "I soon get bored when there is not enough going on" and Item 4 "I get tired of doing the same thing all the time," both items having a similar specific item content which was related to boredom experience. The second correlated error was hypothesized between Item 5 "I avoid busy, noisy places" and Item 10 "I like to have lots of activity around me." Item 5 was a reverse worded item with very low reliability ($SMC_{Girls}$=0.03 and $SMC_{Boys}$=0.00), and Item 10 was a very similar non-reverse worded item. Thus, it was hypothesized, based on previous studies (e.g., Byrne, 1994; Fleishman & Benson, 1987), that the reason for both nonzero error covariances was the similarity in item wording. In addition to these two correlated disturbances, one gender-specific finding also emerged. For boys only, the MIs suggested that a substantial improvement in model fit would be gained by an additional specifying of an error covariance between Item 6 "I am always glad to have someone visit me" and Item 7 "I never spend any time alone if I can help it." Freeing this parameter was theoretically meaningful because both items had a specific item content: both items describe a wish to have other people around. Moreover, these two items may not be totally locally independent. Thus, the hypothesized model was respecified to include these three additional parameters and then reestimated. As indicated in Table 8, the overall fit of Model DC2 was good. Also the 90 percent confidence interval for RMSEA suggested that the model was a good population model. For girls this confidence interval was from 0.0 to 0.04 and for boys from 0.0 to 0.05.

The results of both estimated models, DC1 and DC2, showed that the reliability of the three reverse worded items, 3, 5 and 8, was low. The SMCs for these items were from 0.00 to 0.27 (Figure 10) and they did not take part in any conceptually meaningful large residuals. Surprisingly, all these items loaded negatively on the Diversive Exploration factor. An exception was Item 5, which had a mild ($t < 1.96$) positive loading on this factor in the girls' group. Two of the five negative loadings were statistically non-significant ($t < 1.96$).

The possibility of dropping the three reverse-worded items was investigated using a method suggested by Hoyle and Lennox (1991). Model DC2 was compared to a model where Items 3, 5 and 8 were fixed to value 0 on the Diversive Exploration factor. This model, DC3, treated the three items as sharing no common variance among themselves or with other items. The $\chi^2$ difference test was significant for both gender groups, $S\text{-}B\Delta\chi^2(3)_{Girls} = 8.63$ and $S\text{-}B\Delta\chi^2(3)_{Boys} = 15.89$, suggesting that the assumption of no common variance was too tight. However, when the items were eliminated one by one, the $\chi^2$ difference test indicated that at least Item 5 had no common variance among the other items, $S\text{-}B\Delta\chi^2(1)_{Girls} = 1.18$ and $S\text{-}B\Delta\chi^2(1)_{Boys} = 0.57$.

Given the findings of three problematic items, namely Item 5, which had no common variance among the other items, and Items 3 and 8, which had a weak negative loading on expected factor, a final model, Model DC4, was constructed in which these items were eliminated from Model DC2. Table 8 shows the fit indices of Model DC4. As expected the fit of this model was very good.

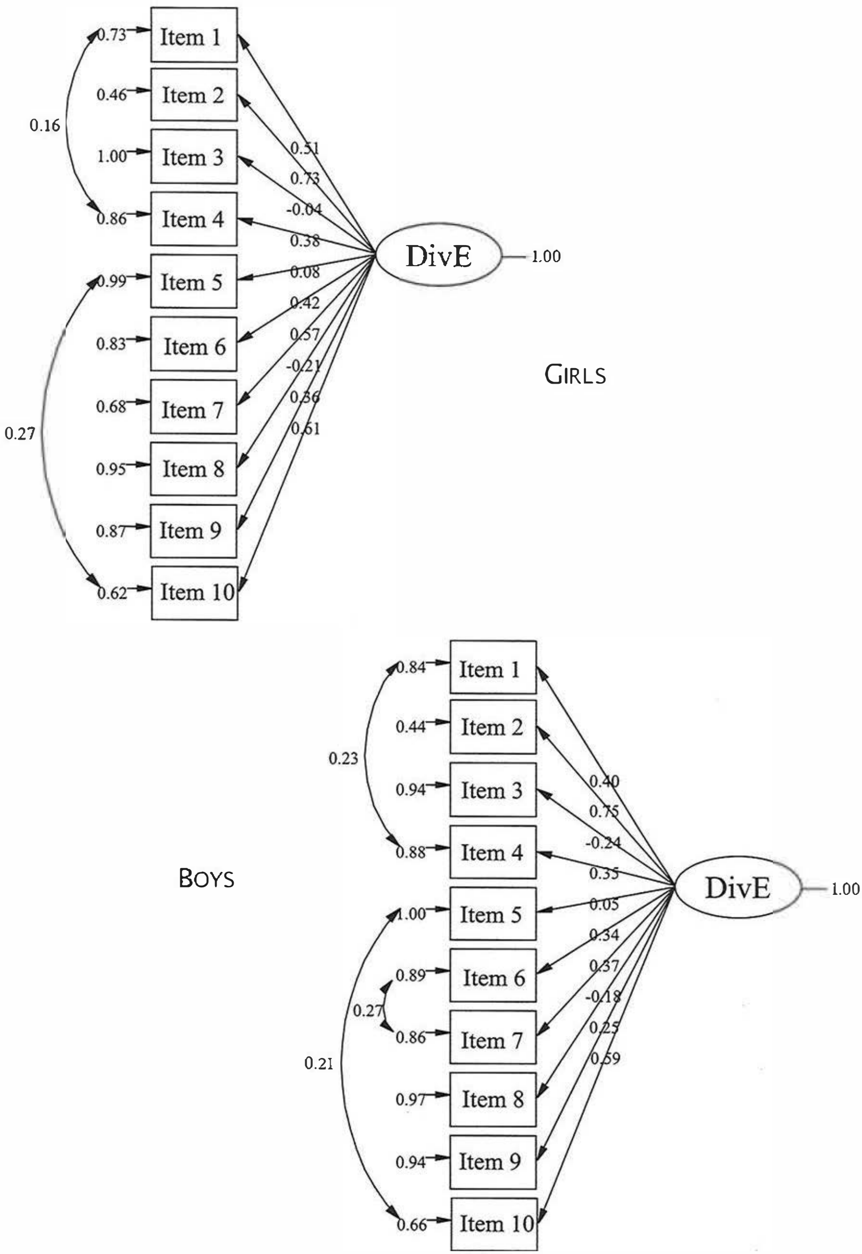



*FIGURE 10.* MODEL DC2 FOR THE OTIMDC SCALE.

### 8.5.4.2 Tests of Invariance

Model DC4 served as baseline model for the invariance testing procedure. Thus, the corresponding multigroup Model $H_{form}$ was constructed first. As expected, the fit of this model was good, S-B$\chi^2$(25) = 39.48, CFI = .96, IFI = .96. Next, Model $H_{\Lambda}$ was constructed by constraining all factor loadings to be equal. This model was compared to the Model $H_{form}$ with the chi-square difference test. The difference in $\chi^2$ was non-significant (S-B$\Delta\chi^2$(6) = 4.83), thus supporting the hypothesis of invariant factor loadings.

Next, the invariance of factor variances was investigated. A comparison of Model $H_{\Lambda\Phi}$ to Model $H_{\Lambda}$ yielded a non-significant chi-square value, S-B$\Delta\chi^2$(1) = 1.56, thus suggesting acceptance of the hypothesis of invariant factor variance. Thus, a valid comparison of factor means could be conducted. However, the fit of Model $H_{\Lambda\nu\kappa}$ was unsatisfactory, S-B$\chi^2$(38) = 84.62, CFI = .89 and IFI = .89. The MIs indicated that the intercept for Item 6 (MI = 33.01) was not invariant across gender. Relaxation of this constraint yielded a Model $H_{\Lambda\nu\kappa^*}$, which according to goodness-of-fit indices fitted the data quite well. Both CFI and IFI indices rose 0.06 to value .95. A detailed evaluation of the results showed that boys scored higher than girls on the Diversive Exploration factor ($t$ = 2.33).

### 8.5.4.3 Conclusion

Taken together the analyses of the OTIMDC scale indicated that the scale contained three problematic items, Items 3, 5, and 8. The main result of OTIMDC-scale analyses was that the reliability of these items is poor. This finding was consistent across gender. Thus, in future studies the content of these items should be modified if used in this population. Because all these three items were reverse-worded, it is suggested that the wording of these items be reversed. This can lead to response sets, but the problem with reverse-worded items is that they are difficult for fifth-graders to understand. Reverse-worded items have been problematic also in other curiosity studies (e.g., Boyle, 1989).

## 8.5.5 THE BROAD C-TRAIT SCALE

To test the hypothesis that four factors account for the observed covariances, an equivalent unrestricted factor model (Model 1u) was constructed for both gender groups. Following Jöreskog's (1979a) suggestion, Item 23 was postulated to load only on Factor 1, Item 25 only on Factor 2, Item 11 only on Factor 3, and Item 22 only on Factor 4. These four reference variables were expected to be pure in their respective factors and very close to the meaning of the concept. Thus, each column of the factor pattern matrix contained three fixed vales of 0 and one fixed value of 1. The other elements of the factor pattern matrix were estimated freely. The factor-covariance matrix contained factor variances on the diagonal. This model was tested separately in both gender groups. However, when the fit indices of Model 1u[1] were compared to the indices of a saturated model (see e.g., James et al., 1982), the hypothesis of the four-factor structure had to be rejected (see Table 9). Thus, the initially postulated,

TABLE 9 *FIT INDICES OF THE BROAD C-TRAIT SCALE*

| Model | df | $\chi^2$ | CFI | PCFI | IFI | RMSEA |
|---|---|---|---|---|---|---|
| Girls | | | | | | |
| 1 Model 1u | 626 | 1108 | .81 | .65 | .82 | .05 |
| 2 Model 2u | 590 | 968 | .85 | .64 | .86 | .05 |
| 3 Model 3u | 581 | 872 | .88 | .66 | .89 | .04 |
| 4 Model 4 | 18 | 21 | .98 | .63 | .99 | .03 |
| 5 Model 5 | 167 | 256 | .92 | .81 | .93 | .05 |
| 6 Model 6 | 169 | 256 | .92 | .82 | .92 | .05 |
| Boys | | | | | | |
| 1 Model 1u | 626 | 986 | .86 | .69 | .87 | .05 |
| 2 Model 2u | 590 | 883 | .88 | .67 | .89 | .04 |
| 3 Model 3u | 583 | 795 | .92 | .69 | .92 | .03 |
| 4 Model 4 | 18 | 27 | .96 | .62 | .96 | .04 |
| 5 Model 5 | 165 | 255 | .92 | .80 | .92 | .04 |
| 6 Model 6 | 151 | 268 | .90 | .72 | .90 | .05 |

unrestricted four-factor model was not sufficient to account for the covariances among variables. Given the rejection of the initially postulated four-factor unrestricted model, the next step was to find out whether theoretical, statistical or measurement shortcomings could explain the misfit.

#### 8.5.5.1 Assessment of the Statistical Issues

When the fit statistics of Model 1u were computed, the assumption of multivariate normality was violated. This may have influenced the chi-square statistics and the fit indices. In this kind of situation, Jöreskog and Sörbom (1998) suggest that standard errors and chi-square should be estimated under the non-normality assumption. However, LISREL 8.30 requires an asymptotic covariance matrix in order to take non-normality into account by computing, for instance, Satorra-Bentler scaled chi-square statistics (see e.g., Hu, Bentler & Kano, 1992). Since the number of variables at the beginning of the CFA phase was 40, at least 780 observations (k(k-1)/2, where *k* is the number of variables), would be required to estimate the asymptotic covariances in the present study. Thus, it was not possible to use robust estimation methods and in this way estimate the magnitude of the non-normality effects. However, because some variables showed slight kurtosis and the data was not multivariate normal, it was expected that these features would increase the chi-square values and lead to underestimation of the fit indexes and the standard errors in the parameter estimates (Hu & Bentler, 1995; West et al., 1995).

One possible reason for the unfavorable results of Model 1u may have been outliers. Extreme data points have been shown to have dramatic effects on the results of SEM analyses, even when the remaining data are well distributed (West et al., 1995). Furthermore, a Mahalanobis distance was calculated for each case in order to identify multivariate outliers, that is, cases with an unusual combination of two or more scores. The Mahalanobis distance was evaluated as a $\chi^2$ with df = 40 (i.e., the number of variables). There were 14 cases in the girls' sample and 13 cases in the boys' sample which showed a Mahalanobis distance value of over 73.402 ($p < .001$). In order to determine the combination of variables that made these cases multivariate outliers, an analysis suggested by Tabachnick and Fidell

(1989, p. 69) was carried out. First, a dummy variable was created in which cases with Mahalanobis distances over 73.402 were given a value of one and the remainder of the cases given a value of zero. Next, a discriminant function analysis was conducted in which the dummy variable was used as a grouping variable. In the girls' subsample, the difference between the two groups seemed to be in the scores for items intended to measure Manipulatory Exploration; that is, cases that were the multivariate outliers had systematically lower scores than the remaining sample on these items. In the boys' sample, the results of the discriminant function analysis were not so clear, but were still quite similar. The common feature of the potential multivariate outliers was that these cases had systematically lower scores than did the other cases on items intended to measure Perceptual Exploration. Given these findings, corrective actions for outliers were used (see Gorsuch, 1991; West et al., 1995). However, the effects of outliers on the analysis seemed to be minimal.

#### 8.5.5.2 Test of Alternative Theoretical Models

In several studies, perceptual exploration has been split into visual, tactile, and auditory exploration (e.g., Keller, 1987; Schneider & Unzner, 1994; Schölmerich, 1994). These findings suggested a five- and six-factor alternative model for the model suggested by Olson (1986). Following this approach, first a five-factor unrestricted model (Model 2u) was constructed in which items measuring tactile and auditory exploration were expected to load on the fifth factor (cf. Schneider & Unzner, 1994; Schölmerich, 1994). Item 37 was fixed as a reference variable on the fifth factor, and Items 14, 17 and 38 were also expected to load on this factor. As shown in Table 9, the fit indices of Model 2u gave somewhat contradictory information. The CFI and IFI indices were far from the cut-off criterion, .95, but the RMSEA point estimate was below 0.05, suggesting that Model 2u is a reasonably good population model for both sexes.

Although RMSEA suggested acceptance of Model 2u, the interpretation of this model was problematic. First, the loading patterns of the fifth factor only partially supported the further definition of Perceptual Exploration into Visual Exploration and Auditory-Tactile Exploration. The results of Model 2u revealed that the fifth factor was quite specific. It mainly accounted for an unexpectedly high standardized residual ($5.33_{Girls}$ and $2.33_{Boys}$) between items 37 and 38. An unexpectedly high correlation ($r_{Girls}$ = 0.42 and $r_{Boys}$ = 0.39) and high residual covariance between these same two items cast some doubts on the validity of Item 38. Item 38 may have been confusing, and the subjects may have therefore answered it in the same way as they answered Item 37. Second, a gender-specific finding of the five-factor model was that the two Perceptual Exploration factors correlated highly in the boys' model ($r$ = .89). This result indicated that the two Perceptual Exploration factors may not be distinct.

In the present study, the number of subjects was not sufficient for reliable estimation of a six-factor unrestricted model since, with more than five factors, the number of estimated parameters exceeded the number of subjects, and reliable estimates could not be obtained.[2]

#### 8.5.5.3 Systematic Measurement Effects

Because the five-factor alternative, Model 2u, could not totally explain the misfit of the initial model, Model 1u, the possibility of systematic measurement errors was investigated next. The principles used in constructing the original scale referred to the possibility that such measurement errors do occur in certain items. First, based on previous studies it was hypothesized that reverse-worded items were difficult to understand for some of the fifth graders, and thus they answered them according to a certain pattern (cf. Groves, 1989). Specifically, Items 2, 5 and 32 (see Appendix A), which begin with a negative expression "I ignore" or "I avoid," were expected have confused some fifth graders because they require a lot of cognitive activity. MIs and standardized residuals gave support to this assumption. Thus, a specific method factor or a response style factor (S1) was expected to account for the error covariances among items 2, 5 and 32. A gender-specific post hoc hypothesis was that in the girls' sample Item 3 "I question a lot of things" had a negative loading on this factor. It seemed that some girls found this item confusing and thus they had answered this item in the same way as Item 2 (response bias).

The second reason for the systematic measurement error was expected to be related to the high internal consistency of the scale.

Olson (1986) reported an alpha coefficient of .92 for her inventory. In the present study the alpha coefficient was .88. An internal consistency coefficient of over .90 is quite high for a total scale that measures four dimensions of exploration; in fact, it is almost too high, given that the alpha coefficient underestimates reliability when it is used to measure the reliability of a multidimensional measurement scale. Excessively high homogeneity may indicate redundancy in item construction and narrowness of scale (Boyle, 1991; Cronbach, 1990; Tarkkonen, 1987). Olson's inventory consisted of 40 generally-worded and specifically-worded items. Some of the generally and specifically worded items had very similar item contents. In other words, the same statement was presented by both items, but done so in a slightly different way. Based on this logic, it was hypothesized that redundant item wording is a source of systematic variance and several correlated measurement errors exist.

By comparing MIs and standardized residuals and by carefully reading the contents of items, it was possible to postulate three meaningful specific factors which were expected to be consistent across gender. Two of these three specific factors, S2 and S3, were very similar. They were expected to explain correlated measurement errors among the three items. Specific factor S2 was expected to explain the correlated measurement errors among items 12, 29 and 36, all having something to do with machines. Correspondingly, specific factor S3 was expected to explain correlated measurement errors among items 17, 18 and 28, all having something to do with art. Based on the thinking of Aish and Jöreskog (1990), it was possible to find two alternative interpretations for these two specific factors. The first possible interpretation for specific factors S2 and S3 was that they both reflect item characteristics (method bias). In several studies (e.g., Byrne, 1994; Tomás & Oliver, 1999), similarity in item wording has led to correlated error terms. On the other hand, an alternative interpretation for specific factors S2 and S3 was that they reflect respondent's characteristics. In this case they may represent a small omitted factor which is very content-specific. Rubenstein (1986) has suggested the existence of this kind of curiosity factor. In Rubenstein's study, items which concerned interest in machines loaded on a factor which she named "Interest in Science." Thus, it seems possible that items 12, 29 and 36 had a special meaning for technically-oriented pupils and they interpreted these items differently than the great majority of the students. Moreover, in the study of Rubenstein (1986), a factor which reflected interest in expressive arts (Divergent Thinking factor) also emerged. Thus, it was assumed that in the present study a small subset of students interpreted also Items 17,18 and 28 differently than the great majority of students, as well.

Specific factor S4 was postulated to be a doublet factor that explains the specific variance between Items 15 and 23. These items have a very similar specific item content and they were designed to measure the same common factor, Curiosity for Complexity. In addition to these four specific factors, which were expected to be consistent across gender, gender-specific factors were also postulated. For girls it was possible to find five theoretically meaningful specific factors which were expected to account for correlated errors between Items 6 and 18, 31 and 33, 35 and 36, 33 and 37, and 24 and 40. For boys it was possible to postulate three gender specific factors. Boys' specific factor S5 was very similar to specific factors S2 and S3. It was expected to account for correlated errors among items 11, 24 and 25, all having something to do with information seeking and the pleasure related to that activity. The other two specific factors for boys S6 (doublet) and S7, were assumed to explain correlated errors between Items 3 and 7 and Items 21 and 22. The gender-specific factors were interpreted to account for a systematic measurement error that may derive from characteristics specific either to these items or to the respondents (Aish & Jöreskog, 1990; Byrne, 1994). For example, similar items which were successive or very near each other may not have been totally locally independent (response bias).

Thus, on the basis of MIs, standardized residuals and theoretical relevance, an alternative post hoc model for Model 2u was estimated separately for both gender groups. Following Leskinen (1989), the five-factor model was reparameterized by adding columns to the factor-loading matrix. These columns contained only fixed values of either 0 or 1. Those items expected to form the specific factors in question were fixed as values of 1 in these additional columns. Utilizing this procedure, it is possible to make explicit and partial out the variance of an item into three parts as in second-order factor analysis (cf. Leskinen, 1989; Rindskopf & Rose, 1988; Jöreskog & Sörbom, 1993a). The three sources of variance are the common factor (communality), the specific factor and the measurement error. Together, the first two determine the reliability of the item.

The variances of all specific factors were statistically significant. This result supported the decision to include specific factors into the model (Leskinen, 1989). However, the overall fit of Model 3u was considered to be promising only for boys. As indicated in Table 9, the fit to the girls' data was still rather poor. Thus, Model 3u only partly explained the mismatch of the original model, Model 1u.

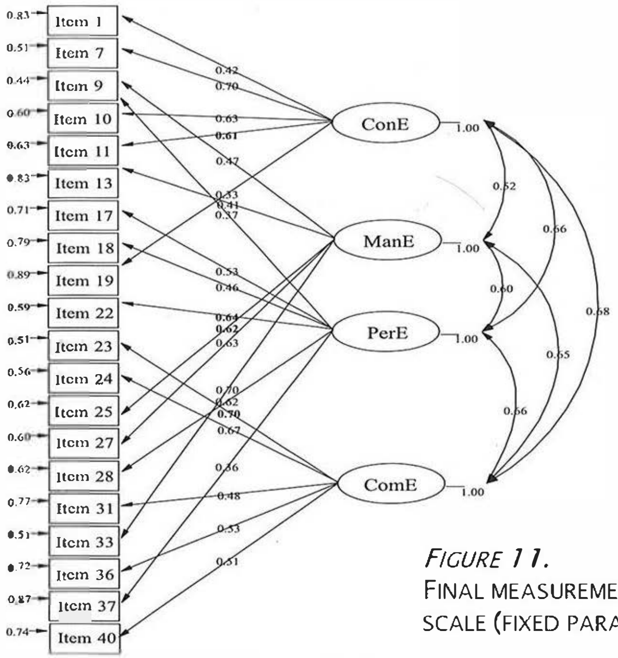
#### 8.5.5.4 Revised Scales

Together, the tested a priori and post hoc models revealed that the 40-item scale has a set of problematic items. In particular, those items which loaded on non-target factors or otherwise had a loading pattern which did not match with the underlying conceptual model were problematic. In this sense, the discriminant and convergent validity of these items seemed to fail. Thus, a reduced scale was needed because carrying out the analysis with all 40 variables had led to a conceptually problematic and unknown model. To find items for the reduced scale, conceptual analysis techniques were carefully applied to compare the contents of items to the underlying theoretical concepts. The original 40-item inventory included many item pairs where the same idea was operationalized with two very similar items, one specifically worded and one generally worded. Thus, very similar and hence redundant items were localized. Also, each of the reverse-worded items had a very similar non-reverse-worded parallel item included among the other items. Thus, the rejection of certain items was not a threat to validity. Following this logic, a new model was constructed.

Applying the suggestions of Hayduk (1996), the reference item loadings were fixed to value 1 and the corresponding error terms to the values of Model 1u. Specifically, the error variances found in the girls' model served as reference values for the fixed error terms in both groups. The error variances for the boys' model were calibrated to the same level as for the girls' model using a method suggested by Hayduk (1987). This procedure confirmed that both gender groups originally had the same concepts underlying the model. Next, Items 40, 13, 19, 17, 31, 27, 1, 28, 36, 33, 7, 37, 24, 9, 10 and 18 were included in the model in separate LISREL runs (see Hayduk, 1996, pp. 29-30). These items were chosen so that they cover all the aspects of the underlying concepts as well as possible. Clear and non-redundant wording and good distributional properties of items were also used as criteria for the selection of the final group of items.

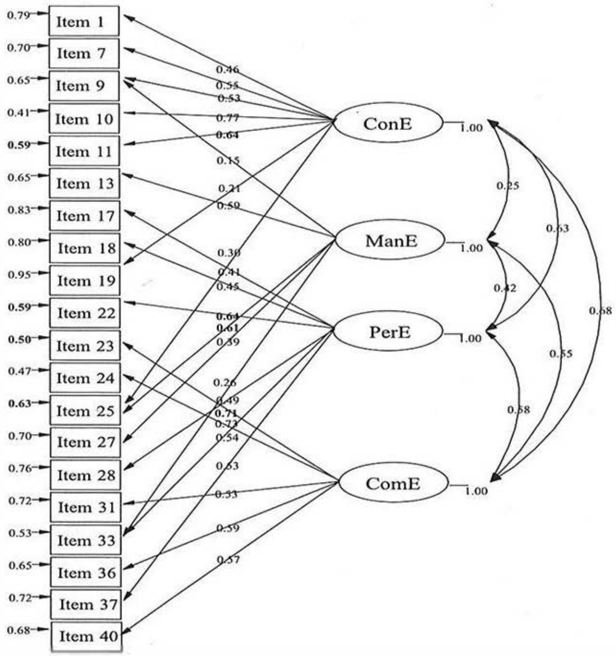
The resulting model contained 20 items (Model 5). As expected, MIs and EPCs of the factor-loading matrix suggested some meaningful moderation effects. These cross-loadings were freed one by one in separate LISREL runs. Items 27 and 33 had a gender-specific secondary loading: for boys, item 27 loaded positively on the Conceptual Exploration factor and item 33 on the Perceptual Exploration factor. Item 9 had also a gender-specific positive secondary loading. For girls it loaded on the Perceptual Exploration factor and for boys on the Conceptual Exploration factor. Thus, items 9, 27 and 33, which had been designed to measure Manipulatory Exploration, seemed to have a secondary meaning for some students, especially for boys. The final model, Model 5, accounted reasonably well for the observed variances and covariances among items in both gender groups.

The values of the RMSEA gave support to the conclusion that Model 5 is a good population model. The point estimate was 0.045 and thus under the value 0.05 which, according to Browne and Cudeck (1993, p. 144), indicates a close fit of the model in relation to the degrees of freedom. The 90 percent confidence interval for RMSEA was from 0.037 to 0.053 for girls and from 0.033 to 0.055 for boys. Since the lower limit of the 90 percent confidence interval was below the recommended 0.05 and the upper confidence limit under 0.08 for both sexes, it was concluded that Model 5 is parsimonious and represents a reasonably close approximation to the population. To get further support for this conclusion, the corresponding model was constructed with the SEPATH program (Steiger, 1995). In addition to the RMSEA estimate, SEPATH also gives the Population Gamma Index and confidence intervals for both indices in all situations. In the present study, the point estimate of the Gamma Index for Model 5 was .967 and the 90 percent confidence interval from .956 to .978. Because the lower limit of the 90 percent confidence interval also exceeded .95, it was concluded that Model 5 has an outstanding fit (Steiger, 1995). Model 5 is displayed in path diagram form in Figure 11.




GIRLS

*FIGURE 11.*
FINAL MEASUREMENT MODEL OF C-TRAIT SCALE (FIXED PARAMETERS IN BOLD FACE).




BOYS

To study the hypothesis that the expected second-order factor lay behind the first-order factors, a second-order factor analysis was conducted (Model 6). In this analysis, the values of the first-order factors ($\lambda$'s and $\theta$'s) were fixed to the values of Model 5 to keep the meaning of the first-order factors constant (see Hayduk, 1996, p. 69). Next, an additional latent variable, Curiosity, was added to the CFA model. It was expected to explain the correlations among the first-order factors, which range from 0.26 to 0.68. Model 6 was tested simultaneously in both gender groups. As shown in Table 9, the statistical fit of this model was acceptable for both samples. The factor loadings of the first-order factors on the second-order factor were all very high and statistically significant for both gender groups. However, the fit for the girls' data was better than for the boys' data. In the boys' analysis, the second-order factor could not explain all the associations among the first-order factors; four MIs of the boys' PSI matrix exceeded the value 5.

8.4.5.5 Testing for the Invariance of the 20-item Scale

A hierarchically nested series of MGCFA was applied to test whether Model 5 is invariant across gender. The initial Model $H_{form}$ tested whether the four-factor Model 5 fit the girls' and the boys' data equally well. The fit of MGCFA Model $H_{form}$ was good, $\chi^2(332) = 508.69$, CFI = .92, IFI = .92, RMSEA = .045. Thus, the next invariance hypothesis, $H_\Lambda$, proposed that the two gender groups have equal factor loadings. As Byrne et al. (1989) noted, in reality most measuring instruments are only partially invariant across groups. Thus, since it was already known that Items 9, 27 and 33 had group-specific cross-loadings, the invariance of these loadings was not tested. All other factor loadings were constrained to be equal across gender. This model was then compared to Model $H_{form}$, in which no equality constraints existed. The result of the $\Delta\chi^2$-test was statistically significant, thereby supporting rejection of the hypothesis of invariant factor loadings, $\Delta\chi^2(13) = 23.71$, $p < .05$. However, the other goodness-of-fit measures remained quite high (e.g., CFI = .92 and IFI = .92). Moreover, RMSEA was .045 and its 90 percent confidence interval from .038 to .053, thus supporting the conclusion that $H_\Lambda$ was a good population model. Examination of MIs suggested that the factor loading of Item 37 was not invariant across gender (MI = 5.2). However, because of the relatively small MIs and to avoid capitalization on chance, the hypothesis of invariant factor loadings was accepted with caution (cf. Steenkamp & Baumgartner, 1998).

Having identified the non-invariant items, the next test of invariance added the restriction that the factor variances and covariances were equivalent across gender ($H_{\Lambda\Phi}$). This model was compared to the previous model, in which only invariant factor loadings were constrained to be equal. The result of the $\Delta\chi^2$-test supported rejection of the hypothesis of equal factor variances and covariances, $\Delta\chi^2(10) = 21.27$, $p < .05$, whereas the other goodness-of-fit measures supported the acceptance of the hypothesis, CFI = .91 and IFI = .91. Moreover, because none of the MIs were over 5 and the point estimate of RMSEA was .046, the hypothesis of equal factor variances and covariances was accepted.

Thus, the next aim was to compare the means of the exploration factors (Model $H_{\Lambda\nu\kappa}$). In terms of both chi-square and goodness-of-fit indices, the fit of this model was unsatisfactory, $\chi^2(358) = 610.95$, $p < .001$, CFI = .89, IFI = .89. The MIs indicated that at least the intercept for Item 36 (MI = 45.13) was not invariant across gender. Relaxing this constraint yielded Model $H_{\Lambda\nu\kappa'}$, which according to goodness-of-fit indices fitted the data quite well, $\chi^2(357) = 564.29$, $p < .001$, CFI = .91, IFI = .91. Comparison of the factor means revealed that there were statistically significant differences between the two gender groups on three factors. On the Manipulatory Exploration factor ($t = -7.03$), boys had a higher mean score than did girls. On the Conceptual Exploration ($t = 3.26$) and Perceptual Exploration factors ($t = 5.99$), girls had statistically higher mean scores than did boys. Finally, the invariance of the second-order factor loadings was tested ($H_{\xi}$ - $H_{\xi\gamma}$). All the loadings were invariant.

### 8.4.5.6 Discussion and Conclusions

When the fit of the initial four-factor model of C-Trait scale was tested, it did not get statistical support. Neither statistical-theoretical not measurement issues were able to explain the misfit of the initial model. A five-factor model supplemented with specific factors (Model 3u) gave the most promising results. Thus, the analyses revealed that it is possible that five rather than four common factors are needed to account for the covariances among the 40 C-trait items. I believe that the postulated specific factors present part of the stable factor structure with this population when the 40-item inventory is used. However, although the statistical fit of Model 3u was almost acceptable, the results of the three unrestricted models revealed that proceeding to a restricted model with all the 40 items was not justified because of serious interpretation problems. Thus, the analysis was continued with a reduced scale.

Hayduk (1996) has argued strongly that only one or two items are needed to measure a factor when both the factor loadings and the error variances of the reference variables are fixed to a certain value. On the other hand, from a scientific perspective, the construct validity of exploration and curiosity is at a preliminary stage. Thus, it was considered unlikely that, at this stage in the construct validation of the inventory, two C-Trait items would be sufficient to capture all aspects of a concept such as exploration aroused by curiosity. Following this rationale, a 20-item model was constructed. Items were chosen for this model one by one in order to ensure that they would cover all aspects of the concepts. However, only one indicator per idea was included in the model. The statistical fit of the 20-item model was reasonably good.

When the metric invariance of the 20-item C-Trait scale was analyzed, the results showed that the scale contained non-invariant factor loadings. According to Carlson and Mulaik (1993), this sort of finding means that girls and boys interpreted contents of C-trait items 9, 27 and 33 differently. Items 9, 27 and 33 concerned manipulations of percept objects. Thus, its not surprising that girls and boys interpreted these items differently. The invariance analyses also showed that on the Conceptual Exploration and Perceptual Exploration factors girls had a higher mean score than boys. On the other hand, on the Manipulatory Exploration factor boys scored higher than girls. The mean difference on Manipulatory Exploration supported previous findings. Several studies (e.g., Keller, Schölmerich, Miranda, & Gauda, 1987; Kreitler et al., 1975, 1984; Schneider, 1987; Voss & Keller, 1986) have reported that boys show more manipulatory exploration than girls, and they do this more often than girls. On the other hand, Manipulatory Exploration was the only dimension of the four dimensions (Manipulatory Exploration, Perceptual Exploration, Conceptual Exploration, and Exploration of the Complex) where Kreitler et al. (1975; 1984) found sex difference.

## 8.5.6 SENSATION SEEKING SCALE

The attempt to test the four-factor measurement model of sensation seeking with the item-level data led to same kind of problems as in the study of Björck-Åkesson (1990, p. 99). The model fitted poorly for both gender groups. Following the thinking of Björck-Åkesson, these results were expected to have been caused by statistical shortcomings, specially the poor measurement proprieties of the dichotomously scored items. Thus, to avoid these problems, the same kind of strategy as that of Björck-Åkesson (1990, pp. 99-100; see also Gorsuch, 1983, pp. 292-295; Reuterberg & Gustafsson, 1992) was used. Following the thinking of Björck-Åkesson scores were summed to form two half scales for each subscale TAS, NES, Act, and Out. The same items as in the study of Björck-Åkesson were used to create the half scales. Imputation of missing data was used. All the other variables were used as matching variables in this process. After imputation of missing data, the total effective sample sizes were 255 for girls and 264 for boys. The eight half scales were treated as continuous variables. None of the new variables had skewness or kurtosis over two, and only one variable had these values over one, namely, variable N1.2. in the boys' data had kurtosis 1.4. Moreover, Mardia's multivariate tests of skewness and kurtosis showed that both samples did not deviate statistically significantly ($p_{Girls}$ = .031 and $p_{Boys}$ = .014) from multivariate normality. However, in spite of this finding both covariance matrix and asymptotic covariance matrix were used as input matrices in the analyses to get Satorra-Bentler scaled Chi-Square statistics.

### 8.5.6.1 The Model With Four First-Order Factors

First, a simple measurement model (Model SS1) was tested where the eight half tests were used as observed variables. They were expected to measure four latent variables (TAS, NES, Act, and Out). Using the same logic as Björck-Åkesson (1990) the half scales were labeled as T1.1, T1.2, N1.1, N1.2, A1.1, A1.2, O1.1 and O1.2. However, the analysis differed in two respects from that of Björck-Åkesson. First, all models were tested separately for girls and boys. Second, a covariance matrix together with an asymptotic covariance matrix was used instead of a plain correlation matrix. As shown in Table 10, Model SS1 fitted the data well in both gender groups. This model is shown graphically in Figure 12.

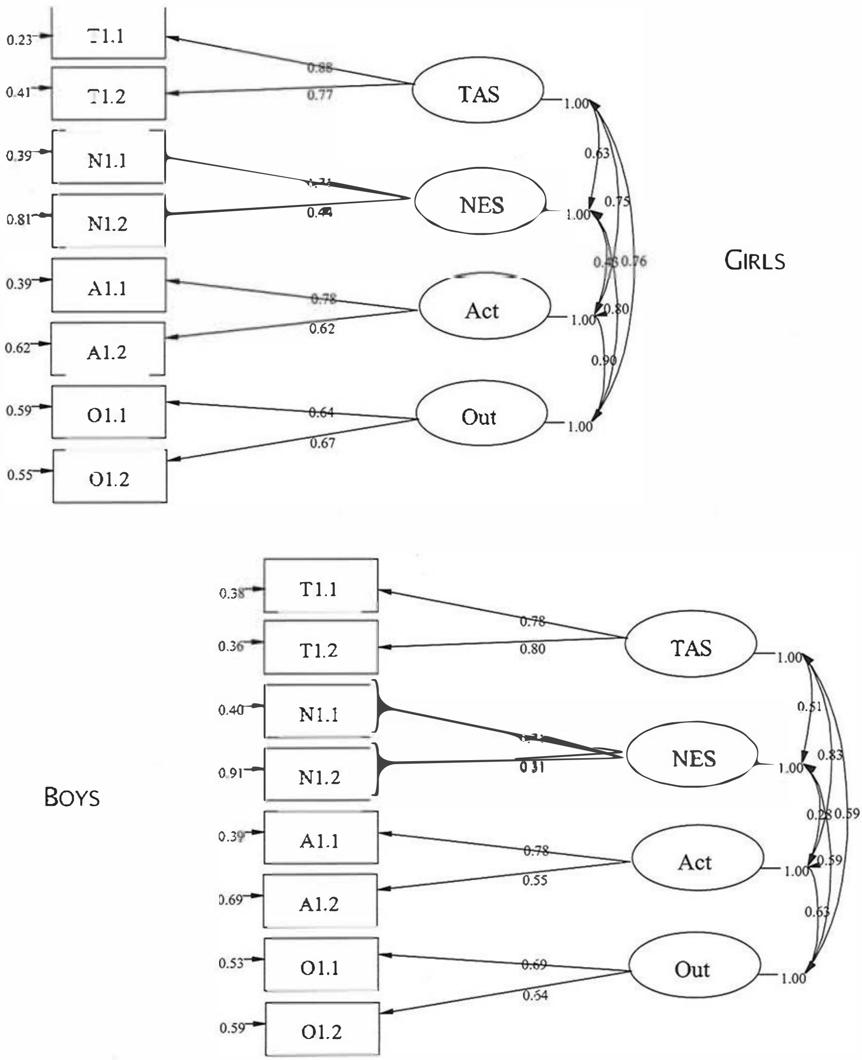



*FIGURE 12.* THE MEASUREMENT MODEL SS 1 WITH FOUR FIRST-ORDER FACTORS (SESE).

Although the overall fit of Model SS1 was good, the detailed analysis of the results revealed discriminant validity problems for the four first-order factors. The correlation between factors Act and Out was .90 in the girls' group and the correlation between TAS and Act factor was .81 in the boys' group. These findings suggested lack of discriminant validity. To test this hypothesis, a method suggested by Schumacker and Lomax (1996; see also Maruyama, 1998) was used. The chi-square value of Model SS1 was compared to the values of two restricted models in which the correlations between the problematic factors was fixed to value 1. This comparison provided evidence of a lack of discriminant validity only between factors Act and Out in the girls' group, S-B$\Delta\chi^2(1) = 1.51$, $p$ = n.s.

### 8.5.6.2 The NF Model with Five Orthogonal Factors

The second model tested was a five-factor model (Model SS2) which was based on the logic of Gustafsson and Blake (1993). Model SS2 contained five orthogonal factors, and one of them (SESE) was assumed to be a broad factor which should relate to all observed variables. The other four factors (TAS, NES, Out, and Act) were expected to be more specific and to account for the covariance only between two observed

TABLE 10 GOODNESS-OF-FIT OF ALTERNATIVE SENSATION SEEKING SCALE MODELS

| Model | df | S-B$\chi^2$ | CFI | PCFI | IFI | RMSEA |
|---|---|---|---|---|---|---|
| Girls | | | | | | |
| 1 Model SS1 | 14 | 20 | .99 | .50 | .99 | .04 |
| 2 Model SS2 | 16 | 39 | .96 | .55 | .96 | .08 |
| 3 Model SS3 | 16 | 39 | .96 | .55 | .96 | .08 |
| Boys | | | | | | |
| 1 Model SS1 | 14 | 23 | .98 | .50 | .98 | .05 |
| 2 Model SS2 | 16 | 39 | .95 | .54 | .95 | .07 |
| 3 Model SS3 | 16 | 39 | .95 | .54 | .95 | .07 |

variables. For instance, latent variable TAS was expected to account for the variance between half scales T1.1 and T1.2. The resulting model is presented in Figure 13. Overall, as indicated by the fit indices of Table 10, the fit of Model SS2 was good.

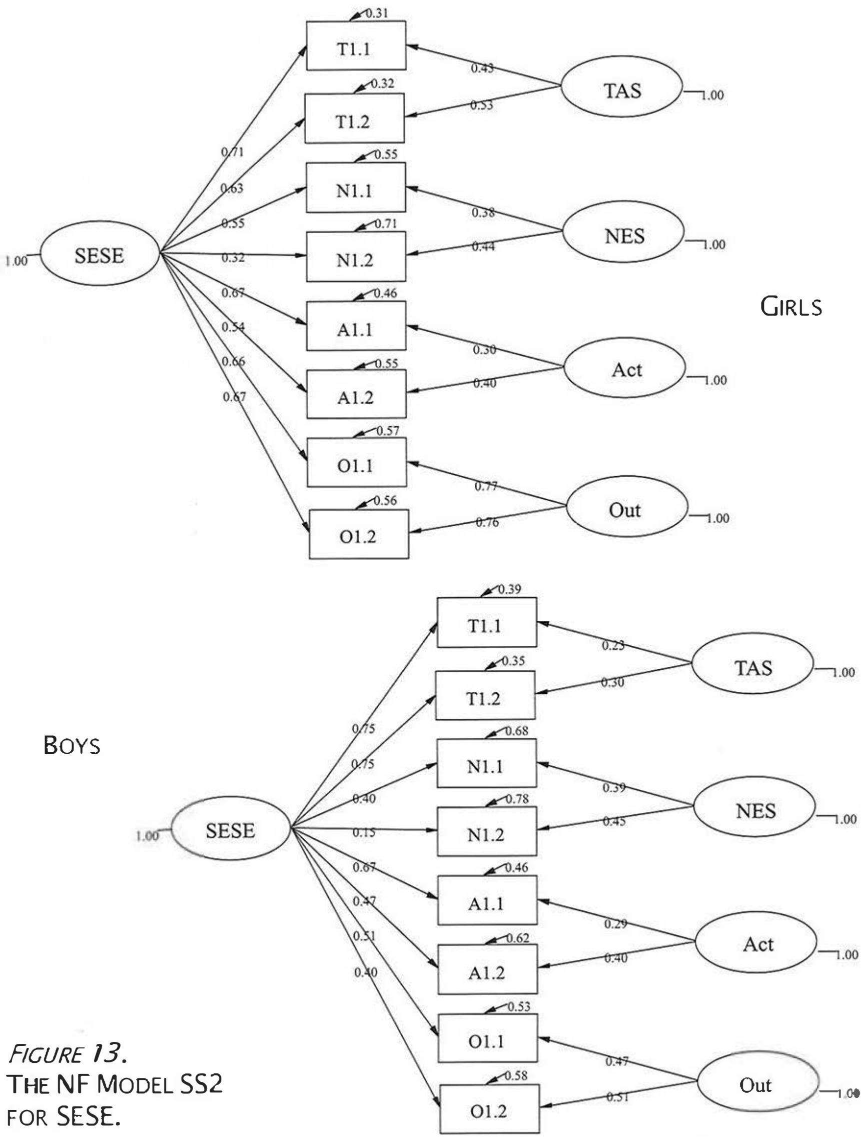
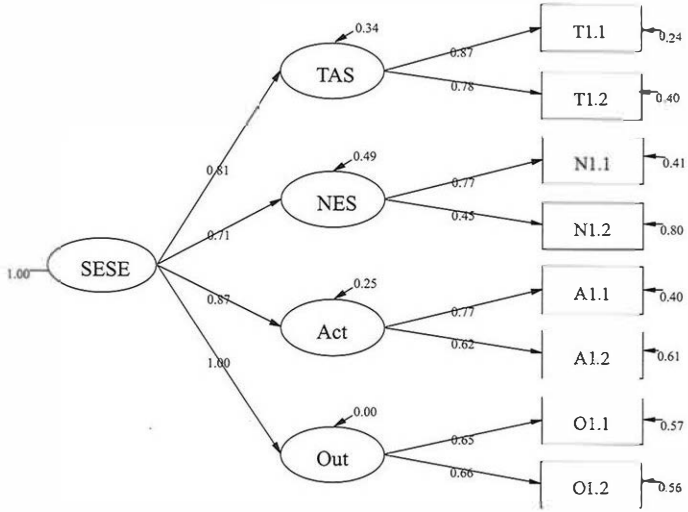



*FIGURE 13.*
THE NF MODEL SS2
FOR SESE.

### 8.5.6.3 The Model with One Second-Order Factor

Model SS3 assumed that a second-order factor Sensation Seeking (SESE), accounts for the intercorrelations among the four first-order factors TAS, NES, Act, and Out. In this sense, the Model SS3 was a special case of Model SS1 (Rindskopf & Rose, 1989). Table 10 shows




GIRLS

BOYS

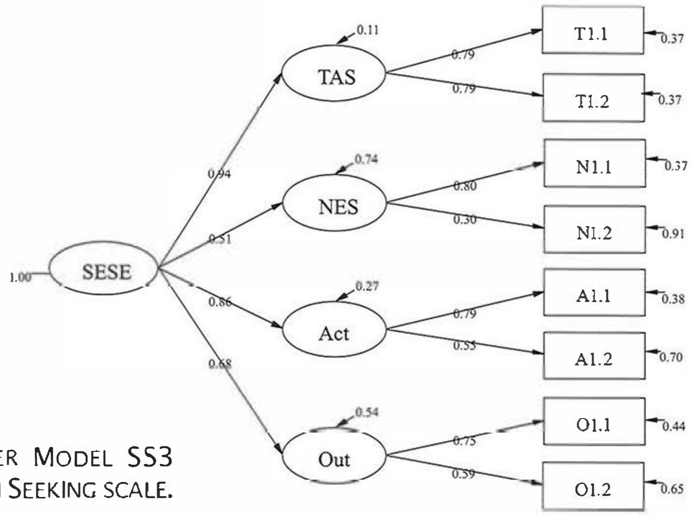



*FIGURE 14.*
SECOND-ORDER MODEL SS3 FOR SENSATION SEEKING SCALE.

that the fit of Model SS3 was good and on the same level as the fit of Model SS2. The results of Model SS3 are presented in Figure 14.

A detailed inspection of the results of the Model SS3 for both gender groups showed that all loadings on the secondary factor were statistically significant. Moreover, variables T1.1, A1.1 and O2.2 had the highest total effects ($t > 9.0$) on the second-order factor in the girls' analysis and half-scales T1.1, T1.2, and A1.1 in the boys' analysis. However, the results included also one unexpected finding. In the girls' group the error variance of the Out factor was 0, while the corresponding SMC was 1.00, thus indicating that the second-order factor SESE accounts for all the variance of the Out factor, and that Out and SESE are identical dimensions. This was a surprising result, although Björk-Åkesson (1990) has argued that the broad Sensation Seeking factor has a larger influence on the Out scale than on the other scales. There are several possible reasons for this finding (see Rindskopf & Rose, 1988). However, according to the results of Models SS1 and SS3, the most plausible reason for this finding may be the excessively high intercorrelation of the first order factors Act and Out. As the results for Model SS1 showed these two factors may be identical dimensions.

#### 8.5.6.4 Testing for the Invariance of the Sensation Seeking Scale

Only the invariance of Model SS1 was tested. The testing process began with the base model, which was a multigroup version of Model SS1. The fit of this model, $H_{form}$, was excellent. However, some reservations must be made to accepting Model $H_{form}$ because the high correlation between factors Act and Out was .90 in girls' sample. As Steenkamp and Baumgartner (1998) noted, one precondition to configural invariance is that the correlations between factors are significantly below unity.

Next, the invariance of factor loadings was investigated. Model $H_{\lambda}$, where factor loadings were determined as invariant, fitted the data very well. Moreover, there was statistically non-significant rise in chi-square value (S-B$\Delta\chi^2(8) = 3.91$). Next, the invariance of factor variances and covariances was tested. The fit of Model $H_{\Lambda\Phi}$ was good, S-B$\chi^2(42) = 52.90$, $p = 0.12$, CFI = .99, IFI = .99, RMSEA = .032. In addition, no statistically significant increase in chi-square was found when Model $H_{\Lambda\Phi}$ was compared to Model $H_{\lambda}$ using the chi-square difference test, S-B$\Delta\chi^2(6)$ = 8.28. The range of correlations among four first-order factors range was .39 to .79 in Model $H_{\Lambda\Phi}$, where the factor correlations were estimated as invariant across gender. The factor correlation between factors Act and Out, which was .90 in girls' Model SS1, was .78 in this model.

Finally, Model $H_{\Lambda\nu\kappa}$ was constructed to test the equality of the factor means. In this model the intercept terms were defined as invariant across gender. According to the fit indices, the model fitted the data quite well, S-B$\Delta\chi^2(36)$ = 86.71, $p < .001$, CFI = .96 and IFI = .97. The model contained one high intercept term, namely for Item A1.2 (MI = 34.87). However, because the RMSEA estimate was .074 and its 90 percent confidence interval from .054 to .094, no modifications were made in Model $H_{\Lambda\nu\kappa}$. A detailed inspection of the results showed that boys scored higher than girls on the TAS ($t$ = 5.40) and Act ($t$ = 5.87) factors, whereas girls scored higher than boys on the NES ($t$ = -1.97) and Out ($t$ = -2.53) factors.

### 8.5.6.5 Discussion and Conclusions

The testing process of SESE was not as simple as the testing process of other scales. The model with four first-order factors was the base model of the Sensation Seeking scale. However, to test the validity of SESE, two other models were also tested. The first test involved a second-order factor model that reflected the thinking of Zuckerman (1979, 1984, 1994). The second model was an orthogonal model with five nested factors. This NF model was based on the methodological thinking of Gustafsson and Blake (1993). Although the interpretation of these two models was different, the statistical fit of both models was on the same acceptable level. This result leads one to ask which is the most plausible model for sensation seeking. Björk-Åkesson (1990) preferred the NF model, but Mulaik and Quartetti (1997) later criticized its starting point. They argued that the orthogonality hypothesis is a very strong assumption which should also be empirically testable. However, when Mulaik and Quartetti tried to do this by freeing up the correlations among the factors, they got "inadmissible solutions," and the LISREL program indicated that the correlation matrix among the factors was not positively definite.

The discriminant validity problems of SESE were not surprising, as several other studies (e.g., Russo et al., 1991, 1993) have also reported similar problems when developing a children's sensation seeking scale. Thus, how many dimensions are needed to describe children's sensation seeking is still open to question. In adults studies, the TAS subscale has had the best cross-gender and cross-cultural replicability (Zuckerman, 1994). The most recent version of the adults' Sensation Seeking scale contains only two subscales, TAS and ES. In the light of recent study, the Björck-Åkesson version of SSS also needs to increase the discriminant validity of the subscales. In order to reach this goal, additional work at both the conceptual and the measurement level is needed.

In Björck-Åkesson's study boys had a higher mean score than did girls on the Thrill and Adventure Seeking and Activity factors. I obtained a similar result; however in contrast to her findings, my data showed a statistically significant difference between the gender groups on the mean structure of the New Experience Seeking and Outgoingness factors. Girls scored higher than boys on these factors. Zuckerman (1994) also reported several studies where boys have scored higher than girls on the Thrill and Adventure Seeking factor.

### 8.5.7 TEACHER RATING SCALE

To test the hypothesis that two factors account for the observed covariances among the five teacher rating scale items, an equivalent unrestricted factor model (Model TeR1u) was constructed for both gender groups. Item 1 was postulated to load only on Factor 1, and Item 4 only on Factor 2. These two reference variables were expected to be pure in their respective factors and very close to the meaning of the concept. This model was tested separately for girls and boys. As shown in Table 11, the fit of Model TeR1u was almost perfect for both gender groups.

Next, a two-factor measurement model (Model TeR1) was constructed and tested separately in both gender groups. The overall fit indices for this model suggested contradictory conclusions (Table 11). That is, CFI and IFI suggested the acceptance of the model in both gender groups. On the other hand, Satorra-Bentler scaled Chi-Square was highly significant, and RMSEA was much over 0.05. Thus, these two statistics suggested the rejection of the model. Comparison

TABLE 11 GOODNESS-OF-FIT STATISTICS OF THE TER SCALE

| Model | df | S-B$\chi^2$ | CFI | PCFI | IFI | RMSEA |
|---|---|---|---|---|---|---|
| Girls | | | | | | |
| 1 Model TeR 1u | 1 | 0 | 1.00 | .10 | 1.00 | .00 |
| 2 Model TeR 1 | 4 | 22 | .94 | .38 | .94 | .14 |
| 3 Model TeR 2 | 2 | 0 | 1.00 | .20 | 1.00 | .00 |
| Boys | | | | | | |
| 1 Model TeR 1u | 1 | 0 | 1.00 | .10 | 1.00 | .00 |
| 2 Model TeR 1 | 4 | 29 | .95 | .38 | .95 | .17 |
| 3 Model TeR 2 | 2 | 2 | 1.00 | .20 | 1.00 | .00 |

of Model TeR1 to Model TeR1u with the chi-square difference test, suggested also rejection of model $S\text{-}B\Delta\chi^2(3)_{Girls} = 21.21$, $p < .001$ and $S\text{-}B\Delta\chi^2(3)_{Boys} = 29.22$, $p < .001$. This result indicated that Model TR1 contained some unjustified restrictions.

The detailed inspection of the results for Model TeR1 showed that two cross-loadings could explain the contradictory findings of the overall fit. MIs suggested that Items 1 and 3 should be freed on the Diversive Exploration factor in both gender groups. According to the EPC matrix, Item 1 "Student reacts positively to new, incongruous or mysterious elements in his environment by moving toward them, by exploring them or by manipulating them" had a positive secondary loading on the Diversive Exploration factor and Item 3 "Student persists in examining and exploring stimuli in order to know more about them" a negative loading on this factor. Because both cross-loadings seemed to be substantially meaningful, a post hoc Model TeR2 was postulated where these two variables were allowed to load on the Diversive Exploration factor. The statistical fit of this model was good (see Table 11). Thus, the two cross-loadings of Items 1 and 3 explained the contradictory results of Model TeR1. However, a detailed inspection of the results of all three models reveled one unexpected finding: the correlation between the two factors was unexpectedly high: 0.67 for girls and 0.53 for boys in Model TeR2. Figure 15 displays the final model of TeR scale.

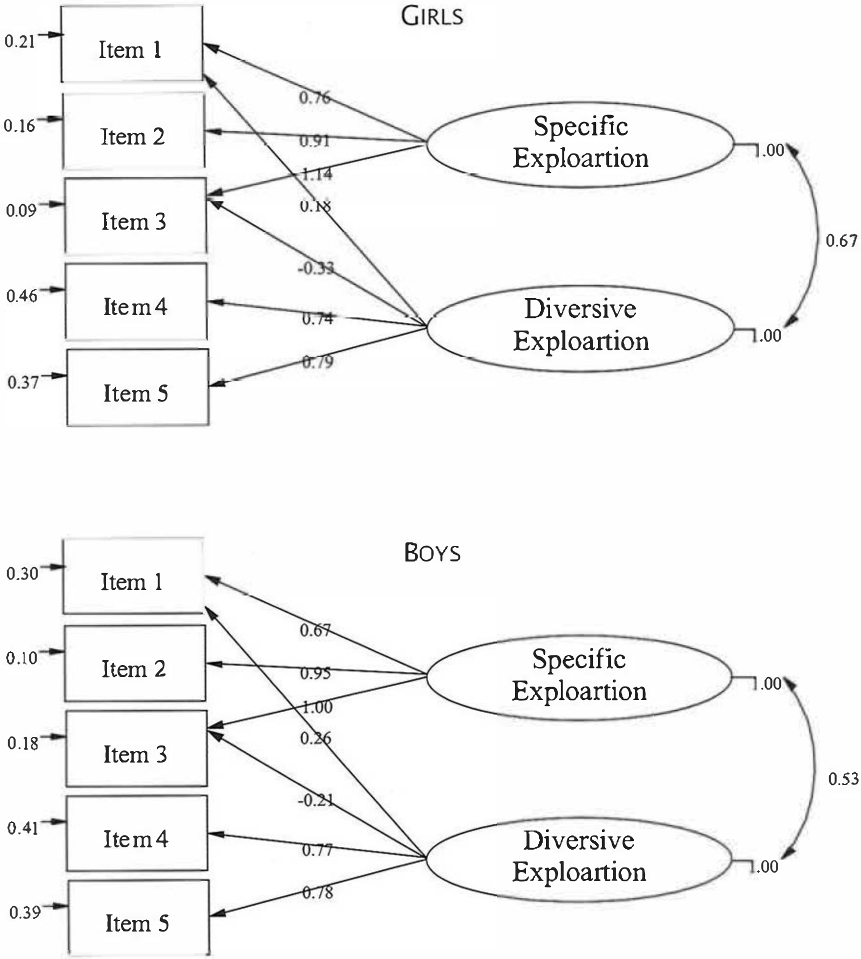



*FIGURE 15.* FINAL MODEL OF TER SCALE.

### 8.5.7.1 Tests of Invariance

The process of testing the hierarchical invariance of the TeR scale began with the construction of a baseline model. Because it is important to have a well-fitting baseline model (see e.g., Byrne, et al., 1989), Model TeR2 was used as a multigroup model. The fit of this Model $H_{form}$ was excellent (S-B$\chi^2$(4) = 1.66, CFI = 1.00, IFI = 1.00). Next, Model $H_{\Lambda}$ was constructed where the factor loadings were set be equal. When the fit of this model was compared to the fit of the baseline Model $H_{form}$, the value of the chi-square test was non-significant (S-B$\Delta\chi^2$(7)=10.81) thus supporting the hypothesis of equal factor loadings.

Having received support for the invariant measurement parameters, the invariance testing process continued with the test of structural parameters. The factor variance-covariance matrix $\Phi$ was set to be equal across gender, and a multigroup Model $H_{\Lambda\Phi}$ was postulated. This model was compared to Model $H_{\Lambda}$ in which only invariant factor loadings were constrained to be equal. The result of the $\Delta\chi^2$-test supported the hypothesis of equal factor variances and covariances, S-B$\Delta\chi^2$(3) =2.57. Thus, the invariance testing process continued with a test of equal factor means. To explore this, Model $H_{\Lambda\nu\kappa}$ was constructed. The fit of this model was excellent, S-B$\chi^2$(15) = 19.98, p = .17, CFI = .99 and IFI = .99. According to the results of Model $H_{\Lambda\nu\kappa}$, boys scored significantly higher than girls on the Diversive Exploration factor ($t$ = 3.15). There was no statistically significant difference between gender groups on the Specific Exploration factor.

### 8.5.7.2 Conclusions

The analyses of the TeR scale revealed that Items 1 and 3, which were designed to measure specific curiosity, have a secondary loading on the Diversive Exploration factor. Moreover, it was also found that the correlation between both exploration factors was unacceptably high. Together these findings suggested that modifications should be made to the wording of the TeR scale items so that the items really have only one meaning. Moreover, the wording must also be changed so that there is a clearer difference between items which measure specific exploration and diversive exploration. As one teacher noted, items in the present form are too similar, and it is difficult to differentiate between the meanings of separate items.

## 8.6 DISCUSSION

In the present study it seemed at first that the factorial validity of IM, OTIMDC, C-Trait and SESE scales failed. The fit of the hypothesized factor models did not have enough statistical support. To determine the reason for the badness of fit of a model, the strategic starting point of the present analysis was to continue in an exploratory mode and systematically investigate the three reason categories first suggested by Cronbach and Meehl (1955): statistical issues, theory, and measurement.

In the present study, Satorra-Bentler scaled chi-square statistics were used as the strategy to eliminate the statistical issues related to the non-normality of the observed variables. However, sample size restrictions prevented the use of the SCALED $\chi^2$ statistic to analyze both C-Trait and SESE scales. This and the poor measurement properties of dichotomously scored items led to serious problems in analyzing the initial SESE models. The initial models fitted the data poorly and convergence problems also emerged. To improve the measuring properties of the SESE items, the same parceling technique was used as in the study of Björck-Åkesson (1990). Each parcel was composed of half of the items of each subscale. Thus, the remaining analyses were done with eight half scales. However, together, these findings suggested a new response format for the SESE. Recently, Zuckerman (1984, 1994) has presented a version of the Sensation Seeking Scale (SSS VI) where the dichotomously scored items are replaced by a three-point Likert-type, weighted scale. The results of the present study suggest that a multipoint response format should also be devised for the present children's version of the Sensation Seeking scale. This multipoint response format can also be similar to Harter's (1981) Intrinsic Motivation scale.

If statistical issues cannot explain the misfit of the initial model, another theoretical model may be better able to account for the covariances among the observed variables. In this study the IM scale and the C-Trait scale were the only scales where it was possible to postulate two alternative conceptual models for the initial model. However, either the statistical fit of the presented alternative models was poor or the models suffered from a lack of discriminant validity. The five-factor model of the C-Trait scale was an exception to this rule. The findings of the present analysis gave some support to the hypothesis that what is called perceptual exploration by Kreitler and Kreitler (1994) may split into at least two dimensions in girls' population.

Method artifacts, such as response sets and redundant item wording, contaminated all scales used. Several studies (e.g., Byrne, 1994; Byrne & Shavelson, 1996; Byrne, Shavelson & Muthén, 1989) with psychological variables have demonstrated that in this kind of situation it is necessary to allow some substantially meaningful measurement errors to correlate in order to get a well-fitting model. Actually, Bentler and Chou (1987) argued that the specification of a model that forces all error parameters to be uncorrelated is "rarely appropriate with real data." Moreover, Byrne et al. (1989) emphasized that error parameter specifications are justified because they usually represent non-random measurement errors due to such method effects as response style or similarity in item wording. The results of the present study gave support to this claim. The results also suggested that some reverse-worded items are a source of appreciable extraneous variance. Because reverse-worded items were difficult to understand for some of the fifth graders, it was suggested that these items lead to "yea-saying" or "nay-saying" response sets which manifested themselves in correlated errors.

When the extra parameters were added to the models of the present study, the estimates of major parameters changed very little. According to Byrne et al. (1989), this is an indication of a robust initial model. However, some correlated errors and specific factors of the present study may also be interpreted as omitting factors. Rubenstein (1986) suggested a distinction between domain-specific and content-specific curiosity. In the present study the analysis of IM and C-Trait scales suggested that one manifestation of domain-specific curiosity may be curiosity which is related to machines. In the present study this dimension was treated as an orthogonal dimension, but in future studies the orthogonality hypothesis must also be tested.

In the present study, analyses of the C-Trait and SESE scales had a special role. Both scales were expected to be four-dimensional measures of curiosity. Moreover, the conceptual models of both scales suggested that a more general second-order factor determined the four first-order factors. Thus to obtain support for the validity of the

C-Trait and SESE scales, the hypothesized second-order structures were also tested. When the higher order structure of the 20-item C-Trait scale was tested, the expected second-order curiosity factor explained the correlations (covariances) among the four first-order factors quite well.

The invariance tests conducted here revealed that inventories were differently valid for girls and boys. The tests followed the logic suggested by Byrne (1998). The analyses of configural invariance ($H_{form}$) revealed that several different factors may account for the covariances among SESE items in the two samples. The correlation between factors Act and Out was very high in the girls' data and the tests indicated that it may be even 1. This finding means that Act and Out may be the same factors in the girls' population. On the other hand, as Bollen and Hoyle (1990) have noted, high or even perfect correlations between two dimensions do not necessarily mean that the concept behind the measurement is unidimensional rather than bidimensional. What this kind of finding means is that empirically the conceptually distinct concepts are almost perfectly correlated.

In the present analysis some scales contained non-invariant measurement intercepts. However, all these intercepts were also theoretically meaningful. By relaxing these intercepts and thus removing the cumulative bias of non-invariant items, it was possible to compare the mean differences between gender groups. Girls scored higher than boys on the Conceptual Exploration, Perceptual Exploration, New Experience Seeking, and Outgoingness factors. Boys scored higher than girls on the Diversive Exploration, Manipulatory Exploration, Thrill and Adventure Seeking, and Activity factors. These findings mostly supported previous findings, although only a few studies systematically tested invariance of the scales used before comparing the means.

Thus, although mild gender effects were found by applying the systematic invariance tests, these effects did not pose a threat to intergroup consistency. However, the results of the invariance tests done here must be taken account when the results of subscale level analyses are inspected. Otherwise, the scales used here produced results that were consistent with the presented hypothesis. The mean differences were in the same directions as in previous studies, and the higher-order models account reasonably well for the correlations among first-order factor. Together, these findings gave partial support to the validity of the scales used.

---

[1]According to Jöreskog (1979a, 1979b), an unrestricted model is equivalent to an exploratory common factor analysis model. To test this, a four-factor ML exploratory factor solution was computed using SPSS for both gender groups. The chi-square values with 626 degrees of freedom were 1035.79 (1108.39) for girls and 924.49 (985.96) for boys. The corresponding LISREL values are in parentheses. The chi-square values from LISREL are slightly higher than chi-square values from SPSS because of different multiplicative factors in front of the ML fit function. In SPSS the (n-(2p+5)) multiplicative factor is used, while in LISREL n is used (p = number of variables and n = N-1, N=number of observations). For instance, if the SPSS chi-square value for boys 924.49 is divided by [(271-1)-((2*40+5)/6)] the result (3.61), is very near the reported Minimum Fit Function value for the unrestricted model fitted in LISREL (see Hartman, personal communication to David Kaplan on SEMNET, April 26, 1996).

[2]In this kind of situation the LISREL program prints a warning, "Total sample size is smaller than the number of parameters. Parameter estimates are unreliable."

[3]Although the final 20-item scale corresponded very well with the initial 40-item inventory, there was one important exception. Neither Item 14 nor Item 38, both of which measured auditory exploration, were included in the model. Thus, a better name for the final perceptual exploration factor may be Visual and Tactile Exploration. The validity of Items 14 and 38 was found to be questionable with this population, and thus rewording of these items is needed.

[4]The fit of the corresponding unrestricted MGCFA model (Model 5u) was good, $\chi^2(232, n_g=258, n_b=271)=307, p<.001$, IFI=.97, CFI=.97, RMSEA=0.036.

[5]Degrees of freedom for the second-order factor model test were reduced by one for each fixed measurement parameter (see Hayduk, 1996, p. 69).

# 9. Subscale-Level Analysis

This chapter focuses on the fourth research problem, namely the dimensionality problem of the subscale-level data. First, the problems related to item level and a subscale-level analysis are compared. Thereafter nine alternative conceptual models are presented in details. Finally, the postulated alternative models are statistically tested.

## 9.1 Methodological Starting Points

Most of the previous studies have investigated the dimensionality problem of curiosity by analysing subscale data (e.g., Ainley, 1987; Langevin, 1971; Spielberger & Starr, 1994). Rubenstein (1986) has been an exception to this rule. She analyzed both the item level and subscale-level data of five curiosity scales. However, the results of the item level data are questionable because the sample size for her analysis was 155 and the factor analysis contained 212 variables. Thus, the sample size-variable ratio of her analysis was miserable and below all recommendations (see e.g., Gorsuch, 1983; Nunnally, 1978). Moreover, several researchers (e.g., Byrne, 1988; Cattell & Burdsal, 1975; Comrey & Lee, 1992; Kishton & Widaman, 1994) have argued that single items are too unstable for precise factor analytical work, especially for studies designed to develop the scientific taxonomy of personality. In this kind of work, single items have the following problems: 1) they are unreliable, 2) they contain unique variance since they are affected by idiosyncratic wording, 3) usually they are not normally distributed, 4) they have low intercorrelations, and 5) analyses using single items as measured indicators for the factors often make it necessary to estimate a large number of parameters.

In the present analysis, the use of item-level data would have been meant that 125 items would have specified 12 factors each with five to 15 loadings and then the first-order factors would have specified one to four second-order factors. In practice, reliable estimation of this kind of model with a huge number of free parameters would have required a very large sample size (c.f. Bentler & Chou, 1987, p. 91). Moreover, based on the thinking of Floyd and Widaman (1995, p. 293), item-level analysis in the present investigation would have led to overwhelming difficulties in specifying correlated residuals and disturbances among pairs of items, as well.

To reduce the idiosyncratic characteristics of individual items, several authors (e.g., Cattell & Burdsal, 1975; Comrey & Lee, 1992; Floyd & Widaman, 1995; Schau, Stevens, Dauphinee & Vecchio, 1995) have suggested the use of summed item scores as variables in factor analysis. Cattell and Burdsal (1975) called these summed scale scores "parcels," and Comrey (1978) called them FHIDs (Factored Homogenous Item Dimensions). Both Cattell and Burdsal (1975) and Comrey (1978; see also Comrey & Lee, 1992) have proposed their own methods of finding suitable items for summed scale scores. Cattell and Burdsal (1975), for instance, have suggested that factor analysis should be used to find suitable items for parcels, after which the parcels are factored. However, common to all systems used to construct parcels is the need to find internally consistent and unidimensional subsets of items which then are summed as new variables (Kishton & Widaman, 1994). Thus, following the principle of parsimony and to avoid the idiosyncratic characteristics of individual items, the starting point of the present investigation was that summed subscale scores represent the major constructs that are needed to provide the basis for scientific development in the field of curiosity.

The item-level analyses of the present study at least partly confirmed the dimensionality hypothesis of the 15 subscales. The internal consistency and invariance hypotheses also received partial support. Thus, on the basis of item-level analysis, summed subscale scores (direct sums) were formed separately for girls and boys. Before testing the statistical models, the corresponding conceptual models were presented. Several alternative models were constructed. The subscale-level models presented here and the mediation effects of gender were tested using confirmatory factor analysis. The testing procedure started with an unrestricted model. Depending on the results with the unrestricted models, the analysis continued with tests of more restricted models. If an unrestricted model failed to adequately account for the variances and covariances in the observed data, then the other, more restricted model was rejected (Mulaik & Millsap, 2000). The same fit criteria as in the item-level analyses were used.

## 9.2 CONCEPTUAL MODELS FOR THE SUBSCALE DATA

Langevin (1971) was the first researcher who empirically investigated the question "Is curiosity a unitary construct?" Langevin's approach to the problem was inductive. Thereafter, Ainley (1987), Boyle, (1989), Olson and Camp (1984), and Rubenstein (1986) also approached the dimensionality problem of curiosity, using inductive methodology. Based on this earlier empirical research and some theoretical analyses (e.g., Berlyne, 1960; Day & Berlyne, 1971; Livson, 1967), it is now possible to construct alternative conceptual models of curiosity. The starting point of the model construction was the same as in modern trait theory (Johnson, 1997; Wiggins, 1997; Wiggins & Trapnell, 1997). The 15 subscales (see Appendix B) were expected to present outer or behavioral traits. They described forms of exploration that have been related to curiosity in several other studies (e.g., Ainley, 1987; Kreitler & Kreitler, 1994). The postulated curiosity factors were expected to be inner cognitive-emotional traits which generate the outer traits and thus explain the covariances among the 15 subscales (see Figure 16). All the conceptual models presented here are based on factor analytical thinking (cf. Hayduk, 1996).

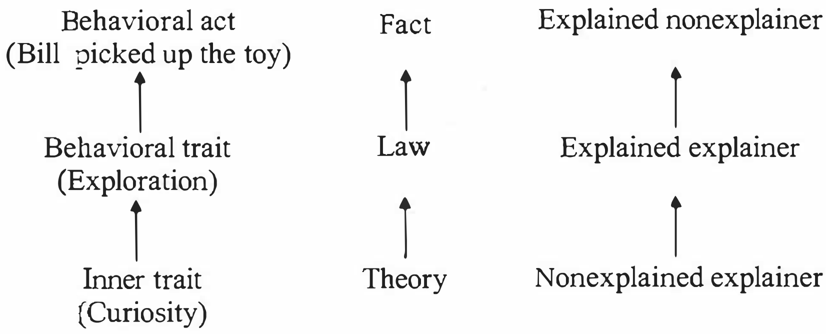



*FIGURE 16.* CURIOSITY AND EXPLORATION AS INNER AND OUTER TRAITS.

### 9.2.1 ONE-DIMENSIONAL MODEL

The one-dimensional model of the present investigation was based on the study of Olson and Camp (1984). In that study a broad General Curiosity factor appeared. Later, Boyle (1989) also suggested the existence of a global Curiosity factor. However, what Boyle meant by curiosity on the conceptual level seems not to be as broad a concept as that of Olson and Camp. Aspects which have been related to diversive curiosity (Day, 1971) and sensation seeking (Zuckerman, 1978; 1994) are lacking in the conceptualization of Boyle. However, the first subscale model of the present investigation argues that the 15 subscales used are congeneric measures of curiosity, and that, one general curiosity factor thus accounts for the covariances among the 15 subscales (Model 1). Figure 17 presents the corresponding conceptual path model.

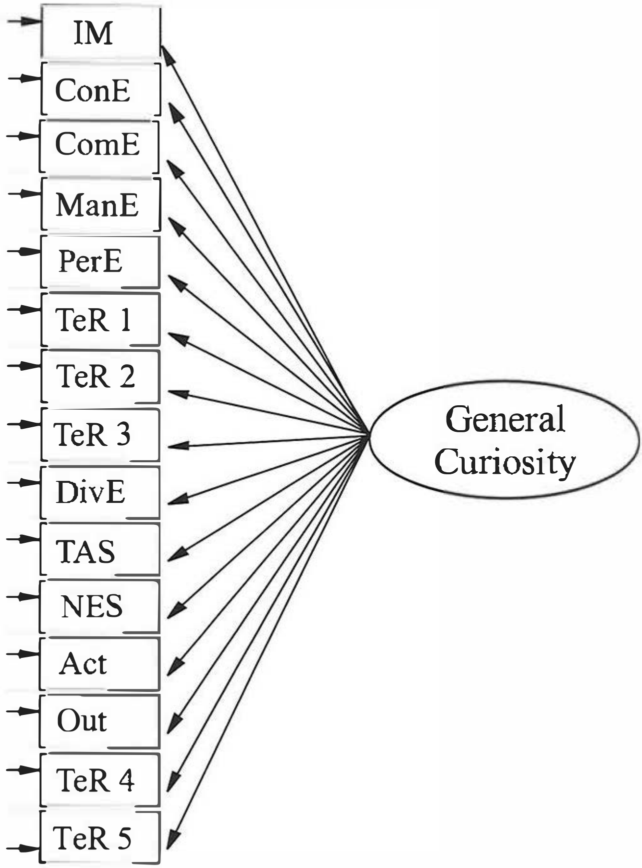



*FIGURE 17.* CONCEPTUAL MODEL OF GENERAL CURIOSITY. (NOTE: DIVE IS THE SUM OF SEVEN OTIMDC ITEMS, SEE TEXT.)

## 9.2.2 Two-Dimensional Model

Both theoretical and empirical findings support the two-dimensional model. Berlyne (1960) first made an implicit distinction between specific and diversive curiosity. Later, Berlyne and Day (1971) made this distinction explicit. Based on factor analytical findings, both Langevin (1971, 1976) and Ainley (1987) also suggested a two-dimensional model. The present investigation and the investigation of Ainley have some scales in common, namely Test of Intrinsic Motivation (IM), Sensation Seeking Scale (SSS), and Diversive Curiosity Scale (OTIMDC). In Ainley's study IM loaded highly ($\lambda$=.78) on the factor called "Depth-of-interest Curiosity Style." Pearson's (1970) Novelty Experiencing Scale – Internal Sensations (NESIS), Internal Cognitions (NESIC) and External Cognitions (NESEC) together with Day's (1971) Ontario Test of Intrinsic Motivation – Specific Curiosity Scale (OTIMSC), and Naylor's Melbourne Curiosity Inventory – Trait Form (MCIT) were other scales that loaded on the depth-of-interest curiosity factor (see Table 2). Ainley called the second factor of her study the "Breadth-of-Interest Curiosity Style." The four subscales of the Sensation Seeking Scale (Zuckerman, 1979) together with Pearson's Novelty Experiencing Scale – External Sensations (NESIS) and Day's Ontario Test of Intrinsic Motivation – Diversive Curiosity Scale (OTIMDC) loaded on this factor. Langevin (1971) noted that the Breadth and Depth Curiosity factors resemble the Specific-Diversive distinction. Later, Giambra, Camp and Grodsky (1992) called the two dimensions *Information Seeking* and *Stimulation Seeking,* and Spielberger and Starr (1994) *Information Seeking* and *Experience seeking*.

Compared to current research, neither the studies of Langevin nor the study of Ainley contained scales which measured Kreitler et al.'s (1974, 1975) curiosity concept. However, the measuring instruments which Kreitler et al. used to measure manipulatory, perceptual, conceptual and complexity exploration were very similar to those used by Berlyne. Thus, it is expected that these subscales will load on the same factor as IM, namely depth-of-interest curiosity style or specific curiosity. On the whole the two-dimensional model of the present investigation (Model 2) suggested that two, almost orthogonal, latent variables account for the variances among the 15 subscales. The two factors were named *Curiosity* and *Sensation Seeking* (cf. Boyle, 1989; Langevin, 1976). In principle, it would have been possible to name the second factor *Experience Seeking* (cf. Olson & Camp, 1984; Spielberger & Starr, 1994), but then this name would have referred to a broad concept which include both sensation seeking and diversive curiosity. However, Zuckerman (1994) used the same term in a more limited sense to refer only to one sub-dimension of sensation seeking. Figure 18 presents the details of the corresponding conceptual model, Model 2.

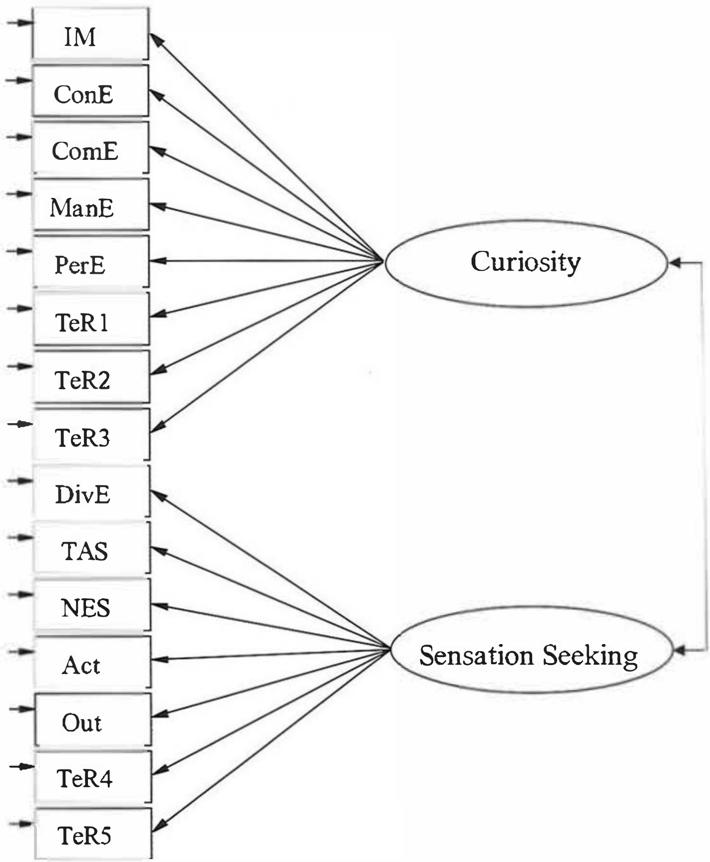



*FIGURE 18.* TWO-DIMENSIONAL MODEL OF CURIOSITY.

## 9.2.3 THREE-DIMENSIONAL MODELS

Both theoretical and empirical findings of previous studies suggested several three-dimensional models. Based on these findings, three alternative models are presented.

*Model 3a.* There is evidence for distinguishing between sensory and cognitive types of curiosity. Thus, if compared to Model 2, the difference between Models 2 and 3a was that the Curiosity factor was expected to split into two parts. This logic is based on the distinction first made by Berlyne (1960, 1963, 1965). According to this distinction, specific curiosity can be divided into Perceptual Curiosity and Epistemic Curiosity. Perceptual Curiosity can take the form of *receptor adjusting*, *locomotory exploration* or *investigatory behavior*. Epistemic curiosity results from *conceptual conflict*. Berlyne (1965) divided epistemic behavior into three categories, namely epistemic *observation*, which includes different kinds of experimental and observational techniques; *consultation*, which includes asking other people questions or consulting books; and *directed thinking*.

Recently, Trudewind and Schneider (1994) have argued that *curiosity motive* can be described with three factors, namely *Epistemic Curiosity*, *Perceptive and Manipulative Curiosity* and *Searching for Stimulating Events*. Accordingly, it was hypothesized that Manipulatory Exploration and Perceptual Exploration subscales should load on the factor that Berlyne called Perceptual Curiosity. Subscales Conceptual Exploration and Complexity Exploration should load on the factor named Epistemic Curiosity together with scale IM. The third factor was hypothesized to be formed from Diversive Exploration and Sensation Seeking scales. This factor is called Sensation Seeking. Figure 19 presents the corresponding conceptual model.

*Model 3b.* Henderson and Moore (1979) and Olson and Camp (1984) identified a factor that the former called "Venturesomeness" or "Venturesomeness-Timidity." Olson and Camp found that two subscales of SSS loaded on this factor, namely Thrill and Adventure Seeking ($\lambda = .63$) and Experience Seeking, which had a secondary loading on this factor ($\lambda = .31$). The Proverbs Test (Maw & Maw, 1975) was the third measurement which loaded highly ($\lambda = .48$) on the Venturesomeness factor in the study of Olson and Camp. Later,

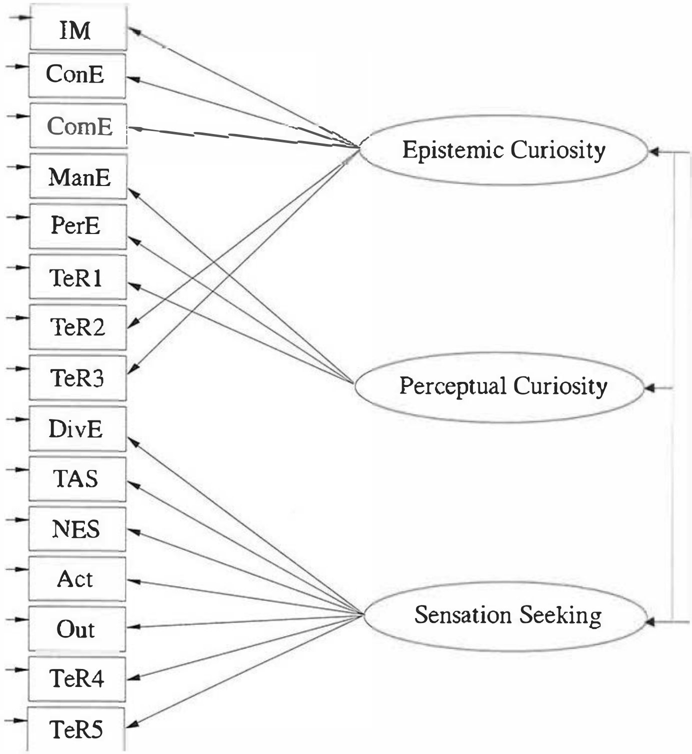



*FIGURE 19.* CONCEPTUAL MODEL 3A.

Byman (1993) reanalyzed the data of Ainley (1987). Byman found a narrow factor which resembled the Venturesomeness factor of Olson and Camp and the Physical Thrill-seeking factor of Kohn, Hunt and Hoffman (1982). In Byman's study, TAS and the subscale of Novelty Experiencing named External Sensation (see Pearson, 1970) loaded on the Venturesomeness factor. Recently, Zuckerman (1994) has noted that Venturesomeness is nearly equivalent to Thrill and Adventure Seeking. Thus, Model 3b of the present study hypothesized that three latent variables, namely Curiosity, Experience Seeking and Venturesomeness (cf. Olson & Camp, 1984), account for the variances among the 15 subscales (see Figure 20).

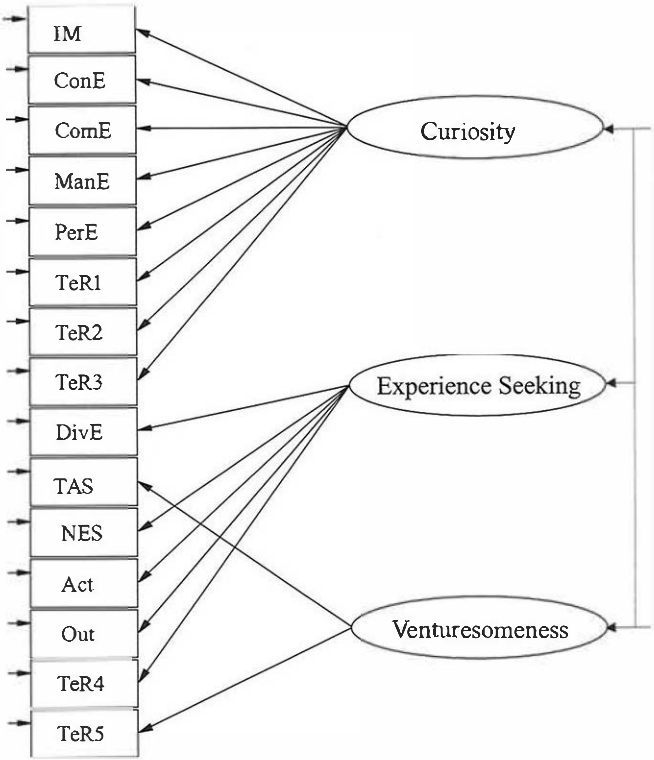



*FIGURE 20.* CONCEPTUAL MODEL 3B.

*Model 3c.* Campbell and Fiske (1959) were the first to argue that construct validation requires both convergent and discriminant validity. Campbell and Fiske also were the first to present a multitrait-multimethod (MTMM) design to test convergent validity, discriminant validity and method effects of measurement operations. In the factor analytical context, the logic of the MTMM approach means that different measures of the same construct should load on the same factor (convergent validity), and different constructs of the model should be independent (discriminant validity). Moreover, comparing models with and without a method factor gives information regarding the presence of method factors.

Bollen and Paxton (1998) have shown that the benefits of the MTMM design are also available under "more relaxed conditions." According to Bollen and Paxton, subjective ratings typically contain systematic errors (method factors). Method variance is a classical example of this kind of nonrandom error variance, which Messick (1993) also included in his concept of construct-irrelevant variance. Several studies (e.g., Coie, 1974; Henderson, 1994; Langevin, 1971; Maw & Maw, 1965, 1975) have shown that teacher ratings of curiosity can be biased by intelligence. That is, teachers describe intelligent pupils as curious. However, Trudewind and Schneider (1994) have noted that this finding is not surprising, because teachers observe their pupils' exploratory behavior in classroom settings where learning and problem-solving situations are typical. Based on this logic, Trudewind and Schneider argued that the shared variance between curiosity and intelligence does not invalidate the assessment of curiosity with teacher ratings. However, taken together, these findings suggest that a method factor, Intelligence Biased Teacher Ratings, is needed to explain some of the covariances among Teacher Rating scale items. Thus, Model 3c was conceptually similar to Model 2, but the third factor was expected to be a method factor. Based on the logic of the MTMM approach, this factor was expected to be orthogonal to the two other factors (see e.g., Marsh & Grayson, 1995). Figure 21 presents the logic of this conceptual model.

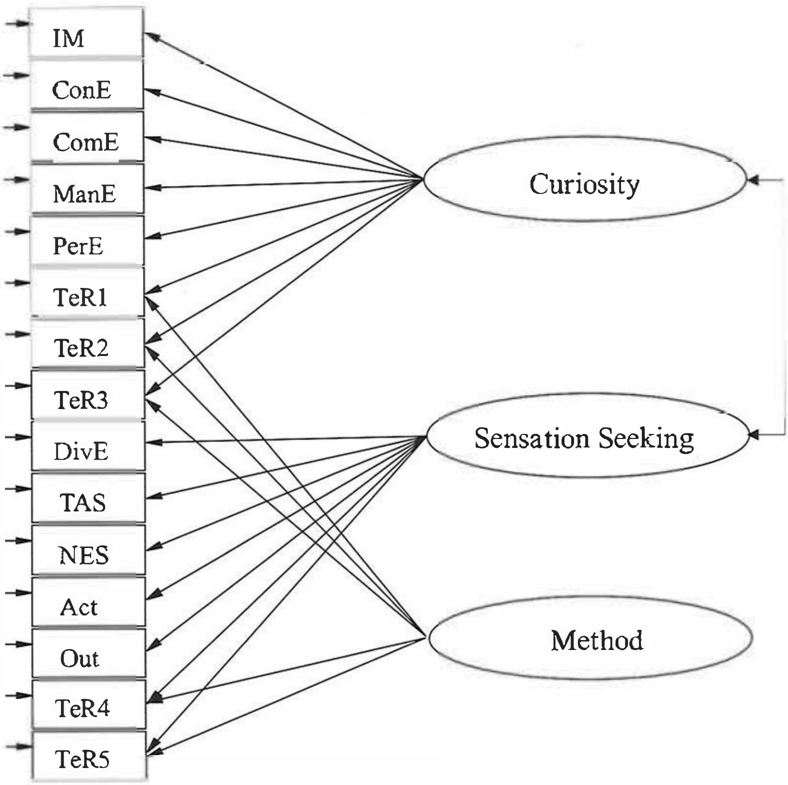



*FIGURE 21.* CONCEPTUAL MODEL 3C.

### 9.2.4 FOUR-DIMENSIONAL MODEL

The results of previous studies suggested four alternative four-dimensional models.

*Model 4a.* This model was otherwise similar to Model 3a, but an extra method factor, Intelligence-Biased Teacher Ratings, was postulated.

*Model 4b.* This model was a modification of Models 3b and 3c. The three trait factors were expected to be Curiosity, Experience Seeking and Venturesomeness. In addition a orthogonal method factor was postulated.

*Model 4c.* This model was a combination of Models 3a and 3b. That is, four factors, namely Perceptual Curiosity, Epistemic Curiosity, Experience Seeking, and Venturesomeness, are expected to reproduce the data.

*Model 4d.* This model was similar to the classical 2 x 2 MTMM design. Two trait factors (Curiosity and Sensation Seeking) and two method factors were expected to account for the covariances among the 15 variables. This model was exploratory in that sense that it was not possible a priori interpret the second method factor which was expected to account for the method effects of self-rating instruments.

Several researchers (e.g., Bollen & Paxton, 1998; Marsh & Grayson, 1995) have reported serious identification problems in models similar to 2 x 2 MTMM. Thus, as an alternative to Model 4d, a two-dimensional model was suggested in which two correlated trait factors were presented (cf. Model 2) but method effects were inferred from correlated uniqueness among method variables based on the same method. Marsh and Grayson (1995) called this a correlated traits/ correlated uniqueness model (CTCU).

## 9.3 PRELIMINARY ANALYSIS

To construct subscale scores, the corresponding items were summed according to the results of item-level analysis. Items 3, 5 and 8 were eliminated from the diversive exploration scale because according to the item-level analyses they had very low reliability and they also loaded negatively on the expected latent variable. Table 12 presents the names of the subscales and the item-level variables which were summed to it.

All the resulting subscale variables were treated as continuous variables. The PRELIS program showed that none of the sum variables showed excessive skewness or kurtosis. However, the Mardia's multivariate tests of skewness and kurtosis showed again that both samples deviated significantly ($p < .001$) from multivariate normality. Thus, to get Satorra-Bentler scaled chi-square statistics, the Maximum Likelihood method (ML) and the asymptotic covariance matrix were used in all analyses.

TABLE 12 *CONSTRUCTION OF THE SUBSCALE-LEVEL DATA*

| Subscale | | |
|---|---|---|
| Summed Items | | |
| IM (15) | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 | |
| C-Trait (20) | | |
| ConE | 1, 7, 10, 11, 19 | |
| ComE | 23, 24, 31, 36, 40 | |
| ManE | 9, 13, 25, 27, 33 | |
| PerE | 17, 18, 22, 37, 28 | |
| DivE (7) | 1, 2, 4, 6, 7, 9, 10 | (OTIMDC items) |
| SESE (55) | | |
| TAS | T1.1, T1.2 | |
| NES | N1.1, N1.2 | |
| Act | A1.1, A1.2 | |
| Out | O1.1, O1.2 | |

## 9.4 THE ONE-FACTOR MODEL

The hypothesis of one General or Global Curiosity factor (Model 1) was tested by congeneric factor analysis. Thus, all 15 variables were allowed to load freely on the General factor. However, as shown in Table 13, Model 1 fitted the data poorly. Thus, the fit indices of Model 1 suggested that a single-factor model was insufficient to account for the variances and covariances among the observed variables.

TABLE 13 *GOODNESS-OF-FIT STATISTICS FOR THE FIRST SUBSCALE-LEVEL MODELS*

| Model | df | S-B$\chi^2$ | CFI | PCFI | IFI | RMSEA |
|---|---|---|---|---|---|---|
| Girls | | | | | | |
| 1 Model 1 | 90 | 1245 | .26 | .22 | .27 | .24 |
| 2 Model 2u | 76 | 895 | .45 | .33 | .46 | .22 |
| 3 Model 3u | 63 | 169 | .92 | .55 | .93 | .09 |
| 4 Model 4u | 51 | 47 | .97 | .47 | .97 | .00 |
| Boys | | | | | | |
| 1 Model 1 | 90 | 1044 | .27 | .23 | .28 | .22 |
| 2 Model 2u | 76 | 194 | .46 | .33 | .47 | .08 |
| 3 Model 3u | 63 | 72 | .88 | .53 | .88 | .03 |
| 4 Model 4u | 51 | 29 | .93 | .45 | .93 | .00 |

## 9.5 TESTS OF THE UNRESTRICTED MODELS

The four-step logic suggested by Mulaik was used in testing the multidimensional subscale-level models. According to this view, testing should start from an unrestricted factor model and then proceed to a more restricted factor models. The first-step model tests whether the a priori postulated number of latent variables can account for the covariances among the subscales (Mulaik & Millsap, 2000). According to Jöreskog (1979a, p. 23), all unrestricted models with the same number of factors will yield the same fit to the data. Thus, it was necessary to construct only one two-, three-, and four-factor unrestricted model in the present study. Each of these alternative models tested whether the corresponding factor number can account for the covariances among the 15 subscales.

First, a two-factor unrestricted model was constructed (Model 2u). According to Ainley (1987), the IM scale was fixed as reference variable to the first factor, which was expected to be Curiosity. The second factor was labeled as Sensation Seeking. Ainley (1987) and Rubenstein (1986) chose the OTIMDC scale (DivE) as a reference variable for this factor. Loadings of these reference variables were fixed to the value of 1 on the corresponding factor and to the value of 0 on the other factor. Model 2u was estimated separately for both gender groups. However, as shown in Table 13, the results of Model 2u revealed that a two-factor model could not explain the covariances among the 15 subscales.

Next, a three-factor unrestricted model was constructed (Model 3u). According to conceptual Models 3a and 3b, the two first factors were expected to be the same as in Model 2u, but the third factor was hypothesized to be a Venturesomeness factor or a Teacher Rating factor (method factor). Thus, variable TeR5 was chosen as a reference variable for the third factor. The same fixing logic as in the Model 2u was used. The factor variance-covariance matrix was set free. As indicated in Table 13, the fit of Model 3u was promising. Satorra-Bentler scaled Chi-Square was non-significant, as was also RMSEA, and its 90 percent confidence interval suggested acceptance of Model 3u. However, CFI and IFI indices were both under .95, indicating the unsatisfactory fit of the model. The reason for this was expected to be the covariances among measurement errors. The girls' results contained 10 and boys' solution 16 MIs over 10.

Finally, a four-factor unrestricted model was constructed (Model 4u). This model was based on the assumption that the first factor would be Epistemic Curiosity, the second factor Perceptual Curiosity, the third factor Sensation Seeking, and the fourth factor Teacher Ratings (method). The corresponding reference variables were IM, PerE, DivE and TeR5. Theoretically, Model 4u was based on the thinking of Day and Berlyne (1971) and Langevin (1971). As Table 13 shows, the statistical fit of Model 4u was very good for both gender groups.

To sum up, the results of the four models tested here indicated that it is not sensible to continue the testing procedure to the second phase with one- and two-factor models. Instead, both three- and four-factor unrestricted models fitted so well that it was sensible to construct and test also the corresponding, more restricted models. In practice, the three-factor model and the four-factor model may be very near each other because the detailed investigation of the results of Model 4u showed that factors one and two correlated highly.

## 9.6 Tests of the Measurement Models

The testing procedure began with three alternative measurement models.

*Model 3a.* According to conceptual Model 3a, three factors, namely Epistemic Curiosity, Perceptual Curiosity and Sensation Seeking were expected to reduce the data in both gender groups. The reference variables were expected to be Conceptual Exploration (ConE) for Epistemic Curiosity factor, Perceptual Exploration (PerE) for the Perceptual Curiosity factor, and Thrill- and Adventure Seeking (TAS) for the Sensation Seeking factor. These variables were fixed to the value of 1 in corresponding factors and to the value of 0 in other two factors. Otherwise the measurement of Model 3a followed the previously presented conceptual model. As shown in Table 14, the statistical fit for Model 3a was poor. Moreover, a detailed investigation of the results shows that the correlation between Epistemic Curiosity factor and Perceptual Curiosity factor was very high in both gender groups ($\phi_{Girls}$ = .94 and $\phi_{Girls}$ = .87). This finding suggested a lack of discriminant validity.

*Model 3b.* According to conceptual Model 3b, the three factors, Curiosity, Experience Seeking, and Venturesomeness, were expected to explain the covariances among the 15 subscales in both gender groups. Following this thinking, a three-factor CFA model was constructed separately for both gender groups. The reference variables were expected to be IM for the Curiosity factor, NES for the Experience Seeking factor, and TAS for the Venturesomeness factor. These variables were fixed to the value of 1 in corresponding factors and to the value of 0 in other two factors. Otherwise the constructed CFA model followed the previously presented conceptual Model 3b. As Table 14 presents, the statistical fit for Model 3b was poor. In addition, a detailed investigation of the results of Model 3b revealed that this model has same kind of problems as Model 3a. Correlation between the factors Experience Seeking and Venturesomeness was very high in both gender groups ($\phi_{Girls} = .95$ and $\phi_{Boys} = .90$).

*Model 3c.* Conceptually Model 3c resembled Model 2a. This model contained two trait factors: Curiosity and Sensation Seeking. In addition to this, one orthogonal method factor, Teacher Ratings, was postulated in Model 3c. To get a scale for the parameters of the model, the diagonal elements of the PHI matrix, that is the factor variances, were fixed to value one. Then all covariances among the trait factors were set free. To get an orthogonal method factor for the model the covariances between the two trait factors and the method factor were fixed to value 0.

As indicated in Table 14, the overall fit statistics of Model 3c were promising. If compared to other three-dimensional models, Models 3a and 3b, which included only trait factors, namely the results of Model 3c indicated that there was a notable improvement in fit when the method factor was included in the model. On the other hand, when Model 3c was compared to the corresponding unrestricted model using the chi-square difference test, the results indicated that the model was too restrictive (S-B$\Delta\chi^2(42) = 382.47$, $p < .01$).

A detailed analysis of the results of Model 3c showed that this model had the qualities of a good construct. First, the factor loadings on the expected factors were all statistically significant ($t > 1.96$), except for variables TeR1, TeR2, and TeR3 on the Curiosity factor in the girls' results. All three reference variables (marker variables), namely IM, DivE and TAS, had a very high loading ($t > 3.0$) on expected factors. Taken together, these findings gave support to the convergent validity of the model (Schumacker & Lomax, 1996). A second finding which supported the construct validity of the model especially its discriminant validity, was the low correlation between the two trait factors ($\phi_{Girls}$ = .24 and $\phi_{Boys}$ = .01). Thus, the two factors were almost orthogonal.

*Model 4a.* Construction of CFA Model 4a was otherwise the same as for Model 3a, but an extra orthogonal method factor was added to the model. As shown in Table 14, the overall fit statistics for this model were also promising. However, the results of Model 4a showed again that the Epistemic Curiosity factor and the Perceptual Curiosity factor had a very high intercorrelation ($\phi_{Girls}$ = .94 and $\phi_{Boys}$ = .88), which suggested a post hoc hypothesis that Epistemic Curiosity and Perceptual Curiosity are perfectly correlated. To test this post hoc hypothesis, a model was constructed where the correlation between the Epistemic Curiosity factor and the Perceptual Curiosity factor was fixed to value 1 (Gorsuch, 1983; Schumacker & Lomax, 1996). The resulting chi-square was compared to the chi-square statistics for Model 4a using a chi-square difference test. The value of chi-square difference test was non-significant for both gender groups, (S-B$\Delta\chi^2(1)_{Girls}$ = 0 and S-B$\Delta\chi^2(1)_{Boys}$ = 1.92). This evidence gave support to the presented post hoc hypothesis and to the rejection of Model 4a.

*Model 4b.* This model otherwise resembled Model 3b, but an extra orthogonal method factor was constructed to explain some of the covariances among the Teacher Rating items. As we can see from Table 14, the overall fit statistics for Model 4b were on the same level as for Models 3c and 4a. However, Model 4b had same kind of problems as Model 4a. The interfactor correlation between Experience Seeking and Venturesomeness was high, 1.12 (Hawood Case), in the girls' model and .93 in the boys' model. To test whether these two factors are perfectly correlated, a model was constructed where the correlation between Experience Seeking and Venturesomeness was constrained to equal 1.0. The chi-square estimate for this model was subtracted from the chi-square estimate for Model 4b. The chi-square difference test estimates were clearly non-significant indicating that the two dimensions are perfectly correlated, S-B$\Delta\chi^2(1)_{Girls}$ = 1.47 and S-B$\Delta\chi^2(1)_{Boys}$ = 0.77. Thus, it seemed that Model 4b empirically reverted to Model 3c.

*Model 4c.* The four-factor model, in which Perceptual Curiosity, Epistemic Curiosity, Experience Seeking, and Venturesomeness were expected to reproduce the data, was not constructed because of the discriminant validity problems of Models 4a and 4b.

*Model 4d.* Neither the 2 x 2 MTMM-model nor the CTCU-model could be calculated because of identification problems.

TABLE 14 *GOODNESS-OF-FIT STATISTICS FOR THE SUBSCALE-LEVEL MODELS*

| Model | df | S-B$\chi^2$ | CFI | PCFI | IFI | AIC | RMSEA |
|---|---|---|---|---|---|---|---|
| Girls | | | | | | | |
| 1 Model 3a | 87 | 939 | .42 | .35 | .43 | 1005 | .21 |
| 2 Model 3b | 87 | 936 | .42 | .35 | .43 | 1002 | .21 |
| 3 Model 3c | 84 | 237 | .88 | .70 | .88 | 309 | .09 |
| 4 Model 4a | 82 | 236 | .88 | .69 | .88 | 312 | .09 |
| 5 Model 4b | 82 | 236 | .88 | .69 | .88 | 312 | .09 |
| 6 Pos Hoc Model | 79 | 143 | .95 | .71 | .95 | 223 | .06 |
| Boys | | | | | | | |
| 1 Model 3a | 87 | 790 | .44 | .36 | .45 | 856 | .19 |
| 2 Model 3b | 87 | 805 | .43 | .36 | .44 | 871 | .19 |
| 3 Model 3c | 84 | 257 | .84 | .67 | .85 | 329 | .10 |
| 4 Model 4a | 82 | 244 | .86 | .67 | .86 | 320 | .10 |
| 5 Model 4b | 82 | 257 | .84 | .66 | .85 | 333 | .10 |
| 6 Pos Hoc Model | 81 | 160 | .93 | .72 | .93 | 238 | .07 |

To sum up, Model 3c was the best-fitting model of the five CFA models tested. The two identified four-factor Models 4a and 4b suffered from a lack of discriminant validity. However, despite being promising, Model 3c was not acceptable. Thus, the next aim of the study was to investigate how Model 3c could be modified to fit the data better.

## 9.7 POST HOC ANALYSIS

What Jöreskog (1993) called a model-generating (MG) situation can be data- or theory-driven. Because no previous studies with the same variables as in the present study existed, the data-driven approach was the only possibility in the present investigation. However, when using the data-driven approach, it is also extremely important that only substantially meaningful modifications are made. Thus, MIs associated with each fixed parameter served only as a guide in the search for a better-fitting model.

The result that the estimate of the chi-square difference test was highly significant when Model 3c was compared to Model 3u indicated that Model 3c contained some restrictions that did not hold. To investigate this possibility, the MIs and EPCs were examined parameter by parameter. This overview indicated that a substantial improvement in model fit would be gained with the additional specification of one cross-loading (ManE on the Sensation Seeking factor) and two correlated measurement errors (between TeR1 and TeR5, TeR4 and TeR5). These findings were consistent across gender. Additionally, the MIs indicated that for girls the fit of the model could be improved substantially by allowing the Act variable to have a negative cross-loading on the Curiosity Factor and by specifying a correlated error between DivE and Act. For boys only, the MIs suggested a correlated error between TeR1 and TeR5.

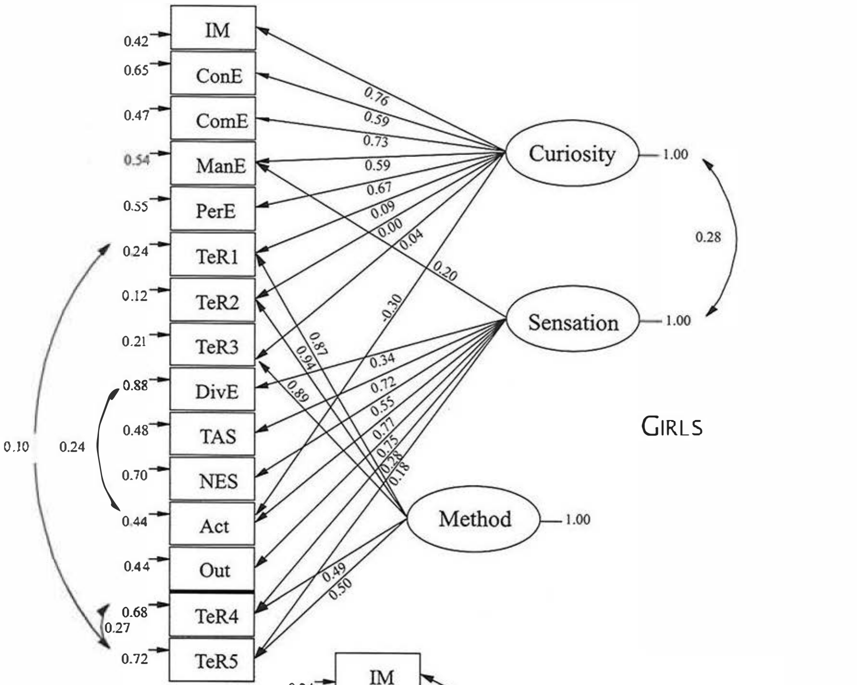
The cross-loading of variable ManE indicated that this subscale, in addition to measuring curiosity, appears to measure sensation seeking. The negative cross-loading of variable Act on the Curiosity factor suggested that for girls Activity is negatively related to curiosity. Because the Act subscale measures a "preference for activity and affiliating with people" (Björck-Åkesson, 1990, p. 103), this seems to be a rational gender effect. The correlated measurement errors among teacher rating scale items may reflect response bias or correlating specific item content. Items TeR4 and TeR5 are successive items which were designed to measure the same factor. For boys only, the pattern of correlated error terms referred also to the possibility of omitting the specific group factor (Items TeR1, TeR4 and TeR5). The wordings of Items TeR1, TeR4 and TeR5 have some similar characteristics. For instance, the word "mysterious" is common to both Items TeR1 and TeR5.

Based on this logic, Model 3c was respecified with these parameters to be freely estimated. However, because in Model 3c the correlation between the two trait factors was very low in both gender groups, an alternative model was also estimated in which the correlation between the two trait factors was fixed to value 0.

As shown in Table 14, the statistical fit for the respecified post hoc model was acceptable. For both sexes, the cross-loadings and correlated errors were substantially and statistically significant ($p < .001$). Although the value of S-B$\chi^2$ was highly significant, the other overall fit statistics supported acceptance of the models. For instance, the CFI values were over .90 for both gender groups, thus suggesting that the models can explain over 90% of the covariation in the data. This result is acceptable for studies operating with behavioral science variables (Byrne, 1994, 1995).

For both gender groups, the chi-square difference test was used to compare the orthogonal model to the model where trait factors were allowed to correlate. In the girls' sample this comparison yielded a chi-square value that was statistically significant (S-B$\Delta\chi^2(1) = 8.76$, $p < .01$), thus indicating that the fit of the orthogonal model was worse than that of the model where the trait factors were allowed to correlate. However, in the boys' sample the result was the opposite. The fit of the orthogonal model was not significantly worse than that of the oblique model, S-B$\Delta\chi^2(1) = 0.98$, $p$ = *n.s.*. Thus, because the orthogonal model is more parsimonious than the oblique model, it was chosen as a final model for the boys' sample.

When the final models were compared to Model 3c using the chi-square difference test, the value of $\Delta\chi^2$ indicated a substantial improvement in model fit for both girls (S-B$\Delta\chi^2(5) = 94$) and boys (S-B$\Delta\chi^2(5) = 103$). On the other hand, when the final model was compared to the corresponding unrestricted Model 3u, the $\Delta\chi^2$-test still supported rejection of the model, thus indicating that the model still has some unjustifiable restrictions, S-B$\Delta\chi^2(16)_{Girls} = 26.31$, $p < .01$ and S-B$\Delta\chi^2(18)_{Boys} = 87.42$, $p < .001$. However, because all additional respecifications suggested by MIs were theoretically questionable, no further respecifications were made. In addition, RMSEA and its 90 percent confidence interval suggested that the respecified Model 3c is a good population model for both gender groups. The final models, including factor loadings in standardized form, are illustrated in Figure 22.



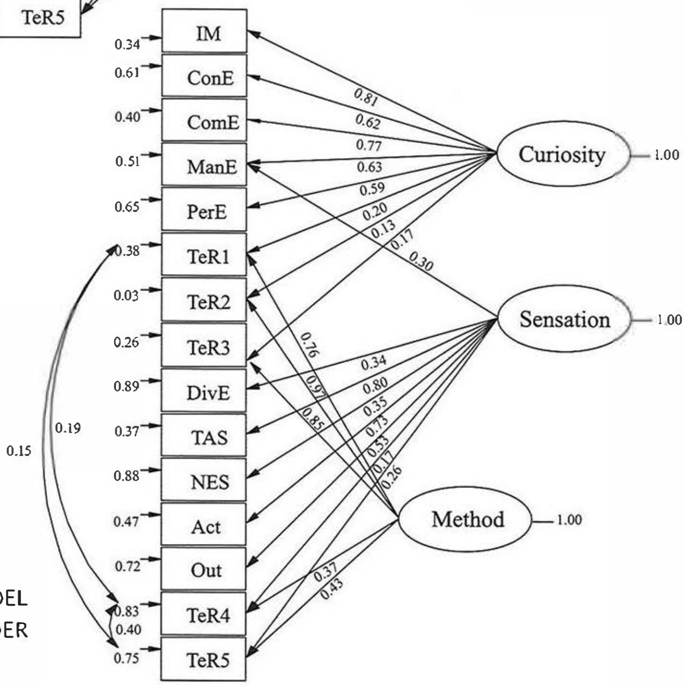




*FIGURE 22.*
POST HOC MODEL FOR BOTH GENDER GROUPS.

When post hoc model modifications are to be made, Tabachnick and Fidell (1996) have proposed that a correlation coefficient should be calculated between the hypothesized model parameter estimates and the parameter estimates from the final model. In the present study these correlations were .998 for the girls' models and .988 for the boys' models. These results indicated that the parameter estimates changed very little during the modification process.

## 9.8 DISCUSSION

The aim of the subscale-level analysis was to find out if some of the nine suggested alternative models of curiosity could account for the covariances among the fifteen subscales. The suggested conceptual models were constructed using the results of previous studies and some theoretical analyses. However, none of these models fitted the data acceptably. One- and two-factor models were insufficient to account for the covariances among the subscales, and two of the three three-factor models as well as all postulated four-factor models suffered from discriminant validity problems. The best-fitting model was a three-factor model which contained two almost orthogonal trait factors and one method factor. The trait factors were called Curiosity and Sensation Seeking. Using post hoc CFA it was possible to modify this model so that it fit the data acceptably.

The respecification phase also confirmed some moderation effects of gender. First, in the girls' group teacher rating variables TeR1-TeR3 did not have a statistically significant loading on the Curiosity factor. This finding means that in the present study the teacher rating items TeR1-TeR3 measured only the method factor which, according to previous studies, may have something to do with intelligence. This finding seems to give support to the previous studies (e.g., Coie, 1974; Henderson, 1994; Langevin, 1971) according to which teachers have difficulties in distinguishing between intelligence and curiosity. On the other hand, when the teachers in the present study estimated the curiosity of boys, the three items worked better, the estimations seemed to be only intelligence-biased. This finding supports the claim of Trudewind and Schneider (1994, p. 156) according to which the shared variance (correlation) between curiosity and intelligence does not invalidate the assessment of exploratory behavior with teacher ratings. However, the present study differed from previous studies (e.g., Coie,1974; Langevin, 1971) in one important respect, namely the teacher rating scale first developed by Maw and Maw (1964) was not handled as a unitary scale. Berlyne (1963) and Day (1968, 1971) have noted that the scale contains descriptions of two different forms of exploratory behavior, namely specific and diversive exploration. If these two exploration modes are not differentiated, it confuses the raters. Thus, in the present study these two modes of exploration were differentiated as four items. In addition to this, one sensation seeking item was added to the rating scale. This item was expected to define general sensation seeking. Both Björck-Åkesson (1990) and Zuckerman (1994) have reported that the General Sensation Seeking scale and some subscales of SSS, especially the Experience Seeking subscale, have had low but statistically significant correlations with intelligence and cognitive ability measures. On the other hand, several researchers (e.g., Berg & Sternberg, 1985; Voss & Keller, 1983; Mayer, Caruso, Zigler & Dreyden, 1989) have implicitly or explicitly related intelligence to curiosity. The relationship between curiosity and intelligence is present even in some terms used to describe curiosity. Lloyd and Barenblatt (1984), for instance, called their curiosity construct Intrinsic Intellectual Motivation. Thus, the method factor of the present study may also have something to do with intelligence. However, this is only a hypothesis which should be confirmed through new research which measures also intelligence, as well.

The final model of the present study is based on the factor analytical studies of Langevin (1971, 1976), Ainley (1985, 1987) and Rubenstein (1986). Compared to these studies, the final model is very near that suggested by Langevin. Including teacher ratings, Langevin used different methods to investigate curiosity-related exploration. Thus, Langevin's results first revealed that teacher ratings of curiosity are intelligence-biased. Both Ainley and Rubenstein used only questionnaires to investigate curiosity. In principle the present study and the study of Ainley had three common scales, namely a Intrinsic Motivation scale (IM), a Diversive Curiosity scale (OTIMDC) and a Sensation Seeking scale (SSS). However, the subscales of SSS were not the same in these two studies. From the standpoint of interpretation these two studies had only one common subscale, namely the Thrill and Adventure Seeking (TAS) subscale. On the other hand, according to Zuckerman (1994, p. 36) the Outgoingness subscale of the present study contains both Experience Seeking (ES) and Disinhibition (Dis) type items from the original SSS. However, according to Zuckerman, the Activity subscale of the present study and the Boredom Susceptibility subscale of the original SSS are distinct factors. Thus, it was possible to use only three common subscales, namely IM, DivE and TAS, as marker variables in the present investigation. In the final model all three of these marker variables had statistically significant loadings on expected factors. Moreover, also subscales NES, Act and Out loaded on the same factor as the marker variable TAS. In this sense, subscales NES, Act and Out worked the same way as subscales ES, Dis and BS did in the study of Ainley. Thus, it seems that the two trait factors of the present study are parallel to Ainley's Depth Curiosity and Breadth Curiosity factors.

Compared to Rubenstein's study, the present study had only one common subscale, the Diversive Curiosity scale. However, the Specific Curiosity scale of the OTIM, which was one of the subscales used by Runbenstein, correlates highly with the Intrinsic Motivation scale. In Beswick's (1971) study, the correlation between these two scales was .70 and in Ainley's (1985) study .60. Thus, it is reasonable to argue that the Specific Curiosity and Intrinsic Motivation scales should load on the same factor. If this logic is accepted, the two trait factors of the present study are very similar to the two Curiosity factors in Rubenstein's study. As in the present study, the two Curiosity factors in Rubenstein's study were almost orthogonal. Thus, a two-factor model with two almost orthogonal factors seems to be a valid classification of questionnaire-type curiosity scales. However, the question of what term should be used to describe these two trait factors is more difficult. In the present study, the two dimensions were named a priori Curiosity and Sensation Seeking. However, according to previous studies (e.g., Ainley, 1987; Giambra, Camp & Grodsky, 1992; Olson & Camp, 1984; Spielberger & Starr, 1994), it seemed possible that the Sensation Seeking factor could also have been called Breadth of Interest Curiosity, Stimulation Seeking, Novelty Seeking or Experience Seeking. However, the empirical findings of the present study support the hypothesis that the second dimension should be named General Sensation Seeking. All subscales of the Sensation Seeking scale had a statistically highly significant loading ($t > 3.58$) on this expected factor. In a completely standardized solution the loadings ranged from .35 to .80 on the Sensation Seeking factor. On the other hand, the OTIMDC subscale had a loading of only .34 on this factor. Thus, OTIMDC did not figure sufficiently prominently ($SMC_{Girls}$ = .12 and $SMC_{Boys}$ = .11) in the second factor for us to conclude that the factor content is something else than General Sensation Seeking. Moreover, because Experience Seeking is also a subdimension of Sensation Seeking, using this term in a broader meaning seems to be confusing. On the other hand, it seems possible to call the Sensation Seeking factor of the present study also Novelty Seeking, inasmuch as Zuckerman (1994) has shown that Novelty Seeking is nearly equivalent to Sensation Seeking.

[1]Compared to Zuckerman (1994), the term Experience Seeking is used here in a broader meaning.

# 10. General Discussion

This final chapter contains a discussion of the ideas raised in the theoretical part and the results obtained in the empirical part. The main focus will be on conceptual and psychometric issues.

## 10.1 The Nature of the Two Trait Factors

At the beginning of the present work Wittgenstein's (1980, 1981) concepts of language game and family resemblance were stated as essential for the present investigation. The ways different researchers use the term curiosity, both in their theories and operationalizations, give meaning to the concept. Applying this idea, I first studied curiosity at the theoretical level from the viewpoint of different psychological theories. Guided by theoretical definitions, researchers have designed different ways to measure curiosity. The most common way to measure curiosity as a trait has been to operationalize it as a response to questionnaire items. Thus, the second aim of my study was to find out how four very popular curiosity questionnaires overlap and differ.

The present study has some meta-analytical characteristics in that it attempted to synthesize the plethora of existing definitions and measurements. My goal was to take a major step toward clarifying the role of curiosity in individual's life and especially in schoolwork. The method of the present study differed from those of previous studies: deductive methodology was used instead of inductive methodology. Although it was possible according to previous studies to construct nine different conceptual models of curiosity, only one model received statistical support in the present study. Findings from the present study appear to support the contention that curiosity and sensation seeking incline a person to exploration. Berlyne (1971b) referred to these two traits as "needs of the brain."

The subscales which loaded on the Curiosity factor seemed to share a tendency to active information seeking or acquisition of knowledge in common. McReynolds (1971) referred to this kind of motivation as "cognitive motivation" and Maslow (1970) called it the "need to know." Roger (1969) must have meant this when he argued that persons should free up their curiosity and open everything to questioning and exploration. However, what Hunt (1963) called "motivation inherent in information processing and action" seems to be more narrow concept. On the other hand, Hunt's definition is based on the distinction first made by Berlyne (1960), according to which curiosity should be split into perceptual and epistemic curiosity. What Hunt meant by motivation seems to be epistemic curiosity. Surprisingly, in the present study the distinction between epistemic and perceptual curiosity did not receive empirical support. Of course this result does not mean that this distinction is theoretically useless.

Based on the interpretation of the Curiosity factor, I suggest that curious persons manifest curiosity as a state more easily and more strongly than others, they remain curious longer than others and they are more willing to explore or acquire new information than others. Curious persons are also more sensitive to novelty, complexity and ambiguity than normal people. They react to these kinds of stimuli by exploring. This exploration can take the form of observation, manipulation, thinking, or consultation. Typical of curious people is that they try to restore and enhance the clarity of their cognitive system. In Piaget's (1977) words, curious persons try to restore and increase their cognitive equilibrium. Thus, curious persons also actively seek and choose situations and objects which are novel, complex or ambiguous.

The second trait factor of the present study was called Sensation Seeking. If compared to the first trait factor, the main difference may be seen on the inspective-affective continuum first suggested by Wohlwill (1981). Essential to the second trait factor seem to be sensations or experiences, not information. Wohlwill (1981) called this kind of exploration "affective exploration." The aim of this kind of exploration is to maintain an optimal level of hedonic activity. Where persons who are curious choose situations which are complex, sensation seekers choose situations where there is possible to experience, for instance, feelings of excitement. The aim of a curious person is to seek information and the aim of a sensation seeker is to feel various kinds of sensations. Zuckerman (1994, p. 95) formulated this difference nicely by arguing that "sensation seeking seems to be limited to sensations and does not include the need for cognitive experience to the same degree. Immediate sensations, whether from external or internal sources, are generally more arousing than cognitions." Also Björck-Åkesson (1990, p. 193) implicitly referred to this difference when she emphasized that a "preference versus nonpreference for arousing experiences" is essential to her conception of sensation seeking. Moreover, Berlyne (1971b) made the difference explicit by arguing that sensitiveness to problems and a tendency to notice "anomalies" affect the probability of specific exploration. On the other hand, the probability of diversive exploration depends on such things as how adventurous or open to new experiences the person is and how ready he or she is to take a risks which accompany such an exploration. There is also some evidence that the two trait factors of the present study have a different neuropsychological basis (see Zuckerman, 1994).

But why do curious people prefer situations that may arouse an aversive state of uncertainty? This is maybe because of the fact that although curiosity as a state is aversive, the process of satisfying curiosity is pleasurable. During that process a person can feel such positive feelings as competence and mastery (see, e.g., White, 1959; Loewenstein, 1994). However, it must be remembered that what White meant by competence motivation covers both what I have described here as curiosity and sensation seeking (cf. Loewenstein 1994, p. 84). The competence perspective highlights the fact that both traits involve one's self-concept.

What Zuckerman (1994) stated about sensation seekers seems be applicable to the interpretation of the second trait factor of this study. Thus, I have proposed that high sensation seekers are persons who seek varied, novel, complex and intense sensations and experiences. They are also ready to take physical and other kinds of risks for the sake of such experiences. Björck-Åkesson (1990) defined risk as the chance of injury, damage or loss. The risk can also be social or psychic. However, as Zuckerman (1994) noted, the risk is not an end in itself. Actually, sensation seekers try to minimize the risk involved in the activity. On the other hand, they do not consider what they are doing as very risky, because they trust their ability and equipment. Although risk is the central element of the Thrill and Adventure Seeking subscale of the general sensation seeking factor, sensation seeking as trait has also other elements. Actually sensation seekers enjoy many kinds of experiences, for instance in music, art, and food, that are not at all risky. Drawing on Björck-Åkesson (1990), I suggest that sociability, playfulness and avoidance of constancy are other elements of the sensation seeking trait in children. By constancy Björck-Åkesson meant "unchangeableness." Sociability refers to the fact that sensation seekers get most of their everyday stimulation in social settings. Several authors (e.g., Nunally, 1981; Zuckerman, 1994) have also related playful behavior and activity directed toward amusement to sensation seeking. In Björck-Åkesson's sensation seeking scale for children, all items which included the word "fun" referred to this aspect of sensation seeking.

Björck-Åkesson (1990) has suggested that in school, sensation seekers best learn in situations where they can actively take part in the creation of the learning situation. Björck-Åkesson emphasizes that in such situations there "may be a match between sensation seeking tendency, preferred learning style and cognitive ability" (p. 188). On the other hand, in learning situations which are highly structured and teacher-centered, high sensation seekers are easily distracted from their schoolwork by other activities. Ryan and Deci (2000, p. 59) have also proposed that autonomy-supportive teachers, as opposed to restrictive teachers, promote their students intrinsic motivation and curiosity. Several studies (see e.g., Deci & Ryan, 1985) have shown that learning that is based on this kind of self-determined motivation is of high quality. It, for instance, results in greater conceptual understanding, and children also really learn to think. It must also be remembered, as Deci (1992) observes, that the same factors that decrease curiosity also dampen creativity.

## 10.2 THE TWO TRAIT FACTORS AS PARTS OF PERSONALITY

Broad consensus exists among personality psychologists that all trait dimensions can be reduced at the broadest level of hierarchy to approximately five basic bipolar categories: (1) extraversion-introversion, (2) neuroticism, (3) openness to experience, (4) agreeableness-antagonism, and (5) conscientiouness-undirectedness. According to McCrae and Costa (1997), these "Big Five" dimensions gather together a broad constellation of traits with cognitive, affective, and behavioral manifestations. It must be asked how and where the two trait factors of the present study can be placed on this taxonomy? The answer to this question is not a truism. McCrae and Costa suggest that both curiosity and sensation seeking are aspects of openness to experience, a category that other supporters of the Big Five model have also called *inquiring intellect*, *culture, intelligence culture, intellectance* and *intellect* (see John, 1989). However, the empirical findings of the present study do not support this claim. The final three-factor model of the present study contained two trait factors which were orthogonal in boys' sample and almost orthogonal in girls' sample. Because there is very little common variance between these two trait factors, it is not possible that the same higher-order construct lies behind both factors. Zuckerman (1994) has reported results that support this conclusion. According to Zuckerman, only the experience-seeking (ES) part of sensation seeking is related to the Openness to experience construct, while the other parts of sensation seeking, particularly Dis, are negatively related to conscientiousness and agreeableness in the big five taxonomy.

According to Zuckerman (1994, pp. 97-98), "there is little in common between sensation seeking and neuroticism or extraversion dimensions of the 'big five'." Nevertheless, in the present study the children's version of the sensation seeking scale was used. Except for the interpretation of the Thrill and Adventure Seeking dimension, the interpretation of the subdimensions of this scale were different from those of Zuckerman's subfactors. The adult version of the sensation seeking scale does not include the subscales Activity (Act) and Outgoingness (Out). According to the interpretation of Act and Out subfactors, it seems rationale to argue that the subscales Activity and Outgoingness are more related to the Extroversion dimension than the original subscales of the SSS. Moreover, Zuckerman's studies have shown that Venturesomeness and Thrill and Adventure seeking are almost equivalent constructs. According to Watson and Clarke (1997), Venturesomeness is one of the component traits of Extroversion. On the other hand, Zuckerman (1994) has recently argued that sensation seeking is part of a broader personality trait called *Impulsive Unsocialized Sensation Seking* (ImpUSS).

If the interpretation of the curiosity factor of the present study is compared to the conception of curiosity presented by McCrae and Costa (1997), it seems plausible that this kind of curiosity is an important aspect of the Openness to experience trait. According McCrae and Costa, "open people are not the passive recipients of a barrage of experiences they are unable to screen out; they actively seek out new and varied experiences. Openness involves motivation, needs for variety, cognition, sentience, and understanding. This activity pursuit of experience can be seen in all the facets of Openness (p. 839)." Thus, what McCrae and Costa mean by experience seems to have a cognitive connotation. This is also the central feature of the curiosity factor of the present study. Beswick (1974, p. 23) formulated this feature by arguing that the more highly curious persons "tend to prefer the more intellectually challenging situations which offer more opportunities for originality, which build upon already established skills, and which provide maximum variety of experience."

## 10.3 CURIOSITY, SENSATION SEEKING AND INTEREST

One of the starting points of the present study was the claim made by Iran-Nejad, McKeachie, and Berliner (1990) that interest and curiosity are essential concepts of a unified theory of learning. Thus, one may ask what the relationship is between interest and the two trait factors of the present study. Krapp's (1994) theory seems to give the best framework for this discussion. According to Krapp, "an individual interest is a unique relationship between a person and an object, or object domain, found in that person's environment. This relationship must be of some duration, and does not refer to one-time, unrepeated forms of engagement" (p. 84). Krapp proposes that interest is the concept which can explain the direction or "content dimension" of both diversive and specific exploration. Krapp emphasizes that diversive curiosity and exploration are not directed randomly to whatever object or action is available but that individual interests direct (play a decisive role in) the process. Zuckerman's (1984) study supports this claim. According to the results of this study, the correlations between past and desired sensation seeking experiences is high (see also Day, 1971). This indicates that people tend to engage in activities that are familiar to them and that they have engaged in before (individual interest). Moreover, Csikszentmihalyi (1997) has argued that curiosity is at first diffuse and generic but later becomes specific and focuses on a specific domain of interest. In addition, White (1959) has argued that the competence motivation of children is first undifferentiated but later may lead to life-long, specific, exploratory interest.

Krapp (1994) noted that "general exploratory tendencies" have a direct or indirect influence on the development of interest. Day and Maynes (1972) also specified that curiosity can contribute to the development of interest, especially to the development of scientific and mechanical interest. Based on my results, I suggest that the exploratory tendencies which Krapp referred to are two traits, curiosity and sensation seeking. Curious persons seek situations where they can meet cognitive challenges like complexity and novelty, whereas sensation seekers look for situations where they can feel sensations. Whether an enduring individual interest is developed depends on experiences of success or failure in these kinds of situations.

According to Hidi (1990), individual interest develops slowly and tends to have long-lasting effects on a person's knowledge and values. Schiefele (1991) drew a conceptual distinction between a *latent* (disposition) and an *actualized* individual interest. A latent individual interest is a relatively enduring preference for certain topics, subject areas, tasks, contexts, or activities. Moreover, Schiefele suggested that interest is a content-specific concept as well as a directive force, and that it consists of two kinds of valences: feeling-related and value-related valences. Feeling-related valences are feelings that are associated with a topic or an object, for instance feelings of enjoyment and involvement. Value-related valences refer to the attribution of personal significance to an object. Thus, some objects of interest are preferred because involvement with them creates, for instance, strong feelings of excitement, whereas other objects of interest are preferred because they may have high personal relevance. According to his valence distinction, Schiefele (1992, p. 154) reinterpreted interest "as a domain-specific or topic-specific motivational characteristic of personality, which is composed of intrinsic feeling-related and value-related valences." Schiefele also emphasized that interest has an intrinsic character which he termed "self-intentionality."

One possible way to distinguish curiosity from interest and intrinsic motivation seems to be to see whether they are externally or internally controlled. This view also reflects the contradiction which is typical of the difference between behaviorism and cognitive psychology. Interest and intrinsic motivation are mostly seen as self-determined and self-regulated (see, e.g., Deci, 1992; Krapp, 1994), whereas curiosity is typically seen as intensive and compelling, thus possessing the basic characteristics of a primary drive (see Loewenstein, 1994; von Wright, 1986). However, as has been already noted, voluntary curiosity also exists, depending on the conceptualization of curiosity.

## 10.4 PSYCHOMETRIC AND METHODOLOGICAL ISSUES

Validity is nowadays considered to be a unitary concept (Cronbach, 1990; Messick, 1993). Specifically, all validation is construct validation. On the other hand, as Cronbach (1990, p. 159) has stated, every test is to some degree impure, which means that there are always unavoidable sources of variance in all test scores. Thus, as Maruyama (1998, p. 81) noted, one part of the construct validation process is to identify impurities in the measurement "so that the actual effects of theoretical variables can be clearly observed." Impurity can mean that the test scores are corrupted by systematic sources of error, such as method variance or measure-specific variance. It is even possible that the test measures an extra theoretical variable. In the present study, the results of the item-level analyses showed that test scores always include impurities. However, the analysis also showed that CFA methodology provides tools to make the extra systematic variance explicit. One objective of the present study was to answer the methodological question of what to do when an unrestricted factor model fails. Figure 23 sums up the strategy suggested by the results of this study.

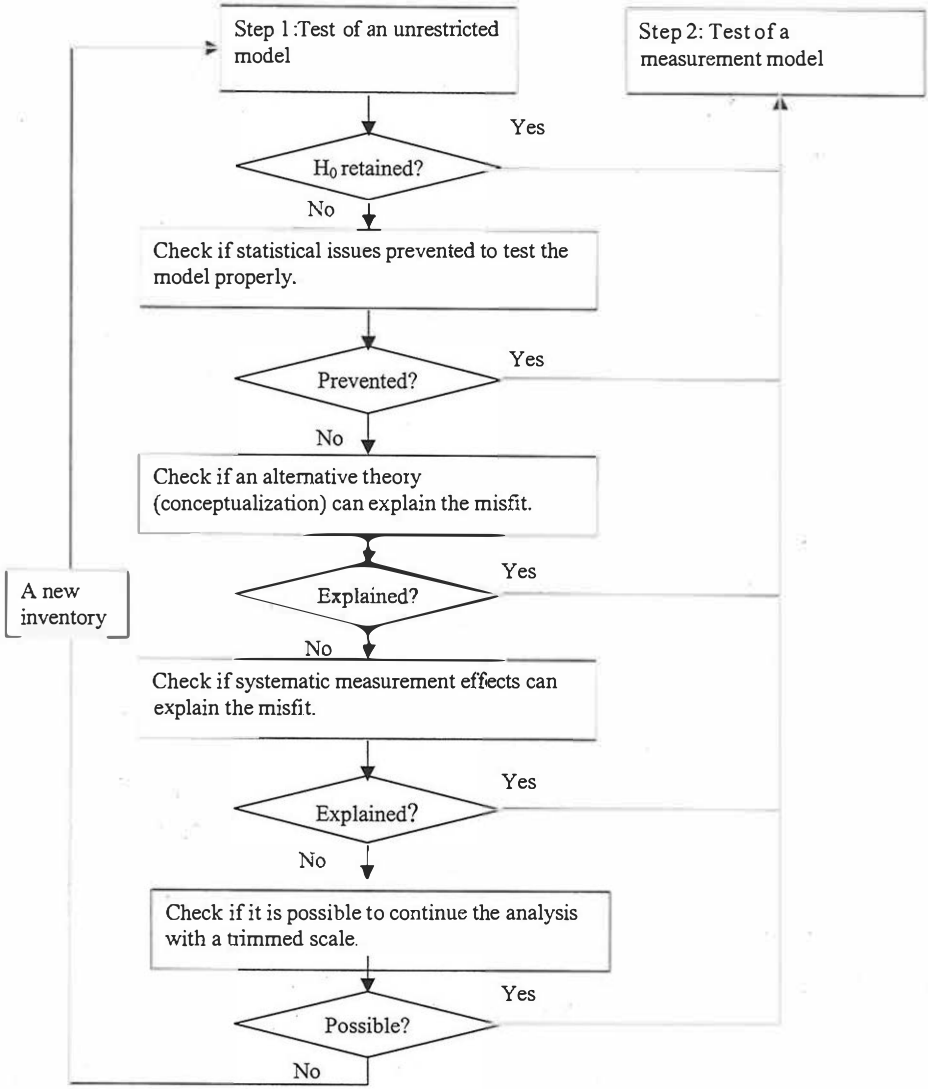



*FIGURE 23.* STEPS TO TAKE WHEN A FIRST-STEP UNRESTRICTED MODEL FAILS.

What have been called "statistical issues" in the present investigation constitute a group of possible reasons which can explain the unfavorable results of an analysis. Sample size, for instance, has always been a major concern in the application of SEM because it has been shown that goodness-of-fit indices are affected by sample size (see e.g., Bollen, 1989). Moreover, violation of the multivariate non-normality assumption and sample size may also have some interaction effects. Thus, several authors (e.g., Hoyle & Panter, 1995; West et. al, 1995) have suggested such goodness-of-fit indices as CFI and IFI, which have only a small downward bias (3% to 4%), even under severely non-normal conditions. However, the relations between all the factors that affect power in first-step model testing are complex and need further clarifying.

SCALED $\chi^2$, WLS (ADF) estimation method, and bootstrapping have been demonstrated to be remedies for non-normality (West et al., 1995). The problem in the LISREL context is that both the estimation of SCALED $\chi^2$ and use of the WLS method require that the asymptotic covariance matrix is first offered to the program, and if the number of variables is large, the estimation of the asymptotic covariance matrix requires large samples. Moreover, when the sample size is small or moderate, the WLS procedure produces statistics that are too high (Hu et al., 1992). Thus, reliable estimation of robust test statistics is not possible because of the sample size restriction.

If statistical issues cannot explain the misfit of a model, the model needs modification. According to Jöreskog (1993), this model generation process can be either data- or theory-driven. Purely data-driven model generation can mean that one uses common factor analysis or additional unrestricted models to find out the "true number of factors." However, data-driven model generation has all the risks of an inductive analysis technique (see Mulaik, 1987). Several studies (see e.g., MacCallum, 1995; MacCallum, Roznowski & Necowitz, 1992) have showed that because of risk of capitalization on chance factors, a purely inductive approach is unlikely to induce the correct population model. The price of this inductive strategy is also the loss of degrees of freedom resulting from seeing the data. Thus, the reported significance levels for the final model are off to an unknown degree (Bollen, 1989, p. 61; Hayduk, 1996). This data-driven process corresponds to what Cronbach (1990) called the weak approach to construct validation.

A strong approach to construct validation tries to resolve crucial uncertainties. The heart of the strategy is the recognition and testing of plausible rival hypotheses. According to Cronbach (1990, p. 184), "one tries to think of challenges to the interpretation that might be made by the advocate of another theory or of a competing test." In the present study this deductive strategy was chosen when the possibility of an alternative conceptualization was investigated. The strategy included a test of alternative hypotheses which were induced from a previously presented conceptualization of the exploration concept in question.

Following the thinking of Cronbach and Meehl (1955), the third possible reason for invalid construct validation might be a problem in measurement. This seems to mean that the test scores are corrupted by systematic sources of error. A popular way to test the possibility of correlated errors has been to look at the MIs and free up from the corresponding error terms the model. However, use of an unrestricted model soon leads to serious identification problems. Thus, the use of this strategy requires a restricted factor model. If this strategy is to be used, one must first fix some additional loadings, for instance the next best reference items, and then free up the substantively meaningful error terms. An alternative approach to test the possibility of systematic sources of error variance is to postulate factors that are limited to one to three items. Both strategies were used in the present study.

Use of questionnaires based on language and require a common understanding of language. However, misunderstanding of the meaning of questionnaire terms occurs more often than is commonly realized (see e.g., Groves, 1989; Goldberg & Kilkowski, 1985; Johnson, 1997). An obvious way to eliminate this kind of error is to choose words with unambiguous meaning. However, as Groves (1989) noted, this goal is almost impossible to reach. Wittgenstein (1980, 1981) emphasized that language is not a tightly regulated activity. According Wittgenstein, as peoples' language and knowledge are based on private experience, people can attach quite different meanings to the same sets of words. Johnson (1997) referred to this in arguing that when personality is assessed with questionnaires, both pragmatic and semantic misunderstandings are encountered. Thus, different subcultures of the same population may use different pragmatic rules when they interpret questionnaire items.

The analyses of the item-level data showed that method effects are associated with item wording, especially for both negatively and similarly worded items. Findings of the present study showed that writing very similar indicators of a single concept guarantees high internal consistence reliability but also leads to several correlated disturbances and thus may not be the best approach (Clark & Watson, 1995). A better way is to write items which tap different aspects of the concept and correlate only moderately with each other. Thus, as Floyd and Widaman (1995, p. 293) have noted, item-content overlap increases the potential for correlated disturbances and may make confirmatory factor analysis more difficult. On the other hand, as Maruyama (1998, p. 89) noted, the use of a shared method variance is not necessarily undesirable, as method variance may be a necessary byproduct of a researcher's efforts to tap substantive dimensions of interest. However, what is important in this kind of situation is to make method effects, such as wording effects, explicit so that they do not confound the substantive factors. As a whole, the results of the present study also highlight the need to study possible method effects of all personality inventories explictely. Moreover, methodologically, the present study also provides a useful application that demonstrates the flexibility of CFA for purpose of construct validation.

## 10.5 LIMITATIONS OF THE PRESENT STUDY

The present study has certain limitations. First, the nature of the factors depends chiefly on the variables included in the analysis. Thus, although a broad variety of curiosity inventories were included in this study, certain conceptual aspects are lacking at the empirical level. Recently, Giambra, Grodsky, and Camp (1992) have broadened the conceptualization of curiosity by suggesting that the concept should also include *interpersonal curiosity*. Moreover, Moch (1987) has also proposed that what he called *socioemotional curiosity* should be included in the scientific concept of curiosity. Second, there is no guarantee that the results obtained with the Finnish version of the scales used can be generalized to those obtained with other languages. Finnish versions of the scales may not measure the same constructs as English versions. In addition, some of the items used were translated too clumsily and mechanically from English to Finnish. Thus, the wording of the Finnish versions of the scales used should be improved. Third, most of the inventories used here were self-report instruments. This kind of operationalization may tap only certain parts of the curiosity concept. Fourth, the data of the present study had a hierarchical multilevel structure. The sample was made up from 529 pupils from 24 classes, 14 schools and four towns. Thus, different models might have resulted if multilevel analysis techniques had been used in the present study (see Heck & Thomas, 2000; Hox, 1995; Goldstein, 1995). Nowadays it is possible to use such programs as STREAMS (Gustaffson & Stahl, 1999) and Mplus (Muthén & Muthén, 1998) as a preprocessor for standard SEM software.

## 10.6 FURTHER RESEARCH ON CURIOSITY

The proposed factor interpretations of the present study are only tentative. Factorial validity is only one part of construct validation, and its use cannot guarantee that the suggested interpretations are correct. What is needed to reinforce the interpretations presented here is a broader model where the factors are connected to other variables. Cronbach and Meehl (1955) referred to this task as the creation of a *nomological network*. Thus, the next step in the present research project is the development of a broader theoretical model where the three factors of the final model are connected to such concepts as state-like epistemic curiosity and success in school. The aim is also to empirically test the postulated theoretical model with SEM. However, the debate surrounding the placement of sensation seeking and curiosity in the personality sphere should continue.

## 10.7 CONCLUSIONS

Using a statistical method that has not previously been used in connection with curiosity measures, we have come a step further in understanding what curiosity scales measure by obtaining a more diversified knowledge of the dimensionality problem of curiosity. By clarifying the conceptual content of various measures used, the results obtained here should help investigators to choose measures appropriate to their conceptual interests. The findings of the present study help to clarify what is measured in the general psychometric domain of curiosity. These findings appear to support the conclusion that curiosity and sensation seeking are completely different traits. However, the term intrinsic motivation seems better suited to describe the motivation (or motivational aspects) inherent in both traits. The next step is to develop a full model of the structural relationships between curiosity and other concepts. The results may also help to link together such important concepts as curiosity, interest and intrinsic motivation.

The item-level analyses of the present study showed that the verbal content of personality inventories is not understood in a consistent way in peoples' minds: Each answerer makes his or her own interpretations. As Wittgenstein observed, using language is not a tightly regulated activity, and people play different language games. Thus, both common and unique meaning components exist. This is a fact that researchers should account for better when they construct inventories for personality research.

A common solution in data-driven analyses has been a split-sample approach and cross-validation (Cudeck & Browne, 1983). While nothing can substitute for cross-validation with a completely new sample, the price of cross-validation is a reduced sample size. Moreover, what one really gets from cross-validation with two randomly split halves seems to be information about the sampling stability (Browne & Cudeck, 1989; Murphy, 1983). The same information, however, is also available with a single undivided sample using sample theory. In the present study, Steiger's (1990, 1995) RMSEA index and the Population Gamma Index were used to estimate how well the final three-factor model fit the statistical population.

According to these indices and their confidence intervals, the final model is a good population model. However, because the final model was a result of a partly data-driven model-generation phase, it still is tentative and needs verification with a new independent sample.

## REFERENCES

Ainley, M.D. (1985). *Exploration in curiosity: Breadth and depth of interest curiosity styles.* Unpublished doctoral dissertation, University of Melbourne, Australia.

Ainley, M.D. (1987). The factor structure of curiosity measures: Breadth and depth of interest curiosity styles. *Australian Journal of Psychology, 39,* 53-59.

Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of strange situation.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Aish, A.-M., & Jöreskog, K.L. (1990). A panel model for political efficacy and responsiveness: an application of LISREL 7 with weighted least squares. *Quality & Guantity, 24,* 405-426.

Anastasi, A. (1988). *Psychological testing.* New York: Macmillan Publishing Company.

Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411-423.

Aronoff, J. (1962). Freud's conception of the origin of curiosity. *Journal of Psychology, 54,* 39-45.

Bandalos, D.L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behavioral research, 28,* 351-374.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Benson, J., & Bandalos, D. L. (1992). Second-order confirmatory factor analysis of the reactions to tests scale with cross-validation. *Multivariate Behavioral Research, 27,* 459-487.

Bentler, P.M. (1992). On the fit of models to covariances and methodology to the *Bulletin. Psychological Bulletin, 112,* 400-404.

Bentler, P.M., & Chou, C.P. (1987). Practical issues in structural model via parsimony: A rationale based on precision. *Sociological Methods & Research, 16,* 78-117.

Bentler, P.M., & Wu, E.J.C. (1993). *EQS/Windows user's guide.* Los Angeles: BMDP Statistical Software.

Berlyne, D. E.(1957). Conflict and information-theory variables as determinants of human perceptual curiosity. *Journal of Experimental Psychology, 53*, 399-404.

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

Berlyne, D.E. (1960). *Conflict, arousal, and curiosity.* New York: McGraw - Hill.

Berlyne, D.E. (1963). Motivational problems raised by exploratory and epistemic behavior. In S. Koch (Ed.), *Psychology: A study of science (Vol.5)* (pp. 284-364). New York: McGraw-Hill.

Berlyne, D.E. (1965). Curiosity and education. In J.D. Krumboltz (Ed.), *Learning and the educational process* (pp. 67-89). Chicago: Rand McNally.

Berlyne, D.E. (1971a). *Aesthetics and psychobiology.* New York: Appleton-Century-Crofts.

Berlyne, D.E. (1971b). What next? Concluding summary. In H.I. Day, D.E. Berlyne, & D.E. Hunt (Eds.), *Intrinsic motivation: A new direction in education* (pp. 186-196). Toronto: Holt, Rinehart & Winston.

Berlyne, D.E. (1975). Behaviorism? Cognitive theory? Humanistic psychology? To Hull with them all. *Canadian Psychology Review, 16*, 69-80.

Berlyne, D.E. (1978). Curiosity and learning. *Motivation and Emotion, 2*, 97-175.

Berlyne, D.E., & Day, H.I. (1971). Intrinsic motivation. In G.S. Lesser (Ed.), *Psychology and educational practice* (pp. 294-335). Glenview, Illinois: Scott, Foresman and Company.

Beswick, D.G. (1971). Cognitive process theory of individual differences in curiosity. In H.I. Day, D.E. Berlyne, & D.E. Hunt (Eds.), *Intrinsic motivation: A new direction in education* (pp. 156-170). Toronto, Montreal: Holt, Rinehart and Winston of Canada.

Beswick, D.G. (1974). Intrinsic motivation in senior secondary students. *Education Research Perspectives, 1*, 15-25.

Björck-Åkesson, E. (1982). *Development of a sensation-seeking scale for preadolescent children.* Reports form the Department of Education, University of Göteborg, 1982:4.

Björck-Åkesson, E. (1990). *Measuring sensation seeking*. Göteborg Studies in Educational Sciences 75. Göteborg: Acta Universitatis Gothoburgensis.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons, Inc.

Bollen, K.A., & Paxton, P. (1998). Detection and determinants of bias in subjective measures. *American Sociological Review, 63*, 465-478.

Bollen, K.A., & Hoyle, R.H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces, 69*, 479-504.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York: Basic Books.

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation*. The Hogarth Press.

Boykin, A.W., & Harackiewicz, J. (1981). Epistemic curiosity and incidental recognition in relation to degree of uncertainty: Some general trends and intersubject differences. *British Journal of Psychology, 72*, 65-72.

Boyle, G.J. (1983). Critical review of state-trait curiosity test development. *Motivation and Emotion, 7*, 377-397.

Boyle, G.J. (1989). Breadth-Depth or State-Trait curiosity? A factor analysis of state-trait curiosity and state anxiety scales. *Personality and Individual Differences, 10*, 175-183.

Boyle, G.J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences, 12*, 291-294.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Beverly Hills: Sage Publications, Inc.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park: Sage Publications, Inc.

Bruner, J.S. (1966). *Toward a theory of instruction*. Cambridge, Mass.: Belknap Press of Harvard University.

Burisch, M. (1984). Approaches to personality inventory construction. A comparison of merits. *American Psychologist, 39*, 214-227.

Byman, R. (1993). From two-dimensional to three-dimensional curiosity: A reanalysis of depth-breadth-factor model. *Australian Journal of Psychology, 45,* 155-160.

Byman, R. (1995). Curiosity, interest, and intrinsic motivation: A conceptual analysis. In P. Kansanen (Ed.), *Discussions on some educational issues VI.* (Research Report No. 145), University of Helsinki, Department of Teacher Education.

Byrne, B.M. (1988). Measuring adolescent self-concept: Factorial validity and equivalcy of the SDQ III across gender. *Multivariate Behavioral Research, 23,* 361-375.

Byrne, B.M. (1991). The Maslach Burnout Inventory: Validating factorial structure and invariance across intermediate, secondary, and university educators. *Multivariate Behavioral Research, 26,* 583-605.

Byrne, B.M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslac Burnout inventory. *Multivariate Behavioral Research, 26,* 289-311.

Byrne, B.M. (1995). One application of structural equation modeling from two perspectives: Exploring the EQS and LISREL strategies. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 138-157). Thousand Oaks, California: Sage Publications, Inc.

Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Byrne, B.M., Baron, P., & Campbell, T.L. (1993). Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. *Journal of Research on Adolescence, 3,* 127-143.

Byrne, B.M., & Shavelson, R.J. (1996). On the structure of social self-concept for pre-, early, and late adolescents: A test of Shavelson, Hubner, and Stanton (1976) model. *Journal of Personality and Social Psychology, 3,* 599-613.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issues of partial measurement invariance. *Psychological Bulletin, 105,* 456-466.

Campbell, D.T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Carlson, M., & Mulaik, S.A. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research, 28*, 111-159.

Cattell, R.B. (1950). *Personality. A systematic theoretical and factual study.* New York: McGraw-Hill Book Company, Inc.

Cattell, R.B. (1957). *Personality and motivation. Structure and measurement.* New York: World Book Company.

Cattell, R.B., & Kline, P. (1977). *The scientific analysis of personality and motivation*. London: Academic Press.

Chou, C.-P., & Bentler, P.M. (1995). Estimates and tests in structural equation modeling. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 37-55). Thousand Oaks, California: Sage Publications, Inc.

Clark, L.A., & Watson, D. (1995). Construct validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46*, 648-659.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd. ed.). Hillsdale, N J: Lawrence Erlbaum Associates, Inc.

Cronbach, L.J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollinsPublishers.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety.* San Francisco: Jossey-Bass.

Csikszentmihalyi, M. (1997). *Creativity: Flow and the psychology of discovery and invention*. New York : HarperCollinsPublishers.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147-167.

Cuttance, P. (1987). Issues and problems in the application of structural equation models. In P. Cuttance, & R. Ecob (Eds.), *Structural modeling by example: Applications in educational, sociological, and behavioral research.* New York: Cambridge.

Danziger, K. (1997). *Naming the mind. How psychology found its language.* London: SAGE Publications.

Day, H.I. (1968). The role of specific curiosity in school achievement. *Journal of Educational Psychology*, *59*, 37-43.

Day, H.I. (1971). The measurement of specific curiosity. In H.I. Day, D.E. Berlyne, & D.E. Hunt (Eds.), *Intrinsic motivation: A new direction in education* (pp. 99-112). Toronto: Holt, Rinehart and Winston.

Day, H.I. (1982). Curiosity and the interested explorer. *Performance and instruction*, *21*, 19-22.

Day, H.I., & Berlyne, D.E. (1971). Intrinsic motivation. In G.S. Lesser (Ed.), *Psychology and educational practice* (pp. 294-335). Glenview, Illinois: Scott, Foresman and Company.

Day, H.I., & Maynes, F. (1972). The teacher and curiosity: A third way. In N. Byrne & J. Quarter (Eds), *Must school fail? The growing debate in Canadian education*. Toronto: McCelland and Stewart Limited.

Deci, E.L. (1975). *Intrinsic motivation*. New York: Plenum Press.

Deci, E.L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K.A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (43-70). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

Deci, E.L., & Ryan, R.M. (1994). Promoting self-determined education. *Scandinavian Journal of Educational Research*, *38*, 3-14.

Deci, E.L., & Ryan, R.M. (2000). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In C. Sansone & J.M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation. The search for opitimal motivation and performance* (pp. 13-54). San Diego: Academic Press.

de Vaus, D.A. (1991). *Surveys in social research*. Singapore: Allen & Unwin.

Dewey, J. (1913). *Interest and effort in education*. New York: Houghton Mifflin.

Eysenck, M.W., & Keane, M.T. (1990). *Cognitive psychology*. Hove, UK: Lawrence Erlbaum Associates, Publishers.

Flavell, J.H. (1963). *The developmental psychology of Jean Piaget*. London: Van Nostrand Company.

Fleishman, J., & Benson, J. (1987). Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurement, 47*, 925-939.

Floyd, F.J., & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.

Fowler, H. (1965). *Curiosity and exploratory behavior.* New York: The Macmillan Company.

Freud, S. (1969). *Johdatus psykoanalyysiin* [A general introduction to psychoanalysis]. Jyväskylä: Gummerus.

Freud, S. (1971). *Seksuaaliteoria* [Three essays on the theory of sexuality]. Jyväskylä : Gummerus.

Geen, R.G. (1997). Psychophysiological approach to personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 387-414). San Diego, California: Academic Press.

Gerbing, W.G. & Hamilton, J.G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling, 3*, 62-72.

Giambra, L.M., Camp, C.J., & Grodsky, A. (1992). Curiosity and stimulation seeking across the adult life span: Cross-sectional and 6- to 8-year longitudinal findings. *Psychology and Aging*, 7, 150-157.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Gorsuch, R.L. (1983). *Factor analysis* (2nd. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gorsuch, R.L. (1988). Exploratory factor analysis. In J. R. Nesselroade, & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 231-258). New York: Plenum Press.

Gorsuch, R.L. (1991). *UniMult Guide* [Computer software]. Altadena, CA: UniMult.

Gottfried, A.E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology, 77*, 631-645.

Gottfried, A.E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82*, 525-538.

Groves, R. M. (1989). *Surveyerrors and survey costs*. New York: John Wiley & Sons, Inc.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*, 407-434.

Gustafsson, J.-E., & Stahl, P.E. (1999). *Streams user's guide*. Vs. 2.0. Mölndal, Sweden: MultivariateWare.

Görlitz, D. (1987). Concluding observations about curiosity and play. In D. Görlitz, & J.F. Wohlwill (Ed.), *Curiosity, imagination, and play* (pp. 350-374). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Harman, H. (1976). *Modern factor analysis*. Chicago: The University of Chicago Press.

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in classroom: Motivational and informational components. *Developmental Psychology, 17*, 300-312.

Harter, S., & Jackson, B.K. (1992). Trait vs. nontrait conceptualizations of intrinsic/extrinsic motivational orientation. *Motivation and Emotion, 16*, 209-230.

Hayduk, L.A. (1987). *Structural equation modeling with LISREL*. Baltimore: The Johns Hopkins University Press.

Hayduk, L.A. (1996). *LISREL issues, debates, and strategies*. Baltimore: The Johns Hopkins University Press.

Haywood, H.C. (1971). Individual differences in motivational orientation: A trait approach. In H.I. Day, D.E. Berlyne, & D.E. Hunt (Eds.), *Intrinsic motivation: A new direction in education* (pp. 113-127). Toronto: Holt, Rinehart and Winston.

Hebb, D.O. (1955). Drives and the C.N.S. *Psychological Review, 62*, 243-245.

Heck, R.H., & Thomas, S.L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Heckhausen, H. (1987). Emotional components of action: Their ontogeny as reflected in achievement behavior. In D. Görlitz, & J.F. Wohlwill (Eds.), *Curiosity, imagination, and play* (pp. 326-348). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Henderson, B.B. (1984a). Parents and exploration: The effect of context on individual differences in exploratory behavior. *Child Development, 55*, 1237-1245.

Henderson, B.B. (1984b). Social support and exploration. *Child Development, 55*, 1246-1251.

Henderson, B.B. (1994). Individual differences in experience-producing tendencies. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration.* (pp. 213-225). Heidelberg: Springer-Verlag.

Henderson, B.B., & Moore, S.G. (1979). Measuring exploratory behavior in young children: A factor analytic study. *Developmental Psychology, 15*, 113-119.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*, 549-571.

Hidi, S., & Anderson, V. (1992). Situational interest and its impact on reading and expository writing. In K.A. Renninger, S. Hidi, S. & A. Krapp (Eds.) *The role of interest in learning and development* (215-238). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Hidi, S., & McLaren, J. (1990). The effect of topic and theme interestingness on the production of school expositions. In H. Mandl, E. De Corte, N. Bennet, & H.F. Friedrich (Eds.), *Learning and instruction: European research in an international context* (Vol. 2, pp. 295-308). Oxford: Pergamon.

Hox, J.J. (1995). *Applied multilevel analysis.* Amsterdam: TT-Publikaties.

Hoyle, R.H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 1-15). Thousand Oaks, California: Sage Publications, Inc.

Hoyle, R.H., & Lennox, R.D. (1991). Latent structure of self-monitoring. *Multivariate Behavioral Research, 26*, 511-540.

Hoyle, R.H., & Panter, A.T. (1995). Writing about structural equation models. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 158-176). Thousand Oaks, California: Sage Publications, Inc.

Hoyle, R.H., & Smith, G.T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62*, 429-440.

Hu, L., & Bentler, P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 76-99). Thousand Oaks, California: Sage Publications, Inc.

Hull, J.G., Tedlie, J.C, & Lehn, D.A. (1995). Modeling the relation of personality variables to symptom complaints: The unique role of negative affectivity. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 217-235). Thousand Oaks, California: Sage Publications, Inc.

Hunt, J.McV. (1963). Piaget's obsrevations as a source of hypotheses concerning motivation. *Merrill-Palmer Quarterly, 9*, 263-275.

Hunt, J.McV. (1971a). Intrinsic motivation: Information and circumstance. In H.M. Schroder & P. Suedfeld. *Personality theory and information processing* (pp. 85-130). New York: The Ronald Press Company.

Hunt, J.McV. (1971b). Intrinsic motivation and psychological development. In H.M. Schroder & P. Suedfeld. *Personality theory and information processing* (pp. 131-177). New York: The Ronald Press Company.

Hunt, J.McV. (1981). Experimental roots of intention, initiative, and trust. In H. I. Day (Ed.), *Advances in intrinsic motivation and aesthetics* (pp. 169-202). New York: Plenum Press.

Hutt, C. (1970). Specific and diversive exploration. In H.W. Reese & L.P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 5, pp. 120-180). New York: Academic Press.

Hutt, C. (1981). Toward a taxonomy and conceptual model of play. In H.I. Day (Ed.), *Advances in intrinsic motivation and aesthetics* (pp. 251-298). New York: Plenum Press.

Iran-Nejad, A. (1987). Cognitive and affective causes of interest and linking. *Journal of Educational Psychology, 79*, 120-130.

Iran-Nejad, A., McKeachie, W.J., & Berliner, D.C. (1990). The multisource nature of learning: An introduction. *Review of Educational Research, 60*, 509-515.

James, L.R., Mulaik, S.A., & Brett, J.M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.

Johnson, J.A. (1997). Units of analysis for the description and explanation of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 73-93). San Diego, California: Academic Press.

Jöreskog, K.G. (1979a). A general approach to confirmatory maximum likelihood factor analysis. In J. Magidson (Ed.), *Advances in factor analysis and structural equation models* (pp. 21-43). Cambridge, Massachusetts: Abt Books.

Jöreskog, K.G. (1979b). Simultaneous factor analysis in several populations. In J. Magidson (Ed.), *Advances in factor analysis and structural equation models* (pp. 189-206). Cambridge, Massachusetts: Abt Books.

Jöreskog, K.G. (1993). Testing structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park: Sage Publications, Inc.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS, Inc.

Jöreskog, K.G., & Sörbom, D. (1993a). *LISREL 8: User's reference guide.* Chicago,IL: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1993b). *PRELIS2: User's reference guide.* Chicago,IL: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1999). LISREL 8.30 [Computer software]. Chicago: Scientific Software International.

Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology, 22,* 51-66.

Kaplan, D. (1993). Statistical power in structural equation modeling. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 100-117). Thousand Oaks, California: Sage Publications, Inc.

Keller, J.A. (1987). Motivational aspects of exploratory behavior. In D. Görlitz, & J.F. Wohlwill (Eds.), *Curiosity, imagination, and play* (pp. 24-42). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Keller, H. (1994). A developmental analysis of exploration styles. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration.* (pp. 199-212). Heidelberg: Springer-Verlag.

Keller, H., Schneider, K., & Henderson, B. (1994). Preface: The study of exploration. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration* (pp. 1-13). Heidelberg: Springer-Verlag.

Keller, H., Schölmerich, A., Miranda, D., & Gauda, G. (1987). The development of exploratory behavior in the first four years of life. In D. Görlitz, & J.F. Wohlwill (Eds.), *Curiosity, imagination, and play* (pp. 24-42). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Keller;,J.A. (1987). Motivational aspects of exploratory behavior. In D. Görlitz, & J.F. Wohlwill (Eds.), *Curiosity, imagination, andplay*(pp. 24-42). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Kirkland, J. 1976. Interest: Phoenix in psychology. *Bull. Br. psychol. Soc 29*, 33-41.

Kishton, J.M., & Widaman, K.F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement, 54*, 757-765.

Kohn, P.M., & Annis, H.M. (1975). Validity data on a modified version of Pearson's novelty experiencing scale. *Canadian Journal of Behavioral Science, 7*, 274-278.

Kohn, P.M., Hunt, R.W., & Hoffman, F.M. (1982). Aspects of experience seeking. *Canadian Journal of Behavioral Science, 14*, 13-23.

Krapp, A.. (1994). Interest and curiosity. The role of interest in a theory of exploratory action. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration* (pp. 79-99). Heidelberg: Springer-Verlag.

Krapp, A., Hidi, S. & Renninger, K.A. (1992). Interest, learning and development. In Renninger, K.A., Hidi, S., & Krapp, A. (Eds.) *The role of interest in learning and development* (pp. 3-25). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Kreitler, S., & Kreitler, H. (1986). Types of curiosity and their cognitive determinants. *Archiv für Psychologie, 138*, 233-251.

Kreitler, S., & Kreitler, H. (1990). *The cognitive foundations of personality traits*. New York: Plenum Press.

Kreitler, S., & Kreitler, H. (1994). Motivational and cognitive determinants of exploration. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration* (pp. 259-284). Heidelberg: Springer-Verlag.

Kreitler, S., Kreitler, H., & Zigler, E. (1974). Cognitive orientation and curiosity. *British Journal of Psychology, 65*, 43-52.

Kreitler, S., Kreitler, H., & Zigler, E. (1975). The nature of curiosity in children. *Journal of School Psychology, 13*, 185-200.

Kreitler, S., Zigler, E., & Kreitler, H. (1984). Curiosity and demographic factors as determinants of children's probability-learning strategies. *The Journal of Genetic Psychology, 145*, 61-75.

Langevin, R. (1971). Is curiosity a unitary construct? *Canadian Journal of Psychology, 25*, 360-374.

Langevin, R. (1976). Construct validity of sensation seeking and curiosity measures of normal and psychotic subjects. *Canadian Journal of Behavioral Science, 8*, 251-262.

Leahey, T.H. (1991). *A history of modern psychology*. Englewood Cliffs: Prentice-Hall.

Leherissey, B.L. (1972). Validation of a measure of state epistemic curiosity in a computer-assisted learning situation. *Proceeding of the 80th Annual Convention of the American Psychological Association, 7*, 523-524.

Leskinen, E. (1987). *Faktorianalyysi* [Factor analysis]. Jyväskylä: University of Jyväskylä. Department of Statistics.

Leskinen, E. (1989). *Spesififaktoreiden mallintamisesta ja identifioituvuudesta kovarianssirakennemalleissa*. Jyväskylä: University of Jyväskylä. Department of Statistics.

Leuba, C. (1955). Toward some integration of learning theories: The concept of optimal stimulation. *Psychological Reports, 1*, 27-33.

Livson, N. (1967). Toward a different construct of curiosity. *The Journal of Genetic Psychology, 111*, 73-84.

Lloyd, J., & Barenblatt, L. (1984). Intrinsic intellectuality: Its relations to social class, intelligence, and achievement. *Journal of Personality and Social Psychology, 46*, 646-654.

Loehlin, J.C. (1987). *Latent variable models*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*, 75-98.

Long, J.S. (1983). *Confirmatory factor analysis*. Newbury Park, California: SAGE Publications, Inc.

Looft, W.R., & Baronowski, M.D. (1971). An analysis of five measures of sensation seeking and preference for complexity. *Journal of General Psychology, 85*, 307-313.

MacCallum, R. (1995). Model specification: Procedures, strategies, and related issues. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 16-36). Thousand Oaks, California: Sage Publications, Inc.

MacCallum, R., Roznowski, M., Mar, C.M., & Reith, J.V. (1994). Alternative strategies for cross-validation of covariance structure models. *Multivariate Behavioral Research, 29,* 1-32.

MacCallum, R., Roznowski, M., & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490-504.

MacCrae, R.R., & Costa, P.T. (1997). Conceptions and correlates of opennes to experience. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 825-847). San Diego, California: Academic Press.

Madsen, K.B. (1974). *Modern theories of motivation.* Copenhagen: Munksgaard.

Madsen, K. B. (1981). Berlyne's theory: A metascientific study. In H. I. Day (Ed.), *Advances in intrinsic motivation and aesthetics* (pp. 19-38). New York: Plenum Press.

Marsh, H.W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 177-198). Thousand Oaks, California: Sage Publications, Inc.

Maruyama, G.M. (1998). *Basics of structural equation modeling.* Thousand Oaks, California: Sage Publications, Inc.

Maslow, A. (1968). *Toward a psychology of being.* New York: Harper

Maslow, A. (1970). *Motivation and personality,* 2nd ed. New York: Harper.

Maw, W.H., & Maw, E.W. (1961). Establishing criterion groups for evaluating measures of curiosity. *Journal of Experimental Education, 29*, 299-306.

Maw, W.H., & Maw, E.W. (1964). *An exploratory investigation into the measurement of curiosity in elementary school children.* U.S. Office of Education, Cooperative Research Report No. 801. University of Delaware.

Maw, W.H., & Maw, E.W. (1965). *Personal and social variables differentiating children with high and low curiosity.* Cooperative Research Project No. 1511, University of Delaware.

Maw, W.H., & Maw, E.W. (1975). Contrasting proverbs as a measure of attitudes of college students toward curiosity-related behaviors. *Psychological Reports, 37*, 1085-1086.

Maw, W.H., & Maw, E.W. (1977). Nature and assessment of human curiosity. In P. McReynolds (Ed.), *Advances in psychological assessment.* Vol. 4. San Francisco: Jossey-Bass.

McAdams, D.P. (1997). A conceptual history of personality psychology. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 3-39). San Diego, California: Academic Press.

McArdle, J.J., & Cattell, R.B. (1994). Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research, 29*, 63-113.

McReynolds, P. (1971). The three faces of cognitive motivation. In H.I. Day, D.E. Berlyne, & D.E. Hunt (Eds.), *Intrinsic motivation: A new direction in education* (pp. 33-45). Toronto: Holt, Rinehart and Winston.

Mehrabian, A., & Russell, J.A. (1973). A measure of arousal seeking tendency. *Environment and Behavior, 5*, 315-333.

Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 105-146). Phoenix: Oryx Press.

Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment, 10*, 1-9.

Mikulincer, M. (1997). Adult attachment style and information processing: Individual differences in curiosity and cognitive closure. *Journal of Personality and Social-Psychology, 72*, 1217-1230.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the structure of behavior.* London: Holt, Rinehart and Winston.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Moch, M. (1987). Asking questions: An expression of epistemological curiosity in children. In D. Görlitz, & J.F. Wohlwill (Eds.), *Curiosity, imagination, and play* (pp. 198-211). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22,* 267-305.

Mulaik, S. A. (1988). Confirmatory factor analysis. In J.R. Nesselroade & R.B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 259-288). New York: Plenum Press.

Mulaik, S. A. (1993). Objectivity and multivariate statistics. *Multivariate Behavioral Research, 28*, 171-203.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N. Lind, S., & Stillwell, C. D. (1989). An evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430-445.

Mulaik, S.A., & Millsap, R.E. (2000). Doing the four-step right. *Structural Equation Modeling, 7,* 36-73.

Mulaik, S.A., & Quartetti, D.A. (1997). First order or higher order general factor? *Structural Equation Modeling, 4*, 193-211.

Murphy, P. K., & Alexander, P.A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology 25,* 3-53.

Muthén, L.K., & Muthén, B.O. (1998). *Mplus. The comprehensive modeling program for applied researchers*. Los Angeles: Muthén & Muthén.

Naylor, F.D. (1981). A state-trait curiosity inventory. *Australian Psychologist, 16,* 172-183.

Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill Book Company.

Nunnally, J.C. (1981). Explorations of exploration. In H.J. Day (Ed.), *Advances in intrinsic motivation and aesthetics* (pp. 341-364). New York: Plenum.

Olson, E. (1986). *Measurement of curiosity in junior high school students*. Unpublished doctoral dissertation, Iowa State University, Iowa.

Olson, K., & Camp, C. (1984). Factor analysis of curiosity measures in adults. *Psychological Reports, 54*, 491-497.

Pearson, P. (1970). Relationships between global and specified measures of novelty seeking. *Journal of Consulting Psychology, 11*, 199-204.

Penny, R.K., & McCann, B. (1964). The children's reactive curiosity scale. *Psychological Reports, 15*, 323-334.

Penny, R.K., & Reinehr, R.C. (1966). Development of a stimulus-variation seeking scale for adults. *Psychological Reports, 18*, 631-638.

Pervin, L.A. (1984). *Personality: theory and research* (4th ed.). New York : John Wiley.

Piaget, J. (1971). *The psychology of intelligence*. London: Routledge & Kegan Paul Ltd.

Piaget, J. (1977). *The development of thought. Equilibration of cognitive structures*. Oxford: Basil Blackwell.

Piaget, J. (1981). *Intelligence and affectivity: Their relationship during child development.* Palo Alto, Califomia: Annual Reviews Inc.

Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.

Rappoport, L., Peters, G.R., Downey, R., McCann, T., & Huff-Corzine, L. (1993). Gender and age differences in food cognition. *Appetite, 20*, 33-52.

Reuterberg, S.-E., & Gustafsson, J.-E. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement, 52*, 795-811.

Rigby, C.S., Deci, E.L., Patrick, B.C., & Ryan, R.M. (1992). Beyond the intrinsic-extrinsic dichotomy: Self-determination in motivation and learning. *Motivation and Emotion, 16*, 165-185.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*, 51-67.

Rogers, C.R. (1969). *Freedom to learn.* Columbus, Ohio: Charles E. Merrill Publishing Comapnay.

Rubenstein, A. S. (1986). *An item-level analysis of questionnaire-type measures of intellectual curiosity*. Unpublished doctoral dissertation, Cornell University.

Rummel, R.J. (1970). *Applied factor analysis.* Evanston: Northwestern University Press.

Russo, M.F., Lahey, B.B., Christ, M.A.G., Frick, P.J., McBurnett, K., Walker, J.L., Loeber, R., Stouthamer-Loeber, M., & Green, S. (1991). Preliminary development of a sensation seeking scale for children. *Personality and Individual Differences, 12,* 399-405.

Russo, M.F., Stokes, G.S., Lahey, B.B., Christ, M.A.G., McBurnett, K., Loeber, R., Stouthamer-Loeber, M., & Green, S. (1993). A sensation seeking scale in children: Further refinement and psychometric development. *Journal of Psychopathology and Behavioral Assessment, 15,* 69-86.

Ryan, R.M., & Deci, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology 25,* 54-67.

Saris, W.E.A., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C.C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105-130). Washington, D.C.: American Sociological Association.

Schafer, R. (1968). *Aspects of internalization*. New York: International University Press.

Schau, C., Stevens, J., Dauphinee, T.L., & Del-Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55,* 868-875.

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*, 299-323.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K.A Renninger, S, Hidi., & A. Krapp, (Eds.), *The role of interest in learning and development* (pp. 183-212). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Schneider, K. (1987). Subjective uncertainty and exploratory behavior in pre-school children. In D. Görlitz, & J.F. Wohlwill (Eds.), *Curiosity, imagination, and play* (pp. 179-197). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Schneider, K. & Unzner, L. (1994). Preschoolers' exploratory behavior: The influence of the social and physical context. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration* (pp. 177-197). Heidelberg: Springer-Verlag.

Schumacker, R.E., & Lomax, R.G. (1996). *A beginner's guide to structural equation modeling. Mahwah*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Schölmerich, A. (1994). The process and consequences of manipulative exploration. In H. Keller, K. Schneider & B. Henderson (Eds.), *Curiosity and exploration* (pp. 241-257). Heidelberg: Springer-Verlag.

Slife, B.D., & Williams, R.N. (1995). *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. Thousand Oaks, CA: SAGE Publications.

Spielberger, C.D. (1979). *Preliminary manual for the state-trait personality inventory (STPI)*. Unpublished manuscript, University of South Florida, Tampa.

Spielberger, C.D., Edwards C.D., Lushene, R.E. Montuori, J., & Platzek, D. (1972). *Manual for the state-trait anxiety inventory for children*. Palo alto, CA: Consulting Psychologist Press.

Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. (1970). *Manual for the state-trait anxiety inventory*. Palo alto, CA: Consulting Psychologist Press.

Spielberger, C.D., & Starr, L.M. (1994). Curiosity and exploratory behavior. In H.F. O'neil, & M. Drillings (Eds.), *Motivation: Theory and research* (pp. 221-243). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Steenkamp, H.B., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173-180.

Steiger, J.H. (1995). Structural equation modeling. In STATISTICA release 5 [Computer software] (pp. 3539-3687). Tulsa: StatSoft, Inc.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics*. New York: HarperCollins.

Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park: Sage Publications, Inc.

Tarkkonen, L. (1987). *On reliability of composite scales: An essay on the structure of measurement and the properties of the coefficients of reliability - an unified approach.* Helsinki: Finnish Statistical Society. Statistical Studies 7.

Tepper, K., & Hoyle, R.H. (1996). Latent variable models of need for uniqueness. *Multivariate Behavioral Research, 31,* 467-494.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*, 37-54.

Tomas, J.M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling, 6*, 84-98.

Trudewind, C., & Schneider, K. (1994). Interindividual differences in the development of exploratory behavior. Methodological considerations. In H. Keller, K. Schneider, & B. Henderson (Eds.), *Curiosity and exploration* (pp. 154-176). Heidelberg: Springer-Verlag.

Vando, A. (1970). A personality dimension related to pain tolerance (Doctoral dissertation, Columbia University, 1969). *Dissertation Abstracts International, 31,* 2292B-2293B (University Microfilms No. 70-18,865).

Vando, A. (1974). The development of the R-A Scale: A paper-and-pencil measure of pain tolerance. *Personality and Social Psychology Bulletin, 1*, 28-29.

Voss, H.-G. & Keller, H. (1983). *Curiosity and exploration. Theories and Results*. New York: Academic Press, Inc.

Voss, H.-G., & Keller, H. (1986). Curiosity and exploration: A program of investigation. *The German Journal of Psychology, 10*, 327-337.

Voss, J.F., & Schauble, L. (1992). Is interest educationally interesting? An interest-related model of learning. In K.A Renninger, S, Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 101-120). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equating modeling* (pp. 56-75). Thousand Oaks, California: Sage Publications, Inc.

White, R.W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297-333.

Wiggins, J.S. (1997). In defense of traits. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 95-115). San Diego, California: Academic Press.

Wiggins, J.S., & Trapnell, P.D. (1997). Personality structure: The return of big five. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 737-765). San Diego, California: Academic Press.

Wittgenstein, L. (1980). *Sininen ja ruskea kirja* [The blue and brown books: preliminary studies for the "Philosophical investigations"]. Juva: WSOY.

Wittgenstein, L. (1981). *Filosofisia tutkimuksia* [Philosophical investigations]. Juva: WSOY.

Wohlwill, J.F. (1981). A conceptual analysis of exploratory behavior: The "specific-diversive" distinction revisited. In H.J. Day (Ed.), *Advances in intrinsic motivation and aesthetics* (pp. 341-364). New York: Plenum.

Wohlwill, J.F. (1987). Varieties of exploratory activity in early childhood. In D. Görlitz, & J.F. Wohlwill (Ed.), *Curiosity, imagination, and play* (pp. 59-77). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Woodworth, R.S. (1918). *Dynamic psychology*. New York: Colombia University Press.

Woodworth, R.S. (1958). *Dynamics of behavior*. New York: Holt, Rinehart and Winston.

Zuckerman, M. (1971). Dimensions of sensation seeking. *Journal of Consulting and Clinical Psychology, 36*, 45-52.

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimum level of arousal*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Zuckerman, M. (1984). Sensation seeking: A comparative approach to a human trait. *The Behavioral and Brain Sciences*, 7, 413-471.

Zuckerman, M. (1987). A critical look at three arousal constructs in personality theories: Optimal levels of arousal, strength of the nervous system, and sensitivities to signals of reward and punishment. In J. Strelau & H.J. Eysenck (Eds.), *Personality dimensions and arousal*. (pp. 218-232). New York: Plenum Press.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York: Cambridge University Press.

# APPENDIX A

## ITEMS OF THE INTRINSIC MOTIVATION SCALE (SEE BESWICK, 1974, FOR FURTHER INFORMATION)

1. I visit a library to read material not directly related to my class work.
2. I would like to watch an astronomer calculate the age of a star.
3. If I read something which puzzles me, I keep reading until I understand it.
4. Complicated machinery is fascinating to look at.
5. When I don't know the answers to a question on a test I look up the answers when the test is completed.
6. I read for enjoyment during a large part of my spare time.
7. I am interested in mathematical procedures possible with new calculating machines.
8. I like to look at pictures which are puzzling in some way.
9. I read several magazines regularly.
10. It is interesting to try to figure out how an unusual piece of machinery works.
11. I like to look at rocks which are made of many kinds of minerals.
12. I have had experiences which inspired me to write a poem or a story, make up a humorous tale or paint a picture.
13. I think about how strange plants grow.
14. At times I have focussed on something so hard that I went into a kind of benumbed state of consciousness.
15. If I come across something interesting I drop everything and study it. It is never a waste of time.

Note: Original Item 11, "Some truths can only be expressed in paradoxical statements", was eliminated from the scale (see text).

## ITEMS OF THE DIVERSIVE CURIOSITY SCALE (SEE DAY, 1971, FOR FURTHER INFORMATION)

1. I soon get bored when there is not enough going on.
2. I like to go somewhere different nearly every day.
3. I like to eat the same kind of food most of the time.
4. I get tired of doing the same thing all the time.
5. I avoid busy, noisy places.
6. I am always glad to have someone visit me.
7. I never spend any time alone if I can help it.
8. I like a place better the more I am around it.
9. I often feel restless.
10. I like to have lots of activity around me.

## ITEMS OF THE C-TRAIT INVENTORY (SEE OLSON, 1986, FOR FURTHER INFORMATION)

1. I like to ask about things that I do not fully understand.
2. I ignore objects around me.
3. I question a lot of things.
4. I wonder what makes electricity work.
5. I avoid picking up objects to inspect them.
6. New events capture my attention.
7. I want to find out things.
8. I don't care how television works — I prefer just to watch it.
9. I enjoy handling new objects to explore them.
10. I like to observe things that are going on in my environment.
11. I like to seek out things to find out their meanings.
12. Computers interest me because they seem so complex.
13. When I see knobs or dials on things, I want to turn them.
14. When I hear strange sounds, I like to find out what is making them.

15. I like to create puzzles and games in my own mind.
16. The complex is more exciting than the simple.
17. I like to touch paintings and works of art.
18. Bright colors capture my attention.
19. I dislike looking up words in the dictionary.
20. I like to study things that are easy.
21. I enjoy playing with silly putty, clay, and other things that can be shaped with my hands.
22. I like to discover patterns in designs.
23. I like to think about problems and try to solve them in my head.
24. I like to study objects that are puzzling and unusual.
25. I like to take objects apart to find out more about them.
26. School is boring.
27. I would rather handle things than just look at them.
28. It's fun to look at unusual art.
29. If I see a new machine in the room, I am likely to touch it.
30. When I hear sudden claps of thunder, I like to look at the sky.
31. I would rather solve a problem myself than be told how to do it by someone else.
32. I avoid complex situations.
33. I learn about new objects by touching them.
34. I like to notice everything that goes on around me.
35. I like to explore things to find out information about them.
36. When I see a complex machine, I want to know how it works.
37. It's interesting to handle seashells of different shapes and sizes.
38. I like to sit quietly and listen to the birds sing and the cars pass by.
39. I wish I know everything in books.
40. I look at complex objects longer than I do simple objects.

## ITEMS OF THE SENSATION SEEKING SCALE (SEE BJÖRCK-ÅKESSON, 1990, FOR FURTHER INFORMATION)

1 A Sometimes I think it is fun to do dangerous things.
B A sensible person avoids doing dangerous things.

2 A I like to explore new places myself.
B I'd rather that someone goes with me and shows me around when I come to a new place.

3 A I think it feels safe to see the same old people every day.
B I think it is boring to see the same old faces every day.

4 A I would like to live in the woods where it is peaceful and quite.
B I would like to live in the city where something always happens.

5 A It is dangerous to bike real fast downhill.
B It is exciting to bike real fast downhill.

6 A It is always fun to go to a party.
B Sometimes it is boring to go to a party.

7 A I like to eat food that I have tasted before and that I know taste well.
B I think it is exciting to taste food I have never tasted before.

8 A I like friends who are ingeneous with an intense temper.
B I like friends who are calm with an even temper.

9 A It is fun to act in a theater play.
B I don't like to act in a theater play.

10 A I like to go on excursion where the route I decided beforehand.
B On an excursion I don't want the route to be decided beforehand.

11 A It is terrific to go real fast in a racingboat.
B It is lovely to toad canoe on a calm evening.

12 A When I am going to swim in a cold lake I dip myself gradually to get used to the cold water.
B When I am going to swim in a cold lake I plunge or dive directly into the cold water.

13 A I rather not contact people I don't know.
B It is fun to get in contact with people you don't know.

14 A I think that we always shall follow schedule in school.
B It is fun to do things that are not on schedule.

15 A I don't like to read about violence.
B I think it is exciting to read about violence.

16 A It would be exciting to join a lion-safari in Africa.
B It seems dangerous to join a lion-safari in Africa.

17 A I like to walk alone in the woods.
B I like to walk in the woods with others.

18 A I would like to move often and live in many different places when I get older.
B I would rather live in the same place all the time when I get older.

19 A I don't care if I have to change plans.
B I get irritable if I must change plans.

20 A I would like to join a travelling circus.
B It seems like hard work to join a traveling circus.

21 A I like candy that is medium strong.
B I like candy that is so strong that you get chock when you taste it.

22 A modern art in strong colors appeal me.
B I like art where you easily can trace the motive.

23 A I would not like to be hypnotized.
B It would be exciting to be hypnotized.

24 A It is fun to have a speech in front of people.
B I don't like to speak in front of many people.

25 A I prefer when there are many people around.
B I prefer when there are not so many people around.

26 A Even when I am engaged in something fun I often think of what is going to happen next.
B When I experience something fun I don't think of other things.

27 A I often day-dream of doing dangerous and exciting things.
B I most often day-dream of doing ordinary things.

28 A Punks are terrific.
B I don't understand punks.

29 A I prefer to be with friends that I know.
B I think it is exciting to be with new people.

30 A I often dream of being a mountain-climber.
B I don't understand people who risk life trough climbing mountains.

31 A I prefer people that dress neat and tidy.
B I prefer people who dares to dress differently.

32 A I like wild parties.
B I prefer calm parties where you talk to each other.

33 A I don't like when something unexpected happens.
B It is fun when something unexpected happens.

34 A It seems dangerous to waterski.
B It seems fun to waterski.

35 A It is fun to do practical jokes with Butterick's articles.
B It is silly to do practical jokes with party novelties.

36 A I prefer lessons with many activities in the classroom.
B I prefer calm lessons when everybody works with the same task.

37 A I have certain favourite records that I always listen to.
B I like to listen to new music.

38 A I like to take a trip in the UFO at Liseberg.
B I don't like to take a trip in the UFO at Liseberg.

39 A It is fun to dress in new clothes.
B I have certain favorite-clothes that I always want to wear.

40 A I don't like discussions where people get angry.
B I like "hot" discussions.

41 A Everything that is fun is either illegal or immoral.
B Most things that are fun are both legal and moral.

42 A I think it is fun to be with people that do extraordinary things.
B I don't think it is fun to be with people that do extraordinary things.

43 A It is a wonderful feeling to run real fast on a motorbike.
B It is stupid to risk life through running real fast on a motorbike.

44 A I prefer a job where you see a lot of people.
B I prefer a job where you work a lot by yourself.

45 A It seems silly to fly in a hang glider.
B I would like to try flying in a hang glider.

46 A It is fun to dress up in unusual clothes.
B I'd rather be dressed like everybody else.

47 A I'd rather take the same route every time I go to school.
B Sometimes I take different routes to school for a change.

48 A I think it is fun to ride a roller coaster.
B I don't like to ride a roller coaster.

49 A It is thrilling to try new kinds of candy, food and drink.
B I always by try same kind of candy, food and drink.

50 A I always prefer to shop in the same store.
B It is fun to shop in new stores.

51 A I never care for chocking people.
B Sometimes I do strange things just to chock people.

52 A A lion-tamer has an exciting job.
B Being a lion-tamer is a dangerous job.

53 A I prefer jazz and classical music to popmusic.
B I prefer popmusic to jazz and classical music.

54 A I'd rather have our ordinary teachers all the time.
B It is fun to have substitute teacher for a change.

55 A It is risky and dangerous to climb a tall tree.
B It is thrilling to climb a tall tree.

3[a] A It is fun to go to a discoteque.
B It is too noisy on discoteque.

| Half-scale | Items |
|---|---|
| T1.1: | 1, 5, 16, 17, 30, 42, 43, 52 |
| T1.2: | 2, 12, 23, 27, 40, 45, 55, |
| N1.1: | 7, 13, 20, 34, 38, 47, 49 |
| N1.2: | 10, 14, 19, 26, 33, 37, 48, 50 |
| A1.1: | 8, 11, 15, 25, 32, 41, 54, |
| A1.2: | 3, 4, 6, 21, 28, 36 |
| O1.1: | 9, 18, 44, 46, 51, 53 |
| O1.2: | 22, 24, 29, 31, 35, 39 |

[a]Eliminated form the scale (see text).

## ITEMS OF THE TEACHER RATINGS SCALE (SEE DAY, 1971 AND ZUCKERMAN, 1984, FOR FURTHER INFORMATION)

1. Student reacts positively to new, incongruous or mysterious elements in his environment by moving toward them, by exploring them or by manipulating them.

2. Student exhibits a need or desire to know more about himself and/or his environment (asks questions, thinks over, reads nonfiction books spontaneously etc.).

3. Student persists in examining and exploring stimuli in order to know more about them.

4. Student resists monotony badly and gets easily bored. He continuously scans his surroundings, seeking new experiences.

5. Instead of choosing what familiar and safe, the student choose what is novel and mysterious, even though this choice can be risky or dangerous (physical injury, embarrassment, shame etc.)

# APPENDIX B

Sum Variables in Subscale -level Analysis

**IM** = *Test of Intrinsic Motivation* (Beswick, 1974), 15 items. Base on the "cognitive process" theory of curiosity (Beswick, 1971, 1974). According to this theory, curiosity is a predisposition to create, maintain and resolve conceptual conflicts. By means of investigator acts (exploration), assimilation and accommodation, a person tries to resolve conceptual conflicts and, in this way, to restore the balance of his or her category system.

**ConE** = *Conceptual Exploration* represents active information seeking through asking questions and analytically checking commonly accepted concepts, focusing on what is implied. Recently, Kreitler and Kreitler (1994) have noted that Conceptual Exploration corresponds to what Keller et al. (1987) have called "verbal exploration." Subscale of the broad C-Trait scale (Olson, 1986).

**ComE** = *Exploration of the Complex or Ambiguous* as a factor refers to a preference for a complex stimulus when both a simple and a complex stimulus are present. In describing Exploration of the Complex, Kreitler and Kreitler (1986) emphasized the subjective or internal aspect of perceptions, namely sensations, feelings and emotions. Subscale of the broad C-Trait scale (Olson, 1986).

**ManE** = *Manipulatory Exploration* is elicited by objects which are new in some respect and which can be explored in order to find out how they operate and what their tactile and thermal qualities are. Subscale of the broad C-Trait scale (Olson, 1986).

**PerE** = *Perceptual Exploration* refers to activity whose aim is to reveal the referent's sensory qualities and sensations, mainly what is seen and heard, often to find out what something is made of. Subscale of the broad C-Trait scale (Olson, 1986).

**DivE** = Subscale of OTIM (Day, 1968, 1971), namely the *diversive curiosity* scale (OTIMDC). The diversively curious person tries consciously look for new, amusing or exciting stimuli in order to raise the level of arousal to the optimal plane. Thus the aim of exploration is not to resolve or mitigate the uncertainty but to increase the level of activation or to provide stimulation. The diversive curiosity scale contained 10 items which described different forms of diversive exploration. DivE is the sum of seven OTIMDC items (see text).

**TAS** = *Thrill and Adventure Seeking* reflects the preference for extreme risk and challenge. Subscale of Sensation Seeking scale (SESE).

**NES** = *New Experience Seeking* refers to novelty, variation and a positive attitude towards being out of control. Subscale of Sensation Seeking scale (SESE).

**Act** = *Activity* involves the preference for being an active part of the youth culture and the preference for social interaction. Subscale of Sensation Seeking scale (SESE).

**Out** = *Outgoingness* pertains to non-conformity with generally accepted norms, being the center of attention, preference for extreme appearance and emphasis on social feedback. Subscale of Sensation Seeking scale (SESE).

**TeR 1** = First item of the teacher rating of curiosity. Item 1 was expected to measure what Berlyne (1960) called *specific exploration*.

**TeR 2** = Second item of the teacher rating of curiosity. Item 2 was expected to measure what Berlyne (1960) called *specific exploration*.

**TeR 3** = Third item of the teacher rating of curiosity. Item 3 was expected to measure what Berlyne (1960) called *specific exploration*.

**TeR 4** = Fourth item of the teacher rating of curiosity. Item 4 was expected to measure what Day (1968) labeled *diversive exploration*.

**TeR 5** = Fifth item of the teacher rating of curiosity. Item 5 was expected to describe behavior which Zuckerman (1971) called *General Sensation Seeking*.

# APPENDIX C

Correlation matrix: Girls' Sample (N = 223)

| | IM | ConE | ComE | ManE | PerE | TeR1 |
|---|---|---|---|---|---|---|
| IM | 1.00 | | | | | |
| ConE | 0.42 | 1.00 | | | | |
| ComE | 0.57 | 0.46 | 1.00 | | | |
| ManE | 0.48 | 0.33 | 0.49 | 1.00 | | |
| PerE | 0.51 | 0.44 | 0.44 | 0.46 | 1.00 | |
| TeR1 | 0.08 | 0.06 | -0.01 | 0.03 | 0.01 | 1.00 |
| TeR2 | 0.03 | -0.01 | -0.07 | -0.02 | -0.07 | 0.81 |
| TeR3 | 0.05 | 0.06 | -0.04 | -0.04 | -0.02 | 0.78 |
| DC7 | 0.18 | 0.03 | 0.11 | 0.25 | 0.13 | -0.05 |
| TAS | 0.14 | -0.04 | 0.21 | 0.27 | 0.00 | 0.10 |
| NES | 0.28 | 0.16 | 0.24 | 0.24 | 0.20 | 0.00 |
| Act | -0.06 | -0.08 | 0.04 | 0.13 | -0.13 | 0.03 |
| Out | 0.13 | 0.11 | 0.21 | 0.23 | 0.06 | -0.01 |
| TeR4 | -0.02 | 0.01 | 0.02 | 0.11 | -0.01 | 0.50 |
| TeR5 | -0.05 | 0.01 | -0.01 | 0.06 | -0.02 | 0.55 |

| | TeR2 | TeR3 | DC7 | TAS | NES | Act |
|---|---|---|---|---|---|---|
| TeR2 | 1.00 | | | | | |
| TeR3 | 0.84 | 1.00 | | | | |
| DC7 | -0.07 | -0.11 | 1.00 | | | |
| TAS | 0.05 | -0.03 | 0.14 | 1.00 | | |
| NES | -0.13 | -0.09 | 0.17 | 0.43 | 1.00 | |
| Act | 0.00 | -0.07 | 0.44 | 0.54 | 0.25 | 1.00 |
| Out | -0.13 | -0.16 | 0.28 | 0.51 | 0.47 | 0.54 |
| TeR4 | 0.47 | 0.33 | 0.10 | 0.19 | 0.05 | 0.23 |
| TeR5 | 0.48 | 0.34 | 0.04 | 0.16 | -0.01 | 0.16 |

| | Out | TeR4 | TeR5 |
|---|---|---|---|
| Out | 1.00 | | |
| TeR4 | 0.18 | 1.00 | |
| TeR5 | 0.13 | 0.59 | 1.00 |

Correlation matrix: Boys' Sample (N = 221)

| | IM | ConE | ComE | ManE | PerE | TeR1 |
|---|---|---|---|---|---|---|
| IM | 1.00 | | | | | |
| ConE | 0.50 | 1.00 | | | | |
| ComE | 0.65 | 0.47 | 1.00 | | | |
| ManE | 0.44 | 0.36 | 0.53 | 1.00 | | |
| PerE | 0.51 | 0.33 | 0.38 | 0.47 | 1.00 | |
| TeR1 | 0.11 | 0.21 | 0.20 | 0.06 | 0.03 | 1.00 |
| TeR2 | 0.03 | 0.13 | 0.19 | 0.05 | 0.01 | 0.77 |
| TeR3 | 0.06 | 0.17 | 0.17 | 0.11 | 0.05 | 0.68 |
| DC7 | 0.04 | 0.04 | 0.09 | 0.23 | 0.13 | -0.11 |
| TAS | -0.07 | -0.16 | 0.03 | 0.23 | -0.05 | 0.07 |
| NES | 0.06 | 0.08 | 0.13 | 0.16 | 0.06 | 0.03 |
| Act | -0.14 | -0.20 | 0.00 | 0.12 | -0.07 | -0.01 |
| Out | 0.01 | -0.06 | 0.06 | 0.11 | 0.16 | 0.02 |
| TeR4 | -0.14 | -0.09 | 0.00 | 0.01 | -0.06 | 0.46 |
| TeR5 | -0.01 | -0.04 | 0.07 | 0.08 | 0.00 | 0.49 |

| | TeR2 | TeR3 | DC7 | TAS | NES | Act |
|---|---|---|---|---|---|---|
| TeR2 | 1.00 | | | | | |
| TeR3 | 0.85 | 1.00 | | | | |
| DC7 | -0.10 | -0.10 | 1.00 | | | |
| TAS | 0.00 | 0.00 | 0.18 | 1.00 | | |
| NES | -0.08 | -0.12 | 0.18 | 0.33 | 1.00 | |
| Act | -0.02 | -0.03 | 0.25 | 0.59 | 0.15 | 1.00 |
| Out | -0.03 | -0.07 | 0.18 | 0.38 | 0.31 | 0.42 |
| TeR4 | 0.35 | 0.20 | 0.03 | 0.11 | 0.05 | 0.15 |
| TeR5 | 0.42 | 0.29 | 0.02 | 0.27 | 0.01 | 0.17 |

| | Out | TeR4 | TeR5 |
|---|---|---|---|
| Out | 1.00 | | |
| TeR4 | 0.20 | 1.00 | |
| TeR5 | 0.15 | 0.60 | 1.00 |

KASVATUSALAN TUTKIMUKSIA – RESEARCH IN EDUCATIONAL SCIENCES

1. *ARTO JAUHIAINEN, RISTO RINNE & JUHANI TÄHTINEN (TOIM.)*
   KOULUTUSPOLITIIKKA SUOMESSA JA YLIKANSALLISET MALLIT
   (EDUCATIONAL POLICY IN FINLAND AND GLOBAL MODELS)

2. *RAIJA HUHMARNIEMI, SIMO SKINNARI & JUHANI TÄHTINEN (TOIM.)*
   PLATONISTA TRANSMODERNISMIIN – JUONTEITA IHMISYYTEEN, IHMISEKSI KASVAMISEEN, OPPIMISEEN, KASVATUKSEEN JA OPETUKSEEN
   (FROM PLATO TO TRANSMODERNISM – REFLECTIONS ON HUMANITY, GROWING UP TO BE A HUMAN BEING, LEARNING, EDUCATION AND TEACHING)

3. *ERKKI OLKINUORA, MIRJAMAIJA MIKKILÄ-ERDMAN, SAMI NURMI & MARIA OTTOSON*
   MULTIMEDIAOPPIMATERIAALIN TUTKIMUSPOHJAISTA ARVIOINTIA JA SUUNNITTELUN SUUNTAVIIVOJA
   (RESEARCH-BASED ASSESSMENT OF MULTIMEDIA LEARNING MATERIAL AND MAIN LINES OF PLANNING)

4. *SARI HUSA & JARMO KINOS*
   AKATEEMISEN VARHAISKASVATUKSEN MUOTOUTUMINEN
   (FORMATION OF ACADEMIC EARLY CHILDHOOD EDUCATION)

5. *REIJO BYMAN*
   CURIOSITY AND EXPLORATION: A CONCEPTUAL OVERVIEW AND STRUCTURAL MODELING