Validating developmental sequences in the domain of astronomy using latent trait techniques

<table>
<thead>
<tr>
<th>Item type</th>
<th>text; Thesis-Reproduction (electronic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Schwarz, Richard, 1955-</td>
</tr>
<tr>
<td>Publisher</td>
<td>The University of Arizona.</td>
</tr>
<tr>
<td>Rights</td>
<td>Copyright © is held by the author. Digital access to this material is made possible by the University Libraries, University of Arizona. Further transmission, reproduction or presentation (such as public display or performance) of protected items is prohibited except with permission of the author.</td>
</tr>
<tr>
<td>Downloaded</td>
<td>26-Feb-2016 20:25:44</td>
</tr>
<tr>
<td>Link to item</td>
<td><a href="http://hdl.handle.net/10150/277220">http://hdl.handle.net/10150/277220</a></td>
</tr>
</tbody>
</table>
INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Validating developmental sequences in the domain of astronomy using latent trait techniques

Schwarz, Richard D., M.A.
The University of Arizona, 1989
VALIDATING DEVELOPMENTAL SEQUENCES IN THE DOMAIN OF
ASTRONOMY USING LATENT TRAIT TECHNIQUES

by

Richard Schwarz

A Thesis Submitted to the Faculty of the
DIVISION OF EDUCATIONAL FOUNDATIONS AND ADMINISTRATION
In Partial Fulfillment of the Requirements
For the Degree of
MASTER'S OF ARTS
WITH A MAJOR IN EDUCATIONAL PSYCHOLOGY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1989
STATEMENT BY THE AUTHOR

This thesis has been submitted in partial fulfillment of the requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided the accurate acknowledge of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgement the proposed use of the material in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: [Signature]

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

[Signature]
John R. Berrigan
Professor of Educational Psychology
ACKNOWLEDGEMENT

I wish to extend my appreciation to Dr. Glenn Nicholson who provided me with the necessary foundation in statistics and his encouragement. I would also like to thank Dr. Charles Brainerd, whose scholarship serves as an example to us all, and for his unsurpassed advice and support. I wish to express my deepest appreciation to Dr. John Bergan, who has become one the most influential people in my life and for his encouragement in adversity and his humanity. Finally, I wish to thank Rose, my wife, for her patience and love.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>5</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>6</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>7</td>
</tr>
<tr>
<td>THEORETICAL BACKGROUND</td>
<td>15</td>
</tr>
<tr>
<td>Concept Learning Research Related to the Domain of Astronomy</td>
<td>15</td>
</tr>
<tr>
<td>Children's Use of Causal Reasoning Principles in the Domain of Astronomy</td>
<td>19</td>
</tr>
<tr>
<td>Piaget's Contribution</td>
<td>21</td>
</tr>
<tr>
<td>Recent Research in Childhood Animism</td>
<td>23</td>
</tr>
<tr>
<td>Mechanism, Causal Chains and Familiarity as Causal Principles</td>
<td>24</td>
</tr>
<tr>
<td>Item Response Theory</td>
<td>27</td>
</tr>
<tr>
<td>Item Response Models</td>
<td>31</td>
</tr>
<tr>
<td>Validating Developmental Sequences</td>
<td>33</td>
</tr>
<tr>
<td>Structural Equation Models</td>
<td>34</td>
</tr>
<tr>
<td>METHOD</td>
<td>35</td>
</tr>
<tr>
<td>Subjects</td>
<td>37</td>
</tr>
<tr>
<td>Instrument</td>
<td>38</td>
</tr>
<tr>
<td>Item Construction</td>
<td>38</td>
</tr>
<tr>
<td>Procedure</td>
<td>40</td>
</tr>
<tr>
<td>Validation of Hypothesized Hierarchies</td>
<td>41</td>
</tr>
<tr>
<td>RESULTS</td>
<td>42</td>
</tr>
<tr>
<td>Knowledge About Earth Model Comparison Results</td>
<td>42</td>
</tr>
<tr>
<td>Knowledge About Motion Model Comparison Results</td>
<td>44</td>
</tr>
<tr>
<td>Knowledge About Light Model Comparison Results</td>
<td>46</td>
</tr>
<tr>
<td>Confirmatory Factor Analysis Using LISREL</td>
<td>47</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>58</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>61</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

1. Hierarchy 1 Models and Estimated Parameters for Knowledge about Earth ........................................ 49
2. Hierarchy 1 Model Comparisons ........................................ 51
3. Hierarchy 2 Models and Estimated Parameters for Knowledge about Motion ........................................ 52
4. Hierarchy 2 Model Comparisons ........................................ 53
5. Hierarchy 3 Models and Estimated Parameters for Knowledge about Light ........................................ 54
6. Hierarchy 3 Model Comparisons ........................................ 55
7. Factor Loadings For One Factor ........................................ 56
8. Factor Loadings For Three Factors ........................................ 57
ABSTRACT

The present study was a systematic investigation of developmental skill sequences in the early science domain. Three developmental sequences in the area of astronomy were investigated; knowledge about earth, light and motion. Test items were developed reflecting developmental sequences based on the cognitive processes that are necessary for understanding each task. Data were collected from 1595 kindergarten children from six geographically diverse areas. Latent trait models were constructed to reflect the hypothesized developmental sequences by allowing discrimination and difficulty parameters to vary or by constraining them to be equal. Preferred models were obtained by statistical comparison with other models. The knowledge about light and motion were in the hypothesized developmental sequence. Astronomical events that contradicted personal experience, required causal explanations and consisted of extended causal chains were the most difficult for kindergarten children to understand. Investigations concerning the mechanism for conceptual change are necessary.
INTRODUCTION

There is a growing recognition that scientific literacy is an important component of economic growth and effective citizenship. Despite the recognized importance of scientific literacy, a recent survey (Miller, 1989) found that only one in eighteen adults had a general understanding of how science is conducted and how science impacts society. Taking astronomy as an example, only half of the respondents were familiar with the Big Bang theory. One in four adults did not know that light travels faster than sound. Less than half of the respondents knew that the earth orbited the sun in one year! One might conclude from these finding that science education in this country is fundamentally lacking in many respects. One meaningful long term solution might be to examine how scientific concepts develop such that effective classroom instruction can be designed to meet the needs of children and society.

The present study focused exclusively on young children’s knowledge and conceptions in the domain of astronomy. Astronomy was selected because few studies have been conducted that examine children’s ideas in this area. Secondly, astronomical events such as night and day, the reason for the seasons and tides play an integral part of everyday life. Thirdly, much like past societies and civilizations, how children view astronomical events reveals how they perceive their world and their place in it. Finally, the history of astronomy closely parallels the history of science. Modern science is said to have begun with Copernicus (Kuhn, 1957). It might be interesting to speculate to what extent children’s beliefs in astronomy parallel the view embodied by Ptolemy and the mythologies of ancient
A child's knowledge of the world is based primarily on his or her personal experience (Kuhn, 1989). Like the scientist, the child explores the world about him or her, constructs a model as a basis for comprehending it, and revises the model as new data becomes available. Most early mythologies of astronomy attempted to explain everyday events (e.g. night/day, the seasons, moon phases) using essentially the same information that is available to children. Much of this personal experience or sensory information, however, contradicts expert or scientific knowledge. Many phenomena particularly in physics directly contradict personal experience. Furthermore, daily experience affords few opportunities for a child or adult to confirm scientific theories.

Astronomical events that are directly congruent with experience are probably the easiest for a child to understand. An example of an astronomical event that is directly congruent with experience might be the association of the sun with the occurrence of daylight. Simply observing the event is usually enough to comprehend the phenomena. Unfortunately, there are very few astronomical events that have a simple one-to-one correspondence and are directly congruent with personal experience.

Most astronomical phenomena call for inferences beyond direct experience. For example, a child might erroneously believe that stars are only as big as a shoe box. The child's direct perception that stars are small points of light rather than unbelievably massive celestial objects is reinforced by direct experience. To understand that stars are massive objects requires children to be able to infer that things that far away often appear small.

Astronomical phenomena are also difficult to understand because they invariably
contradict everyday experience. The belief that the earth is flat or at least the part on which we are living is flat is one that children might reasonably harbor. Many of the child's experiences might reinforce this notion (i.e. a long walk on flat ground or a car ride across the midwest). A child might have heard that the earth is round but he or she might interpret this to mean round like a pancake is round thereby assimilating this information into a flat earth concept. The fact that earth is spinning on its axis while rapidly moving through space completely contradicts perception. Another example of an astronomical phenomena that contradicts experience is the gravitational force of massive celestial objects. A young child might believe that you have to be on top of the earth otherwise you might fall off into space if you were on the side or the bottom of the earth.

Many additional factors could also affect the difficulty of a task. Some tasks are more difficult because of language. Many of the child's ideas about phenomena are thought to arise from the natural use of language. For example, the colloquial use of force is very different from the way a physicist conceptualizes force. A novice describing a tossed coin might refer to the upward motion of the coin as a force because they confuse the scientific use of the term "force" with the colloquial expression "thrown with force". The problem for a child who is trying to learn a new concept is how does he or she separate the colloquial term from the scientific meaning.

Suppose children are asked to explain a difficult task that is just beyond their ability to grasp but not one that they are completely unfamiliar with (e.g. the reason for night and day). They might invent what appears to them to be a plausible explanation for the event. In other words, if the information is not in the repertoire of the child then the child might make up an answer, answer in an unpredictable way or not answer at all. Knowing what the
cognitive demands are for each task enables us to construct a developmental sequence based on the relative difficulty of the tasks in the progression. Task demands are those things a child has to know in order to answer a question correctly. When the tasks are related to each other then a developmental sequence is created. A developmental sequence shows the change in cognitive structure needed to understand a given task or concept. In order to assess the difficulty of tasks, it is necessary to hypothesize how children might conceptualize a given task and how cognitive change is effected. Concept learning and causal reasoning are two theoretical perspectives that are relevant to the interpretation of how children might conceptualize a task and how cognitive structures can change.

Concept learning encompasses studies of the qualitative differences between how experts and children or novices explain or view scientific phenomena in domain-specific areas and emphasize the importance of prior knowledge in subsequent learning (Eylon & Linn, 1988). Learning or conceptual change is only effected in the concept learning perspective when naive beliefs have been confronted and students have been taught coherent alternatives. Some naive beliefs are probably more prevalent than others at certain ages.

The examination of causal reasoning provides a framework for determining which occurrences are causes and which are effects and restricts the type of information that is deemed as sufficient and necessary for explaining an event. Causal principles can be used as a framework in attempting to understand reasoning processes. Children may incorrectly attribute the cause for physical events such as night and day to the actions of humans or endow inanimate objects with life. If adult reasoning is constrained by these principles and children’s reasoning is not, how does the child’s view differ from that of an adult.
Cognitive development can be conceptualized as being hierarchical where complex skills evolve from simpler skills (Gagne', 1962; Glaser, 1963; Bergan, 1985). Development from this perspective can be thought of as a qualitative change in the way an individual conceptualizes a concept, where individuals replace simple rules with more complex ones. The difficulty of a skill in a developmental sequence is defined by its task attributes. Paths can be defined and validated in relation to the item parameters in a latent trait model. Latent ability or developmental level may be thought of as a composite of the specific competencies that individuals possess, (Bergan, 1988). Early approaches to (Gagne, 1962) constructing developmental sequences assumed that there were superordinate and subordinate skills. Each skill in the sequence subsumed the skills below it. It is generally recognized now that development can take place in a variety of different ways. Furthermore, Gagne' (1962) made no use of the concept of ability.

For purposes of this study and to facilitate the examination of task attributes, astronomy was further broken down into three separate but not completely unrelated knowledge areas. The knowledge areas included in this study were knowledge about earth, light and the motion of celestial objects. Each of these knowledge areas is believed to comprise a fundamental aspect of astronomy and forms the basis for later learning of astronomical concepts.

Knowledge about earth was deemed to be fundamental to learning almost every other astronomical concept. The knowledge about earth skills in order of hypothesized difficulty were; identifying earth's shape, identifying earth as the place we live, identifying earth as a planet, knowing that earth is suspended in space, understanding earth's gravitational pull, and identifying earth's size versus other objects. Identifying earth's shape was hypothesized to be
the easiest skill because all that is required is the ability to recognize the earth. Knowing that earth is the place we live is more difficult because the child has to physically identify the earth and recognize that the earth is not a far off place. Since the term "earth" is often associated with the names of other celestial objects such as the other planets, a child might believe that earth is some distant object. Identifying earth as a planet requires the child to recognize the earth and have some understanding of what is meant by the term planet. Recognizing that the earth is suspended in space is more difficult than the preceding tasks for several reasons. In a child's everyday experience, he or she does not see objects suspended in space. On the contrary, all the solid objects a child sees are either held up by something or they are self-propelled such as airplanes. Young children probably do not understand the gravitational attraction of massive objects like the sun on objects like earth. A child who might have previously been able to identify the earth might be drawn off by distractors that suggest the earth is held up by strong rope or has ground underneath it. Understanding earth's pull is hypothesized to be more difficult than the preceding tasks because the gravitational force of massive celestial objects contradicts experience. A young child might believe that you have to be on top of the earth otherwise you might fall off into space if you were on the side or on the bottom of the earth. Finally, identifying earth's size versus other objects was hypothesized to be the most difficult task due primarily to children's incomplete understanding of scale especially the scale of astronomical objects. As stated previously, stars might be perceived to be small objects while houses or mountains are perceived to be much larger objects based on experience.

The knowledge about light assessed the perception of celestial objects. Skills in hypothesized order of difficulty were; identifying a star, knowing that the sun is the source of daylight, distinguishing which celestial objects can produce their own light, and recognizing
the cause for night and day. Identifying a star is a very easy skill. A child merely has to identify the object, but can also respond correctly based upon the object shape (i.e. star shaped). As mentioned previously, knowing that the sun is the source of daylight is directly congruent with a child’s personal experience, making this a fairly easy task. Distinguishing which celestial objects can produce their own light requires a child to know that the sun creates its own light. A child might believe that the moon creates its own light not realizing that moonlight is reflected light or that the earth somehow generates its own light. Since the magnitude of sunlight is often perceived to be very bright, it is less likely that a child would be confused about the source of daylight. Knowing the reason for night and day, on the other hand, constitutes a very difficult task. In order to answer this question correctly, a child would have to know that the sun lights the earth, the earth is round and turns on its axis creating night and day. Since a young child would have to integrate a number of concepts in order to understand this skill, there is ample opportunity for a child to introduce his or her own interpretation. For instance, a child might believe that the cause for night and day is due to human intervention or adopt a geocentric view. This skill is also difficult because it directly contradicts experience. To a child, the sun probably appears to move across the sky while the earth remains stationary.

The motion procedure assessed understanding of simple Newtonian mechanics. The skills in hypothesized order of were identifying which bodies go around the sun, identifying the orbit of the earth around the sun, recognizing the moon’s orbit around the earth. Identifying which bodies go around the sun required the child to determine which objects are acceptable celestial objects. The child had to distinguish between topological shapes, a flat earth versus the earth and the sun and recognize that stars do not orbit the sun. Identifying the orbit of the earth around the sun is hypothesized to be the next most difficult skill. In order to
understand this skill, a child would have to know that the natural movement for many objects is not a zigzag and that earth orbits the sun in an elliptical path. Gravity would not be exerting a constant force if an object moved in a zigzag or an angular fashion. Identifying that the moon orbits the earth is the most difficult skill in this procedure because it requires extensive knowledge on the part of the child and also directly contradicts experience. The moon’s orbit of the earth is particularly difficult since the moon changes phase and is only visible at night or is not visible at all. A young child may not even be aware that the moon exists or may believe that the different phases of the moon represent different objects. In addition, the child has to have some knowledge of orbits.
CHAPTER 2

THEORETICAL BACKGROUND

This chapter presents a literature review and an overview of the statistical models used to validate the hierarchical skill sequences. It includes a review of the concept learning and causal reasoning literature relevant to astronomy. In addition, the manner in which latent variable models are applied in validating hierarchical skill sequences are also described.

Concept Learning Research Related to the Domain of Astronomy

Studies from the concept learning perspective have primarily cataloged how learners of different ages conceptualize scientific phenomena. Concept learning research has consistently shown (Gunstone & White, 1981; Champagne & Klopfer & Anderson, 1980; Osborne & Gilbert, 1980) that prior to formal instruction children's and novice adult's conceptions of many physical phenomena are firmly embedded and highly resistant to cognitive change. Secondly, these conceptions are often inconsistent with the principles of expert science. These misconceptions or alternative conceptions occur even after instruction has taken place. Gunstone and White (1981) found that students who had completed two years of high school physics could not apply this knowledge to the everyday situations. In many instances, students knew the mathematical equations but applied them in inappropriate ways.

Champagne, Gunstone and Klopfer (1985) suggest that pre-instructional
conceptions of physics are often poorly differentiated due to the imprecision of everyday language. Many of the terms used by physicists are the same words that are used in everyday life (e.g. force, velocity, speed, acceleration). Secondly, students or children’s ideas may be imprecise because of their attempts to inappropriately formulate general rules from their everyday experiences (e.g. more force means more speed). Finally, Champagne et. al (1985) suggest that explanatory schemata used by children and novices are often situation specific. Many student’s explanations in one situation are directly contradicted by their explanations in another situation where the same principle should have been applied. For example, a child might not understand that an object in free fall and an object sliding down a slide are governed by the same set of constraints.

Chi, Feltovich, and Glaser (1981) used think aloud protocols of expert physicists and novices to characterize their differences in the solution of physics problems. Chi et al. (1981) found that these differences characterized themselves in a number of different ways. Experts used physical principles such as Newton’s laws or conservation of momentum in the solution of a problem. Novice problem solving was based more on the surface features (i.e. a block on an inclined plane or a spring) than a scientific principle. Chi et al. suggest that the expert’s thinking is hierarchically arranged along a dimension from abstract to concrete while the knowledge of novices or children lacks this integration. Kuhn (1989) suggests that the coordination of theories and evidence is a fundamental faculty of scientists and that these skills are not exhibited by children or novices. Kuhn (1989) found that a major strategy of children and novices was to reduce the inconsistencies in evidence by distorting it and attending to it in a selective manner in order to bring it in line with their personalized theory.
The science of astronomy, being a branch of the physical sciences, consists of a well developed and highly specialized body of knowledge. However, like many phenomena in the area of physics (Gunstone and White, 1981; Champagne, Klopfer, & Anderson, 1980; White, 1983), a child's understanding of astronomical events are often in conflict with expert knowledge. Dijksterhuis (1969) states that 

"to this day every student of elementary physics has to struggle with the same errors and misconceptions which then had to be overcome, and on a reduced scale, in the teaching of this branch of knowledge in schools, history repeats itself every year. The reason is obvious: Aristotle merely formulated the most commonplace experiences in the matter of motion as universal scientific propositions, whereas classical mechanics, with its principal of inertia and its proportionality of force and acceleration, makes assertions which not only are never confirmed by everyday experience, but whose direct verification is fundamentally impossible .." (pp.30)

di Sessa (1982) and White (1982) suggest that novices in physics hold theories that resemble those of Aristotle more than the theories of Newton. Contrasts between Aristotelian versus Newtonian thinking is widespread throughout the concept learning literature. Piaget (1930) likened the shift from Aristotelian mechanics to post-Galilean mechanics as being equal to the shift from concrete to formal operations.

To date, there have been several studies in the concept learning area that pertain to astronomy (Nussbaum and Novak, 1976; Klein, 1979; Sneider and Pulos, 1983). Klein (1982) was interested in the differences in the kinds of explanations given by Hispanic and Anglo-children in the area of astronomy. Klein assessed eight concepts; we live on earth,
the earth is round, the earth is in space, objects appear different from different perspectives, the sun is larger than the earth, night and day are caused by the rotation of the earth, sunrise occurs at different times at different geographic locations because of the earth's rotation and the earth makes one rotation every 24 hours. Cautious interpretation suggests that over half of the 24 second grade children in study believed that there was ground below the earth. The second grader's explanations for night and day ranged from precausal to the concept of night and day caused by the earth's rotation.

Nussbaum and Novak (1976) interviewed second grade children concerning their concept of earth. Based on their interviews, Nussbaum and Novak formulated five different notions of earth based on their sample. The first notion is that we live on a flat earth. The second notion of earth is that earth is spherical but objects on the southern hemisphere would fall off the earth. The distinction between notion two and notion three is somewhat unclear. To distinguish between them, Nussbaum and Novak drew water falling out of a bottle positioned at the south pole. The authors suggest that notion two child would respond that the water would fall to the ground underneath the earth. Notion three child would respond by saying that the water falls into space. Notion four children demonstrate some understanding of all of the elements of the earth concept. Notion five children have a satisfactory notion of the three aspects of the earth concept: (1) a spherical planet, (2) surrounded by space (3) and things are attracted to its center. Nussbaum and Novak (1976) concluded that learning the concept of earth is accomplished by a series of steps rather that a single conceptual leap. While Nussbaum and Novak's five notions of earth are intuitively appealing, they offered no validation for their developmental hierarchy.

The concept learning perspective can furnish us with some general
predictions on the nature of young children ideas and thinking. Prior to any formal instruction and often afterwards, children develop a view of the world that is often in disagreement with expert knowledge. These ideas are often strongly held since they appeal to common sense and everyday experiences. The expectation is that young children's concept of the earth and the solar system will be based on their everyday experiences rather than the Newtonian view of classical mechanics. For example, a child's understanding of motion might be stated loosely as; "The natural state of objects is rest." This can be contrasted with how the problem is stated in Newtonian terms; "Every body persists in its state of rest of uniform motion in a straight line unless it is compelled to change that state by forces impressed on it." Examples in the domain of astronomy might include the omission of gravity in explaining planetary motion or the behavior of objects on earth or the expression of a geocentric view of the solar system. A child expressing a geocentric view would not take into account the mass of objects when formulating this view. Everyday experience, however, would suggest that the earth is stationary while the sun moves across the sky. Piaget (1930) argued that the heliocentric view is completely beyond children's conceptualization of the earth-sun relation and that it would be quite pointless to attempt to teach young children this view. Many young children will be attracted by appeal of these alternative ideas or naive notions of astronomy.

Children's Use of Causal Reasoning Principles in the Domain of Astronomy

Our understanding of the physical world hinges, in part, on our ability to group phenomena into coherent units. One basis for organizing the world is our tendency to perceive cause and effect relationships. Our understanding of causal relationships underlies our perception of the physical world. Since causal reasoning is fundamental to how we
perceive the world, it provides an alternative way of examining children's ideas about astronomy. The development of causal reasoning structures has played an important role in investigating children's understanding of physical phenomena (Kun, 1978; Shultz & Ravinsky, 1977; Koslowski, 1981; Shultz, Pardo, & Altman, 1982). A causal schema denotes a relatively firm belief about how certain causes combine to produce certain effects (Kelly, 1972). Consideration of children's causal reasoning gives us additional insight into what they perceive as the cause for various astronomical phenomena.

Causation or the relation between events has been the subject of debate among philosophers for thousands of years. The importance of causality to our thinking is demonstrated by the numerous verbs of causation (e.g. to rust, to squeeze, to starve, to boil, to ignite, to stimulate, to shove, to develop, to shine) that we use in everyday life. Hume (1739) believed that we "organize" the world out of sensory impressions which convince us to accept the regularity of causal relationships. Hume in his 1739 book, *A Treatise on Human Nature* specified causality in terms of four principles. Hume's theories of causation largely consisted of an exposition of these four causal principles.

Priority states that causes must necessarily precede or occur simultaneous with their effects and presumes that events are caused. This causal principle has been alternatively referred to as temporal precedence. For example, the little boy pulled the dog's tail and this caused the dog to bite the little boy.

Cause and effects must also be contiguous in space and time or at the minimum linked by an intervening chain of contiguous events. This principle is called mechanism. For example, the explosion of gasoline in the piston of a car causes it to move
and constitutes the primary mechanism of the internal combustion engine. Likewise, the fusion of hydrogen atoms in the sun is the mechanism for the production of light and heat.

The causal principle of determinism asserts that events must be caused. In principle, at least, causes can be found for every event that occurs in the world (e.g. climatic change, economic collapse, or the changing of leaves).

Covariation or temporal contiguity states that causes and their effects must systematically covary. For example, it was a home run because the batter swung hard and connected with a fast ball. This principle is weaker than the preceding principles since causes can be discontiguous with their effects.

Many researchers have used these causal principles as a framework in attempting to understand reasoning processes. If adult reasoning is constrained by these causal principles and children's reasoning is not, the fundamental developmental question becomes how and when these principles develop.

Piaget's Contribution

Researchers, most notably Piaget (1963), have characterized children's causal reasoning as being fundamentally different from the causal reasoning of older children or adults. Adult causal reasoning was characterized by Piaget as being fundamentally naturalistic, mechanical and logical. In his tenure as the director of the Rousseau Institute, Piaget conducted a series of studies mostly during the 1920's on children's causal thinking. During this period Piaget coined three terms that he believed characterized precausal thinking;
finalism, artificialism, and animism. Piaget theorized that precausal principles emerged at precisely the same time since they all emerged from the same mechanism, namely egocentrism. Piaget maintained that these precausal principles emerged from egocentrism because young children do not readily distinguish between themselves and the external world.

Piaget defined finalism as the belief that every event, however insignificant, must have a specifiable cause. A child exhibiting finalism would not readily accept the view that events often can have random or accidental causes. Laurendeau and Pinard (1962) use the example of a young child explaining the cause for night as being useful for man's sleep as an instance of finalistic reasoning.

Artificialism is a concept that is closely allied to finalism. It refers to the child's belief that everything in the world has been created by human industry or God intentionally for human use and according to some grand design. Anthropomorphic action always precedes the event in question in artificialism. For example, a child might believe that the sun was made by man or God to give us light. Like finalism, artificialism makes no allowance for stochastic events. This view, however, is not one that is held expressly by children.

Of the three concepts, animism has no doubt provoked the greatest interest among researchers. Animism is defined as the attribution of life to inanimate objects. Piaget (1929/1969) suggested that young children initially assimilated physical events into cognitive structures evolved primarily for understanding human actions and relationships. Piaget broke the child's concept of animism down into four substages. The first stage
consists of children who believed that anything that affects humans is alive. For example, a child might believe the sun is alive because it makes us warm. Laurendeau and Pinard (1962) give the example of children who have noticed the invariant relationship between the sun's disappearance and nightfall and believe that the sun goes to bed just like humans. The next substage consists of children who believe that anything that is capable of movement is alive (e.g. a car or bike). The next substage is characterized by children who believe that anything that moves spontaneously is alive. A child at this stage might believe that the sun is alive because it moves spontaneously across the sky. Finally, in the last stage, life is confined to biological organisms.

Recent Research in Childhood Animism

Tunmer (1985) maintains that the development of causal reasoning depends on the child's ability to distinguish between physical events and the actions of humans. This viewpoint maintains that children frequently express ideas about the physical world in terms of cognitive structures that more accurately express human actions and social interaction. In other words, children frequently overgeneralize from social causality to physical causality. A fundamental characteristic of over generalizing from social to physical causality is the ascription of the verbs "want" and "know". Tunmer (1985) gives the following examples in accounting for the type of responses that children frequently give: An object floats because it doesn't want to sink; The sun goes across the sky because it wants to gives us warm sunshine; A marble rolls down hill because it knows you are there. Tunmer (1985) found support for Piaget's (1929) findings that while many older children no longer attributed life to inanimate objects, they continued to ascribe inanimate objects with intentionality.
Bullock (1985) suggests that Piaget overestimated the degree of animistic thinking in children by relying heavily on children's understanding of the word "alive" and by testing children about relatively complex and unfamiliar events. Bullock (1985) assessed pre-schoolers knowledge of animism by showing films that showed animate and inanimate objects moving in different ways. Children as young as three showed clear evidence that they could make the distinction between animate and inanimate objects. However, three-year-old children were not as consistent as four or five year-old children in perceiving this distinction.

It is conceivable that there is no general animistic factor or that children first attribute human qualities to all objects, as Piaget concluded. Animism may simply be a function of uncertainty or an unfamiliarity with the precise properties of an object. Gelman and Spelke (1981) argue that animate objects in the world are more easily distinguished and that children interact and are afforded more opportunities to explore which properties apply to animate objects and which do not apply. Children certainly are not afford many opportunities to interact with astronomical phenomena. Preliminary research conducted by the author suggests that young children are very receptive to animistic statements concerning causation in astronomy. However, the validity of language bound measures of causality are somewhat suspect when dealing with young children.

Mechanism, Causal Chains and Familiarity as Causal Principles

The causal principle that has considerable significance for the child's causal reasoning for physical events is mechanism. Mechanism is the assumption that causes produce their effects by the transfer of a causal force. This requires a child to look for logical
precursors that he or she knows or suspects in explaining a causal occurrence. Bullock (1979) used a Jack-in-the-Box and rolling marbles to examine children's mechanistic reasoning. Bullock coded the children's responses into three categories; mechanistic responses, non-naturalistic and phenomenistic. Non-naturalistic responses included no answers, animistic or magical explanations. Phenomenism was characterized by children who believe that co-occurrence of two events is sufficient for causation and whose thinking consisted of merely restating the events without connecting them causally. Bullock found that 90% of the 5-year-old children, 50% of the four-year-old children and 10% of the three-year-old children gave mechanistic explanations. The remaining children tended to give phenomenistic explanations. Only a few children gave non-naturalistic or animistic explanations.

Shultz, Pardo and Altmann (1982) examined three and five-year-old children use of transitive inference in causal chains. Causal chains (i.e. mechanism) are causal sequences in which the initial cause and terminal effect are mediated by an intermediate factor. In one procedure, Shultz et al. (1982) used light, the cause, to reflect off a mirror, the intermediate factor, to hit onto a target, the terminal factor. Children of both ages were capable of understanding that initial cause and eventual cause can be mediated by a third factor. The investigators concluded that three- to five-year-old children not only understand direct causation but also mediate causation. However, the five-year-old children were considerably more adept (P < .005) than the three-year-old children in their causal judgments.

Direct and mediate causation can be thought of as a comprising a developmental sequence. First, children learn to recognize causal chains containing two elements (i.e. direct causation). Next, children learn that causation can be mediated by an
intermediate factor. Mediate causation makes more cognitive demands than direct causation. As more links are added to the causal chain the more difficult it is to recognize the source of the causal chain and to map all elements of complex chains. Knowing that the sun is the source of daylight exemplifies direct causation. As a result this skill is easy for young children. Knowing the reason for night and day is far harder because there are several causal links for a child to map in order to understand this skill.

Berzonsky (1971) examined the extent that familiarity plays in children’s explanations of physical causality. Piaget’s (1930) theory of causal development had postulated that the reasoning a child employs is independent of the task. Berzonsky found support for the contention that young children resort to non-naturalistic explanations when confronted with unfamiliar objects or remote events. His subjects ranged in age from 6 years three months to 7 years 5 months. Three different types of tasks were used; verbal, water-level apparatus and teeter-totter tasks. In each of the three tasks children were asked, "What makes something happen?". The probes on the verbal task dealt either with remote (i.e. What makes the moon change shapes?) or familiar (i.e. What makes a car go?) objects. Berzonsky (1971) concluded that a child’s familiarity with an object determined that type of causal reasoning he or she used. Children were found to revert more often to non-naturalistic explanations when confronted with unfamiliar objects.

Several hypotheses emerge from these diverse studies in children’s causal reasoning ability. Piaget (1974) concluded that pre-operational children were incapable of understanding causal relationships. Studies less steeped in the Piagetian tradition have found that pre-operational children are capable of causal reasoning. This discrepancy can probably be accounted for by Piaget’s heavy reliance on verbal explanations. Young children
are probably capable of understanding numerous causal relations, but would probably be hard pressed to verbally explain the relational elements of a causal event. Piaget's verbal protocols underestimated children's causal reasoning abilities resulting in suppressed age norms for this class of tasks.

It is hypothesized by the author that tasks requiring cause explanations will be difficult for young children. Extensive item analysis might reveal that children of lower ability respond at a higher frequency to distractors that show animistic explanations for astronomical phenomena than children of higher ability. It is also reasonable to believe that variation in the familiarity of a task can affect the accuracy of causal attributions. Sophian and Huber (1984) suggest early causal reasoning is governed primarily by the concrete features of the events in question, while later reasoning can be characterized as more logical and involving greater reliance on abstract causal principles to interpret unfamiliar events. They hypothesize that early causal reasoning is based on concrete events while later reasoning involves greater reliance on abstract causal principles. It is conceivable that children's first causal knowledge may consist of specific events that have high spatial and temporal contiguity, are familiar, consist of short causal sequences, and are based on the concrete features on the events involved. Events such as these may well be in the causal domain of even the youngest children. Causal reasoning may well progress from simple explicit relationships to situations that have numerous elements that interact to form complex systems. Later reasoning may rely on concrete events in reasoning about more abstract and remote mechanisms such as planetary motion or gravity.

**Item Response Theory**

Item response theory (IRT) or latent trait theory as it is variously called is one of the most significant advances in psychometrics. Item response theory maintains
that an examinee's performance on a set of test items can be explained by one or more latent traits or abilities. These traits or abilities are constructs. Item response theory provides a stochastic method of linking examinee's responses to latent traits. Item response theory typically focuses on fitting a single latent variable based on the responses of examinees to a set of test items.

Item response theory origins can be traced back to Richardson (1936) Ferguson (1942) and Lawley (1943) but Lord (1952) is considered by some to be the father of IRT although many other researchers played significant roles in its development. Richardson (1936) provided a method for obtaining item response parameters estimates. Lawley (1943) refined and produced new procedures for parameter estimation. Lazarsfeld's (1950) work on latent and manifest variables, led to two branches of investigation, the latent class and latent trait models. Lord (1952) in his doctoral dissertation described the two-parameter normal ogive model. Rasch (1960) developed a one parameter model. Birnbaum (1968) substituted the logistic model for the normal ogive which made parameter estimate more tractable. Lord (1974) introduced his parameter estimation method called logist. Wright (1977) became the major proponent of the Rasch model and contributed to its understanding by many researchers.

The most common use of IRT is on multiple choice tests. If we were to administer parallel tests to a group of subjects, we might find that some subjects scored consistently low while others scored consistently high. We might explain this consistency in performance as a trait. Since this trait has no physical referent it is called a latent trait.

The range of ability for latent traits is usually designated by the Greek
letter theta, $\Theta$ with a mean of zero and a standard deviation of one. The shape of this
distribution is not known but it may be assumed to be normal. Theta determines where a
child lies on a dimension of ability. Theta can be thought of as a composite of specific
competencies that an individual possesses. The purpose of a test is to measure this
composite of competencies. Theta as it pertains to this study is the child’s conceptual
understanding of astronomy.

Latent trait models assume a unidimensional latent continuum. Local
independence constitutes a necessary condition for this assumption. The assumption of
local independence asserts that there is mutual independence among items for examinees at
any particular point on the ability continuum. When the assumption of local independence
is met, the probability of an item response vector for any point on the ability distribution can
be given by;

$$P_1 \times (1-P_2) \times P_3 \times (1-P_4) \tag{1}$$

$P_i$ is the probability that the examinee will respond correctly to item $i$, and $1 - P_i$ is the
probability that the examinee will respond incorrectly.

The item characteristic curve (ICC) links the probability of answering an
item correctly with the ability of the individual. The probability of an examinee responding
correctly depends on the shape of the ICC curve and is independent of the number of
examinees at a given ability level. This characteristic of item response models called the
invariance property is one of its most attractive aspects.
Each ICC function belongs to a family of functions of the same general type. The function must always rise because as ability increases the probability of getting an item correct should increase. The shape of a particular ICC function varies based upon three parameters; its intercept (i.e. difficulty) and slope (i.e. discrimination) and lower asymptote (i.e. guessing). The difficulty parameter is denoted by a "b", the discrimination parameter by an "a" and the guessing parameter is designated by a "c".

The inflection point of the logistic ogive is midpoint between the lower and upper asymptotes and is called the difficulty or b-parameter. In other words, the b-value is located at the point on the ability scale where the slope of the ICC is a maximum. The b-parameter represents the location on the ability scale where an examinee has a \( \frac{1 + c}{2} \) probability of correctly answering item i. A negative or low b-value represents an easy item while a high while a positive b-value represents a more difficult item. As the ICC function is shifted to the right, the b-parameter assumes a higher value. When the b-parameter is large the probability decreases that an individual with low ability will answer the item correctly.

The discrimination index, the a-parameter, signifies the slope of the relationship between the probability of responding correctly to a test item and the ability of the individual. A steep slope implies that the discrimination of the test item is greater although the range of its discrimination is more restricted. As the discriminating power of an item improves, the a-parameter acquires a higher value and the slope becomes steeper. The a-parameter reflects the speed with which the probability of a correct response changes as theta increases. A low or gradual slope means that an item is discriminating over a wide range of thetas but does not differentiate very well at any particular ability level.

If an examinee is of low ability and an item is difficult, he or she will
have to guess. The guessing parameter or \( c \)-parameter corresponds to the lower asymptote of the ICC curve and coincides with an examinee of low ability guessing the correct response. Lord (1974) suggests that estimates of guessing are frequently less than we might expect because individuals of low ability are often attracted to one or more attributes of distractors.

**Item Response Models**

The most common latent trait models are the one-, two, and three parameter models. The most straightforward latent trait model is the Rasch or one-parameter model. The one-parameter model is a special case of the three-parameter model.

George Rasch, a Danish mathematician, developed the one-parameter logistic model independently of other item response models in the early sixties. The ICC curve for the Rasch model is a one-parameter logistic. The Rasch model constrains the slope (e.g. discrimination) to be one, guessing is assumed to be zero while the difficulty parameter is allowed to vary.

The equation of the item characteristic curve for the one parameter model can be written as

\[
P(x_j = 1 | \Theta) = \frac{1}{1 + e^{(\Theta - b_j)}}
\]

The \( P(x_j = 1 | \Theta) \) corresponds to the probability that an examinee with ability (theta) will answer item \( j \) correctly. The difficulty parameter, \( b_j \), represents the location on the ability scale where an examinee has a \((1 + c_j)/2\) probability of answering item \( i \) correctly.
The Rasch model is attractive to practitioners due to its simplicity and economy. Since this model has fewer parameters to estimate, it is easier to work with and there are fewer problems with parameter estimation. However, the Rasch model in many cases may not provide an acceptable fit for the data, (Birnbaum, 1968; Hambleton & Traub, 1973; Lord, 1968; Ross, 1966; Traub, 1983). Most criticisms of the Rasch model center around the appropriateness of the two assumptions that this model makes. Many individuals would disagree with the assumption that guessing plays no role on test taking behavior during multiple choice tests. Second, few researchers would disagree with the notion that items differ in the degree which they correlate with the underlying trait.

The two parameter logistic model assumes that items differ in difficulty and in the extent to which they discriminate between individuals. The guessing parameter is fixed at zero. As a result, the lower asymptotes of the two-parameter ICC curves are zero. Consequently, low-ability individuals have almost no chance of making a correct response to difficult items. The equation for the two-parameter model can be given as:

\[
P(x_j = 1|\Theta) = 1 + e^{a/(\Theta - b)}
\]

The three parameter logistic model was developed by Birnbaum (1968). The three parameter model can be obtained from the two-parameter model by adding the guessing parameter. The mathematical form of the three-parameter logistic curve can be given as:

\[
P(x_j = 1|\Theta) = C + 1 + e^{a/(\Theta - b)}
\]
Lord (1980) includes $D$, a scaling factor of 1.7, in order to maximize the similarity between the logistic and normal curves.

**Validating Developmental Sequences**

Latent ability or developmental level score may be conceptualized as the composite of the specific abilities or competencies that individuals possess (Bergan, 1988). A developmental sequence is a set of ordered competencies, (Bergan, 1988). Developmental sequences are defined by the characteristics of each task. The idea of constructing assessment instruments that empirically validate developmental sequences can be traced to criterion-referenced assessment (Glaser, 1963) and continues to be linked to this assessment approach. Recent advances in latent trait models (Bergan, 1988; Thissen & Steinberg, 1988), have made it possible to validate developmental sequences empirically by referencing a child’s position using test items that have been ordered by difficulty.

Until recently, latent trait models did not provide a method to test the validity of hypotheses about the ordering of skills within a knowledge area (Bergan, 1988). Suppose that two items are nearly alike in difficulty. Previously, there was no way to use item response models to determine whether there would be any significant decrease in the fit of a model when two items were constrained to be equal. Thissen (1986) developed a program called **MULTILOG** that was applicable to a broad range of problems and allowed item parameters to be constrained. By placing restrictions on the difficulty and discrimination parameters, (Bergan, 1988) was able to test a set of hierarchically related latent trait models. Two models are said to be hierarchically related if one model contains all the parameters that
the other model contains plus one additional parameter.

Restrictions on the difficulty and discrimination parameters are essential to testing different hypotheses regarding developmental sequences. The difficulty parameter plays a central role in defining developmental sequences. The ordering of a developmental sequence is based on the relative difficulty of the tasks. Constraining the difficulty parameters permits hypotheses about the difficulty sequencing of items to be tested. For example, one of the simplest models could be examined is two items that are ordered by difficulty. One model would constrain the two items to be equal while the other model would allow them to vary. The model that allows the two items to vary contains one more estimated parameter and has one less degree of freedom than the model with the equality constraint. Since the models are hierarchical, comparisons between the two models can be made. Accepting the constrained model says that the two items are equal in difficulty and has the added benefit of being more parsimonious than the other model. Accepting the model that allows the items to vary suggests that the two items differ in difficulty.

Models constraining the discrimination parameter may be constructed in the same fashion as models that constrain difficulty. Restrictions on the a-parameter are of interest because they allow us to determine whether the discriminating power of the items are the same throughout the range of abilities. Constraining the slope indicates that the probability of getting an item correct is preserved throughout the range of abilities. When item slopes vary, there is the possibility of a reversed item difficulty for persons of different ability. In these situations, determining the difficulty of items becomes complicated.

The likelihood ratio chi-square statistic is used to test the fit of the
model to the data. The degrees of freedom are obtained by subtracting the number of estimated parameters plus one from the number of score patterns with score counts greater than zero (Bock & Atkin, 1981; Bergan, 1988).

Two models are said to be hierarchical, when one model contains all the estimated parameters plus one additional estimated parameter. The model with the fewer degrees of freedom is subtracted from the model with the larger number of degrees of freedom. The resulting $\chi^2$ can then be referenced to the chi-square distribution to determine the preferred model. A model with fewer estimated parameters, hence more degrees of freedom, is considered to be more parsimonious. A preferred model is one that is more parsimonious than the other models while providing an adequate fit for the data. The selection of a preferred model is the objective of hierarchical model testing.

Structural Equation Models

In addition to latent trait models, LISREL 7 (Linear Structural Relations; Joreskog and Sorbom, 1988) can be used to factor analyze test items. The basic goal behind factor analysis is to find one or more latent factors that are fewer in number than the observed variables. These latent factors are hypothesized to account or explain the intercorrelations among the variables.

There are three basic structural equations within the Lisrel model: (1) the full structural equation model and (2) the measurement model for $x$ and (3) the measurement model for $y$. LISREL 7 makes fundamental distinctions between independent and dependent variables in addition to latent and manifest variables. The measurement model
specifies two kinds of observed variables, exogenous variables denoted as x variables and endogenous variables denoted as y variables. The measurement model for x independent variables was used in this study and can be expressed as:

\[ x = \Lambda_x \xi + \delta \]  \tag{5}

The x is an observed variable that imperfectly represents the latent variable. Lambda represents the paths from the latent to the observed variables while theta delta is the variance-covariance matrices of residuals.

Hierarchical comparisons (Hayduk, 1987) between various structural equations models are made in the same manner as the latent trait models. Models are evaluated by statistically comparing the fit offered by one model against the fit afforded by alternative models. Confirmatory factor analysis also serves as a test of unidimensionality. Unidimensionality is an important assumption of latent trait theory. Both latent trait models and structural equation modeling are techniques that allow us to validate developmental sequences and determine what the factor structure is for a latent trait.
CHAPTER 3

METHOD

The present study was part of a larger investigation following Head Start children into kindergarten. The study was conducted in six different sites: Arizona, California, New Mexico, Iowa, Louisiana, and Mississippi; which included seven school districts and 21 schools. Seven schools were in rural areas and 14 schools were in urban areas. Rural areas were defined as having less than 2500 inhabitants, and urban areas were specified as populations greater than 2500. The number of children in each site was not equivalent. The teachers who participated were certified and had comparable levels of experience and education. The teachers were each given a brief training on how to administer the test.

Subjects

The subjects were 1595 public school kindergarten children. Forty-five percent of the children were male and 55% were female. The mean age of the children was 68.81 months. The sample was composed of the following ethnic groups: Blacks (26%); Hispanic (15%); Caucasian (42%); Asian (8%); Native American (15%); and other ethnicities (.8%).
Instrument

The assessment instrument was the Measurement and Planning System -Kindergarten Scale (MAPS-K). MAPS-K is a path-referenced assessment (Bergan, 1985) device that links measurement theory to developmental phenomena. Path-referenced is based on the notion that learning and development involve sequential changes in ability that reflect successively higher levels of cognitive functioning or ability (Gagne', 1962; Bergan, 1985). Path-referenced assessment uses latent trait techniques to "reference" a child's ability to a position on a developmental path. Thirteen questions off the science scale dealing with astronomy were examined in this study.

Item Construction

Thirteen astronomy questions were written for inclusion on the nature and science scale. Items were developed to reflect hypothesized developmental sequences. Each of the questions was associated with a knowledge area. Each knowledge area was a hypothesized set of processes or actions performed on objects to achieve a goal (Bergan, 1988). The three knowledge areas were developed within the domain of astronomy: knowledge about earth, knowledge about planetary motion, and knowledge about light.

These knowledge areas can be characterized by demand attributes which are associated with what the child might need to know in order to answer a question correctly. They indicate the special cognitive requirements that are imposed on the child, detail the significant characteristics of a procedure and further delineate the task into its component parts. Task attributes affecting difficulty were identified for each procedure. Task attributes indicate what skills are required to respond correctly to an item and decompose the
item into its constituent tasks.

Each skill is assigned an item code that references the test item to the appropriate skill. The skills and item codes that pertain to each question are listed below;

Knowledge about Earth;

1) CLASCEL5 Identifying Earth's shape.
2) CLASEL13 Identifying earth as the place we live.
3) CLASCEL19 Identifying earth as a planet.
4) CLASCEL12 Knowing that earth is suspended in space.
5) CLASCEL10 Understanding earth's gravitational pull.
6) CLASEL11 Identifying Earth's size versus other objects.

Knowledge about Light

1) CLASCEL1 Identifying a star
2) CLASCELZ2 Knowing that the sun is the source of daylight.
3) CLASCEL9 Distinguishing which celestial objects can produce their own light.
4) CLASCEL8 Recognizing the cause for night and day.

Knowledge about Motion;
1) CLASCELZ6 Identifying which bodies go around the sun.
2) CLASCELZ3 Identifying the orbit of the earth around the sun.
3) CLASCEL4 Recognizing the Moon's orbit around the earth.

Procedure

Site managers were employed at each location and administered the test with the help of local school personnel. Each scale was given verbally to one or more children in a single setting on different days. The test was not timed. The children were asked to locate each item using an item locator; and were instructed to bubble in the correct response underneath the appropriate foil. The children were given several practice items prior to the actual administration of the test.

The score sheets were mailed back to the University of Arizona by the site manager and were scored using Thissen's (1986) MULTILOG program. Correct and incorrect responses were transformed dichotomously into ones and zeros, respectively. The responses were then transposed into contingency counts. Item characteristics were obtained for each item using MULTILOG (Thissen, 1986). MULTILOG produces standard scores with a mean of zero and a standard deviation of one. A modified three-parameter latent trait model (Bock & Akin, 1981) was used to validate the earth and light developmental sequences. The difficulty and discrimination parameters were allowed to vary across subjects while a
pseudo-guessing parameter was held constant across examinees. A two-parameter model was used to validate the motion procedure. The difficulty parameter estimated under the model was used to sequence the skills assessed and to estimate the probability that the child had learned the skill reflected in the test item.

Joreskog and Sorbom's, (1988) LISREL 7 program was used to factor analyze the thirteen test items. LISREL 7 has a preprocessing package call PRELIS that was used to produce the polychoric and asymptotic covariance correlation matrix from the dichotomously scored data. The weighted least squares (WLS) method (Joreskog, 1988) was used since the data was dichotomous. WLS is the recommended approach for obtaining unbiased parameter estimates.

Validation of Hypothesized Hierarchies

Models were evaluated by statistically comparing the fit offered by one model against alternative models. The easiest model (M1) to conceptualize is one in which the $a_j$ and $b_j$ parameters are allowed to vary. This model suggests that each item differs from the other items regarding the $a_j$ and $b_j$ parameters. Recall that two models are said to be hierarchically related when one contains all the estimated parameters, plus one additional parameter. The next model (M2) might restrict all the $a_j$ parameters to be equal while allowing the $b_j$ parameters to vary. M2 is the more parsimonious model since it estimates fewer parameters. M2 is the preferred model if it significantly improves on the fit afforded by M1. Numerous models that constrain various combinations of $a_j$ and $b_j$ parameters can be hypothesized.
Chapter 4

RESULTS

The results of the hierarchical model comparisons for the skill sequences are shown in Tables 1 through 6. The estimated item parameters, the degrees of freedom and the likelihood ratio statistic are shown by model by Tables 1, 3, and 5. In addition, the Chi-square statistic is given for the difference between two models, the resulting degrees of freedom and the fit of the model comparison are by shown by Tables 2, 4, and 6. Factors loadings for a one factor model and a three factor model can be seen in Tables 7 and 8.

Knowledge About Earth Model Comparison Results

Tables 1 and 2 show the item parameters estimates and results from model comparisons for knowledge about earth. The knowledge about earth procedure included the following items in order of hypothesized difficulty; identifying earth's shape, identifying earth as the place we live, identifying earth as a planet, knowing that earth is suspended in space, understanding earth's gravitational pull, and identifying earth's size versus other objects. The first model, (M1), examined was a modified three parameter model that held the guessing parameter constant while allowing the difficulty and discrimination parameters to vary. Degrees of freedom are obtained by subtracting the number of estimated parameters plus one for the sample size from the number of score patterns with observed
counts greater than 0. Thirteen parameters were estimated under this model, six $a_j$, six $b_j$, and one $c_j$ parameter. Thus, there were $50 - (13 +1) = 36$ degrees of freedom for this model. The likelihood ratio statistic was 38.1 for this model and provided a very good fit for the model ($p > .50$).

The second model, (M2), imposed constrains on all the item parameters. All six slopes and six difficulty parameters were constrained to be equal. The likelihood ratio was 1341.7 for this model with 46 degrees of freedom. This model does not provide a good fit for the data ($p < .01$). The difference Chi-square for the comparison of M1 with M2 was 1303.6 with 10 degrees of freedom indicating that M1 significantly improved ($p < .01$) on the fit afforded by M2. Since M1 offered a significant improvement over M2, M1 is the preferred model. This model comparison suggests that the items slopes and difficulty parameters are not equal for these items.

The third model, M3, constrained all the slope parameters to be equal while the difficulty parameters were allowed to vary. The likelihood ratio was 44.6 with 41 degrees of freedom. This model provided an acceptable fit for the data ($p > .25$). The difference Chi-square for the comparison of model M1 with M3 was 6.5 with 5 degrees of freedom ($p > .10$). Model 1 did not improve on the fit afforded by M3. Since M3 provides an acceptable fit for the data and is more parsimonious, M3 is still the preferred model. The acceptance of M3 suggests that the items differ only by their difficulty.

Models 4, 5, 6, and 7 placed restrictions on the difficulty parameters designed to test hypotheses about the ordering of items according to their difficulty. This model constrained slope and difficulty parameters for items 4 and 6 to be equal.
likelihood ratio was 367.0 with 38 degrees of freedom, (p < .01). This model did not offer a very good fit for the data. Model 5 constrained slope and difficulty parameters for items 2 and 3 to be equal. The likelihood ratio was 40.8 with 38 degrees of freedom, (p > .50). This model offered a very good fit for the data. The difference Chi-square for the comparison of M3 with M5 was 3.8 with 3 degrees of freedom indicating that M5 did not significantly improve (p > .10) on the fit afforded by M3. M3 is the more parsimonious model and is preferred over the other models. Model 6 constrained slope and difficulty parameters for items 4 and 5 to be equal. The likelihood ratio was 42.1 with 38 degrees of freedom (p > .50). This model also provided an excellent fit for the data. Model 1 was hierarchical to model 6. The difference Chi-square for the comparison of M1 with M6 was 4.0 with 2 degrees of freedom. Model 1 did not offer a significant improvement in fit over M6 (p > .10). Model 6 was also hierarchically related to M3. The difference Chi-square was 2.5 for this comparison with three degrees of freedom (p > .10). Model 6 failed to improve on the fit provided by model 3. Model 7 constrained the slope and difficulty parameters for items two, three, four, and five to be equal. The likelihood ratio was 446.1 with 42 degrees of freedom. This model provided an unacceptable fit for the data (p < .01).

Model 3 is the preferred model. Since the $c_j$ and $a_j$ were equal, this is a one-parameter or Rasch model because the items differed only in difficulty. The skill identifying the earth's shape proved to be much more difficult than hypothesized. Knowing that earth is suspended in space was slightly more difficult than understanding earth's gravitational pull.

Knowledge About Motion Model Comparison Results
Tables 3 and 4 show the item parameters estimates and results from model comparisons for knowledge about motion. The skills in hypothesized order were; identifying which bodies go around the sun, identifying the orbit of the earth around the sun, recognizing the moon's orbit around the earth. The first model, (M1), examined was a two parameter model that allowed the difficulty and discrimination parameters to vary. Six parameters were estimated under this model, three $a_j$ and three $b_j$ parameters. Thus, there was one $(8 - (6+1) = 1)$ degree of freedom for this model. The likelihood ratio statistic was 7.6 for this model. Model 1 does not provide an acceptable fit for the model ($p < .01$).

The second model, (M2), imposed constrains on all the item parameters. Three discrimination and three difficulty parameters were constrained to be equal. The likelihood ratio was 483.2 for this model with five degrees of freedom. Model 2 does not provide a good fit for the data ($p < .01$). The difference Chi-square for the comparison of M1 with M2 was 475.6 with 4 degrees of freedom indicating that M1 significantly improved ($p > .01$) on the fit afforded by M2. Since M1 offered a significant improvement over M2, M1 is the preferred model. This model comparison suggests that the items slopes and difficulty are not equal for these items.

The third model, M3, constrained all the slope parameters to be equal while the difficulty parameter was allowed to vary. The likelihood ratio was 10.4 with 3 degrees of freedom and provided an unacceptable fit for the data ($p < .025$). The difference Chi-square for the comparison of model M2 with M3 was 472.8 with 2 degrees of freedom ($p < .01$). Model 3 improved on the fit afforded by M2. This suggests that the items differ in difficulty. Since M1 and M3 were hierarchically related, these two models can be compared. The difference Chi-square for the comparison of M1 with M3 was 2.8 with 2 degrees of
freedom. Model 1 did not significantly improve (p > .25) on the fit afforded by M3. Neither M1 or M3 provide an adequate fit for the data.

Since M3 has more degrees of freedom than M1 and is therefore more parsimonious, M3 is the preferred model even though it does not provide an adequate fit for the data. Since the $a_j$ parameters were equal for this model, this is a one-parameter or Rasch model because the items differed only in difficulty. The items were in order of hypothesized difficulty.

Knowledge About Light Model Comparison Results

Tables 5 and 6 show the item parameters estimates and results from model comparisons for knowledge about light. The skills in hypothesized order of difficulty were: identifying a star, knowing that the sun is the source of daylight, distinguishing which celestial objects can produce their own light, and recognizing the cause for night and day. The first model, (M1), examined was a modified three parameter model that held guessing to be constant while allowing the difficulty and discrimination parameters to vary. Nine parameters were estimated under this model, four $a_j$, four $b_j$ and one $c_j$ parameter. Thus, there were (16 - 9+1) six degrees of freedom for this model. The likelihood ratio statistic was 15.6 for this model. Model 1 does not provide an acceptable fit for the data (p < .05).

The second model, (M2), imposed constrains on all the item parameters. All four slopes and four difficulty parameters were constrained to be equal. The likelihood ratio was 2332.3 for this model with 12 degrees of freedom. Model 2 does not provide a good fit for the data (p < .01). The difference Chi-square for the comparison of M1
with M2 was 2316.7 with six degrees of freedom indicating that M1 significantly improved (p < .01) on the fit afforded by M2. Since M1 offered a significant improvement over M2, M1 is the preferred model. This model comparison suggests that slope and difficulty parameters are not equal for these items.

The third model, M3, constrained all the slope parameters to be equal while the difficulty parameters were allowed to vary. The likelihood ratio was 20.8 with nine degrees of freedom and provided an unacceptable fit for the data (p < .025). The difference Chi-square for the comparison of model M2 with M3 was 2311.5 with three degrees of freedom (p < .01). Model 3 improved on the fit afforded by M2. This suggests that the items differ in difficulty. Since M1 and M3 are hierarchically related, these two models can be compared. The difference Chi-square for the comparison of M1 with M3 was 5.2 with three degrees of freedom. M1 did not significantly improve (p > .10) on the fit afforded by M3. Neither M1 nor M3 provide an adequate fit for the data.

Since M3 has more degrees of freedom than M1 and is therefore more parsimonious, M3 is the preferred model even though it does not provide an adequate fit for the data. Since the $c_j$ and $a_j$ were equal for this model, this is a one-parameter or Rasch model because the items differed only in difficulty. The items were in order of hypothesized difficulty.

**Confirmatory Factor Analysis Using LISREL**

Confirmatory factor analysis serves as a test of unidimensionality for latent trait models. As mentioned previously, unidimensionality is an important assumption of
latent trait theory. Table 7 shows the factor loadings for a one factor model. The chi-square statistic is used as a goodness-of-fit measure in LISREL. The chi-square for the one factor model was 69.75 with 65 degrees of freedom and provided an excellent fit ($p=.321$) for the model. A three factor model was also investigated. The three factors were the knowledge about earth, light and motion. The factor loadings for this model can be seen in Table 8. The chi-square statistic for the three factor model was 64.46 with 62 degrees of freedom. The three factor model also provided an excellent fit ($p=.391$). Since these models are hierarchically related, they can be compared. The difference chi-square between the one factor and three factor model was 5.29 with three degrees of freedom. The three factor model failed to improve over the fit afforded by the one factor model. The one factor model is the preferred model and fulfills the assumption of unidimensionality.
Table 1
Hierarchy 1 Models and Estimated Parameters for Knowledge about Earth

<table>
<thead>
<tr>
<th>Model</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
<th>a6</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>.617</td>
<td>.385</td>
<td>.255</td>
<td>.298</td>
<td>.298</td>
<td>2.437</td>
</tr>
<tr>
<td>M2</td>
<td>.205</td>
<td>.205</td>
<td>.205</td>
<td>.205</td>
<td>.205</td>
<td>.205</td>
</tr>
<tr>
<td>M3</td>
<td>.342</td>
<td>.342</td>
<td>.342</td>
<td>.342</td>
<td>.342</td>
<td>.342</td>
</tr>
<tr>
<td>M4</td>
<td>.751</td>
<td>.426</td>
<td>.281</td>
<td>.392</td>
<td>.360</td>
<td>.392</td>
</tr>
<tr>
<td>M5</td>
<td>.607</td>
<td>.315</td>
<td>.315</td>
<td>.304</td>
<td>.299</td>
<td>1.326</td>
</tr>
<tr>
<td>M6</td>
<td>.618</td>
<td>.385</td>
<td>.255</td>
<td>.296</td>
<td>.296</td>
<td>2.421</td>
</tr>
<tr>
<td>M7</td>
<td>.856</td>
<td>.227</td>
<td>.227</td>
<td>.227</td>
<td>.227</td>
<td>7.682</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
<th>b5</th>
<th>b6</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>1.405</td>
<td>.333</td>
<td>.672</td>
<td>2.878</td>
<td>2.520</td>
<td>2.858</td>
</tr>
<tr>
<td>M2</td>
<td>2.925</td>
<td>2.925</td>
<td>2.925</td>
<td>2.925</td>
<td>2.925</td>
<td>2.925</td>
</tr>
<tr>
<td>M3</td>
<td>2.109</td>
<td>.298</td>
<td>.443</td>
<td>2.395</td>
<td>2.095</td>
<td>6.839</td>
</tr>
<tr>
<td>M4</td>
<td>1.448</td>
<td>.466</td>
<td>.855</td>
<td>4.392</td>
<td>2.505</td>
<td>4.392</td>
</tr>
<tr>
<td>M5</td>
<td>1.414</td>
<td>.468</td>
<td>.468</td>
<td>2.811</td>
<td>2.497</td>
<td>3.290</td>
</tr>
<tr>
<td>M6</td>
<td>1.404</td>
<td>.333</td>
<td>.671</td>
<td>2.708</td>
<td>2.708</td>
<td>2.861</td>
</tr>
<tr>
<td>M7</td>
<td>1.097</td>
<td>1.489</td>
<td>1.489</td>
<td>1.489</td>
<td>1.489</td>
<td>2.595</td>
</tr>
<tr>
<td>Model</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M6</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>df</td>
<td>36</td>
<td>46</td>
<td>41</td>
<td>38</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>$L^2$</td>
<td>38.1</td>
<td>1341.7</td>
<td>44.6</td>
<td>367.0</td>
<td>40.8</td>
<td>42.1</td>
</tr>
</tbody>
</table>
Table 2

**Hierarchy 1 Model Comparisons**

<table>
<thead>
<tr>
<th>Models</th>
<th>df</th>
<th>$x^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3, M1</td>
<td>5</td>
<td>6.5</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>M2, M1</td>
<td>10</td>
<td>1303.6</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>M3, M5</td>
<td>3</td>
<td>3.8</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>M5, M1</td>
<td>2</td>
<td>2.7</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>M1, M6</td>
<td>2</td>
<td>4.0</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>M3, M6</td>
<td>3</td>
<td>2.5</td>
<td>&gt;.10</td>
</tr>
</tbody>
</table>
Table 3

Hierarchy 2 Models and Estimated Parameters for Knowledge about Motion

<table>
<thead>
<tr>
<th>Model</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>.637</td>
<td>.187</td>
<td>.295</td>
<td>.107</td>
<td>1.246</td>
<td>3.698</td>
</tr>
<tr>
<td>M3</td>
<td>.329</td>
<td>.329</td>
<td>.329</td>
<td>.168</td>
<td>.729</td>
<td>3.313</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>$L^2$</td>
<td>7.6</td>
<td>483.2</td>
<td>10.4</td>
</tr>
</tbody>
</table>
Table 4

Hierarchy 2 Model Comparisons

<table>
<thead>
<tr>
<th>Models</th>
<th>df</th>
<th>$x^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2, M1</td>
<td>4</td>
<td>475.6</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>M2, M3</td>
<td>2</td>
<td>472.8</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>M3, M1</td>
<td>2</td>
<td>2.8</td>
<td>&gt;.01</td>
</tr>
</tbody>
</table>
Table 5

Hierarchy 3 Models and Estimated Parameters for Knowledge about Light

<table>
<thead>
<tr>
<th>Model</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>.301</td>
<td>1.154</td>
<td>.436</td>
<td>2.835</td>
</tr>
<tr>
<td>M2</td>
<td>.366</td>
<td>.366</td>
<td>.366</td>
<td>.366</td>
</tr>
<tr>
<td>M3</td>
<td>.601</td>
<td>.601</td>
<td>.601</td>
<td>.601</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
<th>df</th>
<th>L^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>-6.1</td>
<td>.09</td>
<td>.97</td>
<td>4.58</td>
<td>6</td>
<td>15.6</td>
</tr>
<tr>
<td>M2</td>
<td>25.25</td>
<td>25.25</td>
<td>25.25</td>
<td>25.25</td>
<td>12</td>
<td>2332.3</td>
</tr>
<tr>
<td>M3</td>
<td>-3.41</td>
<td>.13</td>
<td>.80</td>
<td>25.28</td>
<td>9</td>
<td>20.8</td>
</tr>
</tbody>
</table>
Table 6

Hierarchy 3 Model Comparisons

<table>
<thead>
<tr>
<th>Models</th>
<th>df</th>
<th>$x^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3, M1</td>
<td>3</td>
<td>5.2</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>M2, M1</td>
<td>6</td>
<td>2316.7</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>M3, M2</td>
<td>3</td>
<td>2311.5</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>
Table 7

Factor Loadings For One Factor (WLS)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASCEL1</td>
<td>.756</td>
</tr>
<tr>
<td>CLASCEL2</td>
<td>.631</td>
</tr>
<tr>
<td>CLASCEL6</td>
<td>.314</td>
</tr>
<tr>
<td>CLASCEL12</td>
<td>.339</td>
</tr>
<tr>
<td>CLASCEL13</td>
<td>.462</td>
</tr>
<tr>
<td>CLASCEL5</td>
<td>.348</td>
</tr>
<tr>
<td>CLASCEL19</td>
<td>.390</td>
</tr>
<tr>
<td>CLASCEL9</td>
<td>.305</td>
</tr>
<tr>
<td>CLASCEL3</td>
<td>.328</td>
</tr>
<tr>
<td>CLASCEL8</td>
<td>-.050</td>
</tr>
<tr>
<td>CLASCEL11</td>
<td>.116</td>
</tr>
<tr>
<td>CLASCEL4</td>
<td>.068</td>
</tr>
<tr>
<td>CLASCEL10</td>
<td>.456</td>
</tr>
</tbody>
</table>

*Note: Weighted Least Squares is denoted by WLS
Table 8

Factor Loadings For Three Factors (WLS)*

<table>
<thead>
<tr>
<th></th>
<th>EARTH</th>
<th>LIGHT</th>
<th>MOTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASCEL1</td>
<td>.000</td>
<td>.766</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL2</td>
<td>.000</td>
<td>.651</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL6</td>
<td>.000</td>
<td>.000</td>
<td>.482</td>
</tr>
<tr>
<td>CLASCEL12</td>
<td>.348</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL13</td>
<td>.471</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL5</td>
<td>.359</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL19</td>
<td>.397</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL9</td>
<td>.000</td>
<td>.314</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL3</td>
<td>.000</td>
<td>.000</td>
<td>.503</td>
</tr>
<tr>
<td>CLASCEL8</td>
<td>.000</td>
<td>-.051</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL11</td>
<td>.133</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>CLASCEL4</td>
<td>.000</td>
<td>.000</td>
<td>.150</td>
</tr>
<tr>
<td>CLASCEL10</td>
<td>.466</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Note: Weighted Least Squares is denoted by WLS
CHAPTER 5

DISCUSSION

The present study investigated developmental sequences in the domain of astronomy and examined three knowledge areas within astronomy. Developmental sequences were identified, described and validated for each knowledge area. These three areas included knowledge about earth, motion and light. In general, these results show that young children formulate some notions about celestial phenomena with little formal instruction. The assumption that children have no knowledge of a topic before formal instruction has taken place and are a vessel that can be simply filled with "expert" science is not supported by this study. Test item distractors that depicted naive conceptualizations of celestial phenomena were appealing to many children. Children found tasks that involve more complex phenomena with many causal links, such as the reason for night and day or recognizing the moon's orbit around earth, to be quite difficult. In addition, tasks that contradict their everyday experiences, such as the size of celestial objects or that earth is suspended in space, were also very difficult for young children to grasp.

Two of the developmental sequences that comprised knowledge about light and motion areas were in the hypothesized order (i.e. arranged by difficulty). The observed developmental sequence for knowledge about motion in order of difficulty were; identifying which bodies go around the sun, identifying the orbit of the earth around the sun, recognizing the moon's orbit around the earth. The skills in order of observed difficulty for knowledge about light
were; identifying a star, knowing that the sun is the source of daylight, distinguishing which celestial objects can produce their own light, and recognizing the cause for night and day. However, one major inversions and several minor inversions occurred in the knowledge about earth procedure. The knowledge about earth procedure included the following items in order of hypothesized difficulty; identifying earth's shape, identifying earth as the place we live, identifying earth as a planet, knowing that earth is suspended in space, understanding earth's gravitational pull, and identifying earth's size versus other objects. The task hypothesized to be the easiest (i.e. identifying the shape of earth) turned out to be more difficult than anticipated. Simply recognizing the earth versus other objects is probably a fairly easy task. However, the test item reflecting this task required children not only to distinguish that the earth is round but the shape of the continents as well. As might be expected, this was a very difficult task for young children. Knowing that earth is suspended in space was slightly more difficult than the item reflecting knowledge of earth's gravitational pull. However, model 6 constrained these two items to be equal and the model provided an acceptable level of fit. Therefore, these two items are nearly alike in both difficulty and discrimination value.

Some support was also found for Nussbaum and Novak's (1976) five notions of earth. However, Nussbaum (1979) extends these same notions up to the eighth grade. This sequence was validated using kindergarten children, who demonstrated competence with several of these concepts. The results of this study suggest that Nussbaum (1979) is over extending the validity of his various hypothesized notions of earth to much older children, thereby underestimating children's ability in this area.

Glaser (1984) suggests that the development of scientific thinking consists of a progression of partially correct theories within individual conceptual domains. The first task in elucidating
the development of scientific thinking is identifying and describing such sequences of understanding. Studies such as this one fulfill this category of research. The next step is the investigation of the mechanism for conceptual change. How are theories constructed and revised? Are paradigm shifts characterized by weak and strong restructuring as Carey (1986) suggests? Kuhn (1989) suggests that the notion of child-as-scientist does not adequately account for the mechanism of conceptual change and is fundamentally misleading since children and naive adults frequently adjust evidence to fit theories. Developmental research that reliably describes the conceptual change that occurs with knowledge acquisition and provides a mechanism for these changes is essential.
REFERENCES


