THE ASSESSMENT OF THE BIOLOGY ITEMS IN THE 1998-2001 SECONDARY SCHOOL INSTITUTIONS STUDENT SELECTION AND PLACEMENT TESTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ELA AYŞE KÖKSAL

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF SECONDARY SCIENCE AND MATHEMATICS
EDUCATION

APRIL 2002

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Tayfur ÖZTÜRK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science

Prof. Dr. Ömer GEBAN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science

Prof. Dr. Giray BERBEROĞLU
Supervisor

Examining Committee Members

Prof. Dr. Giray BERBEROĞLU

Prof. Dr. Ömer GEBAN

Assist. Prof. Dr. Hüsniye DEMİRCİOĞLU

Assist. Prof. Dr. Ceren TEKKAYA

Dr. Jale ÇAKIROĞLU

ABSTRACT

THE ASSESSMENT OF THE BIOLOGY ITEMS IN THE 1998-2001 SECONDARY SCHOOL INSTITUTIONS STUDENT SELECTION AND PLACEMENT TESTS

Köksal, Ela Ayşe

M.Sc., Department of Secondary Science and Mathematics Education

Supervisor: Prof. Dr. Giray BERBEROĞLU

April 2002, 96 pages

This study aims at assessing the Secondary School Institutions Student Selection and Placement Tests' biology items used between 1998 and 2001 with respect to cognitive processes and subject matters measured and gender performances across cognitive processes observed.

For this purpose, a three phases study was conducted. In the first phase, cognitive processes assessed by the items were evaluated with the aids of expertise in biology education. Second, the empirical study was conducted to assess the grouping of items in the biology tests used. Finally gender performances were evaluated across the sets of items whatever determined in the second phase of the study.

It is found that, the biology items indicated consistency in terms of empirically determined clusters with respect to the cognitive processes measured. Also, significant differences were found across the gender groups in the item clusters in favor of both female and male students.

Keywords: biology test, content specification, factor analysis, and gender difference.

ÖZ

1998-2001 ORTAÖĞRETİM KURUMLARI ÖĞRENCİ SEÇME VE YERLEŞTİRME SINAVLARINDAKİ BİYOLOJİ SORULARININ DURUM TESBİTİ

Köksal, Ela Ayşe

Yüksek Lisans, Ortaöğretim Fen ve Matematik Alanları Eğitimi Bölümü

Tez Danışmanı: Prof. Dr. Giray BERBEROĞLU

Nisan 2002, 96 Sayfa

Bu çalışma 1998-2001 yılları arasında yapılan Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştime Sınavlarında çıkan biyoloji sorularının ölçülen bilişsel süreçler ve konu alanları ve bilişsel süreçlerde gözlenen cinsiyet farklılıklarıyla ilgili durum tesbiti yapmayı amaçlamaktadır.

Bu amaçla, üç aşamalı bir çalışma izlenmiştir. İlk aşamada, sorularla ölçülen bilişsel süreçler biyoloji eğitimi alanındaki uzmanların yardımıyla değerlendirilmiştir. İkinci olarak kullanılan biyoloji testlerindeki soru gruplarını bulmak amacıyla deneysel çalışma yapılmıştır. Son olarak da öğrencilerin cinsiyetlere göre başarı durumları ikinci aşamada belirlenen soru gruplarında değerlendirilmiştir.

Biyoloji sorularının deneysel olarak belirlenen boyutlar bağlamında ölçülen bilişsel süreçler ile uyum gösterdiği bulunmuştur. Ayrıca soru boyutlarında cinsiyetler arasında kız ve erkek öğrenciler lehine önemli farklılıklar bulunmuştur.

Anahtar Kelimeler: biyoloji testi, içerik özellikleri, faktör analizi, cinsiyet farklılığı

To My Parents.

ACKNOWLEDGEMENTS

I primarily wish to express my special gratitude to Prof. Dr. Giray BERBEROĞLU for his valuable suggestions, guidance, and criticism in the realization of this study.

I also wish to thank my jury members Prof. Dr. Ömer GEBAN, Assist. Prof. Dr. Hüsniye DEMİRCİOĞLU and Assist. Prof. Dr. Ceren TEKKAYA, for their valuable remarks and advises. Thanks are also due to my parents for their endless support and patience and to Dr. Jale ÇAKIROĞLU for her patience, help and guidance throughout my thesis.

And thanks to my real friends without whose encouragement; this study could not have been accomplished.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

CHAPTER I

INTRODUCTION

The word "test" originally comes from the Latin "testum" which was a porous cup for determining the purity of metal. Later it was meant procedures for determining the worth of a student's effort or tools of the school efficiency movement as in the USA from 1911 to 1916 (Wiggins, 1989).

There are many usage of tests, such as instructional decisions of diagnosis and grading; guidance decisions in occupational, educational, and personal domains; administrative decisions of selection, classification, and curriculum planning; and research-related decisions, initial evaluation (placement); formative evaluation (prescription); and summative evaluation (attainment) (Gronlund, 1981, Mehrens & Lehmann, 1984, and Bloom, Hastings & Madaus, 1971 as cited in Millman & Greene, 1989, p. 336).

Among them, selection and placement are two important usage for formal schooling especially to give equal chance to each student to enter a further school type. The Secondary School Institutions Student Selection and Placement Test (Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı, SSISSPT) is one of the mentioned tests above. There are total 100 questions in the SSISSPT. It has 4 subtests: Turkish, Mathematics, Science and Social Studies. Its duration is 120 minutes. Since 1998, this test has been used to select and place students in various programs of high schools in Turkey, such as science lycee, Anatolian high schools, some programs of vocational high schools etc. For this reason, "it is said to have high-stakes, when significant educational paths or choices of an individual are directly affected by test performance, such as whether a student is promoted or retained at a grade level, graduated, or admitted or placed into a desired program" according to American Educational Research Association, American Psychological Association and National Council on Measurement in Education (Brennan, Kim, Wenz-Gross, & Superstein, 2001).

Sometimes test results are being used for both the national assessment of education in some countries and the international evaluation in many countries such as International Association for the Evaluation of Educational Achievement (IEA); the International Assessment of Mathematics and Science (IAEP); and OECD's International Educational Indicators Project (INES). As an example of the first usage, the National Assessment of Educational Progress (NAEP) has been continuing its studies about science education since 1969, the Sputnik's hurl (the first rocket sent into space) in the United States. The NAEP program were administered by the US

Department of Education but now conducted by the Educational Testing Service (ETS). Science tests in the NAEP program are distributed every four-five years to a nationally representative sample of students aged 9, 13, and 17. The science test is composed of life science, chemistry and nature of science at all three age groups. However, physics, including earth and space science, is conducted at the last two age groups. It measures students' knowledge of everyday facts, understanding of simple principles, application of basic information, analysis of procedures and data, and integration of specialized information related to science. Besides answering the test questions, students are also asked for background information about themselves, their families, and their schools. The main aim to use these types of tests is to investigate where American education is in science education and where it is going (See Rakow, Welch & Hueftle, 1984; Alexander & James, 1987 as cited in Askling & Härnquist, 1989; Askling & Härnquist, 1989; TIMSS, 2001).

Alike NAEP science test in USA, in Turkey, some science tests have been administered by the Educational Research and Development Directorate (ERDD) (Eğitimi Araştırma ve Geliştirme Dairesi, EARGED) of the Ministry of National Education (MONE) (Milli Eğitim Bakanlığı, MEB) to grades 5, 8, and 11 in curriculum laboratory schools (Müfredat Laboratuar Okulları, MLO). These tests are piloted to analyze items statistically, and review or eliminate them in order to increase the reliability of test and to give evidence to its validity. These steps make the tests standardized (EARGED, 1995, Çalışkan, 2000).

When compared to other sciences (physics and chemistry), although biology has been thought of a verbal and descriptive for many years, it has been appeared quantitative with the latest developments in genetics, ecology and taxonomy (Weir, 1972 and Herreid, 1977 as cited in Marsh & Anderson, 1989).

This trend in Biology, which is favoring advancements of quantitative skills, has been started to be implemented in teaching. For example, the Committee on the Undergraduate Program in Mathematics published a report in this context in 1970. The committee's panel on Mathematics for the Life Sciences recommended students to take at least two semesters of calculus, which was consisted of linear algebra and probability and statistics. In its report, The Royal Society-Institute of Biology's Biological Education Committee stressed biology teachers' out-of-date usage of mathematics terms and concept and the need to train them by mathematics teachers. The mathematical competencies required from students were also listed in this report (Marsh & Anderson, 1989).

Molitor and George (1976) indicated that the development of science process skills has been aimed in elementary education since 1960 in the USA. These skills are composed of collecting, analyzing, and acting upon data used in test questions. Observation is an activity of collecting data. It is use of senses and an ability to select information. Classification, inference, prediction, verification, and formulating of hypothesis are activities of data analysis. They are abilities to process information. Inference means the ability to make a judgment about the non-observable properties of an object or event based on the observable properties of that object or event. On the other hand, verification is the ability to recognize observations. Observations are necessary and sufficient to establish the validity of an inference. The verification of an inference requires that those properties that have been inferred be actually observed. The importance of science process skills thereafter resulted with the studies to develop new measures, unlike achievement and aptitude tests evaluating concept attainment and fact learning, to evaluate the students' level of attainment it.

There are two kinds of science process skill tests. One of them is to evaluate complex skills, i.e., the Sequential Tests of Educational Progress: Science, the TAB Science Test, the Questest, and the Balance Problem Test. The second one is to evaluate single skill rather than multi-trait, i.e., the Science Problem-Solving Skills Test (Molitor & George, 1976).

According to Piaget, the development of science process skills depends on cognitive development of students. In the process of adaptation to the environment, intelligence constructs cognitive structures given in advance, neither in human mind nor in the environment as it is perceived and organized. Although there are four stages of cognitive development, elementary school students can demonstrate the characteristics of the last three:

The first stage is called preoperational. It covers the period from 2-3 years to 7-8 years. In this stage, the student starts logical thinking and holds the idea of conservation of matter. For example if a glass of water poured into a cup, it is still same. But his/her thought process is based on perceptual clues, and he/she is unaware of contradictions and thinks that because soap is small, it floats and because a piece of iron is thin, it sinks.

Concrete operational period comes later. It can be seen in years from 7-8 to 12-14. Student develops logical thought linked to concrete objects and things independent from perceptual clues.

In the last stage that is called formal operational period, student begins dealing logically with multifactor situations and reasons from hypothetical to concrete situations. Students older than 14 can show the characteristics of this stage (Gredler, 1992).

Piaget's formal operational period (formal cognitive functioning) was described by Flavell as "*orientation towards problem solving: organizing data (combinatorial analysis), isolation and control of variables, the hypothetical, and logical justification and proof.*" (p. 211 as cited in Lawson & Wollman, 1976).

As the Educational Policy Commission points in 1961, the development of problem-solving processes is the main aim American Education. Known as 'rational powers', 'problem-solving processes' include recalling and imagining; classifying and generalizing; comparing and evaluating; analyzing and synthesizing; and deducing and inferring (Lawson & Wollman, 1976).

On the other hand, the previous elementary science curriculum, which was effective between the years of 1992-2001, concentrated on some learning objectives in Turkey. These are students' knowledge of facts, terminology, concepts, classifications, sequences, techniques and procedures, principles and laws, and theories; comprehension of liveliness, science process skills of observing; investigating description of observations; designing of experiment; analyzing data and generalization; application of information to daily life and ability to use materials as well as critical and creative thinking skills. New developments on communication, information and technology in the global conjecture resulted in restructuring of education in Turkey and the elementary science curriculum was renewed in 2000. The increasing amount of information made elementary curriculum, which was assumed to help students how to make science, emphasize on science process skills (MEB, 1995 and 2000).

Science process skills are one of the psychological constructs that some of the science tests aforementioned measure. For example, biology items within Science subscale of the Secondary School Institutions Student Selection and Placement Test (SSISSPT) measure 8th grade students' understanding of biology contents and higher order thinking skills (HOTS). Even though the test is designed to select and place students in the secondary school institutions, because of the bulk of the group taking the test, it is also possible to assess the level of biology achievement of the students at the beginning of the secondary education within the context of biology subscale of the science test. To assess both content and HOTS of biology items in a high stake test such as SSISSPT will help us to understand current status of biology education in Turkey in terms of student achievement.

In the present study, using the test forms and data obtained at the Secondary School Institutions Student Selection and Placement Test, the aim is to assess the content of the test forms with respect to cognitive processes measured, find empirical evidence for the existence of different cognitive processes measured by the test items, and investigate the gender differences across the different groups of items measuring different cognitive processes. Even though each item has its own cognitive and content specifications given by the item writers of the MONE, in this study additional evidence will be collected for the content specifications of the items by consulting biology education researchers. Similarly, clusters of items are studied by the MONE routinely for each test forms. However, these studies are based on the whole forms rather than the subcomponents. Thus, evaluating the Biology items separately will give more detailed information for this particular subcomponent. Gender differences are generally not observed at this grade level. However, no specific study was aimed at assessing gender differences for the subcomponents of the biology items.

CHAPTER II

ABOUT RESEARCH STUDY

2.1 – Purpose of the Study

The primary concern of this study is to assess the Biology items within the science subscale of the Ministry of National Education's (MONE) Secondary School Institutions Student Selection and Placement Test forms through the years of 1998 to 2001. This assessment will be in terms of

(1) Evaluation of content specifications (subject matters and cognitive processes) of the biology items,

(2) Factorial structure of the biology items, and

(3) Gender performances in different clusters of items in the test forms.

2.2 – Research Questions

1. What are the content specifications of the biology items within the science subscales of the 1998-2001 SSISSPTs?

a. Which subject matters are measured by the biology items within the science subscales of the 1998-2001 SSISSPTs?

b. Which cognitive processes are measured by the biology items within the science subscales of the 1998-2001 SSISSPTs?

c. Are there any differences in the subject matters of the biology items within the science subscales through 1998-2001 SSISSPTs?

d. Are there any differences in the cognitive processes of the biology items within the science subscales through 1998-2001 SSISSPTs?

2. What are the factorial structures of the biology items within the science subscales of the 1999-2001 SSISSPTs?

3. Are there any differences in the gender performances in different clusters of biology items in the 2000 SSISSPT?

2.3 – Significance of the Study

1. This study will investigate students' performance levels in different taxonomic stages of cognitive processes across gender groups in the biology items of SSISSPT.

2. This assessment will provide feedback to item writers in MONE for further testing attempts.

3. This work will also give information to the educators for carrying out comparative studies in line with some international assessment projects such as the Third International Mathematics and Science Studies (TIMSS) about Turkish students' achievement in Biology. The results might have some impacts on curriculum development attempts in science in Turkey.

2.4 – Definition of the Key Terms

Cognitive Processes are (1) knowledge, (2) comprehension, (3) solving problem, and (4) science process (scientific inquiry, conducting inquiry) skills that measured by the items. These categories were determined according to skills categories of The Taxonomy of Educational Objectives by the MONE. Cognitive Processes were translated into English by using Bloom, Hastings, and Madaus' Handbook of Formative and Summative Evaluation and Klopfer's chapter on Science in the same book and Wolfinger (1984), and Mechling et al. (1983) (as cited in Smith & Welliver, 1990). Cognitive Processes are represented in Appendix E on page 83.

1. **KNOWLEDGE:** The knowledge of fact, term, concept, principle and law, classification and sequencing and technique and procedure which student learned by heart and will recall when face with them are in this category.

2. **COMPREHENSION:** The skills to explain in their own words, make summary, give a new and original example (exemplify), predict, translate into a different form of knowledge related to science concepts, principles, laws, etc. are in this category.

3. **SOLVING PROBLEM:** The skills to bring knowledge, laws, and formulas together in order to solve the problem, use units correctly, make transition, and show the answer in required form are in this category. It is used in problem solving procedures and algorithms often in our schools.

4. **SCIENCE PROCESS (SCIENTIFIC INQUIRY):** The skills to observe, recognize problem, formulate a hypothesis, propose a procedure, verify the hypothesis, design an experiment, analyze data and reach a generalization are in this category. It requires students to use higher order thinking skills.

Subject Matter means a specific Biology topic which is shown in the elementary curricula and student's understanding of which is assesses by an item (Rakow et al., 1984). Subject matter was measured with a checklist using the elementary science curricula of the MONE and represented as Appendix F on page 84.

Content Specification is the union of both subject matter and cognitive process. It is table of test specifications or test-blue-print.

Student performance is the success of the students on the test.

CHAPTER III

RELEVANT LITERATURE

3.1 Research Studies on Content Specifications of Science Tests

The studies included in this part of the literature review describe the way science tests are constructed and analyzed according to their subject matters and cognitive processes.

In their study, Rakow et al. (1984) compared 1977 and 1982 data of national progress in science. First they clustered the items of 1977 science test into groups of subject matter (1) Content which is knowledge of terminology and about Biology, Physical Sciences, Earth Science and Integrated Science, and (2) Science, Technology and Society (STS) that is understanding of science and self, societal problems, and science and technology and cognitive processes (3) Inquiry which needs higher order thinking skills compared to content knowledge and grouped as process and decision-making. Then they administered it to a sample of students in 1982. Analyses were made according to p value that is item difficulty index, number of students in percentage correctly answering the given item. The t-test which was undergone to find significant differences between the 1977 and 1982 data found that there was no significant difference between the last and previous achievement in science among the age groups except 17-year-old one, which was statistically lower than the previous assessment. When increasing student enrollment in high school was considered, 17-year-olds achievement in science could be based on their less science knowledge. In the cluster level, 9-year old students performed significantly better in STS. This increase was related to stressing of these concepts in media, because the research studies showed less importance on those concepts in textbooks and instruction. This study conclude that students were still experiencing learning problems in science and some initiatives should be taken in order to improve this situation, i.e., for improving STS, textbooks should cover this subject.

SSISSPT is one of the tests, which has gained a lot of interest from Turkish researchers nowadays in terms of their content specifications. For example, in his research on the relationship between science curriculum and science subtest items of the SSISSPTs through 1998-2001, Çoban (2001) studied the subject matters of the science items according to physics, chemistry and biology disciplines by listing the units of science curriculum from grades 4 to 8 and giving the percentages of biology items according to test years. According to him, Reproduction and Inheritance in Living Things, Living Things and Life, Knowing Our Body, and Diversity of Living

Things are the biology units for grades 4 to 8 of the previous science curriculum which were asked in the SSISSPTs through 1998-2001.

On the other hand, cognitive processes of the science items in both the 1998-2000 SSISSPTs and teacher-made tests in the fall 2000-2001 in Trabzon were studied by Çepni et al. (2001) by using Bloom taxonomy via content analysis. The researchers found that there is no item placed into knowledge category, but an increasing number of items placed into comprehension (12-28%), application (16-20%) and a complex of analysis, synthesis and evaluation (56-68%) categories when considering the SSISSPTs. On the contrary, teacher-made tests have an opposite trend in terms of number of items per category. Here knowledge items are dominant with 49%. Comprehension, application, and the complex skills of analysis, synthesis and evaluation follow it with 32%, 13%, and 4% respectively.

3.2 Research Studies on Factorial Structure of Science Tests

A test is consisted of some subcomponents (dimensions, measures) for measuring different attributes (constructs). As defined by Cronbach and Meehl (1959, cited in Hoste, 1977) a construct is "some postulated attribute of people assumed to be reflected in test performance. In test validation, the attribute about which we make statements is a construct." For example, the test called CSE physics consists of a practical test measuring practical physics and a theory paper measuring physics theory. If a researcher wants to give evidence to the construct validity of the CSE physics test, he/she should use factor analysis techniques for his/her validation of the skills and attributes underlying the test. Factor Analysis takes a number of measures of a group of skill and ability and searches for relationships between them. The constructs here are not the types of the test, practical versus theoretical, as validated above, but skills and attributes validating performance the construct measures. Factor analysis starts with a correlation matrix between measures. When the correlations between the components of a test factor analyzed, components measuring aspects of the same attribute will significantly load on the same factor. For example, factor analysis of the CSE showed that theoretical papers with short and extended answer papers and practical paper loaded on the same factor suggesting that they were testing the same attribute (Hoste, 1977).

The principal component analysis is the most used extraction method in factor analysis of "factorial validity" referring construct validity. It is performed on subtest (component, factor) correlation matrices using the "Eigenvalues over 1 criterion" (Gorsuch, 1983 as cited in Powell, 1993).

In his MSc thesis, Aslan (2000) studied the construct validity of the 1998 SSISSPT. He examined the whole test and each one of the four subtests of that year through factor analysis. He found that there were both verbal consisting of Turkish and social studies items, and quantitative consisting of science and mathematics items dimensions on the whole test and three dimensions consisting of physics, chemistry, and biology items within the science subscale respectively.

3.2.1 Science Process Skills

It has been stated since Gagne (1965) that science courses should help students to have science process skills. Although science educators tried to achieve this goal with many curricular designs in the US in 1970's, this objective has been recently placed in the elementary science curriculum with the approval of the new science curriculum in Turkey since September 2001. This new science curriculum also aimed to teach students how to get information, learn, and develop solving problem skills (Kaptan & Korkmaz, 2001).

As stated above, the change in the aim of science education on behalf of developing science process skills turned the attention of science education researchers into the studies of Piaget, who theorized the development of human cognitive stages (Ateş et al., 2001).

In this part of the literature review cognitive development and science process skills test will be described because of the close relationship of students' cognitive development level and their science process skills.

3.2.1.a Research Studies on Cognitive Development Tests

The cognitive development tests determine students' level of cognitive development. These tests were originally in interview format and later developed into multiple-choice type to make them simpler. Therefore, American science teachers used them widely (Shayer & Wylam, 1981).

Science Reasoning Tasks Test (SRTs) is one of the cognitive development tests, which was constructed by Shayer & Wylam (1981). These researchers used Piaget and Inhelder's studies (1958) to describe the tasks of six cognitive development stages, which are preoperational (1), early concrete operational (2A), late concrete operational (2B), transitional to formal operational (2B/3A), early formal operational (3A), and late formal operational (3B). After administering SRTs to 200 students, Shayer and Wylam took into account correctly answered items. For example, a student with 4 out of 4 2A items, 4 out of 6 2B items, and 1 out of 5 3A items was categorized as 2B on the task. They determined the discrimination level of an item with a graph showing the percentages of students in each stage that answered the item correctly. It was observed that, the graph of a high discriminating item was sharp. It also meant that students were successfully placed into stages, thus supported content validity of the test. For the reliability issue, test-retest correlation and Kuder-Richardson 20 (K-R 20) formulas were used. K-R 20 is a method to estimate reliability of dichotomously scored items. It is simply the ratio of the sum of item covariances to the total observed score variance. (Crocker & Algina, 1986 p.153) Validity of the test, on the other hand, was assessed by its correlation with interviews. The correlation between students' level of cognitive development and their score on a science test with 50-60 items was well enough to give an evidence of predictive validity of the SRTs. Nevertheless, some content validity studies showed that there were some correlations between tasks, which mean that Piaget and

Inhelder's stages of cognitive development have not been confirmed yet. On the other hand, factor analysis of within task clusters showed a unique factor responsible for the 90.6 % of the total variance (Shayer and Wylam, 1981).

While the focus of science teachers was on students' level of cognitive development of Inhelder and Piaget (1958 cited in Padilla et al., 1983), science curriculum specialists paid special attention to only formal operational level abilities of it because of the relation of achievement with this psychological construct. It was also hypothesized that formal operational abilities relate with science process skills (Padilla, 1979 cited in Padilla et al., 1983). While process skills are required in science or laboratory experiments, their usage can depend on formal thinking abilities.

Padilla et al. (1983) assessed science process skills and logical (formal) thinking abilities of 492 students at grades 7-12 in Atlanta, Georgia and tried to establish relationships between these constructs. The Test of Integrated Science Process Skills (TIPS) consisted of 36 items with 4 choices representing Hypothesizing, Identifying Variables, Operationally Defining, Designing Investigations and Graphing and Interpreting Data subscales. The Test of Logical Thinking (TOLT) consisted of 10 items representing Proportional Logic, Controlling Variables, Probabilistic Logic, Correlational Logic and Combinatorial Logic. Statistical analyses showed that there is a positive relationship between these tests, and that they measure the same construct. It means if science process skills are developed, so do logical thinking abilities.

3.2.1.b Research Studies on Science Process Skills Tests

Science process skills, on the other hand, were assessed with unique tests focusing on only a few specific or complex skills independent of cognitive development tests. Molitor and George's study (1976) can be given as an example to the former, and Smith and Welliver's (1990) to the latter.

Molitor and George (1976) developed a test of science process skills of inference and verification for grades 4 through 6. Classification and inference are two of the abilities to process information in data analysis. While inference means making a judgment about the non-observable properties of an object/event, verification is observing that property. When developing the Science Process Skills Test (SPST), Molitor & George (1976) wrote items to be content-free because students' skill performance could confound with their knowledge on the conceptual material. They accepted multiple-choice test format with illustrations for easy and group administration, and confounding reading ability elimination. These researchers developed 10 items for each skill and piloted them to students at grades 4 and 6 to check their understanding. They then asked the experts for consistency in the definition of skills and performance criteria. After administering of the revised test for item difficulty and discrimination, the final form consisted of 9 items for each skill. In order to answer the question about the association between a correct item response and behavior consistent with the performance criteria, Molitor & George (1976) interviewed suburban students at grades 4-6 (N=120) with the randomly selected items and found a positive association between response and behavior. Items equal to or less than 0.025 were judged as valid. They gave evidence to the construct validity of the test with the student's t-test for independent means by finding that students having experience with the curriculum performed better than the other students. To answer the last question of reliability of the test among grade levels, Molitor & George (1976) administered the test to groups of randomly selected suburban students from grades 5 (N=96) and 6 (N=100). They used K-R 20 formula to estimate the reliability of each skill items, which are considered to be homogeneous subtests. These researchers found the difficulty level of items between 0.40-0.60. They used corrected biserial correlations between the subtests and their items and found that the items have satisfactory power of discrimination.

Smith and Welliver (1990), on the contrary, developed a science process skills test consisting of complex skills. Their test, which is called Science Process Assessment (SPA), assesses students' science process skills required by the Recommended Science Competency Continuum for Grades K to 6 for Pennsylvania Schools. The SPA included observing, classifying, inferring, predicting, measuring, communicating, using space/time relations, defining operationally, formulating hypotheses, experimenting, recognizing variables, interpreting data, and formulating models. Smith and Welliver (1990) brought together a group of 10 people who were either classroom teacher or science educators to prepare this science test. These researchers wanted the group to construct items representing at least one science process skill, consisting of one (or more) sentence and idea, and having 4-choices. After the group developed 65 items, Smith and Welliver (1990) wanted the group to check the items for appropriateness to the predetermined criteria for content validity issues such as (1) having an idea or a problem, clarity and grammatically true, (2) appropriateness to that science process skill and that goal listed in the continuum, (3) measuring the knowledge of student gained at grade 4 or before, (4) readability, (5) comprehensibility, (6) clarity of the figures clear and understandability and write again. They also wanted the group to make sure that the items were reflecting the skill and there were enough number of items in each skill. The developed 61 items were asked to a jury of science educators in order to determine their appropriateness to the goals and skills, or their content validity. The items then were piloted on 9 4th grade classes in West Pennsylvania. K-R 20 reliability coefficient was found 0.70. Item-total correlation (point biserial) was found with Program Tessia and when the items with a correlation of equal to and less than 0.20 were omitted, the left 40 items were retained as pilot 2. Pilot 2 was administered twice to 113 4th grade students from 7 classes in 3 school districts with two-week-period. While reliability coefficient was found 0.80 in the first administration of the pilot 2, it was 0.82 in the second. Pearson product moment correlation coefficient was 0.77 between the administrations.

3.3 Research Studies on Factors Influencing Science Achievement

3.3.1 The Effect of Gender on Science and Biology Achievement

The first factor influencing science achievement is gender. Although gender difference is said to be disappeared in 10 TIMSS participating countries, Australia, Colombia, Cyprus, Flemish Belgium, Ireland, Romania, the Russian Federation, Singapore, South Africa and the United States of America (Peak, 1997 as cited in TIMSS, 2001), gender can be a determining factor on student achievement in science depending on the age-grade level of the student. For example, 1996 National Assessment of Educational Progress (NAEP) science achievement showed that only at grade 12, American students differed in score for being female or male. No difference among 4 and 8 graders was found in terms of gender (TIMSS, 2001).

When gender differences are considered in the overall science achievement, male students seem to perform better than their female counterparts. But this cannot be said for the all three discipline of science, physics, chemistry, and biology. In their meta-analytic study on gender difference in motivation related to achievement in science, Steinkamp & Maehr (1984) found that motivational orientation of female students in biology and chemistry is higher than male students except physical and general science. Female students also tend to enroll in biology courses. The researchers related female students' more positive orientation in biology to their anticipation of maternal role, verbal inclination, and early preference not affected by stereotyping (Kelly, 1978 as cited in Steinkamp & Maehr, 1984) On the other hand, male students' more positive orientation in physical and general science related to their learning outside of the class by extracurricular activities and contacts with knowledgeable males when comparing to school-based biology and chemistry. Steinkamp and Maehr (1984) did not accept the definition of cognitive superiority of male students by Hyde (1981) and Macoby & Jencklin (1982) as valid because verbal superiority of female students is more conducive to the development of positive attitude toward science, because verbal ability is more easily measured and rewarded than cognitive one.

3.3.2 The Effect of Cognitive Processes on Science Achievement

Cognitive domain is consisted of cognitive processes of knowledge, comprehension, solving problem and science process, which are in order, and in which a higher order process such as science process follows a lower order one such as comprehension. The effect of cognitive processes on science achievement was studied by Gürçay et al. (2001). The researchers sampled 222 university students and gave them the Academic Competency Test (FBAYT). FBAYT consisted of physics, chemistry, and biology subcomponents and measured students' comprehension, solving problem, and science process skills. Descriptive statistics of the students' answers by SPSS showed that comprehension questions were the most answered area by the students. Questions related to science process skills area were answered the least. This ordering of cognitive processes was found also by the ITEMAN, items related to science process had the highest difficulty and the items related to comprehension skills had the lowest difficulty.

3.3.3 The Effect of Learning Difficulties on Science Achievement

Water transport, osmosis and osmoregulation, respiration and photosynthesis, energy cycles (ATP and ADP), cell respiration, Mendel's laws of genetics are said to be the difficult subjects among high school students (Johnstone & Mahmoud, 1980, Steward, 1982, and Finely et al. 1982 as cited in Lazarowitz & Penso, 1992).

Bahar et al. (1999) surveyed 207 freshmen Biology students with a questionnaire asking their views regarding difficulty level of 36 biology subjects and the subjects of mono and dihybrid crosses, gametes, alleles and genes were defined as difficult.

Learning difficulties in these subjects were attributed to biological level of organization and abstract level of concepts, which will be described in the following paragraphs.

The biological level of organization: When students are learning biology on the organism, population and community level, they do not have a difficulty. Because these are macro and tangible level thoughts and can easily be sensible. But when students are learning biology on molecular, cellular, tissue and organ level, they have difficulty. Because these are micro or sub-micro and molecular level thoughts and cannot be easily sensible but with representational (symbolic) (Klinckman, 1970 as cited in Lazarowitz & Penso, 1992, Johnstone, 1991 as cited in Bahar et al. 1999).

The abstract level of concepts: As mentioned earlier in the cognitive development, when students are learning abstract (formal) concepts such as photosynthesis, respiration, dominance and co-dominance, and sex linkage, their level of cognitive development plays as a determining factor on their understanding. For example students at the concrete operational level are able to understand 30% of the concepts requiring concrete operations but very few or none of the concepts requiring higher cognition i.e., formal reasoning (Inhelder & Piaget, 1958, Lawson, 1975, Lawson & Renner, 1975, and Walker et al., 1980 as cited in Lazarowitz & Penso, 1992).

3.3.4 The Effect of Misconceptions on Science Achievement

Although the presence of misconceptions should be determined via a three-tier-test, item analysis of multiple-choice tests can give which misconceptions students have. Yürük & Çakır (2001) investigated high school students' misconceptions on aerobic and anaerobic respiration with a 20-item-test. When developing the test they reviewed the literature on misconceptions related to their research subject and gave them as alternatives in their test. They decided to a misconception using the point biserial correlation coefficient value of an item. If the test item was difficult enough, and discriminated well high and low achievers, but had a positive point biserial value on any of the alternatives, that alternative was thought to be a misconception.

3.3.5 The Effect of Other Factors on Science Achievement

Beside the above-mentioned factors influencing students' science and (or) Biology achievement, there are some other factors that cannot be manipulated in order to make a student achieve. According to the United States' national education reports there are some factors other than gender and biological and physiological status of students affect science achievement. They are quality and quantity of science courses taken, family income and education level (referred as socio-economic status), and school characteristics (TIMSS, 2001). But the research concern of the present study is not dealing with these factors.

As the review of literature shows science tests aims to assess students' not only knowledge but also science process skills because of its own nature. The literature also explain that science tests are constructed after a series of steps consisting of writing test items on specific objectives with the help of subject matters and cognitive processes, consulting science educators for content and construct validity purposes, analyzing reliability, factors and items after administering or piloting the test, and if necessary revising or eliminating some of its items. As some of the literature points, it is also possible to follow a way opposite to test construction and study validity of the tests after a series of steps similar to test construction except writing test items. Moreover students' performances among subject area and science process skills are measured by these tests as in the literature.

Although there is one for the 8-graders' science achievement in curriculum laboratory schools (EARGED, 1995), another with the construct validity of the 1998 test done by Aslan (2000) and the other on subject matters through content analysis (Çoban, 2001), the review of the literature showed that there is no sufficient study to evaluate biology items in the SSISSPT in Turkey. This study has another importance for evaluating the outcomes of the Turkish elementary science curriculum, which was replaced with a new one. The new version has been effective since the beginning of the 2001-2002 academic year.

The study is intended to fill the information gap in Turkish elementary science curriculum mentioned above through the assessment of biology items in terms of their content specifications, factorial structures, and student performances on different clusters of biology items across gender groups.

CHAPTER IV

METHODOLOGY

4.1 – Population and Subjects

There are both a growing interest to take the SSISSPT and the number of students studying at grade 8 because of the extension in compulsory education in Turkey since 1998. Therefore, the number of the students who took the test increased much overtime.

The data for the SSISSPTs from 1999 through 2001 were gathered from the MONE. From each of the student populations 10.000 students were taken randomly as subjects of the years 1999 and 2000. On the other hand, only 5.000 students were taken randomly as the subject of 2001. Subjects were taken randomly from the whole group took the test. Even though the sample sizes seem small compared to the whole group sizes, it is assumed that randomization made the samples representative of the whole groups.

Table 1
Number of students who took the SSISSPT and were placed in the secondary schools

	Year			
	1998	1999	2000	2001
Number of students taking the SSISSPT	300.624	309.474	419.216	568.474
Number of students placed	89.276	90.193	83.954	94.355

Note: From EĞİTEK, 1998, 1999, and 2000.

4.2 – Procedures

The content specifications and dimensions of biology tests were investigated in the present study. Also the gender effect on the students' performances was investigated in the study.

Content specifications of the biology items were determined via content analyses while subject matters of the biology items were determined using the previous elementary science curriculum by the researcher. Expert opinion was taken in classification of items according to the educational objectives of the MONE. Interrater agreement was taken into account for accuracy of the judgments. Factor analyses to show the dimensionality of the biology tests were done by using the SPSS program. Also gender performances in different clusters and whole scale of the biology tests were analyzed with MANOVA by using the same program.

4.3 – Data Collection and Analysis

The content specifications i.e., subject matters and cognitive processes of the tests were found through content analysis. These results were shown as frequency distributions.

When subject matters were being considered, the recent Turkish elementary science curriculum was used to determine what each item designed to measure. The titles and subtitles of grades 4-8 biology units in the elementary science curriculum helped to develop the subject matter checklist, and their aims and student behaviors given in the curriculum helped to classification of items into these subject matters. For example there are five items dealing with living things in the 1998 SSISSPT. The first one, which is also the first item of the test, was easily placed into fungi, one-celled organisms, and bacteria subject of the Diversity of Living Things unit (see number 17 in Appendix J on page 88). The second item was asking students to translate some characteristics of birds into schematic form (see number 18 in Appendix J on page 88). This item was considered in animals subject of the same unit because it is considered to be aiming "be able to comprehend the diversity of animals" with related "classification of animals according to similarities and differences" behavior. On the other hand, the third item asking students to make a comment on herbivorous and carnivorous animals was placed into life in animals subject of the Living Things & Life unit because of aiming "be able to comprehend feeding of animals" and related with the objectives in the knowledge and comprehension categories mentioned in the curriculum (see number 19 in Appendix

J on page 88). The fourth item asking to relate the length of life in a butterfly with some living things was placed as the cell & cellular activities subject of the Living things and Life unit since it is aiming the comprehension of both photosynthesis and respiration (see number 21 in Appendix J on page 89). The last item asking students to comment on the advantages of human heart with separated ventricles over a turtle heart with half-separated ventricles was considered as animals subject of the Diversity of Living Things unit (see number 23 in Appendix J on page 89).

Cognitive processes were determined via the Taxonomy of Educational Objectives used by the MONE. Cognitive processes are consisted of four categories: knowledge, comprehension, solving problem and science process skills (see Appendix E on page 83).

This ability list was developed into a checklist to use in the content analyses. In order to make content analyses, thirty-four Biology questions were classified into groups representing the units of the elementary science curriculum. Then, they were given to four biology educators in Science Education department at METU to determine cognitive processes with a written instruction on the aim of the study and how and where to code each one of the 34 items according to the educational objectives by MONE. After the experts completed the task, the results were examined for consistency by using the SPSS package program (for experts' ratings see Appendix H on page 86). The experts were defined as variables 1 to 4. The codes of each item made by an expert were entered into the cells of the column representing that expert. In order to give evidence to the interrater reliability, which is consistency between experts, central tendency and restricted variability, severity, and percentage of exact agreement were examined according to the definitions in the study of Harwell (1999). For the reliability purpose, the Spearman correlations between experts were also found. The experts with high reliability in between were selected for expert agreement step of content analysis via the expert opinion in giving evidence to both content and construct. The experts were tried to agree on the items that took different ratings from them at this task. When doing this second rating together, experts considered similar type items and gave them the same rating. For example, item 22 of the 1999 SSISSPT (see item 22 in Appendix K on page 91) and item 9 of the 2000 SSISSPT (see item 9 in Appendix L on page 94) were considered as trying to measure students ability to comment on given knowledge given in the schematic form. On the other hand, the similar biology items 19, 21, and 23 of the 20001 SSISSPT were considered to be processing/analyzing and interpreting data since they were acquiring knowledge other than given in the graph (see Appendix M for these items on pages 95-96). After this step was completed, the items were shown to a measurement and evaluation specialist in science education at METU who also works in the testing center of the MONE as a last step and corrections were made according to the suggestions (see Appendix I for the result on page 87).

Students' scores on the SSISSPTs through 1999-2001 were calculated by the use of SPSS for Windows computer program. Principal Component Analysis was used for studying the clusters of the Biology items within science subscales. The varimax rotated factor solutions found dimensions account for the variances in the subscales. Principal component analysis transformed original variables into new set of linear combinations. Number of factors was defined as either three or four in the varimax rotation and therefore the components with eigenvalues less than 1 were retained.

MANOVA was used to study gender differences across different clusters of the biology items with well-working data. When doing this, each correct answer received one point and wrong answers and no answer received zero. When computing biology achievement score, students' scores in each biology item were added together and divided by the number of items in that year and represented as percentages. For example, there are seven items in the 2001 SSISSPT, students' scores on each item were added together and divided by 7 to find their overall Biology achievement scores. Similarly when computing each cluster score, students' scores in each cluster item were added together and divided by the number of items in that cluster and represented as percentages.

ITEMAN was used to determine difficulties and discriminations of biology items in the SSISSPTs through 1999-2001. ITEMAN provides item statistics in terms of item difficulty denoted as p, item discrimination and point biserial correlations.

Item difficulty (p) is the percentage of students who answered correctly. When the number of correct responses is divided by the number of total responses and multiplied by 100, item difficulty is found. P value is showed as Proportion Correct of item statistics and Proportion Endorsing of the alternative statistics parts in the ITEMAN output.

Index of discrimination (D) is the difference between the proportion of upper group responded the item correctly and the proportion of lower group responded the item correctly. The value of D has some interpretations about the item. The item functions quite satisfactorily when D≥.40, needs little or no correction when D=.30-.39, requires revision when D=.20-.29, or requests elimination or complete revision when D≤.19.

When measuring the degree of linear relationship between item and total scores two correlational indices of item discrimination is used. The first correlational index is called Point Biserial Correlation. It is the degree of relation between performance on an item scored dichotomously as either 0 or 1 and the total test. It assumes a student successfully answering an item will be successfull on the whole test. The second correlational index is called Biserial Correlation Coefficient. It is the degree of relation between a latent variable underlying item performance and a continuously distributed test score. It assumes that the latent variable underlying item performance is normally distributed. The value of biserial correlation is higher than point biserial correlation. The difference between the biserial and point biserial correlation depends on item difficulty, p. When p is moderate, the difference is moderate that is .25-.75, but, when p is below .25 which means item is difficult or more than .75 which means item is easy, the difference is high. Thus, which correlation formula will be taken into account in item analysis is depend on purpose. When either low or high difficult items is the concern, biserial p should be used. When students differ in ability, biserial p should be used. In this case low biserial p means the item is low in discrimination. But point biserial p for a sample of low or high ability means the item is difficult or easy.

It is assumed that an item discriminates high and low ability students and the correct choice has a high and positive biserial p value evidencing only high ability students select that choice or the other choices have low and negative biserial p values evidencing low ability students selected those answers. But sometimes high ability students select the other choice(s) other than the correct one. In this case they have a possible misconception.

The values of both correlational indices are seen in the ITEMAN output.

CHAPTER V

RESULTS

5.1 – What are the content specifications of the biology items within the science subscales of the 1998-2001 SSISSPTs?

5.1.1 Which subject matters are measured by the biology items within the science subscales of the 1998-2001 SSISSPTs?

This question was tried to be answered by a list prepared by using the previous elementary science curriculum. The titles of biology units and their subjects of grades 4-8 were selected from the curriculum and listed as a table. Then, biology items asked in the 1998-2001 Secondary School Institutions Student Selection and Placement Tests were placed into that table according to their subject matters as frequencies (see Table 2 on page 40).

When the subject matters of the biology items across years are taken into account, we can see that some of them took great emphasis in terms of number of questions asked by the item writers, on the contrary some did not.

Living Things & Life is the most asked unit (41%). From 1998 to 2000 about half of the questions were coming from this unit. Diversity of Living Things is the second most asked unit of the SSISSPT with 26%. In 2001, nearly half of the questions (43%) were coming from this unit. Reproduction & Inheritance in Living Things and Knowing Our Body were the less asked units with 14%. There was no question asked about the unit of Human & Environment. There was one item in the category "other" in the 1999 SSISSPT. Interestingly, this item was dealing with science process skills.

Table 2
Subject matters of biology items with respect to test years

Unit	Year 1998	1999	2000	2001
Human & environment				
Pollution types				
Pollution & health				
Humans' effects on environment				
Matter cycle				
Reasons of environmental damage				
Sustainable development				
Reproduction & inheritance in living things				
Cell & cell division				
Reproduction in living things		1		
Inheritance	1			
Inheritance & environment		1	1	1
Knowing our body				
Sensory organs				
Motion system				
Digestion system & feeding			1	
Circulation system		1		
Respiration system				
Excretion system				
Nervous system	1	1		
Hormones& hormone producing organs				
Immune system				1
Diversity of living things				
Plants				
Animals	2		1	
Fungi, one-celled organisms, & bacteria	1			1
Interactions between living things		1	1	2
Biologic wealth of Turkey				
Living things & life				
The world of living things				
Life in plants	2	1	1	1
Life in animals	1		1	
Cell & cellular activities	1	2	3	1
Growth & development in humans				
Other		1		
Total	9	9	9	7

Note: This table is prepared by using "Elementary Science Curriculum" by the MONE, 1992, and İlköğretim Programı, p.50.

5.1.2 Which cognitive processes are measured by the biology items within the science subscales of the 1998-2001 SSISSPTs?

After the four of the biology education faculty at METU determined the cognitive processes of the biology items according to the taxonomy of educational objectives via content analysis, their ratings were used as data to further analysis of interrater agreement in terms of reliability in the SPSS (see Appendix H on page 86 for the experts' ratings).

The reliability coefficients were found to be .6781 for Alpha and .7604 for standardized item alpha (Table 3).

Table 3
Result of reliability analysis

Item statistics	Mean	Minimum	Maximum	Range	Max/Min	Variance
Item Means	2.7406	2.0671	3.4406	1.3735	1.6645	.3184
Item Variances	.8717	.0005	1.5363	1.5358	3312.3090	.4083
Inter-item covariances	.3007	.0066	.8523	.8456	128.8004	.1167
Inter-item correlations	.4425	.3055	.7097	.4043	2.3233	.0185

Note: Method 2 (covariance matrix) was used for this analysis. N=34. Reliability coefficients 4 items.

The correlations between these four raters were ranging from .235 to .616 as illustrated in Table 4.

Table 4
Correlation matrix of experts' ratings

Experts	1	2	3	4
1	1.0000			
2	.246	1.0000		
3	.605**	.165	1.0000	
4	.616**	.354*	.291	1.0000

Note: Spearman's rho correlation coefficient was used for this analysis. Correlations are significant at the .01 (**) and .05 (*) levels (2-tailed).

Central Tendency and Restricted Variability: When the ratings of the second and third experts with lowest correlation with each other were examined, as it is seen from the Table 4, it was found that while the former was grading items mostly in the last cognitive process, the latter was more likely to grade items as if they are measuring comprehension skills of the students (Table 5).

Table 5
Descriptive statistics of the experts' ratings

Experts	Mean	Std Dev
1	2.6524	.9688
2	3.4406	1.0057
3	2.0671	.0215
4	2.8024	1.2395

Note: There are 34 cases. Method 2 (covariance matrix) was used for this analysis

Severity: When the mean rating score of a rater is compared with the others' in terms of difference between mean scores, the pattern of his/her ratings appears. As can be seen from the Table 6, the mean score differences between Expert 2 and others and Expert 3 and the others are very high.

Table 6
Result of severity analysis

Experts	Mean Difference
1	(-0.7908)-(0,5853)
2	(0,6408)-(1,3761)
3	(–0,5853)-(-1,3761)
4	(0,7353)-(-0,6408)

Percentage of Exact Agreement: When the ratings of experts in each item were compared and the same ratings were counted and divided by the number of items the percentage of exact agreement value can be found. The result of this interrater reliability analysis can be seen in Table 7. Expert 1 has agreed most with Experts 3 and 4. Moreover, Expert 3 does not have such degree of agreement with others and Expert 4 has high agreement with Expert 4.

Table 7
Percentage of agreement index

Experts	Agreement
1-2	26,47
1-3	32,35
1-4	32,35
2-3	8,82
2-4	23,52
3-4	2,94

When the results of the reliability analyses were considered, Experts 2 and 3 were eliminated from the analyses and only the remaining two whose ratings have the highest correlation with each other were selected for examining the content validity of the tests. These experts met several times in order to determine the cognitive level of 23 out of 34 biology items since they had already agreed with 11 out of 34 items.

Actually they had some problems with only 10 items out of 23 because when one expert placed an item in comprehension level, the other placed in knowledge, or when an item was thought to be requiring science process skills by the former, it was thought as comprehension level skill by the latter. The final examination of the items on cognitive processes revealed that there are 12 items in science process skills, 1 item in solving-problem, 20 items in comprehension, and 1 item in knowledge.

As a last step, the items were then shown to a measurement and evaluation specialist in science education at METU who also works in the testing center of the MONE. Furthermore, some corrections were made in the light of the suggestions. As a result, it was found that the item rated as knowledge was in the comprehension level, some of the science process skill items were actually in the comprehension level. The result of the expert agreement and specialist's opinion can be found in Appendix I on page 87.

Table 8
<u>Cognitive processes of biology items with respect to test years</u>

Cognitive processes	Year			
	1998	1999	2000	2001
Knowledge				
Facts				
Terminology				
Concepts				
Classifications				
Sequences				
Techniques and procedures				
Principles and laws				
Theories				
Comprehension				
Identification of knowledge in a new context		1		
Explanation	2			
Summarization				
Giving example	2	2		
Identification of criteria in a given information				
Relation	1		3	1
Comment	2	2	2	
Translation of knowledge from one form into another	1		1	3
Prediction				2
Solving problem				
Finding required principle				
Bringing knowledge, law, and principle together		1		
Using formula and algorithms				
Using units correctly and making transition				
Showing the answer in required form				
Science process				
Observing				
Description of observation				
Comparison of results of observation				
Classification of results of observation				
Selection of appropriate measuring instrument				
Recognition of a problem	1	2		1
Relation between elements of problem				
Formulation of a hypothesis to solve the problem			3	
Proposing procedure to verify (test) the hypothesis				
Design of experiment/think/formulate in figure		1		
Collection of data				
Processing/analyzing/interpreting data				
Discussing and evaluating hypotheses according to evidences				
Conclusion according to evidences				
Generalization and proposing new research questions				
Application of observation & research results to daily life or new situation				
Total	9	9	9	7

5.2 – What are the factorial structures of the biology items within the science subscales of the 1999-2001 SSISSPTs?

5.2.1 What is the factorial structure of the biology items in the 1999 SSISSPT?

The nine items of the 1999 Secondary School Institutions Student Selection and Placement Test were factor analysed; four components came up with a total of 53.107 % explained variance. Only one of the components has eigenvalue over 1.

Table 9
Communalities of the biology items in the 1999 SSISSPT

	Communalities	
Items	Initial	Extraction
17	1.000	.758
18	1.000	.907
19	1.000	.406
20	1.000	.417
21	1.000	.525
22	1.000	.419
23	1.000	.278
24	1.000	.387
25	1.000	.682

Note: Extraction method: principal component analysis.

Table 10
Total variance explained of the biology items in the 1999 SSISSPT

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.908	21.195	21.195	1.908	21.195	21.195
2	.988	10.982	32.177	.988	10.982	32.177
3	.952	10.576	42.753	.952	10.576	42.753
4	.932	10.354	53.107	.932	10.354	53.107
5	.918	10.200	63.307			
6	.873	9.701	73.009			
7	.836	9.286	82.295			
8	.815	9.053	91.348			
9	.779	8.652	100.000			

Note: Extraction method: principal component analysis.

Table 11
Rotated component matrix of the biology items in the 1999 SSISSPT

	Component			
Items	1	2	3	4
25	.767	-.264		
24	.507	.305		
23	.497			
19	.455	.282	.318	
21		.705		
20		.600		
17			.851	
22		.290	.547	
18				.945

Note: Extraction method: principal component analysis. Rotation method: varimax with kaiser normalization. Rotation converged in 5 iterations.

When rotated component matrix was taken into account, the items which are measuring comprehension (items 19 and 24) and science process (item 25 and 23) skills were found to be in the first factor; the items which are measuring comprehension skills (items 20 and 21) in the second, the items which are measuring science process (item 17) and comprehension skills (item 22) in the third; whereas the item related problem solving skills (item 18) in the last factors These items can be found in Appendix K on page 90-91.

5.2.2 What is the factorial structure of the biology items in the 2000 SSISSPT?

The nine items of the 2000 Secondary School Institutions Student Selection and Placement Test were factor analysed; four components came up with a total of 53.132 % explained variance. Only two of them have eigenvalues over 1.

Table 12
Communalities of the biology items in the 2000 SSISSPT

	Communalities	
Items	Initial	Extraction
1	1.000	.583
2	1.000	.524
3	1.000	.392
4	1.000	.544
5	1.000	.449
6	1.000	.833
7	1.000	.588
8	1.000	.490
9	1.000	.378

Note: Extraction method: principal component analysis.

Table 13
Total variance explained of the biology items in the 2000 SSISSPT

	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
Component	Total	% of variance	Cum. %	Total	% of variance	Cum. %	Total	% of variance	Cum. %
1	1.816	20.179	20.179	1.816	20.179	20.179	1.278	14.203	14.203
2	1.066	11.842	32.021	1.066	11.842	27.145	1.231	13.682	27.885
3	.970	10.772	42.793	.970	10.772	42.793	1.229	13.657	41.542
4	.931	10.339	53.132	.931	10.339	53.132	1.043	11.590	53.132
5	.917	10.192	63.324						
6	.898	9.977	73.301						
7	.836	9.293	82.594						
8	.806	8.954	91.548						
9	.761	8.452	100.000						

Note: Extraction method: principal component analysis.

Table 14
Rotated component matrix of the biology items in the 2000 SSISSPT

	Component			
Items	1	2	3	4
8	.689			
5	.659			
3	.483	.386		
7		.724		-.205
4		.705		
1			.727	-.215
2			.690	
9	.326	.212	.413	.237
6				.912

Note: Extraction method: principal component analysis. Rotation method: varimax with kaiser normalization. Rotation converged in 5 iterations.

In the varimax rotation performed after the principal factor analysis in order to show the items with their best correlations in the scale, 8, 5 and 3 composed the first factor; 7 and 4 composed the second factor; 1, 2 and 9 composed the third factor; and 6 composed the last factor. When we look at the cognitive processes each items measured, we can see that the first factor included the skill of formulation of a hypothesis to solve a problem (items 3, 5 and 8). The second factor was asking students' comprehension process skills namely relation (items 4 and 7). The third factor included the items, which measure comment (items 1 and 2) and translation of knowledge from one form into another skill of comprehension (item 9). The fourth factor required the relation skill of comprehension as in the taxonomy (item 6). These items can be found in Appendix L on page 92-94.

Table 15
Correlation matrix of the 2000 biology test and its subtests

Subtests	F1	F2	F3	F4	TOTAL
F1	1.0000				
F2	.2056	1.0000			
F3	.2498	.2293	1.0000		
F4	.0258	-.0241	.0473	1.0000	
TOTAL	.7154	.6075	.7151	.2323	1.0000

Note: For subtests, alpha=40.77 and standardized item alpha=3578

As seen from the Table 15, there are low correlations ranging from -.0241 to .2498 between the subtests and low to moderate correlations from .2323 to .7154 between the subtests and the entire scale. These low correlations between subtests, which are possibly because of the small number of the items, i.e. only 1 or 3, make the reliability low. As seen in the same table, these low correlations, on the other hand, mean that subtests are measuring really different cognitive processes of students in the biology test.

5.2.3 What is the factorial structure of the biology items in the 2001 SSISSPT?

After nine items of the 2001 SSISSPT were factor analyzed; three components were found to be responsible from the 54.949% of the variance in the scores. There are two components (factors) which eigenvalues are bigger than 1.

Table 16
Communalities of the biology items in the 2001 SSISSPT

	Communalities	
Items	Initial	Extraction
17	1.000	.422
18	1.000	.831
19	1.000	.334
20	1.000	.984
21	1.000	.403
22	1.000	.388
23	1.000	.485

Note: Extraction method: principal component analysis.

Table 17
Total variance explained of the biology items in the 2001 SSISSPT

	Initial eigenvalues			Extraction sums of squared loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.924	27.491	27.491	1.924	27.491	27.491
2	1.010	14.429	41.920	1.010	14. 429	41.920
3	.912	13.029	54.949	.912	13.029	54.949
4	.865	12.356	67.305			
5	.801	11.443	78.749			
6	.791	11.306	90.055			
7	.696	9.945	100.000			

Note: Extraction method: principal component analysis.

Table 18
Rotated component matrix of the biology items in the 2001 SSISSPT

	Component		
Items	1	2	3
23	.679		
21	.624		
22	.616		
19	.526	.230	
17	.460	.451	
18		.910	
20			.985

Note: Extraction method: principal component analysis. Rotation method: varimax with kaiser normalization. Rotation converged in 5 iterations.

Examination of the rotated component matrix showed that 23, 21, 22, 19, and 17 were loaded into factor 1 representing translation of knowledge to a different form of comprehension skills; 18 were loaded into factor 2 representing science process skills; and 20 was loaded into factor 3 representing lower level comprehension skills, relation. These items can be seen in Appendix M on page 94-95.

5.3 – Are there any differences in the student performances in different subcomponents of the biology items across the gender groups?

This question was answered with MANOVA by using the 2000 SSISSPT data. Because the 2000 data are the only one that in which the items thought as measuring the same cognitive processes loaded into the same factors thus worked well. Students' scores on the biology test (Whole), science process skills (F1, Factor 1), relation skills (in comprehension level) (F2, Factor 2), translation skills (in comprehension level) (F3, Factor 3), and relation skills (in comprehension level) (F4, Factor 4) were taken as dependent variables and students' gender as independent. Girls were defined as 1 and boys as 2 in the outputs.

Table 19
Student performances in the 2000 SSISSPT biology test and subtests across gender groups

Factors	Mean	Std. Deviation	N
F1			
1.00	.4483	.3166	4934
2.00	.4384	.3233	5061
Total	.4433	.3200	9995
F2			
1.00	.5895	.3738	4934
2.00	.5600	.3770	5061
Total	.5745	.3757	9995
F3			
1.00	.3806	.3048	4934
2.00	.4309	.3095	5061
Total	.4060	.3082	9995
F4			
1.00	.1804	.3845	4934
2.00	.1932	.3949	5061
Total	.1869	.3898	9995
Whole			
1.00	.4273	.2085	4934
2.00	.4357	.2125	5061
Total	.4315	.2106	9995

In the 2000 SSISSPT, girls got higher scores than their counterparts in Factor 1 and Factor 2. On the other hand, boys had higher scores in factor 3, factor 4 and total. Students were successful when answering Factor 2, namely relation questions. The second area which students were successful was Factor 1, namely formulating hypothesis skill of science process and Factor 3 that is measuring the skill of translation of knowledge from one from into another in comprehension category. Students were less successful in Factor 4, relation type of comprehension skills.

Table 20
Multivariate test result of gender performances in the 2000 Biology test and subtests

Effect	Value	F	Hypothesis df	Error df	Sig.	Eta squared	Noncent. Parameter	Observed power
Intercept								
Pillai's trace	.813	10836.553	4.000	9999.000	.000	.813	43346.213	1.000
Wilks' lambda	.187	10836.553	4.000	9999.000	.000	.813	43346.213	1.000
Hotelling's trace	4.339	10836.553	4.000	9999.000	.000	.813	43346.213	1.000
Roy's largest root	4.339	10836.553	4.000	9999.000	.000	.813	43346.213	1.000
Gender								
Pillai's trace	.011	28.114	4.000	9999.000	.000	.011	112.456	1.000
Wilks' lambda	.989	28.114	4.000	9999.000	.000	.011	112.456	1.000
Hotelling's trace	.011	28.114	4.000	9999.000	.000	.011	112.456	1.000
Roy's largest root	.011	28.114	4.000	9999.000	.000	.011	112.456	1.000

Note: Computed using alpha = .05. Exact statistic. Design: intercept+gender

Table 21
Test result of between-subjects effects

Source Dependent variable	Type III sum of squares	Df	Mean square	F	Sig.	Eta squared	Noncent. Parameter	Observed power
Gender								
F1	.246	1	.246	2.407	.121	.000	2.407	.342
F2	2.176	1	2.176	15.441	.000	.002	15.441	.976
F3	6.326	1	6.326	67.049	.000	.007	67.049	1.000
F4	.413	1	.413	2.720	.099	.000	2.720	.378
Whole	.173	1	.173	3.914	.048	.000	3.914	.507
Error								
F1	1023.246	9993	.102					
F2	1408.294	9993	.141					
F3	942.843	9993	9.437E-02					
F4	1518.470	9993	.152					
Whole	442.955	9993	4.433E-02					
Total								
F1	2987.556	9995						
F2	4709.750	9995						
F3	2597.000	9995						
F4	1868.000	9995						
Whole	442.955	9995						
Corrected Total								
F1	1023.493	9994						
F2	1410.470	9994						
F3	949.169	9994						
F4	1518.883	9994						
Whole	443.128	9994						

Note: Computed using alpha = .05. R squared = .000 (adjusted r squared = .000). R squared = .002 (adjusted r squared = .001). R squared = .007 (adjusted r squared = .007)

As seen from the MANOVA output, 0.2% of the variances in students' scores in Factor 2 and 0.7% of the variances in students' scores in Factor 3 are due to students' gender. In other words, there are small differences in students' scores in relation skill (Factor 2) favoring female students and in translation skill (Factor 3) in favor of male students. No difference is found in students' scores in Factor 1 and 4, and overall biology scores across gender groups.

CHAPTER VI

CONCLUSION

Every test should be considered in terms of its validity and reliability. Validity means whether it measures what it intends to measure. Reliability means whether it is beyond the measurement error which means target group is consistently perform similar in every administration of the test. Psychological features of the target group affect reliability of the test. They are fatigue, nervousness, and content sampling, answering mistakes, misinterpreting instructions and guessing. These are can be called test-anxiety, self-efficacy and Weiner's affective factors of attribution (motivation, interest, enjoyment, self-concept, and locus of control) to affect reliability. As being understood from the explanations, validity seems to be more related to test itself, but reliability seems to be related to examinee itself. Since to establish and increase reliability is hard, validity is more easy to study because some evidence of validity can be given when developing a test and evaluating it after its administration.

There are three types of validities and reliabilities in measurement and evaluation in science education (Rudner, 1994). The first type of the validity is content validity. It means the degree to which a test represents the skills in the specified subject area. The question of skills or cognitive process and subject matters asked in the test items is answered via the evaluation of validity in terms of content. Preparing a test blueprint before the test development and taking decisions of expert judges are two steps to follow when proving content validity of the test.

The second type of the validity is criterion related validity. Criterion validity refers to the degree of test to infer about future performance of the examinee. The SSISSPT is giving a chance to the well-performed examinee, because he/she is thought to be successful in its specific secondary schools such as science lycee. In another point of view, if we are studying achievement in science among secondary school students, and using elementary school achievement as a criterion, the SSISSPT can be used as a criterion measure to predict secondary school performance.

The third type of validity, the construct validity, refers to the degree to which the test measures the right construct. It tries to answer whether the test measures ability, achievement, aptitude or intelligence. Inter-item correlations and factor analyses are consulted in order to answer whether there is relationship among items and between items and the whole scale. In this study, it is expected that some biology items in the SSISSPT acquiring same cognitive process will cluster in a single factor. In sum, it assessed the SSISSPT's content and construct validities.

Validity of science achievement tests is questionable, because science achievement tests use multiple-choice format. This format has some advantages or disadvantages when measuring cognitive functions (cognitive process or thinking skills) of the students. In his article, Martinez (1999) tried to compare multiple-choice and constructed response type tests by using research done in the science area. Constructed response means "large class of test item formats in which a response must be generated by the examinee rather than chosen from a list of options" (Cronbach, 1984 as cited in Martinez, 1999). Here, Cronbach also explained what multiple-choice test meant in the latter part of the sentence.

One of the differences between usages of these tests is elicitation of cognitive functions. Cognitive functions are memory (knowledge), comprehension, analysis, and evaluation. Multiple-choice tests are said to elicit lower order thinking skills, including the first two categories, knowledge and comprehension. In their studies, Bowman and Feng (1972 as cited in Martinez, 1999) classified 800 advanced psychology examination questions into their cognitive process and found that 80-88% of the items fell on the lower order, 64-77 of the items alone fell on the knowledge category. They later suggested more frequent elicitation of higher-order thinking skills.

Another comparison was made in cognitive range which means "the total set of cognitive functions, including knowledge, procedures, schemas, images, and self-regulatory skills, elicited by an item or descriptive of a domain". Messick (1987 as cited in Martinez, 1999) stated that, multiple-choice tests tend to emphasize knowledge and convergent thinking rather than synthesis and divergent thinking. According to Balch (1964 as cited in Martinez, 1999) multiple choice tests can be rewritten to make them having complex thought processes, i.e., inference.

Thus, while studying the current status of the biology education at the end of the basic education, it is worth dealing with the validity of the biology in terms of the content specifications, and dimensionality of the content with respect to cognitive processes assessed.

This study aimed to give evidence to the content validity of the biology items in the Secondary School Institutions Student Selection and Placement Tests (SSISSPTs) in terms of content and factor analyses. It also aimed to examine gender differences in biology achievement scores of students.

The content analysis in terms of subject matter revealed that this scale consisting of biology items is not representative of biology subjects indicated in the elementary science curriculum. A similar conclusion is made by Çoban (2001) after analyzing the subject matter of the science subtest in the 1998-2001 SSISSPTs with respect to number of lessons, subjects, goals, and objectives of each item required by curriculum. Although this can be considered as a threat to the content validity of the tests at the first hand, the nature of them that is measuring students' ability overcomes this problem. In other words, SSISSPT is an ability rather than achievement test and it is normal for it not to cover all biology subject matter. As appeared in the factor analyses, which gave evidence to the construct validity, the biology tests measure students' not directly measurable constructs, latent traits or attributes that are abilities of comprehension defined as lower level thinking skill, solving-problem and science process skills defined as higher order-thinking skills appropriate to the Martinez's (1999) terminology in this study, they were predicted as students' directly measurable test performances in students' test scores (Womer, 1968, Linn, 1989, Çalışkan, 2000). This nature was also supported by Aslan (2000) in his work of construct validity of the 1998 SSISSPT. Although he used the all science subscale data of 5598 examinees to run the factorial validity of the test, he found that dimensions consisted of verbal Turkish and social studies subscale items, and quantitative mathematics and science subscale items loaded nearly in the different factors. For example most of the Turkish and social studies items loaded in the same factor. But when he did the same thing in depth with the data of science subscale, the difficulty of science subscale items prevented him to name the factors according to the physics, chemistry, and biology dimensions. For example some of the physics and biology items loaded in the same factor.

On the other hand, content analysis in terms of cognitive processes revealed that the scale is measuring students' higher order as well as lower level thinking skills in Biology. The variety of cognitive processes measured by the biology items was also supported by Çepni et al. (2001) in their study. Although they did used the science items of 1998-2000 SSISSPTs as a whole, they found that there is no item measuring students' attainment of knowledge skills but comprehension, application and the complex of analysis, synthesis, and evaluation skills.

The factorial structure of biology items supported the dimensionality of the scale. It was difficult to name components of factor analyses of biology items of SSISSPTs in 1999 and 2001. Because biology items, which were found to be measuring higher order thinking skills, were loading in the same component with the other items classified in lower level cognitive processes. And as a result, these components were named either lower or higher than actual level of their items. That is to say the skills used by students to answer these biology items were somewhat different than those thought by experts. Because perceptions are considered to be different in both student and expert groups. So those factors included items with mixed cognitive processes.

This study is also seemed to support the evidence of cognitive processes on student achievement in 8th grade as in the study of Gürçay et al. (2001) that there is a relation between the order of cognitive domain and in terms of number of questions answered in each domain. The mean biology achievement score on science process skill questions is lower than that on the total comprehension questions, 0.4433 versus 0.3891). This taxonomy also appears in the mean scores of dimensions of comprehension questions. In the taxonomy, relation and comment skills come before translation skills, therefore the mean score on relation is higher than that on the both comment, and translation skills, .5745 versus .4060 and .1869, as it is expected (see Table19 on page 52).

In this study, small gender differences were found on achievement in Biology and its clusters. Boys got higher scores than girls on the Biology test, and its translation and comment clusters of the 2000 SSISSPT. But their score on translation cluster is significant. On the other hand, girls scored higher than boys on science process and relation clusters. But they only outperformed boys in relation cluster significantly. Male students were also found successful when compared to girls in the total science scale as noted by Aslan (2000). But, it should be mentioned that neither girls nor boys were successful in science and the mean difference between the Biology and Science scores of both genders was small. These findings are similar to the other studies of gender differences, but it should be mentioned that although the small gender difference in science is statistically significant, it is not thought to be practically significant. Aslan's study (2000) also supports this assumption, because when he factor analyzed the data of boys and girls separately, none of the groups showed different factorial structure with respect to their science scores evidencing the construct validity of the 1998 SSISSPT.

On the other hand, although this study does not aim to study misconceptions, the item analysis of the biology items within the science subscale of the tests shows some of them held by students at this grade level. The ITEMAN result of the 1998 SSISSPT by Aslan (2000) reveals some misconceptions. The first one is found in the answers of the first biology question. The question asks what the reason for if there are two pans of soups, but one of them gets spoiled easily because it has not any cover unlike the other. The correct answer of this item is A, the microorganisms in the air. While 13.4% of the students in the upper achievement group selected this, 31.9% of the students in the lower achievement group selected the B alternative, the temperature of the environment. But 24.3% and 12.1% of the students in the higher achievement group selected "C", the CO_2 gas in the air and "D", O_2 gas in the air.

18.3% of the students in the lower achievement group could not answer this item. This means students might have a misconception on this subject. They think the CO_2 and O_2 gases in the air can spoil soup (see item 17 in Appendices A and J on pages 75 and 88).

The other misconceptions are shown in the following two items. The first item asks the schematic diagram of a bird who can hold on a tree branch, feed with grains, and not swim. 38.2% of the higher achiever students answered this item correctly. 16.8% of the students in the same group selected D alternative. This means students have either a misconception or difficulty on this translation or both of them (see item 18 in Appendices A and J on pages 75 and 88).

The third misconception is seen in the items asking students to comment on a given skulls of an herbivorous and a carnivorous animals. Among the higher achiever group, 16.8% of the students answered this item correctly, but 40.2% selected A alternative (see item 19 in Appendices A and J on pages 75 and 88).

The last misconception can be the item asking what is the advantage of human heart with separated ventricles when comparing with turtle heart. Among the higher achiever students, 16.6 % answered it correctly, but 18.8% of them selected B alternative. This means students thinks the blood, which cleaned in the lungs, comes to left atrium as well as the blood going to aorta is rich in oxygen (see item 23 in Appendices A and J on pages 76 and 89).

The ITEMAN output of the 2000 test data shows only one misconception. Item 6 asks when four glass tubes are filled with some limewater, and the following living things placed on a perforated plate i.e., a bug, some germinating chicken peas, some chicken peas, and a few mushroom and the mouth of the tube is covered with a plastic cork, and after waiting a while the limewater in the tube with which animal or plant does not become turbid. 18.7% of the higher achiever students answered correctly, the limewater in the tube with chicken peas does not become turbid. But 25.6% of the students in the same group selected the limewater in the tube with germinating chicken peas does not become turbid (see item 6 in Appendices C and L on pages 79 and 92).

When the literature is reviewed considering students' misconceptions about these subjects, only circulatory system in human (Sungur et al., 2001), and respiration and photosynthesis (Çapa, 2000) were found studied. But studies on circulatory system were not dealing with comparing turtle and human hearts and therefore did not provide evidence to a possible misconception but a learning difficulty (Sungur, 2002). On the other hand, there were enough research studies on students' misconceptions about respiration and photosynthesis. It is probable that when answering this last item these students thought (1) "only animals must breathe (exchange gases)" (Wandersee, 1985 as cited in Çapa, 2000) and omitted the alternative A, (2) mushrooms cannot make photosynthesis and omitted the alternative D, (3) plant seeds can not make photosynthesis before germination and omitted the alternative C, which is also correct, and (4) plant seeds convert CO_2 into O_2 by photosynthesis when germinating and selected the alternative B.

The number 1 and 4 are misconceptions that students hold at this grade level. The reason of this misconception is related to instruction by teacher and textbook by Çapa (2000). These factors act like this way; textbooks represent green plants as making photosynthesis, teachers add animals as making respiration and show the relationship between plants and animals in terms of energy cycle by drawing a figure, and give importance to the lower order thinking skills such as the knowledge of formula of both respiration and photosynthesis rather than the understanding of entire process which requires higher order thinking skills because of limited lesson time. They may be remedied by content sequencing and integrating physics and chemistry knowledge with biology in order to students gain a holistic view on photosynthesis and respiration concepts which take place in the cell and cellular activities subjects of the Living Things and Life unit of the elementary science curriculum in Turkey. Because, students who could not know the difference between organic and inorganic materials, energy and food concept, and chemical conventions face difficulty in comprehending photosynthesis and respiration.

6.1 – Implications for Research

The present research was conducted with the data of a representative sample of students taking place in the Secondary School Institutions Student Selection and Placement Tests (SSISSPTs) which aim to select and place them in further school type through the years of 1998-2001 at the end of their 8th grade education. To generalize the results of this study to the local, provincial and national level, it should be better to include all students who will take a biology ability test like the one in the SSISSPT but with more items.

As seen in the example of the both NAEP's science tests administered to 5th, 8th, and 11th grade students representative of local, state and national level in the USA and EARGED's tests administered to these three grade students in curriculum laboratory schools of Turkey, it is suggested that for giving evidence to the predictive validity of the SSISSPT, students' previous and further performance in biology and science should be assessed with another ability tests at the end of the grade 5 and 11 in other words at the end of the first period of elementary and secondary schooling. This type of study will possibly give educators the chance of cross-national comparisons as in the International Mathematics and Science Studies known as the TIMSS.

As the content analysis of biology items within the science subscale of 1998-2001 SSISSTs in terms of subject matter, the only particular biology units i.e. Living Things and Life, and Diversity of Living Things are asked. Therefore the content of the test should be widened and cover other and currently discussed biology units such as Human and Environment.

This research showed that the biology achievement level is low in the 8th grade students. Girls when comparing to boys are able have higher scores on some of the clusters in biology test of the SSISSPT. This situation should not go so far to a possible stereotyping. Girls should not be limited by reading and speaking activities in biology and science classes. On the other hand, boys should not be limited by problem-solving activities in those classes. While female students should be engaged in physics and chemistry classes, male students should be done in biology class. Teachers of biology and science should emphasize role modeling through giving example of famous woman and man scientists.

As the analyses of the biology items in the SSISSPT through the years of 1998-2001, some of the students have misconceptions on one-celled organisms, animals, human circulatory system, and photosynthesis-respiration, new studies should be performed and the ongoing ones on these subjects should be supported. The importance of comparison of national data at different grade levels, which is stated beforehand, will possibly answer which and when a misconception is held. Because achievement and becoming misconception-free in biology and science classes depends on students' cognitive processes and level of cognitive development. Therefore interview technique should be used to understand students' thought processes and remedy their possible misconceptions.

REFERENCES

American Psychological Association (APA). Standards for Educational and Psychological Tests and Manuals. 1966. APA, INC. Washington, D.C.

Askling, B. & Härnqvist, K. (1989). The NAEP Science Report Card: what does it report? Essay Reviews, 87-93.

Aslan, Ö. (2000). The Construct Validity of the 1998 Secondary School Institutions Student Selection and Placement Test. Unpublished MSc Thesis, METU, Ankara.

Ateş, S., Durmuş, S., & Bahar, M. (2001) Bilimsel düşünme yetenekleri ve bu yeteneklerin ölçülmesi. Yeni Binyılın Başında Türkiye'de Fen Bilgisi Eğitimi Sempozyumu Bildiriler. Maltepe Üniversitesi Eğitim Fakültesi, p. 535-538.

Bahar, M., Johnstone, A. H., & Hansell M. H. (1999). Revisiting learning difficulties in biology. Journal of Biological Education, 33(2), 84-86.

Bloom, B. S., Hastings, J.T., & Madaus, G. F. (1971). Handbook of Formative and Summative Evaluation of Student Learning. McGraw-Hill, Inc.

Borst, W. U. (1996). Guidelines for writing in APA style. Unpublished manuscript, Troy State University at Phoenix City. Retrieved January 6, 2002, from http://www.wooster.edu/psychology/apa-crib.html

Brennan, R.T., Kim, J., Wenz-Gross, M., & Superstein, G.N. (2001). The relative equitability of high-stakes testing versus teacher assigned grades: An analysis of the Massachusetts comprehensive assessment system (MCAS). Harvard Educational Review, 71(2), Summer 2001.

Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. Holt, Rinehart and Winston Inc., p: 289-292.

Çalışkan, M. (2000) The Fit Of One-, Two and Three Parameter Models of Item Response Theory (IRT) to the Ministry of National Education-Educational Research And Development Directorate's (MNE-ERDD) Science Achievement Test Data. Unpublished MSc Thesis, METU, Ankara, p. 1-3, 9-12.

Çapa, Y. (2000) An Analysis of the Ninth Grade Students' Misconceptions Concerning Photosynthesis and Respiration in Plants. Unpublished MSc Thesis, METU, Ankara.

Çepni, S., Ayvacı, H. Ş., & Keleş, E. (2001). Okullarda ve lise giriş sınavlarında sorulan fen bilgisi sorularının Bloom taksonomisine gore karşılaştırılması. Yeni Binyılın Başında Türkiye'de Fen Bilgisi Eğitimi Sempozyumu Bildiriler. Maltepe Üniversitesi Eğitim Fakültesi, p. 28-33.

Çoban, A. (2001) Fen bilgisi dersinin ilköğretim programları ve liselere giriş sınavları açısından değerlendirilmesi. Yeni Binyılın Başında Türkiye'de Fen Bilgisi Eğitimi Sempozyumu Bildiriler. Maltepe Üniversitesi Eğitim Fakültesi, p. 144-150.

EARGED. (1995). Eğitimi Araştırma ve Geliştirme Dairesinin Durum Tesbiti Faaliyetleri ve Fen Bilgisi Durum Tesbiti Sonuçları.

EĞİTEK. (1998). 1998 Yılı Merkezi Sistem Sınavları İstatistik Bilgileri. T.C. Milli Eğitim Bakanlığı, Eğitim Teknolojileri Genel Müdürlüğü Ölçme Değerlendirme Daire Başkanlığı. Kasım 1998, Ankara.

EĞİTEK. (1999). 1999 Yılı Merkezi Sistem Sınavları İstatistikleri. T.C. Milli Eğitim Bakanlığı, Eğitim Teknolojileri Genel Müdürlüğü Ölçme Değerlendirme Daire Başkanlığı. Aralık 1999, Ankara.

EĞİTEK. (2000). 2000 Merkezi Sistem Sınavları Sayısal Veriler. T.C. Milli Eğitim Bakanlığı, Eğitim Teknolojileri Genel Müdürlüğü Ölçme Değerlendirme Daire Başkanlığı. Kasım 2000, Ankara.

German, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. Journal Of Research in Science Teaching, 25(8), 689-703.

Gredler, M. E. (1992). Learning and Instruction Theory Into Practice. Second Edition. Macmillan Publishing Company, New York, p: 216-261.

Gürçay, D., Bozkurt, A. İ., Kaptan, F., & Berberoğlu, G. (2001). Öğretmen adaylarının fen derslerinde değişik taksonomik düzeylerdeki başarılarının değerlendirilmesi. IV Fen Bilimleri Eğitimi Kongresi' 2000 Bildiriler. Ankara, MEB Basımevi, 131-134.

Harwell, M. (1999). Evaluating the validity of educational rating data. Educational and Psychological Measurement, 59(1), 25-37.

Hoste, R. (1977). The Construct Validity of Some CSE Physics Examination. School Science Review, 59(206), 127-131.

Kaptan F., & Korkmaz, H. (2001). Mevcut Fen Bilgisi Programı ile 2001-2002 Öğretim Yılında Uygulamaya Konulacak Olan Yeni Fen Bilgisi Programının Karşılaştırılması. Çağdaş Eğitim, Şubat 2001, 273, 33-38.

Klopfer, L. E. (1971) Evaluation of Learning in Science In Handbook Of Formative And Summative Evaluation Of Student Learning. Bloom, B. S., Hastings, J.T., & Madaus, G. F. (1971). McGraw-Hill, Inc. 562-641.

Lawson, A. E., & Wollman, W. T. (1976). Encouraging the transition from concrete to formal cognitive functioning-an experiment. Journal of Research in Science Teaching, 13(5), 413-430.

Lazarowitz, R., & Penso, S. (1992). High school students' difficulties in learning biology concepts. Journal of Biological Education, 26(3), 215-223.

Linn, R. L. (1989). Educational Measurement. Third Edition, American Council on Education and Macmillan Publishing Company, New York p. 147-149..

Marsh, J. F., & Anderson, N. D. (1989). An assessment of the quantitative skills of students taking introductory college biology courses. Journal of Research in Science Teaching, 26(9), 757-769.

Martinez, M. E. (1999). Cognition and the Question of Test Item Format. Educational Psychologist, 14(4), 207-218.

MEB (2000). İlköğretim Okulu Fen Bilgisi Dersi (4, 5, 6, 7 ve 8. Sınıf) Öğretim Programlarının Kabulü. MEB Tebliğler Dergisi, Kasım 2000, 2518, 1000-1105.

MEB (2001). Retrieved July 30, 2001 from http://www.meb.gov.tr/index2.htm

Millman, J., & Greene, J. (1989). The Specification and Development of Tests of Achievement and Ability. In Educational Measurement. Linn, R. L. (1989). Third Edition, American Council on Education and Macmillan Publishing Company, New York. 335-366.

Molitor, L. L., & George, K. D. (1976). Development of a test of science process skills. Journal of Research in Science Teaching, 13, 405-412.

Padilla, J. M., Okey, J. R., & Dillashaw, F. G. (1983). The relationship between science process skill and formal thinking abilities. Journal of Research in Science Teaching, 20, 239-246.

Powell, T. W. (1993). Factorial validity of the test of problem solving. Perceptual And Motor Skills, 76, 753-754.

Rakow, S. J., Welch, W. W., & Hueftle, S. J. (1984). Student achievement in science: a comparison of national assessment results. Science Education, 68(5), 571-578.

Rudner, L. M. (1994). Questions to ask when evaluating tests [Electronic version]. Practical Assessment, Research & Evaluation, 4(2). Retrieved from http://ericae.net/pare/getvn.asp?v=4&n=2

Shayer, M., Adey, P., & Wylam, H. (1981). Group tests of cognitive development ideals and a realization. Journal Of Research in Science Teaching, 15(2), 157-168.

Smith, A. K., & Welliver, P. W. (1990). The development of a science process assessment for fourth grade students. Journal of Research in Science Teaching, 27(8), 727-738.

Steinkamp, M. W., & Maehr, M. L. (1984). Gender differences in motivational orientations toward achievement in school science: a quantitative synhesis. American Educational Research Journal, 21(1), 39-59.

Sungur, S., Tekkaya. C., & Geban, Ö. (2001). Lise öğrencilerinin insanda dolaşım sistemi konusundaki kavram yanılgılarının belirlenmesi. IV. Fen Bilimleri Kongresi' 2000. Bildiriler. Ankara, MEB Basımevi, p.1-4.

Sungur, S. (2002). Personal communication.

Tannenbaum, R. S. (1971). The development of the test of science processes. Journal of Research in Science Teaching, 8(2), 123-136.

T.C. MEB İlköğretim Genel Müdürlüğü. (1995). İlköğretim Okulu Programı. MEB Yayınları 2846, Mevzuat Dizisi 297, Ankara, 293-325.

TIMSS (2001) Women, minorities, and persons with disabilities in science and engineering 1998-Chapter2: Precollege Education Influences On Mathematics And Science Achievement. Retrieved July 6 2001 from http://www.c2s2.htm. #note1

Vural, M. (2000). İlköğretim Fen Dersi Programı. İlköğretim Okulu Programı. 5. Baskı. Yakutiye Yayıncılık, Erzurum, 151-186.

Wiggins, G. (1989). A true test: toward more authentic and equitable assessment. Phi Delta Kappan, May 1989, 703-713.

Womer, F. B. (1968). Kuder-Richardson method. Basic Concepts in Testing. Guidancce Monograph Series Series III: Testing. In Stone S.C. & Shertzer B. editors. Houghton Mifflin Company, Boston, p: 36-37.

Yürük, N. & Çakır, Ö. S. (2000). Lise öğrencilerinde oksijenli ve oksijensiz solunum konusunda görülen kavram yanılgılarının saptanması. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi 18, 185-191.

APPENDIX A

1998 ITEMAN Results

Table A.1
Item analysis for data of the 1998 SSISSPT

	Item Statistics				Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
17	0-1	0.134	0.122	0.077	A	0.134	0.122	0.077	*
					B	0.319	-0.126	-0.097	
					C	0.243	0.018	0.013	
					D	0.121	0.087	0.054	
					Other	0.183	-0.017	-0.011	
18	0-2	0.382	0.266	0.209	A	0.394	-0.210	-0.165	
					B	0.027	-0.268	-0.103	
					C	0.382	0.266	0.209	*
					D	0.168	0.030	0.020	
					Other	0.029	-0.176	-0.070	
19	0-3	0.168	0.086	0.057	A	0.402	0.239	0.189	?
	CHECK THE KEY				B	0.168	0.086	0.057	*
	B was specified, A works better				C	0.061	-0.276	-0.139	
					D	0.312	-0.203	-0.155	
					Other	0.057	-0.080	-0.040	
20	0-4	0.143	0.161	0.104	A	0.166	-0.180	-0.121	
					B	0.143	0.161	0.104	*
					C	0.184	-0.080	-0.055	
					D	0.163	-0.063	-0.042	
					Other	0.344	0.123	0.095	
21	0-5	0.531	0.159	0.127	A	0.531	0.159	0.127	*
					B	0.107	-0.144	-0.086	
					C	0.139	-0.116	-0.075	
					D	0.093	-0.067	-0.038	
					Other	0.130	0.000	0.000	
22	0-6	0.372	0.453	0.355	A	0.156	-0.045	-0.030	
					B	0.136	-0.260	-0.166	
					C	0.372	0.453	0.355	*
					D	0.247	-0.150	-0.110	
					Other	0.089	-0.349	-0.197	

Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
	Item Statistics				Alternative Statistics				
23	0-7	0.166	0.173	0.116	A	0.110	-0.120	-0.072	
					B	0.188	0.058	0.040	
					C	0.166	0.173	0.116	*
					D	0.137	-0.242	-0.154	
					Other	0.399	0.044	0.035	
24	0-8	0.333	0.525	0.405	A	0.333	0.525	0.405	*
					B	0.131	-0.202	-0.127	
					C	0.106	-0.249	-0.148	
					D	0.279	-0.120	-0.090	
					Other	0.151	-0.264	-0.173	
25	0-9	0.322	0.342	0.263	A	0.108	-0.219	-0.131	
					B	0.175	-0.164	-0.111	
					C	0.321	-0.044	-0.034	
					D	0.322	0.342	0.263	
					Other	0.074	-0.173	-0.092	

Note: There were 5598 examinees in the data file. From "ITEMAN analysis results" by Ö. Aslan, 2000. The Construct Validity of 1998 SSISSPT, p.95-96.

APPENDIX B

1999 ITEMAN Results

Table B.1
Item analysis for data of the 1999 SSISSPT

Item Statistics					Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
17	0-1	0.257	0.578	0.426	A	0.118	-0.193	-0.118	
					B	0.257	0.578	0.426	*
					C	0.248	-0.123	-0.090	
					D	0.178	-0.096	-0.066	
					Other	0.199	-0.301	-0.210	
18	0-2	0.633	0.490	0.383	A	0.058	-0.245	-0.121	
					B	0.083	-0.283	-0.157	
					C	0.633	0.490	0.383	*
					D	0.147	-0.289	-0.188	
					Other	0.078	-0.317	-0.173	
19	0-3	0.361	0.678	0.529	A	0.102	-0.142	-0.083	
					B	0.112	-0.160	-0.097	
					C	0.361	0.678	0.529	*
					D	0.147	-0.235	-0.153	
					Other	0.277	-0.430	-0.322	
20	0-4	0.401	0.599	0.473	A	0.068	-0.229	-0.119	
					B	0.078	-0.220	-0.120	
					C	0.241	-0.164	-0.119	
					D	0.401	0.599	0.473	*
					Other	0.212	-0.409	-0.290	
21	0-5	0.603	0.589	0.464	A	0.111	-0.310	-0.187	
					B	0.175	-0.301	-0.204	
					C	0.603	0.589	0.464	*
					D	0.043	-0.248	-0.112	
					Other	0.068	-0.517	-0.270	
22	0-6	0.252	0.633	0.466	A	0.099	-0.194	-0.113	
					B	0.252	0.633	0.466	*
					C	0.103	-0.188	-0.111	
					D	0.263	-0.083	-0.061	
					Other	0.283	-0.318	-0.239	

Item Statistics					Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
23	0-7	0.344	0.578	0.448	A	0.344	0.578	0.448	*
					B	0.137	-0.186	-0.119	
					C	0.070	-0.255	-0.134	
					D	0.305	-0.182	-0.139	
					Other	0.144	-0.324	-0.209	
24	0-8	0.468	0.606	0.483	A	0.468	0.606	0.483	*
					B	0.185	-0.258	-0.178	
					C	0.071	-0.257	-0.136	
					D	0.175	-0.200	-0.135	
					Other	0.101	-0.483	-0.284	
25	0-9	0.306	0.584	0.444	A	0.299	-0.050	-0.038	
					B	0.124	-0.278	-0.172	
					C	0.306	0.584	0.444	*
					D	0.097	-0.165	-0.096	
					Other	0.174	-0.398	-0.270	

Table B.2
Scale statistics

Scale:	0
N of Items	9
N of Examinees	10000
Mean	3.625
Variance	3.778
Std. Dev.	1.944
Skew	0.423
Kurtosis	-0.213
Minimum	0.000
Maximum	9.000
Median	3.000
Alpha	0.527
SEM	1.337
Mean P	0.403
Mean Item-Tot.	0.457
Mean Biserial	0.593

Note: There were 10000 examinees in the data file.

APPENDIX C

2000 ITEMAN Results

Table C.1
Item analysis for data of the 2000 SSISSPT

		Item Statistics			Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
1	0-1	0.601	0.541	0.426	A	0.115	-0.296	-0.180	
					B	0.601	0.541	0.426	*
					C	0.225	-0.390	-0.280	
					D	0.047	-0.275	-0.128	
					Other	0.012	-0.230	-0.066	
2	0-2	0.282	0.619	0.465	A	0.111	-0.140	-0.085	
					B	0.150	-0.161	-0.105	
					C	0.282	0.619	0.465	*
					D	0.182	-0.246	-0.168	
					Other	0.275	-0.241	-0.180	
3	0-3	0.460	0.579	0.461	A	0.460	0.579	0.349	*
					B	0.223	-0.142	-0.102	
					C	0.050	-0.315	-0.149	
					D	0.130	-0.406	-0.256	
					Other	0.137	-0.315	-0.201	
4	0-4	0.529	0.605	0. 483	A	0.154	-0.306	-0.201	
					B	0.131	-0.283	-0.178	
					C	0.089	-0.252	-0.142	
					D	0.529	0.605	0.483	*
					Other	0.097	-0.394	-0.229	
5	0-5	0.468	0.497	0.396	A	0.227	-0.260	-0.187	
					B	0.066	-0.250	-0.129	
					C	0.112	-0.315	-0.190	
					D	0.468	0.601	0.479	*
					Other	0.127	-0.329	-0.206	
6	0-6	0.187	0.337	0.232	A	0.174	-0.091	-0.062	
					B	0.256	0.124	0.091	
					C	0.187	0.337	0.232	*
					D	0.137	-0.164	-0.104	
					Other	0.246	-0.226	-0.165	

		Item Statistics				Alternative Statistics			
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
7	0-7	0.620	0.566	0.444	A	0.072	-0.304	-0.161	
					B	0.127	-0.213	-0.133	
					C	0.620	0.566	0.444	*
					D	0.110	-0.351	-0.211	
					Other	0.070	-0.470	-0.248	
8	0-8	0.402	0.563	0.444	A	0.111	-0.206	-0.124	
					B	0.152	-0.225	-0.148	
					C	0.402	0.563	0.444	*
					D	0.200	-0.264	-0.185	
					Other	0.134	-0.240	-0.152	
9	0-9	0.334	0.633	0.488	A	0.244	-0.145	-0.106	
					B	0.334	0.668	0.515	*
					C	0.128	-0.321	-0.201	
					D	0.121	-0.237	-0.146	
					Other	0.173	-0.323	-0.219	

Note: There were 10000 examinees in the data file.

Table C.2
Scale statistics

Scale:	0
N of Items	9
N of Examinees	10000
Mean	3.884
Variance	3.590
Std. Dev.	1.185
Skew	0.283
Kurtosis	-0.364
Minimum	0.000
Maximum	9.000
Median	4.000
Alpha	0.486
SEM	1.359
Mean P	0.432
Mean Item-Tot.	0.439
Mean Biserial	0.564

APPENDIX D

2001 ITEMAN Results

Table D.1
Item analysis for data of the 2001 SSISSPT

	Item Statistics				Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
17	0-1	0.677	0.704	0.541	A	0.044	-0.379	-0.172	
					B	0.102	-0.351	-0.206	
					C	0.677	0.704	0.541	*
					D	0.107	-0.405	-0.241	
					Other	0.071	-0.596	-0.315	
18	0-2	0.346	0.539	0.418	A	0.138	-0.168	-0.107	
					B	0.346	0.539	0.418	*
					C	0.225	-0.132	-0.095	
					D	0.168	-0.234	-0.157	
					Other	0.123	-0.314	-0.195	
19	0-3	0.432	0.701	0.556	A	0.129	-0.277	-0.174	
					B	0.132	-0.214	-0.135	
					C	0.147	-0.260	-0.169	
					D	0.432	0.701	0.556	*
					Other	0.160	-0.459	-0.305	
20	0-4	0.295	0.545	0.412	A	0.113	-0.137	-0.083	
					B	0.187	-0.063	-0.044	
					C	0.295	0.545	0.412	*
					D	0.228	-0.077	-0.055	
					Other	0.176	-0.469	-0.319	
21	0-5	0.351	0.671	0.521	A	0.296	-0.046	-0.035	
					B	0.090	-0.305	-0.173	
					C	0.114	-0.309	-0.187	
					D	0.351	0.671	0.521	*
					Other	0.149	-0.534	-0.348	
22	0-6	0.381	0.702	0.551	A	0.120	-0.181	-0.112	
					B	0.381	0.702	0.551	*
					C	0.140	-0.204	-0.131	
					D	0.111	-0.231	-0.140	
					Other	0.247	-0.448	-0.328	

Item Statistics					Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
23	0-7	0.586	0.780	0.617	A	0.082	-0.371	-0.205	
					B	0.098	-0.358	-0.208	
					C	0.101	-0.356	-0.209	
					D	0.586	0.780	0.617	*
					Other	0.133	-0.571	-0.361	

Table D.2
Scale Statistics

Scale:	0
N of Items	7
N of Examinees	5000
Mean	3.068
Variance	3.011
Std. Dev.	1.735
Skew	0.134
Kurtosis	-0.730
Minimum	0.000
Maximum	7.000
Median	3.000
Alpha	0.545
SEM	1.171
Mean P	0.438
Mean Item-Tot.	0.517
Mean Biserial	0.663

Note: There were 5000 examinees in the data file.

APPENDIX E

Taxonomy of Educational Objectives by MONE*

1.0 KNOWLEDGE
- 1.01 Knowledge of facts
- 1.02 Knowledge of terminology
- 1.03 Knowledge of concepts
- 1.04 Knowledge of classifications
- 1.05 Knowledge of sequences
- 1.06 Knowledge of techniques and procedures
- 1.07 Knowledge of principles and laws
- 1.08 Knowledge of theories

2.0 COMPREHENSION
- 2.01 Identification of knowledge in a new context
- 2.02 Explanation
- 2.03 Summarization
- 2.04 Giving example
- 2.05 Identification of criteria in a given information
- 2.06 Relation
- 2.07 Comment
- 2.08 Translation of knowledge from one form into another
- 2.09 Prediction

3.0 SOLVING PROBLEM
- 3.01 Finding required principle for solution
- 3.02 Bringing knowledge, law, and principle together
- 3.03 Using formula and algorithms
- 3.04 Using units correctly and making transition
- 3.05 Show the answer in required form

4.0 SCIENCE PROCESS
- 4.01 Observing
- 4.02 Description of observation
- 4.03 Comparison of results of observation
- 4.04 Classification of results of observation
- 4.05 Selection of appropriate measuring instrument
- 4.06 Recognition of a problem
- 4.07 Relation between elements of problem
- 4.08 Formulation of a hypothesis to solve the problem
- 4.09 Proposing procedure to verify (test) the hypothesis
- 4.10 Design of experiment/think/formulate in figure
- 4.11 Collection of data
- 4.12 Processing/analyzing/interpreting data
- 4.13 Discussing and evaluating hypotheses according to evidences
- 4.14 Conclusion according to evidences
- 4.15 Generalization (formulating models) and proposing new research questions
- 4.16 Application of observation and research results to daily life or new situation

(*) It is developed from "Taxonomy of Educational Objectives" by the MONE, EARGED, 1995, 43 using Bloom's terminology.

APPENDIX F

Subject Matter List**

Units
1. Human & Environment
   1.01 Pollution Types
   1.02 Pollution & Health
   1.03 Humans' Effects on Environment
   1.04 Matter Cycle
   1.05 Reasons of Environmental Damage
   1.06 Sustainable Development
2. Reproduction & Inheritance in Living Things
   2.01 Cell & Cell Division
   2.02 Reproduction in Living Things
   2.03 Inheritance
   2.04 Inheritance & Environment
3. Knowing Our Body
   3.01 Sensory Organs
   3.02 Motion System
   3.03 Digestion System ve Feeding
   3.04 Circulation System
   3.05 Respiration System
   3.06 Excretion System
   3.07 Nervous System
   3.08 Hormones& Hormone Producing Organs
   3.09 Immune System
4. Diversity of Living Things
   4.01 Plants
   4.02 Animals
   4.03 Fungi, One-Celled Organisms, & Bacteria
   4.04 Interactions Between Living Things
   4.05 Biologic Wealth of Turkey
5. Living Things &Life
   5.01 The World of Living Things
   5.02 Life in Plants
   5.03 Life in Animals
   5.04 Cell & Cellular Activities
   5.05 Growth & Development in Humans
6. Other

(**) The previous elementary science curriculum was used when preparing this list.

APPENDIX G

Subject Matters of Biology Items

Year	Item	Subject Matter
1998	17	4.03
	18	4.02
	19	5.03
	20	5.02
	21	5.04
	22	3.07
	23	4.02
	24	5.02
	25	2.03
1999	17	6
	18	3.04
	19	2.04
	20	4.04
	21	3.07
	22	5.04
	23	5.04
	24	2.02
	25	5.02
2000	1	4.02
	2	4.04
	3	3.03
	4	2.04
	5	5.02
	6	5.04
	7	5.03
	8	5.04
	9	5.04
2001	17	4.04
	18	5.02
	19	5.04
	20	2.04
	21	4.03
	22	4.04
	23	3.09

APPENDIX H

Cognitive Processes of Biology Items After Experts' Ratings

Year	Item	Experts I	II	III	IV
1998	17	2.02	2.06	2.02	2.01
	18	2.05	4.14	2.07	2.08
	19	2.07	4.12	2.09	2.09
	20	2.04	4.05	2.04	1.04
	21	2.09	4.14	2.09	1.07
	22	2.04	2.09	2.07	2.05
	23	2.07	1.03	2.09	2.07
	24	4.06	4.06	2.07	4.06
	25	2.02	3.02	2.07	2.01
1999	17	4.06	4.04	2.09	4.06
	18	3.02	2.08	2.07	3.02
	19	2.04	2.04	2.04	2.05
	20	2.07	4.14	2.07	2.01
	21	2.04	2.04	2.01	1.04
	22	2.07	4.14	2.07	4.12
	23	4.06	4.13	2.07	4.06
	24	2.04	2.01	2.07	2.01
	25	4.10	4.10	2.09	4.10
2000	1	1.04	4.14	2.01	1.04
	2	2.06	4.14	2.07	2.01
	3	4.10	4.08	2.09	4.10
	4	2.04	2.04	2.04	2.05
	5	4.09	4.10	2.07	4.09
	6	2.09	4.03	2.09	1.07
	7	4.12	4.12	2.07	4.15
	8	4.06	4.06	2.09	4.06
	9	2.08	4.12	2.07	4.12
2001	17	2.09	2.07	2.07	1.03
	18	4.06	4.06	2.07	4.06
	19	2.07	4.14	2.07	4.12
	20	2.06	2.09	2.07	2.05
	21	2.07	4.12	2.07	4.12
	22	4.12	4.12	2.07	4.14
	23	2.07	4.12	2.07	4.12

APPENDIX I

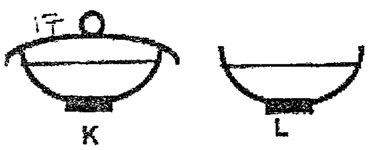
Cognitive Processes of Items After Agreement and Suggestion

Year	Item	Agreement	Suggestion
1998	17	2.02	
	18	2.08	
	19	2.07	
	20	2.04	
	21	2.06	
	22	2.04	
	23	2.07	
	24	4.06	
	25	2.02	
1999	17	4.06	
	18	3.02	
	19	2.04	
	20	2.07	
	21	2.04	
	22	2.07	
	23	4.06	
	24	2.01	
	25	4.10	
2000	1	1.04	2.07
	2	2.07	
	3	4.10	4.08
	4	2.05	2.06
	5	4.09	4.08
	6	2.07	2.06
	7	4.12	2.06
	8	4.06	4.08
	9	2.07	2.08
2001	17	2.09	
	18	4.06	
	19	4.12	2.08
	20	2.06	
	21	4.12	2.08
	22	2.09	
	23	4.12	2.08

APPENDIX J

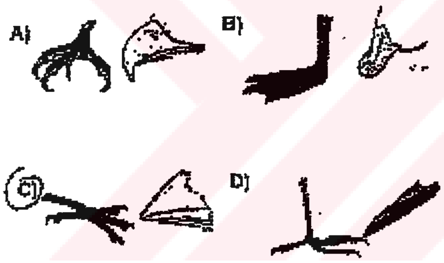
Biology Items In The 1998 SSISSPT

17.

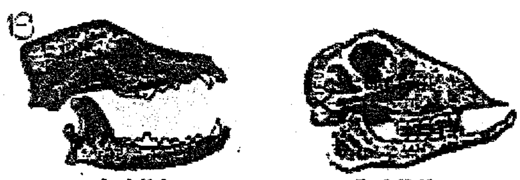


Aynı ortamda bulunan özdeş K ve L tencerelerine aynı çorbadan eşit miktarda konmuştur. Kapaksız L kabındaki çorbanın, kapağı kapatılmış K kabındaki çorbadan daha önce bozulduğu gözlenmiştir.
L kabındaki çorbanın daha çabuk bozulmasının nedeni aşağıdakilerden hangisidir?
A) Havadaki mikroplar B) Ortamın sıcaklığı C) Havadaki CO_2 gazı D) Havadaki O_2 gazı

18. *Ağaç dallarına tutunabilme
*Taneli besinlerle beslenme
*Yüzememe
Yukarıdaki özelliklere sahip bir kuşun gaga ve ayak yapısı aşağıdakilerden hangisi olabilir?



19.

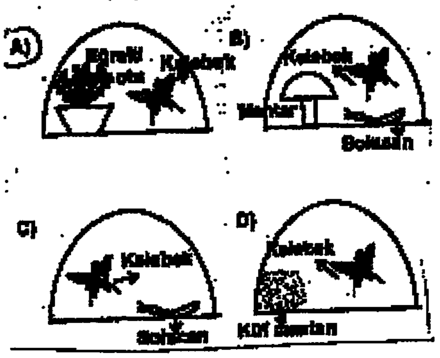


Şekil I Şekil II

Şekil I ve II'de farklı hayvanlara ait çene yapıları verilmektedir. Şekilleri inceleyen bir öğrenci bu hayvanlara ait aşağıdaki tahminlerden hangisini yapamaz?
A) Ayak şekilleri B) Hangi hayvanlar oldukları
C) Beslenme şekilleri D) Midelerinin bölümlü olup olmadığı

20. Bitki hücresinin karakteristik özelliklerini görmek isteyen bir öğrenci, aşağıdakilerden hangisini incelemelidir?
A) Buğdaydaki emici tüy hücresini B) Nohut bitkisinin yaprak hücresini
C) Fasulye bitkisinin taç yaprak hücresini D) Gül bitkisinin taç yaprak hücresini

21 Işıklı ortamda, içerisinde hava bulunan aşağıdaki özdeş cam fanusların hangisindeki kelebek <u>daha uzun</u> süre yaşar?



22. Kalıtsal refleksler bütün insanlarda ortak ve doğuştan olup; şartlı refleksler ise sonradan elde edilen deneyimlerdir. Buna göre aşağıdakilerden hangisi <u>şartlı reflekstir</u>?

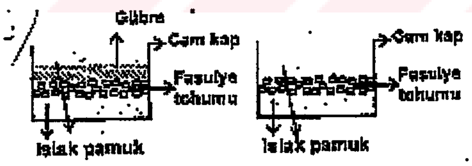
A) Ani bir patlama sesi karşısında sıçrama
B) Bacağa iğne battığında bacağın aniden çekilmesi
C) Çocuğun yanan sobadan kendini sakınması
D)Loş ışıkta göz bebeklerinin büyümesi

23. Aşağıdaki şekillerde, insan kalbinde karıncıkların tam bölme ile ayrıldığı, kaplumbağa kalbinde ise karıncıkların yarım bölme ile ayrıldığı görülmektedir.

Bu durum, insana kaplumbağadan farklı olarak ne sağlar?

A) Aorttaki kanın akış hızının kılcaldamarlardakinden farklı olmasını
B) Akciğerde temizlenen kanın sol kulakçığa gelmesini
C) Aorta giden kanın oksijence zengin olmasını
D) Kanın kalbe toplardamarlarla gelmesini

24.



Bir öğrenci şekildeki deney düzeneklerini hazırlayarak ışıklı bir ortama bırakıp gözlem yapıyor. Öğrenci bu deneyde aşağıda verilenlerden hangisini araştırıyor olabilir?

A) Tohum çimlenirken dışarıdan besin alır mı?
B) Tohum çimlenirken havaya ihtiyaç duyar mı?
C) Suyun çimlenmeye etkisi var mı?
D) Işık çimlenmeye etki eder mi?

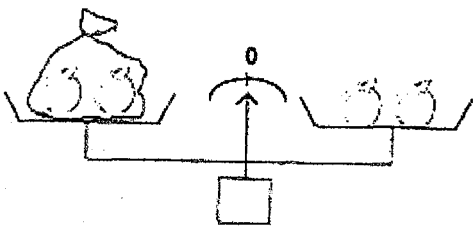
25. Kıvırcık saçlı anne ile kıvırcık saçlı babanın düz saçlı bir çocuğu olmuştur. Çocuğun düz saçlı olmasında aşağıdakilerden hangisinin etkili olduğu söylenebilir?

A) Kıvırcık saçlılığın çekinik olması
B) Düz saçlılığın baskın olması
C) Anne veya babada düz saçlılık geninin bulunması
D) Anne ile babada düz saçlılık geninin bulunması

APPENDIX K

Biology Items In The 1999 SSISSPT

17.



Yukarıdaki terazinin kefelerinde ağırlıkları eşit olmak üzere açıkta ve şeffaf naylon torba içinde elmalar bulunmaktadır. Bir süre bu sistemde gözlem yapan bir kişi aşağıdaki hangi soruya doğrudan cevap veremez?

A) Havayla teması azaltılan besinler daha uzun süre mi dayanır?
B) Havayla teması azaltılan besinler vitaminlerini daha çok mu korur?
C) Havayla teması azaltılan besinler daha mı az kütle kaybeder?
D) Havayla teması azaltılan besinler daha mı az su kaybeder?

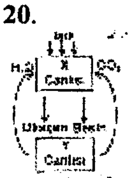
18. Ayşe, Mehmet'e kan verebiliyor. Nalan, Mehmet ve Ahmet'e kan verebiliyor. Ancak, Ahmet, Mehmet'ten kan alamıyor. Buna göre, hangi iki kişinin kan grubu kesinlikle aynı olamaz?

A) Ayşe-Mehmet B) Ahmet-Nalan C) Mehmet-Ahmet D) Nalan-Ayşe

19. Bir canlı türünün farklı ortamlarda yaşayan bireyleri arasında farklılıklar gözlenmektedir. Aşağıdakilerden hangisi buna örnektir?

A) İnsan populasyonunda bazı bireylerin farklı boyda olması
B) Bir köpek populasyonundaki bazı bireylerin farklı desende posta sahip olmaları
C) Bir bitki türü populasyonuna ait bireylerin yüksek bölgelerde kısa, vadilerde uzun boylu olması
D) İnsan populasyonunda, bireyler arasında farklı kan gruplarının olması

20.



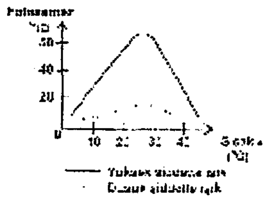
Yukarıdaki şemayı inceleyen bir öğrenci aşağıdakilerden hangisine ulaşamaz?

A) X canlısının üretici olduğuna B) Y canlısının tüketici olduğuna
C) X ve Y canlısının ortak yaşadığına D) Y canlısının otçul olduğuna

21. Aşağıdakilerden hangisi insanda doğuştan gelen bir reflekstir?

A) Daha önceden eli yanan çocuğun sıcak sobadan uzaklaşması
B) Günde üç öğün yemek yiyen insanın, öğün vakti geldiğinde açlık hissetmesi
C) Karanlık ortamdan aydınlık ortama aniden geçildiğinde gözlerin kısılması
D) Keman çalmayı unutmuş olan birinin bir sure sonra tekrar çalabilmesi

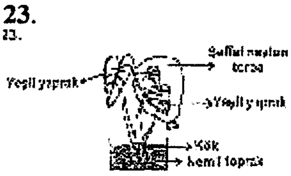
22.



Fotosentez hızının sıcaklığa bağlı değişimi grafikteki gibidir. Yeşil bir bitki aşağıdaki koşulların hangisinde bulunduğunda fotosentez hızı <u>en fazla</u> olur?

A) 10 ^{0}C düşük şiddette ışıkta
B) 25 ^{0}C yüksek şiddette ışıkta
C) 30 ^{0}C düşük şiddette ışıkta
D) 40 ^{0}C yüksek şiddette ışıkta

23.



Şekildeki gibi bir saksı bitkisi üzerine şeffaf naylon torba geçiren bir öğrenci saksıyı bir süre soğuk, sonra ılık ve daha sonra da sıcak ortamda bekletiyor.
Öğrenci bu düzenekle bitkide hangi olayı araştırmaktadır?

A) Ortam sıcaklığının terlemeye olan etkisini
B) Farklı sıcaklıklarda besin yapma hızını
C) Kök ile yaprak arasındaki madde iletim hızını
D) Farklı ortamlarda bitkilerin solunum hızını

24. Aşağıdaki rejenerasyon (yenilenme) olaylarından hangisi üreme olarak kabul edilebilir?

A) İkiye ayrılan deniz yıldızının her bir parçasının kendini yenilemesi
B) Bitkilerde yaşlı kısımların kesilip çıkarılarak bitkinin gençleştirilmesi
C) İnsanda doku tahribi olan karaciğerin büyük ölçüde kendini onarması
D) Kurbağa larvasının yetişkin kurbağaya dönüşmesi

25.

I II

Bir öğrenci kullanılan gübre <u>miktarının</u> bitkilerin büyümesine olan etkisini incelemek istemektedir. Bu öğrenci aynı ortamda bulunan I. saksıya düzenli olarak bir miktar gübre koyarken aşağıdakilerden hangisini yaparsa araştırdığı soruya cevap verebilir?

A) II. saksıya hiç gübre koymadan , her iki saksıyı aynı ölçüde sulayarak.
B) II. saksıya I. saksı ile aynı miktarda gübre koyarak her iki saksıyı aynı ölçüde sulayarak.
C) II. saksıya I. saksıdakinin yarısı kadar gübre koyarak her iki saksıyı aynı ölçüde sulayarak.
D) II. saksıya I. saksının iki katı kadar su ve gübre koyarak.

APPENDIX L

Biology Items In The 2000 SSISSPT

1. Aşağıda hayvanlar alemine ait bir sınıflandırma verilmiştir.

Hayvanlar alemi : Kelebek, kartal, aslan, kuş, alabalık, salyangoz, keçi, kurt

↓

Omurgalılar : Kartal, aslan, kuş, alabalık, keçi, kurt

↓

Memeliler : Aslan, keçi, kurt

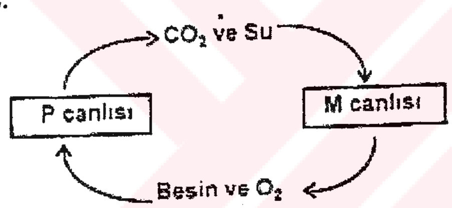
↓

? : Aslan, kurt

Buna gore "?" işareti görülen yere aşağıdakilerden hangisi yazılmalıdır?

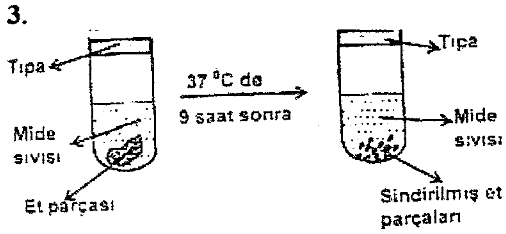
A) Doğurarak çoğalanlar B) Etçiller C) Yırtıcılar D)Büyükbaş hayvanlar

2.



Yukarıdaki şema aynı bölgede yaşayan iki canlı türünün birbiriyle ilişkilerini göstermektedir.
Bu şemaya bakılarak aşağıdakilerden hangisi söylenebilir?
A) P canlısı kesinlikle omurgasız bir hayvandır.
B) P canlısı fayda sağlarken diğeri zarar görür.
C) M canlısı ototroftur.
D) M canlısı sayısı arttığında P canlısı yok olur.

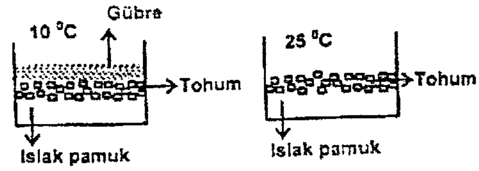
3.



Yukarıdaki şekilde sindirim olayıyla ilgili bir deney düzeneği gösterilmiştir.
Bu deneyi yapan bir öğrenci deneyin dokuz saatten <u>daha kısa</u> bir sürede tamamlanması için hangi değişikliği yapmalıdır?
A) Et parçasını çok ufak dilimler haline getirmelidir.
B) Mide sıvısını kaynatmalıdır.
C) Tüpün içine buz parçası atmalıdır.
D) Ortam sıcaklığını 10 ^{0}C de sabit tutmalıdır.

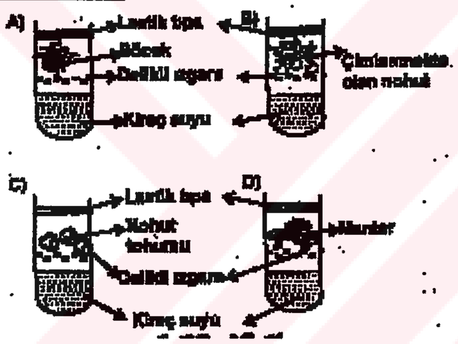
4. Bir canlı türünün farklı ortamlarda yaşayan bireyleri arasında bazı fiziksel özellikler gözlenmektedir.
Aşağıdakilerden hangisi buna örnektir?
A) Martının ayaklarındaki perdenin leyleğinkinden geniş olması.
B) Kutup ayısının daha çok deniz ürünleri ile beslenmesi
C) Balinalarda ön üyelerin yerini yüzgeçlerin alması.
D) Sıcak bölge tilkilerinin soğuk bölge tilkilerinden daha büyük kulaklı olması.

5.



Ortam sıcaklığının çimlenmeye etkisini incelemek isteyen bir öğrenci yukarıdaki düzeneklerde ne yapmalıdır?
A) İki düzeneği de aynı sıcaklıktaki ortama getirmelidir.
B) İkinci düzeneğe daha çok tohum koymalıdır.
C) Birinci düzeneğe su eklemelidir.
D) İkinci düzeneğe de gübre koymalıdır.

6. Karbondioksit kireç suyunu bulandırır.
Aşağıdaki tüplerden hangisi bir süre bekledikten sonra elle çalkalandığında kireç suyunun <u>bulanmaması</u> beklenir?

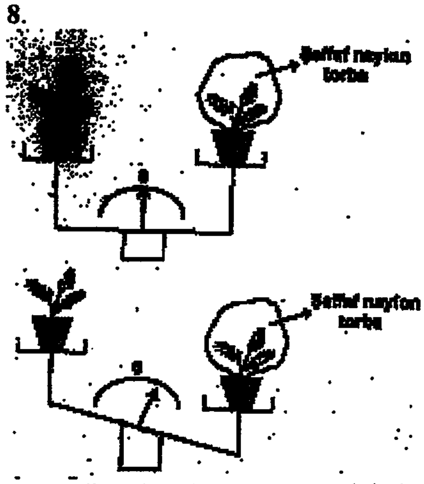


7.

Canlı türü	Üreme sıklığı (yılda)	Bir doğumdaki yavru sayısı (en fazla)	Yaklaşık gebelik süresi (gün)
Ev faresi	7-8	13	21
Tavşan	6-7	6	42
Köpek	2	10	60
Fil	2 yılda bir	1	660

Yukarıdaki tabloda verilen bilgilerle aşağıdaki sonuçlardan hangisine <u>ulaşılamaz</u>?
A) Gebelik süresi büyük vücutlu canlılarda daha uzundur.
B) Çoğalma miktarı küçük vücutlu canlılarda daha fazladır.
C) Üreme sıklığı çevre koşulları ile ilişkilidir.
D) Bir doğumdaki yavru sayısı en büyük vücutlu canlıda en azdır

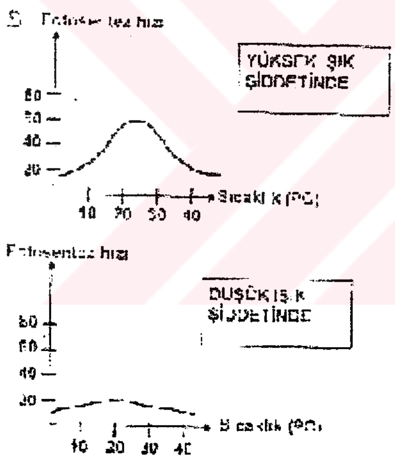
8.



Aynı miktarda sulanmış ve terazinin kefelerinde dengede olan özdeş iki bitkiden bir tanesinin yaprakları şeffaf naylon torba ile kapatılmıştır. Kısa bir süre sonra terazideki dengenin şekildeki gibi bozulduğu gözlenmektedir.
Bu deney aşağıdaki hangi soruya cevap verir?

A) Bitkiler topraktaki mineralleri tamamen kullanır mı?
B) Bitkiler farklı hızda gelişir mi?
C) Bitkiler terleme ile su kaybeder mi?
D) Bitkilerin fotosentez hızları farklı mıdır?

9.



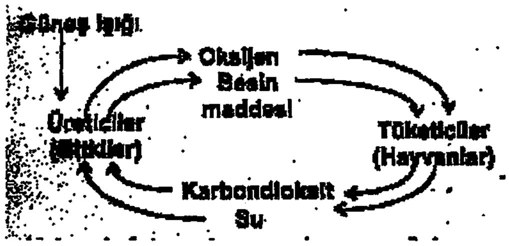
Yukarıdaki grafiklerde farklı ışık şiddetlerinde fotosentez hızı ile sıcaklık arasındaki ilişki gösterilmektedir.
Bu grafiğe göre aşağıdaki koşulların hangisinde bitkinin daha çok fotosentez yaptığı söylenebilir?

A) Öğle saatinde 35 ^{0}C sıcaklıkta
B) Öğle saatinde 25 ^{0}C sıcaklıkta
C) Akşam üzeri 20 ^{0}C sıcaklıkta
D) Akşam üzeri 40 ^{0}C sıcaklıkta

APPENDIX M

Biology Items In The 2001 SSISSPT

17.



Yukarıda fotosentez ve solunumu gösteren döngü verilmiştir.
Burada üreticilerin hepsi yok olursa ne olur?

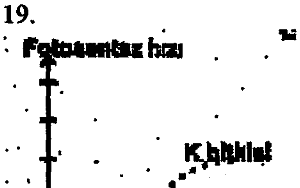
A) Otçul hayvan sayısı sürekli artar.
B) Etçil hayvan sayısı sürekli artar.
C) Besin maddesi üretilemez.
D) Oksijen miktarında artış olur.

18.

Yukarıdaki düzenekleri inceleyen bir öğrenci aşağıdaki hangi soruya cevap verebilir?

A) Çimlenme, ortamın sıcaklığına bağlı mıdır?
B) Çimlenme için hava gerekli midir?
C) Çimlenme, topraktaki nem miktarına bağlı mıdır?
D) Çimlenme, topraktaki gübre miktarına bağlı mıdır?

19.

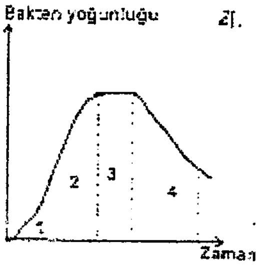


Aynı ortamda bulunan iki bitkinin ışığa bağlı fotosentez hızını gösteren grafik şekildeki gibidir.
Bu grafikle ilgili aşağıdakilerden hangisi söylenemez?

A) K bitkisi ışığı daha çok sever
B) L bitkisi gölge bitkisidir
C) K bitkisi daha çok besin üretmiştir
D) L bitkisi iğne yapraklıdır

20. Bir cins arı yüksek sıcaklıkta büyütülürse açık renkli, gelişebileceği en düşük sıcaklıkta büyütülürse siyah renkli olur. Aşağıdakilerden hangisi bu olayla benzerlik göstermez?
A) Çuha çiçeği bitkisinin 15-20 ^{0}C'de kırmızı çiçek açarken, 30-35 ^{0}C'de beyaz çiçek açması
B) Kuzey Kutbu'na yakın bölgede yaşayan tavşanların kışın ve yazın farklı renklerde olması
C) Afrika'da yaşayan insanların ten renginin siyah, Avrupa'da yaşayan insanların ten renginin açık olması
D) Sivri sineklerin 25 ^{0}C'de tutulan larvalarından kıvrık kanatlı yavruların, 16 ^{0}C'de tutulan larvalarından düz kanatlı yavruların ortaya çıkması

21.



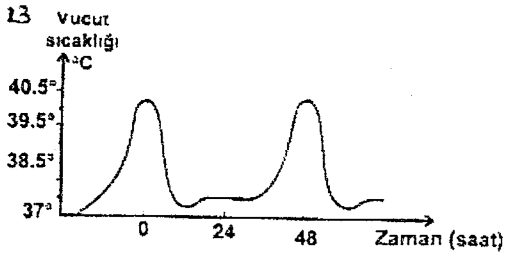
Yukarıda bir besin ortamında gelişen bakterilerin yoğunluğunun zamana gore değişim grafiği verilmiştir. Bu grafiğe gore aşağıdakilerden hangisi söylenemez?
A) 1. zaman diliminde ortamdaki besin en fazladır.
B) 2. zaman diliminde bakteri çoğalma hızı en fazladır.
C) 3. zaman diliminde bakteri sayısı değişmemiştir.
D) 4. zaman diliminde meydana gelen bakteri sayısı ölenden çoktur.

22.

Y türü canlı sayısı	Y türü canlılarla aynı ortamda bulunan canlılar
Artıyor	K, M, N
Azalıyor	M, N, O
Azalıyor	O, K
Artıyor	N, M

K, M, N ve O türleri birbiriyle beslenmeyen türlerdir. Y türü, bu türlerle dört ayrı ortamda bir araya konduğunda, sayısındaki artış ve azalış tablodaki gibi gözleniyor. Buna göre hangi tür canlının Y türü canlı ile beslendiği söylenebilir?
A) K B) O C) N D) M

23.



Sıtma insanda birdenbire ortaya çıkan önemli bir hastalıktır. Yukarıda sıtma hastalığındaki vücut sıcaklığının zamana gore değişim grafiği verilmiştir.
Grafiğe göre aşağıda verilenlerden hangisi yanlıştır?
A) Hastada, vücut sıcaklığı aralıklı olarak yükselmektedir
B) Hastalık belirtileri periyodik olarak tekrarlanmaktadır
C) Belirli aralıklarla vücut sıcaklığı 40 ^{0}C'ye kadar çıkabilmektedir.
D) Hasta birey, hastalık başladıktan 24 saat sonra tamamen iyileşmektedir.