Macrothink Institute™

Journal of Studies in Education
ISSN 2162-6952
2015, Vol. 5, No. 3

Developing an Achievement Test for Astronomy Education[1]

Cumhur TÜRK (Corresponding Author)

Researcher Assistant

Science Education Program, Faculty of Education, Ondokuz Mayıs University

Samsun, 55200, Turkey

Tel: 90-362-312-1919 ext: 5870 E-mail: cturk@omu.edu.tr

Hüseyin KALKAN

Prof.Dr.

Science Education Program, Faculty of Education, Ondokuz Mayıs University

Samsun, 55200, Turkey

Tel: 90-362-312-1919 ext: 5867 E-mail: kalkanh@omu.edu.tr

Received: Feb.22, 2015 Accepted: August7, 2015 Published: August7, 2015
doi:10.5296/jse.v5i3.8099 URL: http://dx.doi.org/10.5296/jse.v5i3.8099

Abstract

The purpose of this study is to develop an Astronomy Achievement Test (AAT) to test students' achievement on basic astronomy concepts and to find out their alternative concepts. In addition, another goal of the study is to fill the gap in the literature of this field. During the development process of the test, studies on basic astronomy concepts were reviewed and after the test items were formed, experts' reviews were examined and the necessary corrections were made. Pretest and pilot studies were conducted with 293 7th graders during the 2013-2014 academic year. The test, which was administered in its last form as a result of the analyses conducted, has 32 multiple questions with 4 choices. The test's KR-20 reliability coefficient in this study was calculated as 0,87. The statistical analyses, which were conducted, show that the AAT is a reliable and valid test.

Keywords: Astronomy education, basic astronomy concepts, achievement test, developing test

[1]This study based on Ph.D. thesis of Cumhur TÜRK.

1. Introduction

Most of the studies on astronomy education were conducted to determine situations such as existing difficulties, lack of practical studies (Swinbank, 1997), confusions caused by illogical concepts (MacRobert, 1995) and the lack of teachers' knowledge in astronomy (Osborne & Simon, 1996). In addition, studies on astronomy education are mostly small scale studies to find out conceptual errors rather than studies to find out large scaled mental models. For example: studies on day and night (Baxter, 1989), the period of the Moon's rotation around the Earth (Skamp, 1994), phases of the Moon (Sharp, 1996) and seasons (Baxter, 1991; Trumper, 2000). There are less large scaled studies conducted to present students' understanding and mental models of astronomy subjects (such as the Sun-Earth-Moon system, Solar system and galaxies) (Sharp, 1996). In the interviews they conducted, Jones & Lynch (1987) found that the students' spatial mental models about the Sun-Earth-Moon system consisted of five different systems and three of these were geocentric while the other two were heliocentric. In addition, they found that the students needed to develop the concepts of day-night, month, year, seasons, phases of the Moon, eclipses and tides in order to understand the relationships with the Sun-Earth-Moon system.

There are very few countries in the world in which astronomy is taught as a separate and compulsory course. Most of the countries do not even have compulsory astronomy units in other courses. On average, it can be seen that students between the ages 10 and 14 are taught basic astronomy subjects (such as explanatory materials about day-night, seasons, phases of the Moon, planets' orbits, planets and stars) within the courses of physics, geography and the like (Percy, 2006). Similarly according to the Next Generation Science Standards (NGSS, 2013), students in Grades 5–8 should have a clear notion about day-night, phases of the moon, daily and seasonal changes in the length and direction of shadows, different positions of the sun, moon, stars at different times of the day, month, and year, eclipses of the sun and the moon, Earth's spin axis, Big Bang.

It is remarkable that studies about astronomy teaching do not date back to old times although astronomy is a very old science. Especially developed countries have made progressions in their science programs by investing in astronomy teaching. Thus, they developed students' attitudes on science and mathematics courses (Türk & Kalkan, 2015).

When the literature is reviewed, concepts such as day-night cycle, time zones, Moon's rotation, Solar and Lunar eclipses, Lunar phases, seasons, Sun overhead at noon, dimensions, distances, center of the universe, shape of the Earth, artificial satellite, constellations are defined as the basic concepts of astronomy by researchers (Baxter, 1989; Bisard, Aron, Francek& Nelson, 1994; Kalkan & Kıroğlu, 2007; Kıroğlu, 2015; Klein, 1982; Mant & Summers 1993; Sharp, 1996; Sneider & Pulos, 1983; Trumper, 2001a; 2001b, 2001c, 2003, 2006a, 2006b; Trundle, Atwood & Christopher, 2002; Türk & Kalkan, 2015; Türk, Şener & Kalkan, 2015; Zeilik, Schau&Mattern, 1998). The research about these concepts of astronomy that we constantly encounter in our daily lives show that the students have alternative conceptions or mental models different from scientific explanations and that as the students' ages-levels of education increase, these do not change too much (Agan& Sneider 2003; Bisard et al., 1994; Danaia & McKinnon, 2007; Frede, 2008; Kavanagh, Agan,

&Sneider 2005; Kavanagh & Sneider 2006a, 2006b; Kikas, 1998; Schoon, 1992; Tsai & Chang, 2005; Vosniadou, 1992; Vosniadou & Brewer, 1992, 1994). In addition when studies conducted in the last thirty years are analyzed, it can be observed that the number of studies involving secondary school students (Baxter, 1989; Bisard, Aron, Francek, & Nelson, 1994; Klein, 1982; Mant & Summers, 1993; Sharp, 1996; Sneider & Pulos, 1983; Trumper, 2001a, 2001b, 2006b; Trundle, Atwood, & Christopher, 2002; Vosniadou, 1992; Vosniadou & Brewer, 1992, 1994) are much more common than those involving college students (Bisard et al., 1994; Kalkan &Kıroglu, 2007; Trumper, 2000, 2001c, 2003, 2006a; Türk, Şener & Kalkan, 2015; Zeilik, Schau, &Mattern, 1998).

In studies about astronomy education, expressions about the reliability and the validity of the achievement or concept tests used are not included much. Researchers generally present the alternative concepts they find out with the help of multiple choice questions they develop. Achievement tests have been developed in a limited number of studies (Bailey, 2008; Hufnagel, 2002; Lindell & Olsen, 2002; Sadler et al. 2010; Zeilik et al. 1998). These tests have focused on only one of the specific subjects such as the Solar system, phases and movements of the Moon, formation of the seasons and stars and few tests have included basic astronomy concepts in general. Thus, the present study developed a test which includes the general basic astronomy concepts in astronomy education and which has undergone necessary analyses. When this study is compared with other astronomy test development studies in literature, it is thought to be important and to contribute to literature since it does not focus only on specific subjects and includes all the basic concepts of astronomy.

The purpose of this study is to develop an assessment instrument to test students' achievement on basic astronomy concepts and to find out their alternative concepts. The fact that there are limited numbers of tests in literature which are developed for basic astronomy concepts in general shows that there is an important gap in the present situation and this also presents the important of the present study. In addition, this study is expected to enlighten teachers and those who want to research in this field.

2. Method and Material

An achievement test was developed in this study to measure the achievements of middle school students on basic astronomy subjects and concepts and to find out their alternative concepts. The test was checked for validity and reliability. While preparing the test questions, the concepts which are defined as basic astronomy concepts by researchers in literature were determined and questions were prepared to test these concepts. Pretest and pilot studies were conducted to develop AAT. For this purpose, 43 multiple questions with four choices were prepared for the test. While preparing the questions, internet resources, study worksheets, leaf tests, "The Astronomy Diagnostic Test" (Hufnagel, 2002) which was developed in the United States of America and which is used in many countries of the world and the achievement tests used in the author's previous studies (Türk & Kalkan, 2015) were used. Table 1 gives the resources for the questions in AAT.

Table 1. The Resources for the items in final AAT

Resources	Item No
Hufnagel (2002)	1,2,5,6,8,17,21,22,28,31
Türk & Kalkan, 2015	12,30
Developed by the researcher	3,4,7,9,10,11,13,14,15,16,18,19,20,23,24,25,26,27,29,32

As can be seen in Table 1, 10 questions of the draft AAT were taken from the questions developed by Hufnagel (2002), while 2 questions were taken from the questions developed by Türk & Kalkan (2015). The remaining 20 questions were developed and placed in the test by the researcher.

2.1. Sample

The sample of the study consists of 293 7th graders (pre-pilot study 110, pilot study 103; final study 80) studying in a school in the Black Sea Region of Turkey during the 2013-2014 academic year. The reason why 7th graders were included in the sample was the fact that the subjects and concepts of astronomy are most extensively taught in 7th grade in Turkish science education programs. Thus, AAT was given to students at the end of the 7th grade, after the students studied the astronomy unit. Otherwise, it was thought that the students would be inclined not to answer the questions in the test, which included subjects or concepts they did not know about. The test was administered at the end of 7th grade to minimize this problem and to make sure that the students answered all the questions.

Pre-pilot and pilot studies were conducted to develop AAT. Pretest and pilot studies were conducted with the 7th graders in a middle school. Table 2 gives the distributions of students who participated in pretest and pilot studies in terms of gender.

Table 2. Numbers of students who participated in the pretest and pilot studies of AAT

	Pre-Pilot		Pilot		Final		Total	
Gender	**f**	**%**	**f**	**%**	**f**	**%**	**f**	**%**
Female	57	51,8	52	50,5	39	48,7	148	50,5
Male	53	48,2	51	49,5	41	51,3	145	49,5
Total	**110**	**100**	**103**	**100**	**80**	**100**	**293**	**100**

Pilot studies of the achievement test's draft were conducted with a total of 293 7th graders, 148 of whom were females and 145 of whom were males.

2.2. Development Process

The procedures during the process of developing AAT are given in items below and in schemes afterwards.

- Before this test was developed, studies on basic astronomy concepts were reviewed and basic astronomy concepts were determined. Thus, the content of the test was determined.

- At least three questions were planned to address each concept but in different ways.
- At least three questions were planned with the same purpose but in different ways on these concepts. While preparing the questions, internet resources, study worksheets, leaf tests, "The Astronomy Diagnostic Test" (Hufnagel, 2002) which was developed in the United States of America and which is used in many countries of the world and the achievement tests used in the author's previous studies (Türk & Kalkan, 2015) were used.
- There were originally 43 questions in the draft of AAT. In addition, when the AAT was analyzed in terms of subjects and conceptual contents, the test can be seen to consist of 5 different foci. These foci are as follows:

– Recognizing celestial bodies and the Solar system
– The mechanical relationship between the Earth and the Sun
– The movements of the Moon and the eclipses
– Constellations and their distances
– Space research and telescopes

Table 3 gives the distribution of the items in AAT in terms of the above mentioned foci.

Table 3. The distribution of items in AAT in terms of foci

Focus	Item Number	Total Item
Recognizing celestial bodies and the Solar system	4, 7, 9, 21, 22, 24, 25, 30	8
The mechanical relationship between the Earth and the Sun	1, 2, 3, 5, 6, 8, 10, 11, 12, 41, 42, 43	12
The movements of the Moon and the eclipses	13, 14, 15, 16, 17, 18, 19, 20, 35, 36, 37, 38, 39, 40	14
Constellations and their distances	26, 27, 28, 29	4
Space researches and telescopes.	23, 31, 32, 33, 34	5

- To ensure the content validity of the test, AAT was analyzed in terms of science, content and format by the experts stated in Table 4. The draft was analyzed by 3 academics and 2 teachers who had at least 10 years of experience. All the experts were informed about the main purpose of the test and they were given the test and the criteria on which they were asked to base their assessments. Necessary corrections were made in light of the assessments of the experts. Table 4 presents the demographic characteristics of the experts whose views were taken about the AAT.

Table 4. The demographic characteristics of the experts whose views were taken about the AAT

Gender	Title	Profession
Male	Prof. Dr.	Science and Astronomy Education
Male	Assoc. Prof.	Educational Sciences
Female	Research Assistant	Science Education
Male	Teacher	Science Education
Female	Teacher	Science Education

- After the experts' views were taken, the draft AAT was conducted on 10 middle schools students and the time they used to answer the test and whether there were any items they had difficulty in understanding were tested. It was observed that the students could answer the test in about 20-25 minutes. In addition, no negative feedback came from the students in terms of whether there were problems in the readability of the test.
- The test's pre-pilot study was conducted. After this, the items were analyzed. In scoring the results of AAT, 1 point was given for correct answers, while no points were given for incorrect answers, unanswered questions or more than one answer for the same item and the total scores of each student were calculated. As a result of the analyses of the pre pilot study, exclusion of 5 items was decided upon.
- After this, the pilot study was conducted. As well as the 5 items excluded from the test, corrections were made in the questions and choices of 14 of the questions. After these processes, it was decided to exclude 6 more questions and to make corrections in the questions and choices of 5 of the questions as a result of the item analysis of AAT which consisted of 38 questions.
- By considering the obligations of measuring each concept at least with one question and including all of the basic astronomy concepts in the test and studies of reliability and validity, some questions were excluded from the test and the number of total questions was determined as 32.

Figure 1 presents the AAT development process in a scheme in order to understand and see the process better.

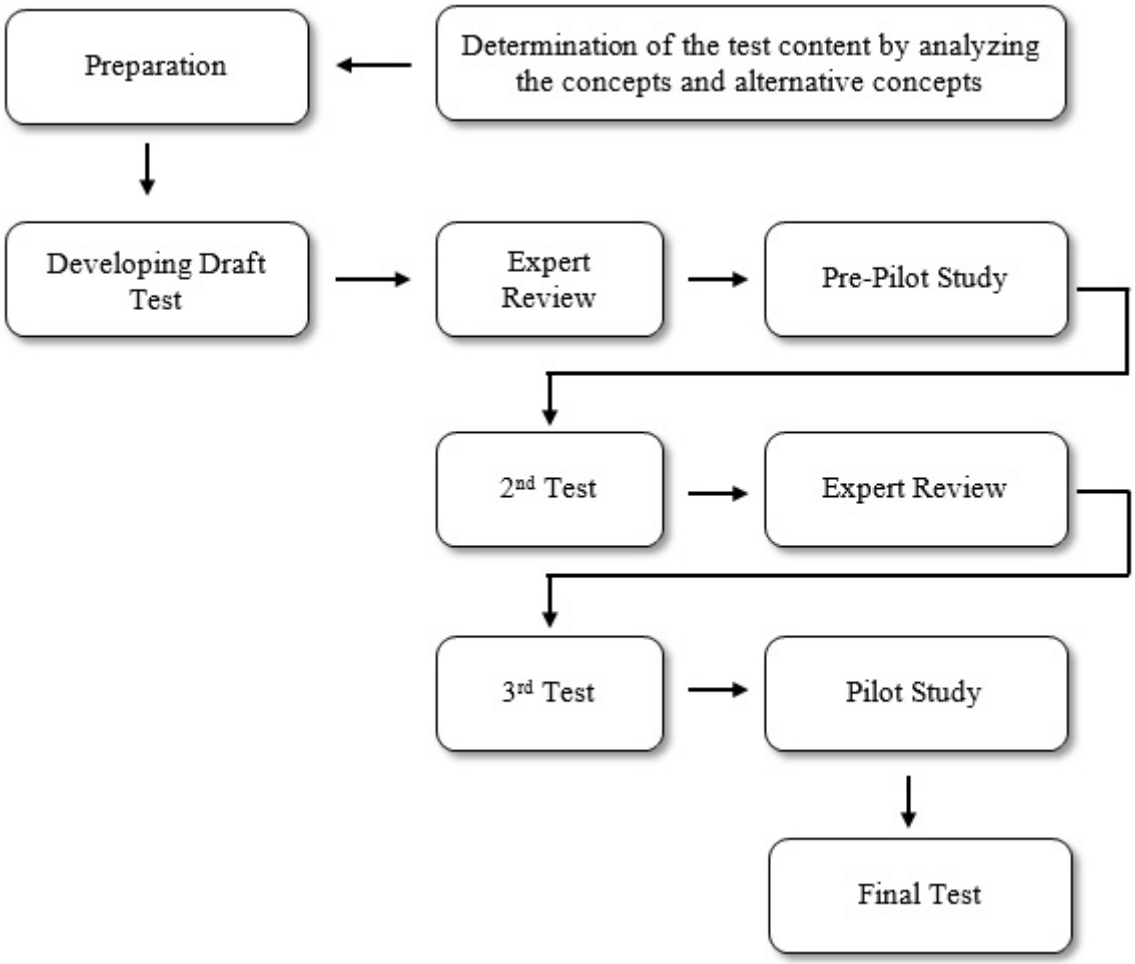


Figure 1. AAT development process

3. Findings and Discussion

In scoring the results of AAT, 1 point was given for correct answers, while no points were given for incorrect answers, unanswered questions or more than one answer for the same item and the total scores of each student were calculated. In addition, item score, which defines how many students answered each question correctly, was calculated for each question. Normality test results of pre-pilot and pilot studies of AAT draft were identified.

Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk) were conducted to test whether AAT had normal distribution. The results of these tests were given in Table 5.

Table 5. The results of normality tests

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Pre-Pilot	,091	110	,200*	,972	40	,426
Pilot	,085	103	,062*	,983	103	,213

When Table 5 is analyzed, a p-value of greater than 0.05 can be commented as the scores did not show a significant (extreme) deviation from the normal distribution in this level of

significance; that is they were suitable (Büyüköztürk, 2005). The results of the Kolmogorov-Smirnov and Shapiro-Wilk tests show that the students' academic achievement test scores did not show a significant difference from normal distribution (p>0.05).

Table 6 gives the arithmetic mean and standard deviation values of the items in the test according to the results of the pre pilot study.

Table 6. Descriptive statistics of the pre-pilot data of AAT

Item No	N	Arithmetic Mean	SD	Item No	N	Arithmetic Mean	SD
1.	110	,6636	,47463	**23.**	110	,2182	,41490
2.	110	,1000	,30137	**24.**	110	,4909	,50221
3.	110	,2000	,40183	**25.**	110	,3455	,47769
4.	110	,1364	,34474	**26.**	110	,3000	,46035
5.	110	,0545	,22813	**27.**	110	,2818	,45194
6.	110	,2091	,40852	**28.**	110	,3727	,48574
7.	110	,2818	,45194	**29.**	110	,1636	,37164
8.	110	,4455	,49929	**30.**	110	,3091	,46423
9.	110	,5818	,49552	**31.**	110	,2455	,43233
10.	110	,3091	,46423	**32.**	110	,3727	,48574
11.	110	,3545	,48056	**33.**	110	,2727	,44740
12.	110	,3182	,46790	**34.**	110	,2182	,41490
13.	110	,2909	,45626	**35.**	110	,3909	,49019
14.	110	,3727	,48574	**36.**	110	,3545	,48056
15.	110	,3000	,46035	**37.**	110	,3091	,46423
16.	110	,5545	,49929	**38.**	110	,4545	,50021
17.	110	,5000	,50229	**39.**	110	,3364	,47463
18.	110	,2273	,42099	**40.**	110	,3909	,49019
19.	110	,3636	,48325	**41.**	110	,2727	,44740
20.	110	,3455	,47769	**42.**	110	,2727	,44740
21.	110	,4091	,49392	**43.**	110	,3273	,47137
22.	110	,2273	,42099				

After the results were found to have normal distribution, item analysis was conducted of the answers given by the students for each question. In the pre pilot study, total scores of the students from the test were calculated and they were ranked from the highest to the lowest. The groups were formed as 27% of the highest scores forming the upper group and as 27% of the lowest scores forming the lower group and the item analysis was conducted. After the upper and lower groups of 27% (N=30) were formed, item analysis was conducted using the Microsoft Excel program.

Items were chosen by taking item discrimination index (rjx) and item difficulty index (p) into consideration. Item difficulty is defined as the division of the number of correct answers to the number of those who answered it correctly while item discrimination is defined as the

degree of an item in discriminating the individuals who have the tested behavior and those who do not (Özçelik, 2010). The formulas used in item analysis are as follows:

Item difficulty has a value between "0" and "1". An item can be said to have low rates of being answered (a difficult item) as the item difficulty approaches 0 and to have high rates of being answered (an easy item) as the item difficulty approaches 1 (Gönen, Kocakaya&Kocakaya, 2011). Item discrimination differs between "-1" and "1". An item discrimination close to 0 means low discrimination while item discrimination close to 1 means high discrimination. An item with negative item discrimination value does not serve the purposes of the test and also lowers the reliability of the test (Gönen et al., 2011).

Difficulty and discrimination indices of each item in AAT were calculated as stated in Table 7. The following values were accepted as the criteria in determining the items to be included in the test.

- Item Difficulty Index (p),

0,00 – 0,19 the item is accepted as a very difficult item

0,20 – 0,34 the item is accepted as a difficult item

0,35 – 0,64 the item is accepted as a moderately difficult item

0,65 – 0,79 the item is accepted as an easy item

0,80 – 1,00 the item is accepted as a very easy item (Özçelik, 2010).

- Item Discrimination Index (rjx),

0,19 and less, the item is unacceptable.

0,20 – 0,29 the item should be corrected.

0,30 – 0,39 it is a good item and it is accepted.

0,40 or greater, it is a very good item and it is accepted (Tekin, 2003).

Table 7 gives the item analysis results of upper and lower group students based on the number of correct answers for the draft AAT.

Table 7. Pre-pilot study item analysis results

	Upper Group Correct Answer Score	Lower Group Correct Answer Score	Difficulty Index p	Difficulty	Discrimination Index rjx	Discrimination
Item 1	21	12	0,55	Easy	0,30	Good
Item 2	5	0	0,08	Difficult	0,17	Very Low*
Item 3	11	3	0,23	Difficult	0,27	Straightened
Item 4	3	3	0,10	Difficult	0,00	Very Low*
Item 5	2	1	0,05	Difficult	0,03	Very Low*
Item 6	11	3	0,23	Difficult	0,27	Straightened
Item 7	11	5	0,27	Difficult	0,20	Straightened
Item 8	18	9	0,45	Average	0,30	Good
Item 9	22	14	0,60	Easy	0,27	Straightened
Item 10	16	6	0,37	Average	0,33	Good
Item 11	17	5	0,37	Average	0,40	Very Good
Item 12	10	9	0,32	Average	0,03	Very Low*
Item 13	15	6	0,35	Average	0,30	Good
Item 14	17	9	0,43	Average	0,27	Straightened
Item 15	16	10	0,43	Average	0,20	Straightened
Item 16	23	15	0,60	Easy	0,27	Straightened
Item 17	19	9	0,47	Average	0,33	Good
Item 18	15	6	0,35	Average	0,30	Good
Item 19	18	7	0,42	Average	0,37	Good
Item 20	18	5	0,38	Average	0,43	Very Good
Item 21	21	12	0,55	Easy	0,30	Good
Item 22	12	5	0,28	Difficult	0,23	Straightened
Item 23	16	6	0,37	Average	0,33	Good
Item 24	24	7	0,52	Easy	0,57	Very Good
Item 25	13	5	0,30	Average	0,27	Straightened
Item 26	16	4	0,33	Average	0,40	Very Good
Item 27	17	5	0,37	Average	0,40	Very Good
Item 28	17	4	0,35	Average	0,43	Very Good
Item 29	10	3	0,22	Difficult	0,23	Straightened
Item 30	13	6	0,32	Average	0,23	Straightened
Item 31	14	5	0,32	Average	0,30	Good
Item 32	17	3	0,33	Average	0,47	Very Good
Item 33	12	5	0,28	Difficult	0,23	Straightened
Item 34	8	5	0,22	Difficult	0,10	Very Low*
Item 35	16	8	0,40	Average	0,27	Straightened
Item 36	15	7	0,37	Average	0,27	Straightened

Item 37	14	5	0,32	Average	0,30	Good
Item 38	19	9	0,47	Average	0,33	Good
Item 39	19	8	0,45	Average	0,37	Good
Item 40	13	3	0,27	Difficult	0,33	Good
Item 41	14	4	0,30	Average	0,33	Good
Item 42	15	4	0,32	Average	0,37	Good
Item 43	13	5	0,30	Average	0,27	Straightened
		Mean	**0,35**	**Average**	**0,29**	**Straightened**

*The items excluded from AAT after the pre-pilot study.

As a result of the item analysis of the pre-pilot study, questions 2, 4, 5, 12 and 34 in the draft AAT were excluded from the test since they had low levels of discrimination. In addition, studies were made to correct the questions which needed correction in terms of the discrimination values. For this, corrections were made in the questions and/or choices and the test was put into its final form before the pilot study. After this, the test, which included 38 remaining questions after the exclusion of 5 questions and corrections was conducted on different 7th graders in the same school.

Table 8 gives the item analysis results of the students in the upper and lower groups (N=28) in terms of the number of correct answers after the pilot study.

Table 8. Pilot study item analysis results

	Upper Group Correct Answer Score	**Lower Group Correct Answer Score**	**Difficulty Index p**	**Difficulty**	**Discrimination Index rjx**	**Discrimination**
Item 1	25	13	0,68	Easy	0,43	Very Good
Item 3	10	3	0,23	Difficult	0,25	Straightened
Item 6	13	4	0,30	Average	0,32	Good
Item 7	11	3	0,25	Average	0,29	Straightened
Item 8	17	8	0,45	Average	0,32	Good
Item 9	20	18	0,68	Easy	0,07	Very Low*
Item 10	15	6	0,38	Average	0,32	Good
Item 11	17	7	0,43	Average	0,36	Good
Item 13	12	3	0,27	Difficult	0,32	Good
Item 14	12	7	0,34	Average	0,18	Very Low*
Item 15	15	11	0,46	Average	0,14	Very Low*
Item 16	19	16	0,63	Easy	0,11	Very Low*
Item 17	14	9	0,38	Average	0,18	Very Low*
Item 18	15	4	0,34	Average	0,39	Good
Item 19	14	6	0,36	Average	0,29	Straightened
Item 20	14	4	0,32	Average	0,36	Good
Item 21	18	8	0,46	Average	0,36	Good

Item 22	9	3	0,21	Difficult	0,21	Straightened
Item 23	15	5	0,36	Average	0,36	Good
Item 24	17	7	0,43	Average	0,36	Good
Item 25	17	6	0,41	Average	0,39	Good
Item 26	13	4	0,30	Average	0,32	Good
Item 27	12	3	0,27	Difficult	0,32	Good
Item 28	14	4	0,32	Average	0,36	Good
Item 29	13	3	0,29	Difficult	0,36	Good
Item 30	14	5	0,34	Average	0,32	Good
Item 31	17	5	0,39	Average	0,43	Very Good
Item 32	13	3	0,29	Difficult	0,36	Good
Item 33	12	7	0,34	Average	0,18	Very Low*
Item 35	15	6	0,38	Average	0,32	Good
Item 36	18	5	0,41	Average	0,46	Very Good
Item 37	16	4	0,36	Average	0,43	Very Good
Item 38	17	8	0,45	Average	0,32	Good
Item 39	17	6	0,41	Average	0,39	Good
Item 40	15	6	0,38	Average	0,32	Good
Item 41	14	3	0,30	Average	0,39	Good
Item 42	11	4	0,27	Difficult	0,25	Straightened
Item 43	17	5	0,39	Average	0,43	Very Good
		Mean	**0,38**	**Average**	**0,32**	**Good**

*The items excluded from AAT after the pilot study.

As can be seen from Table 8, questions 9, 14, 15, 16, 17 and 38 of AAT were found to have very low levels of discrimination. Thus, these questions were excluded from the final test. As a result of the 6 questions excluded from the test after the pilot study, the final test had 32 questions.

The final AAT with 32 questions was conducted on the final sample of the research. Table 9 gives the item analysis results of the students in the upper and lower groups (N=22) in terms of the number of correct answers in the final study.

Table 9. Final AAT Item Analysis Results

	Upper Group Correct Answer Score	Lower Group Correct Answer Score	Difficulty Index p	Difficulty	Discrimination Index rjx	Discrimination
Item 1	22	9	0,70	Very Easy	0,59	Very Good
Item 3	11	1	0,27	Average	0,45	Very Good
Item 6	13	2	0,34	Average	0,50	Very Good
Item 7	14	5	0,43	Average	0,41	Very Good
Item 8	16	6	0,50	Easy	0,45	Very Good
Item 10	17	2	0,43	Average	0,68	Very Good
Item 11	22	5	0,61	Easy	0,77	Very Good
Item 13	22	3	0,57	Easy	0,86	Very Good
Item 18	11	4	0,34	Average	0,32	Good
Item 19	15	2	0,39	Average	0,59	Very Good
Item 20	18	4	0,50	Easy	0,64	Very Good
Item 21	21	16	0,84	Very Easy	0,23	Straightened
Item 22	19	6	0,57	Easy	0,59	Very Good
Item 23	16	3	0,43	Average	0,59	Very Good
Item 24	22	10	0,73	Very Easy	0,55	Very Good
Item 25	21	8	0,66	Easy	0,59	Very Good
Item 26	18	9	0,61	Easy	0,41	Very Good
Item 27	13	3	0,36	Average	0,45	Very Good
Item 28	21	10	0,70	Very Easy	0,50	Very Good
Item 29	10	3	0,30	Average	0,32	Good
Item 30	17	7	0,55	Easy	0,45	Very Good
Item 31	9	2	0,25	Difficult	0,32	Good
Item 32	22	11	0,75	Very Easy	0,50	Very Good
Item 35	20	10	0,68	Easy	0,45	Very Good
Item 36	18	4	0,50	Easy	0,64	Very Good
Item 37	10	4	0,32	Average	0,27	Straightened
Item 38	21	8	0,66	Easy	0,59	Very Good
Item 39	17	2	0,43	Average	0,68	Very Good
Item 40	21	7	0,64	Easy	0,64	Very Good
Item 41	14	4	0,41	Average	0,45	Very Good
Item 42	13	4	0,39	Average	0,41	Very Good
Item 43	16	7	0,52	Easy	0,41	Very Good
		Mean	**0,51**	**Easy**	**0,51**	**Very Good**

As can be seen in Table 9, as a result of the item analysis of the final AAT, the test was found to be easy in general; however, the discrimination of the test was found to be very good.

According to the results of the analyses, item difficulty indices of each item were between

0,25 and 0,75. According to the data in the table, the difficulty index of the most difficult item in the test, item 31, was 0,25. This value shows that 25% of the participants answered the 31st question correctly, while 75% answered incorrectly. The easiest item was the 32nd item with a difficulty index of 75%. In general, the average difficulty index is expected to be around 0,50 in achievement tests (Gönen et al., 2011; Özçelik, 2010; Atılgan, Kan&Doğan, 2009). Thus, the fact that AAT had an average item difficulty of 0,51 is important in terms of meeting the needs of an ideal test.

When the discrimination index results of the items were analyzed, this value was found as between 0,23 and 0,86. In a test, for the items to be accepted as good discriminators, they should have values over 0,30. Accordingly, there were no items in the test which required exclusion except for the corrections made in the items 21 and 37.

In order to determine the internal consistency of AAT, the following formula was used and the KR-20 reliability coefficient was calculated.

$$KR_{20} = \frac{K}{K-1}\left[1 - \frac{\sum pq}{S_x^2}\right]$$

K= Number of test items

p= Item difficulty

q= 1-p

S^2_x= Variance of the test

Table 10 presents the information about the item analyses of pre-pilot, pilot and final AAT and the KR-20 reliability coefficient.

Table 10. Test statistics as a result of item analyses

	Total Item	N	Difficulty p	Discrimination rjx	KR-20
Pre-Pilot	43	110	0,35	0,29	0,52
Pilot	38	103	0,38	0,32	0,67
Final	32	80	0,51	0,51	0,87

When Table 10 is analyzed, the KR-20 reliability coefficient of final AAT was found as 0,87. The lower limit of 0,70 for the assessment instruments in studies shows that the value of 0,87 found in this study is acceptable. The results of the analyses show that AAT is a valid and reliable test in measuring the students' achievement on basic astronomy concepts.

4. Conclusions

With this study, a valid and reliable test was formed to measure the students' achievement on basic astronomy concepts. The average difficulty of the test, which consisted of 32 items with 4 choices, was found as 0,51, the average discrimination was found as 0,51 and the KR-20

reliability coefficient was found as 0,87. Conducting a pre-test before the pilot study is important in terms of finding out the time required for answering the test and finding out whether the questions are understood by the students.

One of the basic elements of a successful education is a successful assessment process. To do this, it is extremely important to find out the existing states of the students at the beginning of the teaching process and to consider the results throughout the process. Teaching materials and strategies as well as the stages of assessment and evaluation should proceed according to the existing situation determined before teaching. For this, the assessment instruments should be valid and reliable. Thus, it is thought that AAT[2] will be important for the assessment activities in astronomy education.

Basic astronomy concepts included by AAT and the alternative concepts AAT tried to find out were determined. These concepts were given in Table 11.

Table 11. Concepts and alternative concepts targeted by the AAT

Concepts targeted by the test	*Alternative concepts targeted by the test*
– Day/night cycle	– Day and night occur as a result of the Earth moving around the Sun
– The direction of the rotation of Earth	– Earth rotates from east to west.
– Apparent motion of the Sun	– The rotation period of the Moon around the Earth is one day / one year.
– Scale of the Solar System	– The Moon will be a full moon during a solar eclipse
– Phases of the Moon	– The Earth closer to the Sun in summer than the in winter
– Linear distance scales	– Seasonal changes result from the Earth's getting closer to and moving away from the Sun.
– Linear size scales	– The Sun is at 90 degrees to the Earth at noon in that spot on the earth
– Seasonal changes	– Jupiter is closer to Earth than the Sun is
– Different seasons in both hemispheres	– Mars the closest planet to the Earth
– Solar eclipse	– Stars are very close to the Earth
– Lunar eclipse	– Polaris is the closest celestial body to the Earth
– General structure of the Universe	– The Earth is larger than Jupiter
– The center of the Universe	– Artificial satellites are very far from the Earth and very close to the Moon.
– Artificial satellites	– The Sun is at the center of the Universe.
– Satellite technology	– The constant orbital speed of the Moon around the Earth
– The change in the Sun's position in the sky during the day	– The Moon having the same phases all the time
– The orbit of the Earth around the Sun	– Constellations are not very far (they are even closer than the Moon).
– Order of the Moon's phases	– The appearance of constellations change when looked at from a high mountain
– Seeing the same side of the Moon all the time	
– Constellations	

[2]Test is provided in the Appendix.

Another important step of the process of developing a test is to determine the questions suitable for the targeted concepts. This study was very careful about the correspondence of questions and concepts, experts were given the concepts and a feedback form and their views were considered. Thus, evidence was collected for content and expert validity. Another factor in determining the validity and reliability of a test may be the difficulties encountered during the process of conducting the test. To overcome this, the final form of AAT was given to 10 students and the possible problems to be encountered were predetermined.

5. Research Limitations and Future Directions

In this section, some recommendations were given to researchers in this field within the context of the experience gained while developing AAT and the results obtained.

- The developed AAT was conducted on 7th graders to find out their achievement and alternative concepts on basic astronomy concepts. However, this test can be conducted for high school and university students and its applicability can be developed.
- It can be used as a data collection tool in the studies in this field.
- The test's reliability can be tested by conducting it on different samples.
- The test can be developed by adding different questions or by revising the questions, after using suitable strategies to measure the reliability and validity of each new item.

References

Agan, L., & Sneider, C. (2003). Learning about the earth's shape and gravity: A guide for teachers and curriculum developers. *Astronomy Education Review*, *2*(2), 90-117. http://dx.doi.org/10.3847/AER2003017

Atılgan H., Kan, A., &Doğan, N. (2009). *Measurement and evaluation in education* (4. Edition). Ankara: Anı Publishing.

Biley, J. M. (2008). Development of a concept inventory to assess students' understanding and reasoning difficulties about the properties and formation of stars. *Astronomy Education Review*, *2*(6), 133-139.

Baxter, J. (1991). A constructivist approach to astronomy in the National Curriculum. *Physics Education*, *26*(1), 38–45.

Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education*, *11*(5), 502-513.http://dx.doi.org/10.1080/0950069890110503

Bisard, W. J., Aron, R. H., Francek, M. A., & Nelson, B. D. (1994). Assessing selected physical science and earth science misconceptions of middle school through university pre-service teachers: Breaking the science 'misconception cycle'. *Journal of College Science Teaching*, *24*(1), 38-42. Retrieved from http://adsabs.harvard.edu/abs/1994JCSTe..24...38B (accessed 17 May 2014).

Büyüköztürk, Ş. (2005). Developing questionnaire. *Turkish Journal of Educational Sciences*, *3*(2), 133-151.

Danaia, L., & McKinnon, D. H. (2007). Common Alternative astronomical conceptions encountered in junior secondary science classes: Why is this so? *Astronomy Education Review*, *6*(2), 32-53. http://dx.doi.org/10.3847/AER2007017

Frede, V. (2008). The seasons explained by refutational modeling activities. *Astronomy Education Review*, *7*(1), 44-56. http://dx.doi.org/10.3847/AER2008005

Hufnagel, B. (2002). Development of the astronomy diagnostic test. *Astronomy Education Review*, *1*(1), 47. http://dx.doi.org/10.3847/AER2001004

Gönen, S., Kocakaya, S., &Kocakaya, F. (2011). A study on developing an achievement test which has reliability and validity on dynamics subject. *Yuzuncuyıl University Journal of Faculty of Education*, *8*(1), 40-57.

Jones, B., & Lynch, P. (1987). Children's conceptions of the earth, sun and moon. *International Journal of Science Education*, *9*(1), 45–53.

Kalkan, H. & Kıroğlu, K. (2007). Science and nonscience students' conceptions of basic astronomy concepts in preservice training for education teachers. *Astronomy Education Review*, *1*(6), 15-24. http://dx.doi.org/10.3847/AER2007002

Kavanagh, C., Agan, L., & Sneider, C. (2005). Learning about Moon phases and eclipses: A guide for teachers and curriculum developers. *Astronomy Education Review*, *4*(1), 19-52. http://dx.doi.org/10.3847/AER2005002

Kavanagh, C., & Sneider, C. (2006a). Learning about gravity Part I. Free Fall. *Astronomy Education Review*, *5*(2), 21-52. http://dx.doi.org/10.3847/AER2006018

Kavanagh, C., & Sneider, C. (2006b). Learning about gravity Part II. Trajectories and orbits. *Astronomy Education Review*, *5*(2), 53-102. http://dx.doi.org/10.3847/AER2006019

Kıroğlu, K. (2015). Students Are Not Highly Familiar with Astronomy Concepts–But What about the Teachers?.*Journal of Education and Training Studies*, *3*(4), 31-41.

Kikas, E. (1998). The impact of teaching on students' definitions and explanations of astronomical phenomena.*Learning and Instruction*, *8*(5), 439-454. http://dx.doi.org/10.1016/S0959-4752(98)00004-8

Klein, C. (1982). Children's concepts of the Earth and Sun: A cross-cultural study. *Science Education*, *65*(1), 95-107. http://dx.doi.org/10.1002/sce.3730660112

Lindell, R. S., & Olsen, J. P. (2002). Developing the lunar phases concept inventory. In *Proceedings of the 2002 Physics Education Research Conference*. New York: PERC Publishing.

Macrobert, A. M. (1995). Understanding celestial co-ordinates. *Sky and Telescope*, *90*(3), 38–39.

Mant, J., &Summers, M. (1993). Some primary-school teachers' understanding of the earth's place in the universe. *Research Papers in Education*, *8*(1), 101-129. http://dx.doi.org/10.1080/0267152930080107

Next Generation Science Standards. *(2013).* Retrieved from

http://nextgenscience.org(accessed 10 September 2014).

Osborne, J., & Simon, S. (1996). Primary science: past and future directions. *Studies in Science Education*, *27*, 99–147.

Özçelik, D. A. (2010). *Test Preparation Guide* (4. Edition). Ankara: PegemA Publishing.

Percy, J. R. (2006). Teaching astronomy: Why and how?.*The Journal of the American Association of Variable Star Observers*, *35*(5), 248-254. Retrieved from http://www.aavso.org/files/webpublications/ejaavso/v35n1/248.pdf (accessed 5 May 2014).

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The astronomy and space science concept inventory: Development and validation of assessment instruments aligned with the K–12 National Science Standards. *Astronomy Education Review*, *8*(1), 010111

Schoon, K. (1992). Students' alternative conceptions of earth and space. *Journal of Geological Education*, 40(3), 209-214. Retrieved from http://adsabs.harvard.edu/abs/1992JGEd...40..209S (accessed 18 July 2014).

Sharp, J. (1996). Children's astronomical beliefs: A preliminary study of year 6 children in south-west England. *International Journal of Science Education*, *18*(6), 631-652.http://dx.doi.org/10.1080/0950069960180601

Skamp, K. (1994). Determining misconceptions about astronomy. *The Australian Science Teachers' Journal*,*40*(3), 63–67.

Sneider, C. I., & Pulos, S. (1983). Children's cosmographies: Understanding the earth's shape and gravity. *Science Education*, *67*(2), 205-221.http://dx.doi.org/10.1002/sce.3730670209

Swinbank, W. (1997). Developing resources for astrophysics at A-level: the TRUMP Astrophysics Project. *Physics Education*, *32*(1), 40–45.

Tekin, H. (2003). *Measurement and evaluation in education* (15. Edition). Ankara: Yargı Publishing.

Trumper, R. (2000). University students' conceptions of basic astronomy concepts. *Teaching Physics*, 35(1), 9-15.

Trumper, R. (2001a). A cross-age study of junior high school students' conceptions of basic astronomy concepts. *International Journal of Science Education*, *23*(11), 1111-1123.http://dx.doi.org/10.1080/09500690010025085

Trumper, R. (2001b). A cross-age study of senior high school students' conceptions of basic astronomy concepts. *Research in Science and Technological Education*, *19*(1), 97-109. http://dx.doi.org/10.1080/02635140120046259

Trumper, R. (2001c). A cross-college age study of science and nonscience students' conceptions of basic astronomy concepts in preservice training for high-school teachers. *Journal of Science Education and Technology*, *10*(2), 189-195. http://dx.doi.org/10.1023/A:1009477316035

Trumper, R. (2003). The need for change in elementary school teacher training—a

cross-college age study of future teachers' conceptions of basic astronomy concepts. *Teaching and Teacher Education*, *19*(3), 309–323.http://dx.doi.org/10.1016/S0742-051X(03)00017-9

Trumper, R. (2006a). Teaching future teachers basic astronomy concepts—seasonal changes—at a time of reform in science education. *Journal of Research in Science Teaching*, *43*(9), 879-906. http://dx.doi.org/10.1002/tea.20138

Trumper, R. (2006b). Factors affecting students' junior high school students' interest in physics. *Journal of Science Education and Technology*, *15*(1), 47-58. http://dx.doi.org/10.1007/s10956-006-0355-6

Trundle, C. K., Atwood, K. R., & Christopher, E. J. (2002). .Preservice elementary teachers' conceptions of moon phases before and after instruction. *Journal of Research in Science Teaching*, *39*(7), 633-658.http://dx.doi.org/10.1002/tea.10039

Tsai, C., & Chang, C. (2005). Lasting Effects of Instruction Guided by the Conflict Map: Experimental Study of Learning about the Causes of Seasons. *Journal of Research in Science Teaching, 42*(10), 1089-1111. Doi: 10.1002/tea.10039

Türk, C., & Kalkan, H. (2015). The effect of planetariums on teaching specific astronomy concepts. *Journal of Science Education and Technology,24*(1), 1-15. http://dx.doi.org/10.1007/s10956-014-9516-6

Türk, C., Şener, N., & Kalkan, H. (2015). Pre-Service Teachers' Conceptions of Specific Astronomy Concepts: A Longitudinal Investigation. *Journal of Social Science Studies*, *2*(2), 56-87.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology*, *41*(4), 347–357. http://dx.doi.org/10.1111/j.1464-0597.1992.tb00711.x

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *41*(1), 45-69. http://dx.doi.org/10.1016/0959-4752(94)90018-3

Vosniadou, S., & Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535–585. http://dx.doi.org/10.1016/0010-0285(92)90018-W

Vosniadou, S., & Brewer, W. (1994). Mental models of the day/night cycle. *Cognitive Science*, *18*(1), 123–183. Doi.org/10.1016/0364-0213(94)90022-1

Zeilik, M., Schau, C., &Mattern, N. (1998). Misconceptions and their change in university level astronomy courses. *The Physics Teacher*, *36*(2), 104–107. http://dx.doi.org/10.1119/1.880056

Appendix

Appendix 1. Astronomy Achievement Test

1. **Which of the following is the reason for the formation of day and night?**
   - **a)** Earth's rotation around the Sun
   - **b)** Sun's rotation around the Earth
   - **c)** Earth's spin on its axis
   - **d)** The Earth going into and getting out of the dark area of the Sun

2. **When the Sun is directly above the flagpole, the pole will not cast shadow. When is this phenomenon observed from where you are (Samsun)?**
   - **a)** Every day at noon
   - **b)** Only on the first day of summer
   - **c)** On both the first days of spring and fall
   - **d)** Never from where you are (Samsun)

3. **What do we call the orbit that forms as a result of the rotation of the Earth around the Sun?**
   a) Axis **b)** Geoid **c)** Ellipse **d)** Meridian

4. **What direction does the Earth rotate on its axis?**
   a) South to North **b)** West to East **c)** North to South **d)** East to West

5. **What is the main reason that summer is warmer than winter?**
   - **a)** The Earth is closer to the Sun in summer
   - **b)** There are less clouds in summer when compared with winter
   - **c)** The Earth's spin axis has a specific tilt relative to its rotational plane around the Sun
   - **d)** The Sun radiates more energy in summer than in winter

6. **What is the main reason for the formation of seasons?**
   - **a)** The Earth's rotation around its axis
   - **b)** The change in the distance between the Earth and the Sun
   - **c)** The fact that the Earth's spin axis has a specific tilt
   - **d)** The fact that the Sun radiates different amounts of energy in each season

7. **The Earth's two hemispheres experience different seasons at the same time. What is the reason for this?**
   - **a)** The Sun's rotation around its axis
   - **b)** The Moon's rotation around the Earth
   - **c)** The Earth's axial tilt
   - **d)** Earth's spin on its axis

8. **How long does it take the Moon to orbit the Earth?**
   a) A day **b)** A week **c)** A month **d)** A year

9. **Which of the following gives the phases of the Moon in order?**
   - **a)** New moon–full moon–first quarter–last quarter
   - **b)** New moon–first quarter-last quarter–full moon
   - **c)** First quarter–new moon–full moon–last quarter
   - **d)** New moon–first quarter – full moon–last quarter

10. What is the main reason why we always see the same side of the Moon?

a) The fact that the Moon always has the same phases
b) The fact that the Moon's orbital speed is constant
c) The fact that the Earth's rotational speed is constant
d) The fact that the Moon's orbital period and the Earth's rotational period are the same

11. Which of the following is caused by the different gravitational force the Moon exerts on different parts of the Earth?

a) Seasons **b)** Tides **c)** Climate **d)** Day and night

12. Which of the following gives the correct order of the Sun, the Earth, Jupiter and the Moon from the smallest to the largest?

a) The Moon-Jupiter-the Earth-the Sun
b) The Moon-Jupiter-the Sun-the Earth
c) The Moon-the Earth-Jupiter-the Sun
d) The Moon-the Earth-the Sun-Jupiter

13. Which of the following is the correct when celestial bodies are ordered from the closest to the Earth to the farthest?

a) Stars-the Moon-the Sun-Pluto
b) The Sun-the Moon-Pluto-stars
c) The Moon-the Sun-Pluto-stars
d) The Moon-Pluton–the Sun-stars

14. Which one of the following is not a unit of length used in astronomy?

a) Astronomy Unit **b)** Light year **c)** Parsec **d)** Day

15. Which of the following matching is incorrect?

a) Saturn---the ringed planet
b) Jupiter---the smallest planet
c) Mars---the red planet
d) The Earth---the planet with life

16. Which of the following is correct when the planets in the Solar system are ordered from the closest to the Sun to the?

a) Mercury-Venus-Earth-Saturn-Uranus-Neptune-Jupiter-Mars
b) Mercury-Venus-Earth-Mars-Jupiter-Saturn-Uranus-Neptune
c) Mars-Venus-Mercury-Earth-Saturn-Jupiter-Uranus-Neptune
d) Mars-Venus-Mercury-Earth-Jupiter-Saturn-Uranus-Neptune

17. When looked from where you are, you can liken the shape of Ursa Major Constellation to a digger. This shape will change when you look from which of the following?

a) The farthest point of Turkey
b) The Moon's surface
c) The Planet Saturn
d) A far away star

18. When the visible movement of the Sun over the course of a year is marked as regards the design of the sky, from how many constellations can it be seen to pass over the course of a year?

a) 7 **b)** 13 **c)** 24 **d)** 30

19. What do we call the groups of stars which are together when we look up at the sky?

a) Halley's Comet **b)** Constellation **c)** The North Star **d)** Venus

20. Which of the following cannot be used to define the star which does not change its place during the day since it is almost in the same direction with the Earth's axis and which is used for finding directions?

a) Halley's Comet **b)** Polar star **c)** Polestar **d)** North star

21. What can be said about the center of the universe according to modern thinking and observations?

a) The universe does not have a center
b) Earth is the center of the universe
c) Sun is the center of the universe
d) Milky way galaxy is the center of the universe

22. Which of the following can be said about the location of Turksat B satellite when the distance between the Moon and the Earth is considered?

a) It is very close to the Earth.
b) It is approximately between the Earth and the Moon.
c) It is over the Moon.
d) It is very close to the Moon.

23. Which of the following is used to observe the sky?

a) Binoculars **b)** Overhead projector **c)** Periscope **d)** Telescope

Ahmet began to keep a lunar calendar at home. The first week of Ahmet's calendar is below. Answer questions 24, 25 and 26 by looking at this calendar.

Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Sunday	
16	19:45	**17**	18:45	**18**	18:55	**19**		**20**	19:30	**21**	19:20	**22**	19:30
						Rainy and Cloudy							

24. On which phase is the Moon on Tuesday?

a) Full Moon **b)** Crescent **c)** First quarter **d)** Last quarter

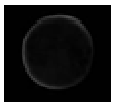
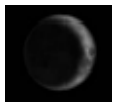
25. On which day did Ahmet most probably draw incorrectly?

a) Tuesday **b)** Wednesday **c)** Friday **d)** Saturday

26. On which phase will the Moon be on Monday, the 23rd of the month?

a)  **b)**  **c)**  **d)** 

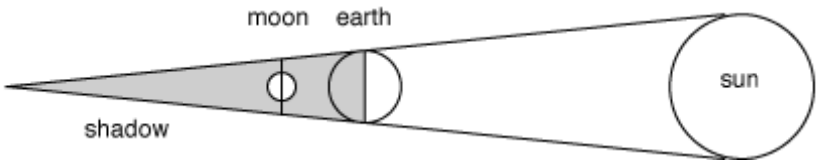
27. Which of the following is the "new moon" phase of the Moon?

a)  **b)**  **c)**  **d)** 

28. In order to observe a total eclipse of the Sun from Earth, in what phase does the Moon need to be?

a) Full Moon **b)** New Moon **c)** First Quarter **d)** Last Quarter

29. The drawing below shows an eclipse that involves the Sun, the Earth and the Moon. What is this eclipse called?



a) Earth Eclipse **b)** Solar Eclipse **c)** Lunar Eclipse **d)** New Moon Phase

30. The Earth follows an elliptical orbit around the Sun, as shown in Fig. I. If the Earth's orbit was changed into a full circle, as shown in Fig. II, the distance between the two bodies would remain the same throughout the year. How would this affect the seasons?

a) There would only be summer and winter.

b) There would only be autumn and spring.

c) There would be no seasons.

d) The seasons would be as they are today.

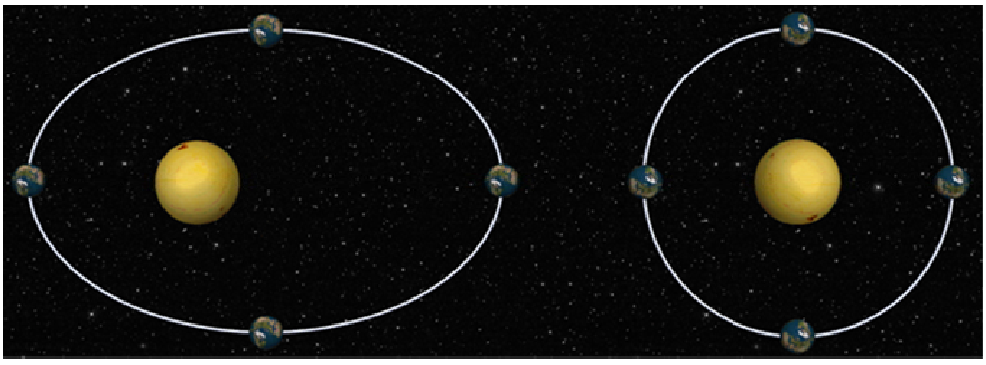


Figure I **Figure II**

31. Beijing, the capital of China, is 90° to the east of Samsun. Considering the direction in which the Earth rotates, when it is noon in Samsun, what time of day is it in Beijing?

a) Midnight
b) Noon
c) Sunrise (morning)
d) Sunset (evening)

32. When a student made observations in the hours of the morning, the afternoon and the evening, he saw the Sun in different positions as in the next figure. Which of the following is the reason why the Sun is seen in different positions?

a) The Sun's rotation around the Earth
b) Earth's spin on its axis
c) The Earth's rotation around the Sun
d) The movement of clouds