วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น ปีที่ 47 ฉบับที่ 4 (เดือนตุลาคม – ธันวาคม 2567) หน้า 90-
Journal of Education Khon Kaen University, Volume 47 No. 4 (October – December 2024) Page xx-xx

บทความวิจัย (Research Article)

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น
https://www.tci-thaijo.org/index.php/edkkuj
ดำเนินการวารสารโดย คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

Evaluating the Effectiveness of the Planets Approximating of Rising and setting Time (*PARST*) Educational Tool

Arthit Laphirattanakul, Suwicha Wannawichian, Pornrat Wattanakasiwich
Department of Physics and Materials Science, Faculty of Science Chiang Mai University, Chiang Mai, Thailand

Received: October 20, 2023 Revised: December 26, 20246 Accepted: December 27, 2024

Abstract

The Planets Approximating of Rising and Setting Time (PARST) educational tool was designed as a hands-on activity to illustrate planets' rising and setting times over long periods, considering the repeating patterns of these events. It was anticipated to enhance students' learning experience about planetary motion. This research aimed to evaluate PARST's learning effectiveness across three student groups in the classroom; these groups are (1) a group using only PARST, (2) a group taught about planetary motion without using PARST, and (3) a group taught about the planet motion and using PARST. The pre-test and post-test, with questions designed based on the elements of Bloom's taxonomy, were used to evaluate the three student groups. The four methods for evaluating pre-test and post-test results are (a) hypothesis testing, (b) effective size, (c) normalized gain, and (d) difficulty level. The four evaluation methods show the same pattern of the highest efficiency in the group of students who use the PARST educational tool only without taking lectures about planetary motion. These results imply that using the PARST educational tool without studying the rising and setting times of planets can encourage students to develop an understanding of planetary motion.

Keywords: educational tool, effectiveness, planetary motion, rising and setting time, Bloom's taxonomy

■ Introduction

Educational or instructional tools can be used in the classroom to assist student learning. Educational tools may be graphics, photos, or demonstration models used to express, simulate, or demonstrate the class content as part of the teaching process to reach specific learning goals. In other words, educational tools are used by teachers and learners to attain particular educational objectives. When using educational tools in the classroom, students are encouraged to take greater ownership by learning the materials and then coming to the session ready to apply and evaluate what they know. Preliminary

*Corresponding author. Tel.:
Email address: arthit.l@cmu.ac.th

research indicates that attendance, learning, and perceived value of education all increase with educational tools (Woods & Rosenberg, 2016). The education tool, which is the main focus of this work, is the Planet Approximating of Rising and Setting Time (PARST) educational tool (Laphirattanakul et al., 2024). As shown in Figure 1 for Jupiter, the PARST is used to estimate rising and setting times by turning the circle mask until the black arrow at the top of the slit points to the date on the cycles for which a user desires to know the rising and setting times. The correct times are indicated by the pattern shown through the slit (Laphirattanakul et al., 2024).

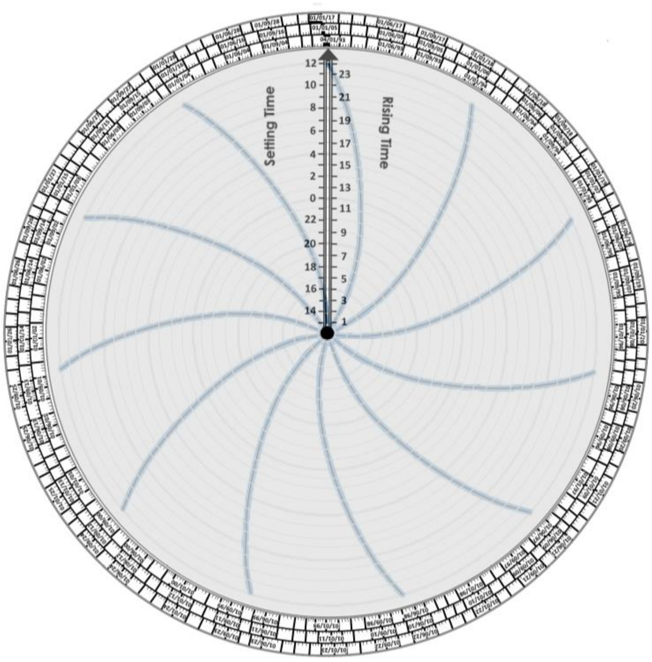


Figure *1*. The PARST educational tool for indicating the rising and setting time of Jupiter.

This educational tool was designed to reveal the repeated cycles and period variations when indicating each planet's rising and setting time. It contains the pattern of changing planets' rising and setting time, corresponding to each planet's synodic event. Therefore, students who use this educational tool are expected to enhance their ability to connect the characteristics of the pattern to the planet's consecutive positions in orbit, which will lead to the understanding of planetary motion and planets' configurations in the classroom.

The evaluation of educational tools' effectiveness is crucial for ensuring optimal learning outcomes in education (Seidel & Shavelson, 2007). Without proper assessment, educators cannot verify if tools like PARST actually enhance student understanding of complex astronomical concepts such as planetary motion and celestial coordinates. Research has shown that students often develop misconceptions about planetary motion when taught through traditional methods alone (Bakas & Mikropoulos, 2003; Stover & Saunders, 2000). The evaluation process serves multiple critical purposes: it enables the improvement of educational resources, validates the tool's effectiveness as a learning aid, and provides objective assessment of teaching methods (Seidel & Shavelson, 2007). Particularly in astronomy education, where abstract concepts are essential, measuring tool effectiveness helps ensure students can correctly understand planetary movements. Without systematic evaluation of educational tools like PARST, educators risk implementing resources that may not effectively support student learning, potentially leading to persistent misunderstandings of fundamental astronomical concepts (Galano et al., 2018).

The effectiveness evaluation can reveal the effectiveness of the educational tool in explaining celestial coordinates, as mentioned above. To evaluate the effectiveness of the educational tool, the students who use this educational tool will be assessed in pre-test and post-test. The pre-test is given to analyze how much the students know about the topic and to help them focus more on the educational tool before using it. After using the educational tool, the post-test is given to evaluate students' learning of the key concepts. For assessment, the students are separated into two groups: (1) a group lectured with the educational tool (treatment group) and (2) a group lectured without the educational tool (control group). The mean scores of the pre-test and post-test of these two groups will be analyzed and compared by statistical analysis methods.

■ Significance and Purposes

1) The effectiveness evaluation of the Planet Approximating of Rising and Setting Time (PARST) educational tool is employed to reveal its capability in assisting students' learning about the rising and setting time of planets and their repeated patterns of synodic period variations in the rising and setting time cycles.
2) The evaluation results will significantly assist teachers in improving their teaching plans about the sequences of motions of celestial objects. This improvement will enhance students' efficiency in learning about planetary motion and planets' configurations in the classroom.

■ Literature Reviews

From the long history of astronomy, teaching planetary motion has employed various types of tools, e.g., Zodiac signs illustrated in mosaics (Cohen, 2022). The demonstration of the celestial coordinate system is a way to specify the positions of celestial objects. They are often implemented in spherical coordinates and projected on a celestial sphere (Ruangsuwan & Arayathanitkul, 2009). Learning about celestial motion requires learners to understand the sequences of motions of celestial objects across the frames of reference. The first frame of reference illustrated by the educational tool is the students' Earth-based perspective. Students' understanding of astronomy from this frame of reference is initially based on their observations of celestial objects in the sky. The explanation of celestial motion phenomena requires students' understanding of the space-based perspective, in which celestial objects are moving in space. Thus, the full explanation of celestial motion phenomena requires understanding both Earth-based and space-based perspectives and the ability to shift between these perspectives to explain why celestial

doi: 10.14456/edkkuj.2020.x

objects appear to move or change as seen from the Earth (Plummer, 2014). An educational tool that describes these celestial motions and the relationships between celestial coordinates should not be too complicated for students. Moreover, while the direct observations with advanced instruments, e.g., Infrared Telescope Facility (Peralta et al., 2023), can give students a clearer perspective about the planet, the budget for the class demonstration would increase significantly.

The preferred analysis method for comparing the difference between pre-test and post-test scores is the paired sample t-test (Xu et al., 2017). However, many researchers believe this method does not provide enough information to indicate the effectiveness of an educational tool (Sullivan & Feinn, 2012). Evaluation of the effect size and the normalized gain are other methods that can be used instead of the paired sample t-test.

Methods

The effectiveness of the PARST educational tool was tested on 57 secondary school students in the science program, which studied the topic of planetary objects in national curriculum in the same high school, in October 2020. The lecture was given by the same instructor. The educational tool was used during an astronomy class about planetary motion. The students were divided into three groups: 18 students in the “Teach Only” group were taught about planetary motion without using the PARST educational tool, 18 students in the “Use Only” group only used the PARST educational without attending a lecture, and 21 students in “Teach and Use” group were taught about planetary motion and used the PARST educational tool. All students took the pretest before and post-test after the procedures. The pre-test and post-test have 15 multiple-choice questions designed to evaluate the first five elements of behavior related to students’ intelligence, followed by Bloom’s taxonomy (Bloom, 1956), as shown in Table 1 and Appendix 1.

In this questionnaire, the highest level of metacognitive knowledge, such as evaluating and creating, was rarely used. Higher levels of taxonomy are generally not used to examine students within the undergraduate level (Karaksha et al., 2014) and high school level in this study. The pre-test and post-test questionnaires were compared and analyzed by the first five elements, as shown in Table 1.

Table *1*. The levels of questionnaires regarding Bloom’s Taxonomy.

Bloom’s taxonomy	Questionnaires
1. Remember	1, 13
2. Understand	7, 11, 14
3. Apply	2, 3, 5, 6, 9
4. Analyze	4, 8,10, 12
5. Evaluate	15
6. Create	-

For ***hypothesis testing***, when the pre-test and post-test scores were obtained, the mean scores were determined and compared as part of hypothesis testing. The most preferable analysis method for comparing the difference between pre-test and post-test scores is the paired sample t-test. Neyman and Pearson (Neyman & Pearson, 1933) developed this test with null and alternative hypotheses. The null hypothesis (H0) states that the means of student scores from pre-instruction and post-instruction are equal. The alternative hypothesis (H1), in which the post-instruction mean score differs from the pre-instruction mean score by a statistically significant amount (higher or lower), is only tenable if sufficient evidence is provided to reject the null hypothesis. The t-test assumes that the population under inspection represents a normal distribution. The significance level, **α**, of a t-test represents the maximum likelihood of rejecting a true null hypothesis. For a 95% confidence interval, the significance level is **α**=0.05. If the probability value p is less than or equal to **α**, a statistically significant difference exists between the means of the two compared populations. Thus, the null hypothesis can be rejected. If p is greater than **α**, the null hypothesis must be retained (Bardar, 2007). The educational tool is expected to be efficient when the mean value of the post-test is higher than the mean value of the pre-test. In other words, the probability value p is less than or equal to **α**. Thus, the educational tools can efficiently be used in the classroom, with the goal of learning progression support for the students.

However, hypothesis testing has been used to test the effectiveness of an educational tool for about a century. Until the beginning of the 21st century, many researchers argued that hypothesis testing does not provide enough information to indicate the effectiveness of an educational tool (Nakagawa & Cuthill, 2007; Barnett & Mathisen, 1997; Coe, 2012). They believed the hypothesis testing was inappropriate and inadequate to interpret whether students can develop a learning process, while the null hypothesis was rejected. Thus, another method, i.e., effect size and normalized gain, can be used instead of a paired sample t-test.

Effect size (d) measures the magnitude of the effect created by an instructional intervention or treatment. It is interpreted as the percentage of non-overlap of the treated group's score with that of the control group without instruction (Rakap, 2015). Mathematically, the effect size relates to the ratio of the difference of mean scores between two groups (treated and controlled groups) to the standard deviation of the treated group's scores as equation (1).

$$d_i = \frac{\bar{x}_i - \bar{x}_j}{SD_i}, \quad (1)$$

where i and j represent those two groups, respectively. Moreover, SD_i is the standard deviation of group i. Effect size can be used for either a single study or multiple studies. It has a practical significance. An effect size of 0.2 is considered small, 0.5 is medium, and 0.8 or greater is largely effective (Cohen, 1988), (Ole, 2020), as shown in Figure *2*.

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 47(4), 90-111
doi: 10.14456/edkkuj.2020.x

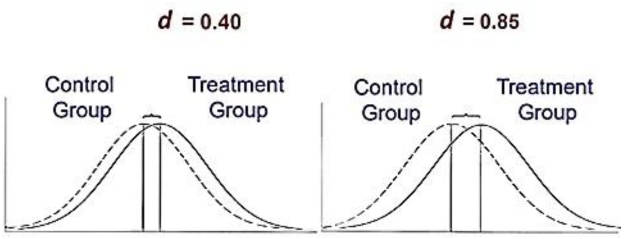


Figure *2*. Comparison between the Effect Size of 0.4 and 0.85.

The normalized gain indicates how much the post-test score increases from the pre-test score. This increasing score only depends on the learning progression of the teaching methods or educational tools. The Normalized Gain $\langle g\rangle$ assesses students' learning progression by comparing the difference between pre-test and post-test scores to the maximum possible gain, as in equation (2).

$$\langle g\rangle = \frac{\langle \text{post}\rangle - \langle \text{pre}\rangle}{(\text{full score}) - \langle \text{pre}\rangle} \tag{2}$$

Where $\langle post\rangle$ and $\langle pre\rangle$ indicate the average post-test and pre-test scores, respectively.

This gain measurement is used to compare pre-test and post-test scores to identify the effectiveness of the teaching methods or educational tools. Hake (Hake, 1998) introduced the definition of this measurement that $\langle g\rangle$ is considered to be a high gain if $\langle g\rangle > 0.7$, a medium gain if $0.7 > \langle g\rangle > 0.3$, and a low gain if $\langle g\rangle < 0.3$. Based on this method, this normalization, by dividing with $(\text{full score}) - \langle \text{pre}\rangle$ in equation (2), is suitable for the analysis of diverse student populations with widely varying scores, corresponding to a comprehensive initial knowledge state. This method is applied based on the general information that the courses with lecture-based instruction had low $\langle g\rangle$, courses with active-engagement instruction primarily had medium $\langle g\rangle$, and following Hake's study, all courses have $\langle g\rangle < 0.7$.

The difficulty level of a test's questions can be measured by comparing the number of students with high test scores who answered that item correctly with the number of students with low scores who answered the same question correctly (Boopathiraj & Chellamani, 2013). The high and low test scores are separated by the mean value, as shown in Figure *3*.

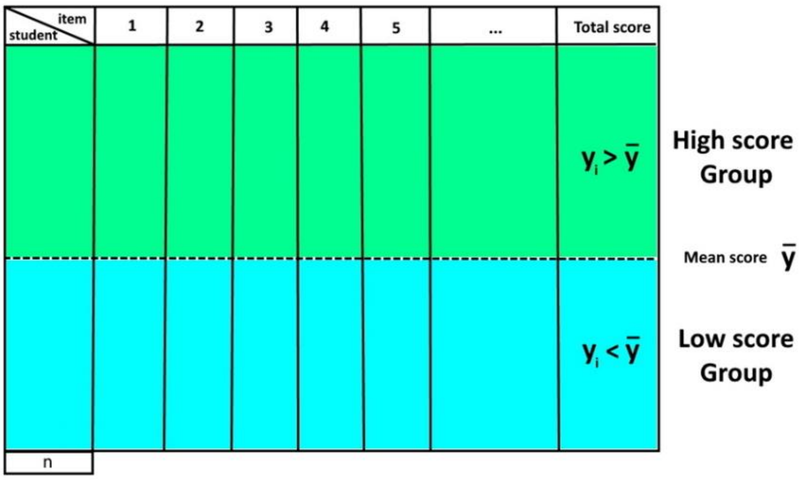


Figure *3*. Students are separated into two groups: students with total scores higher $(y_i > \bar{y})$ and lower $(y_i < \bar{y})$ than the mean score $(\bar{y})$.

The students who scored higher than the mean score were considered to have good performances. In contrast, students who scored lower than the mean score were considered to have bad performance. The difficulty level of each question (P_i) is calculated by counting the number of students with high test scores who answered that question correctly (N_H) plus the number of students with low scores who answered the same question correctly (N_L) divided by half of the total number of students (n).

$$P_i = \frac{N_H + N_L}{n/2} \tag{3}$$

If the difficulty level, P_i, is less than 0.2, the question is considered difficult and can hardly distinguish high-performance students from low-performance students. On the other hand, if P_i is between 0.2 and 0.7, the question is considered of medium difficulty and can classify the student's performance. In addition, if P_i is greater than 0.7, the question is considered easy because students with high and low test scores can answer correctly.

The performance in the pre-test and post-test is conducted to illustrate the percentage of students who obtained pre-test and post-test scores differently from each question. This analysis quantitatively shows the performance comparison between pre-test and post-test of all students. The performance indicates the ability of a student to answer each question correctly after using the educational tool. The number of questions answered correctly and incorrectly in the pre-test and post-test were counted and divided by the number of students.

$$PFM_{(i,j)} = \frac{N_{(i,j)}}{n} \times 100\ . \tag{4}$$

Where $i = 1$ for correct answer in pre-test,
$i = 0$ for wrong answer in pre-test,
$j = 1$ for correct answer in post-test,
$j = 0$ for wrong answer in post-test,
$N_{(i,j)}$ is the number of students in each i and j combination,
n is the number of total students.

From equation (4), the performance in the pre-test and post-test can be classified into 4 cases:

- $PFM_{(1,1)}$ is for the case of consistently good performance (C). These students gave correct answers for both the pre-test and post-test.
- $PFM_{(1,0)}$is for the case of regressive performance (R). These students give correct answers for the pre-test but wrong answers for the post-test.
- $PFM_{(0,1)}$ is for the case of improved performance (I). These students give wrong answers for the pre-test but correct answers for the post-test.

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 47(4), 90-111
doi: 10.14456/edkkuj.2020.x

- $PFM_{(0,0)}$ is for the case of no improvement in the performance (N). These students give wrong answers for both the pre-test and post-test.

These 4 cases will be studied with questions based on Bloom's taxonomy, which classifies five elements of behavior related to a student's intelligence. This method can analyze the performance of students in quantitative values. This method should be able to indicate the development of students' abilities after using the educational tool.

■ Results and Discussion

Fifty-seven students in the science program were separated into three groups: Use Only, Teach Only, and Teach and Use groups. The test score development of the students from these groups is shown in Figure *4*.

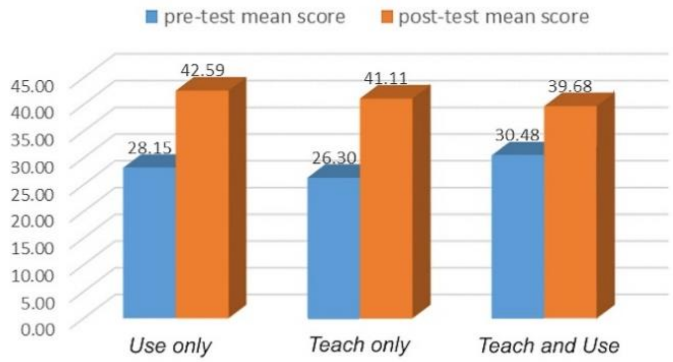


Figure *4*. Percentage of students' development between pre-test and post-test in three testing groups

The pre-test was organized for all three groups. Firstly, both the *Teach and Use* group and the *Teach Only* group were taught about the rising and setting times of the planets. Next, both the *Use Only* group and the *Teach and Use* group were conducted to use the *PARST* educational tools to indicate the rising and setting times of the planets. All student groups took the post-test one week later. The students' pre-test and post-test scores are shown in Figure *4*. There were reasonable increments between the pre-test and post-test, whereby more students from the *Teach and Use* group had high pre-test scores compared to other groups. However, the scores obtained by the *Teach and Use* group increased less than other groups in the post-test. These results suggest that the *Teach and Use* group students had a better background before taking the class. Additionally, the results were analyzed by four methods.

The hypothesis testing consists of a null hypothesis (H_0) and an alternative hypothesis (H_1) described as follows.

H_0: No significant difference in student performance between the pre-test and post-test

H_1: Significant difference in student performance between the pre-test and post-test

For a 95% confidence interval, the significance level is $\alpha = 0.05$. The p-values were determined by paired sample t-test, as shown in Table 2.

Table *2*. The p-values of hypothesis testing by paired sample t-test in three testing groups.

Group	t	n	p-values
Teach Only	-5.06	18	0.0000975
Use Only	-6.65	18	0.0000041
Teach and Use	-5.91	21	0.0000089

From Table 2, all of the p-values were less than 0.05. The null hypothesis can be rejected. These results imply that the total post-test responses were significantly higher than the pre-test responses for all groups. Quantitatively, the students who used only *PARST* educational tools without being taught about the rising and setting time of planets have the largest difference between pre-test and post-test mean scores. In contrast, students who only studied about the rising and setting time of planets without using the *PARST* educational tool have the least difference between pre-test and post-test mean scores.

For ***the effect size***, the difference between the mean post-test and pre-test scores divided by the standard deviation of the treated group was determined by equation (1). The results are shown in Table 3.

Table *3*. The effect sizes of three student groups

Group	Effect size (d)
Teach Only	1.32
Use Only	1.34
Teach and Use	1.24

From Table 3, all effect sizes are bigger than 0.8. However, the effect size of students who used only *PARST* educational tools without studying planets' rising and setting times is quantitatively bigger than the effect sizes of other groups. These differences mean the post-test score of this group is significantly higher than the pre-test score. For ***the normalized gain***, each student's mean pre-test and post-test scores were determined and substituted in equation (2) to evaluate the gain of each group, where a perfect test score is 15. All three groups' average gains are shown in Table 4.

Table *4*. The average gains of three student groups

Group	Average gain $< g >$
Teach Only	0.19
Use Only	0.20
Teach and Use	0.13

From Table 4, the average gains of the three student groups are less than 0.30, considered a low gain. However, students who used only the *PARST* educational tool had the highest gains than other groups. For the difficulty level, this analysis started with arranging students' test scores in ascending order. Next,

doi: 10.14456/edkkuj.2020.x

the numbers of students with scores higher and lower than average were obtained. ***The difficulty*** of each question was calculated and compared with the difficulty difference between the pre-test and post-test for every items, considering the categories of questions according to Bloom's taxonomy, as shown in Table 5. We considered pre-test questions identified as difficult ($P_i < 0.2$) or medium ($0.2 < P_i < 0.7$) that become medium or easy post-test questions ($P_i \geq 0.7$). The questions, whose difficulty levels change from high difficulty in the pre-test to low difficulty in the post-test, are defined as *decrement difficulty level*

ficulty level of each question (Q), along with the levels of questionnaires regarding Bloom's taxonomy. For difficult questions, the difficulty level is 200. For medium questions, the difficulty level is between 0.200 and 0.700. Next, for easy questions, the difficulty level is more than 0.700. The bold cate how the students have progressed in their learning, corresponding to questions in *decrement difficulty level*. The red letters indicate the n *increment difficulty level*.

roup	test	Q 1 Remember	Q 2 Apply	Q 3 Apply	Q 4 Analyze	Q 5 Apply	Q 6 Apply	Q 7 Understand	Q 8 Analyze	Q 9 Apply	Q 10 Analyze	Q 11 Understand	Q 12 Analyze	Q 13 Remember	Q 14 Understand	Q 15 Evaluate
Only	pre-test	1.222	0.778	0.444	0.444	0.667	0.333	0.778	0.333	0.556	0.333	0.444	0.444	0.222	0.111	0.778
		easy	easy	medium	medium	medium	medium	easy	medium	medium	medium	medium	medium	medium	difficult	easy
	post-test	1.778	1.333	0.667	0.222	0.889	0.333	0.667	0.778	0.889	0.778	0.222	0.889	0.556	1.111	1.222
		easy	easy	medium	medium	easy	medium	medium	easy	easy	easy	medium	easy	medium	easy	easy
Only	pre-test	1.444	0.667	0.667	0.333	0.444	0.778	0.667	0.333	0.222	1.000	0.333	0	1.222	0.222	0.111
		easy	medium	medium	medium	medium	easy	medium	medium	medium	easy	medium	difficult	easy	medium	difficult
	post-test	2.000	0.667	0.889	0.667	0.444	1.111	1.111	0.333	0.333	1.444	0.667	0.444	1.222	0.556	0.889
		easy	medium	easy	medium	medium	easy	easy	medium	medium	easy	medium	medium	easy	medium	easy
d Use	pre-test	1.810	0.857	0.286	0.571	0.476	0.19	0.952	0.381	0.381	1.143	0.571	0.762	0	0.286	0.476
		easy	easy	medium	medium	medium	difficult	medium	medium	medium	easy	medium	easy	difficult	medium	medium
	post-test	1.905	1.143	0.667	0.952	0.667	0.476	0.952	0.286	0.571	1.333	0.571	0.762	0.476	0.190	0.952
		easy	easy	medium	easy	medium	medium	easy	medium	medium	easy	medium	easy	medium	difficult	easy

nclusions about the number of questions in the pre-test and post-test were compared in three learning: *Consistent Decrement* and *Increment difficulty*

up	*Consistent difficulty level*			*Decrement difficulty level*			*Increment difficulty level*		
	Easy to Easy	Medium to Medium	Difficult to Difficult	Difficult to Easy	Difficult to Medium	Medium to Easy	Easy to Medium	Medium to Difficult	Easy to Difficult
ch Only	3	5	0	1	0	5	1	0	0
Only	4	7	0	1	1	2	0	0	0
ch and Use	4	5	0	0	2	3	0	1	0

questions. This type of question shows a significant development in students' learning. However, the difficulty levels of some questions changed from low to high. These questions are *increment difficulty level* questions. That means students may be confused about the questions. If the difficulty levels in the pre-test and post-test are similar, these questions are in a *consistent difficulty level*. The questions classified by their types are summarized in Table 6.

The results in Table 6 show that the *Teach* group has six questions in *decrement difficulty level*. This group has more *decrement difficulty level* questions than those of other groups. Moreover, for the *Teach group,* one *increment difficulty level* question was found for the *understood* question type based on Bloom's taxonomy. Compared to the *Teach Only* group*,* the *Teach and Use* group also has one *increment difficulty level* question for *understand* question type, as shown in Table 5. On the other hand, the *Use Only* group has no questions in the *increment difficulty level*, but there are four questions in the *decrement difficulty level*. Moreover, seven questions, the largest among the three groups of students, are at a *consistent difficulty level*.

For ***the performance*** in pre-test and post-test, the pre-test and post-test scores for each student were obtained. The performance in pre-test and post-test based on students' scores are determined by equation (4) for all questions performed in each group. Each question was classified based on an element of behavior related to the student's intelligence, according to Bloom's taxonomy, as shown in Tables 7, 8, and 9. Based on the performance of the *Use Only* group, as shown in Table 7, the number of students with no improvement ($PFM_{(0,0)}$) ranks first. This result implies that the questions are very difficult for this group. However, the largest number of students in this group can correctly answer the *remembered* question. Moreover, the number of students with the performance $PFM_{(0,1)}$ ranks second. This result implies that many students in the *Use Only* group had improved performance. Most students who had consistently good performance ($PFM_{(1,1)}$) in this group can correctly answer the *remembered* question.

For the *Teach Only* group*,* as shown in Table 8, the questions appear to be very difficult for this group, which $PFM_{(0,0)}$ ranks first. The number of students with the performance $PFM_{(0,0)}$ for all question types is that many students in this group have improved, but the overall average value is slightly less than that of the *Use Only* group. The overall average value of regressive performance ($PFM_{(1,0)}$) for the *Teach Only* group is slightly less than that of the *Use Only* group.

For the *Teach and Use* group, as shown in Table 9, the average value of $PFM_{(0,0)}$ for this group is higher than that of the *Use Only* and *Teach Only* groups*.* This result implies that the combination of teaching the planetary topic and using the educational tool (*Teach and Use*) cannot effectively improve students'

doi: 10.14456/edkkuj.2020.x

performance compared to the separated cases of teaching about planetary motion (*Teach Only*) and using the *PARST* educational tool (*Use Only*).

Table *7*. Performance in pre-test and post-test (PFM) of students in the *Use Only* group. The element levels correspond to Bloom's taxonomy.

Use Only group					
Question	Level	$PFM_{(1,0)}$	$PFM_{(0,0)}$	$PFM_{(1,1)}$	$PFM_{(0,1)}$
4	analyze	5.56	61.11	11.11	22.22
8	analyze	16.67	66.67	0	16.67
10	analyze	16.67	11.11	33.33	38.89
12	analyze	0	77.78	0	22.22
average		**9.73**	**54.17**	**11.11**	**25.00**
2	apply	22.22	44.44	11.11	22.22
3	apply	11.11	44.44	22.22	22.22
5	apply	22.22	55.56	0	22.22
6	apply	27.78	16.67	11.11	44.44
9	apply	11.11	72.22	0	16.67
average		**18.89**	**46.67**	**8.89**	**25.55**
15	evaluate	0	55.56	5.56	38.89
average		**0**	**55.56**	**5.56**	**38.89**
1	remember	0	0	72.22	27.78
13	remember	33.33	5.56	27.78	33.33
average		**16.67**	**2.78**	**50.00**	**30.56**
7	understand	16.67	27.78	16.67	38.89
11	understand	16.67	50	0	33.33
14	understand	11.11	61.11	0	27.78
average		**14.82**	**46.30**	**5.56**	**33.33**
Overall average		**12.02**	**41.10**	**16.22**	**30.67**

For the conclusion of the performance, Table 10 clearly shows the highest percentage of the number of students who have no improvement in the performance of each group (i.e., answering wrongly in both pre-test and post-test). The results confirm that the questions are very difficult, even after the students learned about planetary motion from the lecture or using the educational tool. However, the higher percentage of correctly answered pre-test questions in the *Teach and Use* group indicates that the students may have aptitude in this topic before learning the planetary topic and using the educational tool. After using *PARST* and learning the planetary motion, about two-thirds of students in the *Teach and Use* group who answered the pre-test questions correctly also answered the post-test questions correctly. On the other hand, more than two-sevenths of the students in this group who wrongly answered the pre-test questions answered the post-test questions correctly.

Table *8*. Performance in pre-test and post-test (**PFM**) of students in the *Teach only* group. The levels correspond to Bloom's taxonomy.

Teach Only group					
Question	Level	$PFM_{(1,0)}$	$PFM_{(0,0)}$	$PFM_{(1,1)}$	$PFM_{(0,1)}$
4	analyze	16.67	72.22	5.56	5.56
8	analyze	5.56	55.56	11.11	27.78
10	analyze	5.56	55.56	11.11	27.78
12	analyze	11.11	44.44	11.11	33.33
average		**9.73**	**56.94**	**9.72**	**23.61**
2	apply	5.56	27.78	33.33	33.33
3	apply	16.67	50	5.56	27.78
5	apply	16.67	38.89	16.67	27.78
6	apply	11.11	72.22	5.56	11.11
9	apply	11.11	44.44	16.67	27.78
average		**12.22**	**46.67**	**15.56**	**25.56**
15	evaluate	5.56	33.33	33.33	27.78
average		**5.56**	**33.33**	**33.33**	**27.78**
1	remember	5.56	5.56	55.56	33.33
13	remember	5.56	66.67	5.56	22.22
average		**5.56**	**36.12**	**30.56**	**27.78**
7	understand	27.78	38.89	11.11	22.22
11	understand	16.67	72.22	5.56	5.56
14	understand	5.56	38.89	0	55.56
average		**16.67**	**50.00**	**5.56**	**27.78**
Overall average		**9.95**	**44.61**	**18.95**	**26.50**

In the case of improved *performance* (answering pre-test wrongly but correctly answering post-test), the *Use Only* group has more students in this case than the other groups, with 30.67% of all students. On the other hand, there are 26.50% of *improved performance* students in the *Teach Only* group. Moreover, in the *Teach and Use* group, the *improved performance* students make up only 21.03% of all students in this group. This information implies that most students can improve their performance on the test with only the use of the *PARST* teaching tool and without taking the lecture.

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 47(4), 90-111

doi: 10.14456/edkkuj.2020.x

Table *9*. Performance in pre-test and post-test (PFM) of students in the *Teach and Use* group. The levels correspond to Bloom's taxonomy.

Teach and Use group					
Question	Level	$PFM_{(1,0)}$	$PFM_{(0,0)}$	$PFM_{(1,1)}$	$PFM_{(0,1)}$
4	analyze	4.76	47.62	23.81	23.81
8	analyze	9. 52	76.19	9.52	4.76
10	analyze	14.29	19.05	42.86	23.81
12	analyze	9.52	52.38	28.57	9.52
average		**9.52**	**48.81**	**26.19**	**15.48**
2	apply	14.29	28.57	28.57	28.57
3	apply	14.29	52.38	0	33.33
5	apply	4.76	61.9	19.05	14.29
6	apply	9.52	66.67	0	23.81
9	apply	9.52	61.9	9.52	19.05
average		**10.48**	**54.28**	**11.43**	**23.81**
15	evaluate	9.52	42.86	14.29	33.33
average		**9.52**	**42.86**	**14.29**	**33.33**
1	remember	4.76	0	85.71	9.52
13	remember	0	76.19	0	23.81
average		**2.38**	**38.10**	**42.86**	**16.67**
7	understand	19.05	33.33	28.57	19.05
11	understand	19.05	52.38	9.52	19.05
14	understand	14.29	76.19	0	9.52
average		**17.46**	**53.97**	**12.70**	**15.87**
Overall average		**9.87**	**47.60**	**21.49**	**21.03**

Table *10*. Divided cases of all three groups and their performances.

group	pre-test	post-test	case
Use Only	Correct (**28.23%**)	Correct (**16.22%**)	C
		Wrong (**12.02%**)	R
	Wrong (**71.77%**)	Correct (**30.67%**)	I
		Wrong (**41.10%**)	N
Teach Only	Correct (**28.90%**)	Correct (**18.95%**)	C
		Wrong (**9.95%**)	R
	Wrong (**71.10%**)	Correct (**26.50%**)	I
		Wrong (**44.61%**)	N

Teach and Use	Correct (**31.36%**)	Correct (**21.49%**)	C
		Wrong (**9.87%**)	R
	Wrong (**68.63%**)	Correct (**21.03%**)	I
		Wrong (**47.60%**)	N

Cases: C – consistently good performance, R – regressive performance, I – improved performance, N – no improvement in the performance

The effectiveness of the PARST educational tool can be explained through Kolb's experiential learning cycle theory. This theory consists of four stages that describe how students learn through experience: concrete experience, reflective observation, abstract conceptualization, and active experimentation (Abdulwahed & Nagy, 2009). In this research, students in the Use Only group engaged with all four stages through the PARST educational tool. They gained hands-on experience by manipulating the tool to observe planetary positions, reflected on the patterns of rising and setting times they observed, connected these observations to understand planetary motion concepts, and tested their understanding by making predictions about planetary positions. The higher performance of the Use Only group compared to other groups suggests that completing all stages of the experiential learning cycle leads to better learning outcomes. The effect sizes of 1.34 for the Use Only group, compared to 1.32 for the Teach Only group and 1.24 for the Teach and Use group, demonstrate this enhanced effectiveness. These findings suggest important implications for astronomy education. The superior performance of students who used only the educational tool indicates that introducing astronomical concepts through interactive tools before formal instruction may be more effective than traditional lecture-based approaches. Moreover, the lower performance of the Teach and Use group suggests that combining both teaching methods simultaneously may interfere with students' natural learning process. This interference could create difficulties in processing information, potentially reducing the effectiveness of both teaching methods. These results suggest that astronomy educators should consider implementing a tool-first approach, allowing students to explore and learn through educational tools before providing formal instruction about planetary motion concepts.

■ Conclusion

Comparing the effectiveness of the *PARST* educational tool in the *Use Only*, *Teach Only,* and *Teach and Use* groups, four analysis methods for students' performances show the same pattern of highest effectiveness in the *Use Only* group. For the hypothesis testing method, the *PARST* is more effective for the *Use Only* group than other groups because the p-value is less than the significant level, α, and is less than that of the *Teach Only* and *Teach and Use* groups. However, the total post-test responses were significantly higher than the pre-test responses for all three groups. For the results of the second method, the effect size of students who only used the *PARST* educational tool without studying the rising and setting times of planets was quantitatively bigger than the effect size in other groups. However, the effect sizes of all groups were bigger than 0.8, indicating that their efficiencies are significantly increased. For the normalized gain method, the *Use Only* group also has a higher gain than others. However, the gains of all groups are less than 0.30, which is classified as low gain.

For the difficulty level method, the number of questions with a *decrement* in *difficulty level* in the *Use Only* group is lower than in the other groups. However, there is no question with an *increment* in *difficulty level*. These results suggest that the questions do not become more difficult after the students use the *PARST*. For other groups, the *increments* in *difficulty levels* were found, corresponding to a change in difficulty level from low difficulty in the pre-test to high difficulty in the post-test. For analyzing the method of performance in pre-test and post-test, The *Use Only* group may have *improved performance* than the other groups. This result implies that only using the *PARST* educational tool without studying the rising and setting times of planets can encourage students to develop a learning process on planetary motion.

In order to assess the effectiveness, the PARST educational tool was directly assessed by 27 secondary school students who study planetary motions and planets' configurations in the classroom. Upon using this educational tool, students learn to analyze the acquired rising time from using PARST in cooperating with their understanding of the planet's position. According to a post-class survey, 74.1% of students replied that the PARSTs can help them learn about the difference in each planet's rising and setting times. In addition, the survey shows that 63.0% of students replied that the PARSTs can help them learn that the rising and setting times relate to the planets' location in their orbits corresponding to Earth's position. 55.6% of students can identify the times when the planets' rising and setting times repeat, as observed from Earth. Moreover, 77.8% of students agreed that the PARSTs can help them develop the learning progression about planetary motion. The majority of the students comment that the educational tool is easy to use and can be used by the general public.

In conclusion, the students who learn about planetary motion using only the PARST educational tool without taking a lecture can develop an understanding, as shown in every effective evaluation method in this research. This effectiveness can be explained through Kolb's experiential learning cycle, where students progress through concrete experience, reflective observation, abstract conceptualization, and active experimentation while using the tool. Moreover, students who only took the lecture without using PARST developed a learning process to a lesser extent than those who only used PARST. These results show that using PARST educational tools can almost replace teaching about planetary motion, especially when the teacher cannot teach students about this topic. The findings suggest that a tool-first approach, where students interact with educational tools before receiving formal instruction, may be more effective than traditional lecture-first methods. However, careful consideration must be given to the timing and integration of different instructional methods, as simultaneous exposure to both teaching methods might create cognitive overload and reduce learning effectiveness. In addition, the collaboration between PARST educational tools and assigned worksheets can be another alternative teaching plan to help students understand more about planetary motion when properly sequenced.

Finally, in response to the purposes of this study, this work on the effectiveness evaluation of the PARST educational tool reveals its capability to assist students in learning about the planets' rising and setting times corresponding to each planet's synodic period, even though they have not taken a lecture before. In addition, the collaboration between *PARST* educational tools and assigned worksheets can be an alternative teaching plan to help students understand more about planetary motion and planets' configurations in the classroom.

Limitations and Recommendations

Limitations

The educational tool illustrates the planet's rising and setting times on a cardboard chart. The limitation of the size of the tool allows only a specific range of times to be revealed on the PARST. More information forces the font size of the number to be smaller and harder for the students to read.

Recommendations

Although the effectiveness evaluation results have shown that this educational tool can be used as a substitute for teaching planetary motion topics, some results indicate cases of regressive performance or no improvement in student performance. Furthermore, in terms of difficulty level, there are also some cases where the difficulty level of the post-test increased compared to the pre-test or did not change. Therefore, this educational tool may need further development to assist students in learning sufficiently. Improvement of the design of PARST could help students obtain progressive performance when they learn about planetary motion.

References

Abdulwahed, M., & Nagy, Z. (2009). Applying Kolb's Experiential Learning Cycle for Laboratory Education. Journal of Engineering Education, 98.

Bakas, C., & Mikropoulos, T. (2003). Design of virtual environments for the comprehension of planetary phenomena based on students' ideas. International Journal of Science Education, 25, 949 - 967.

Bardar, E. M. (2007). First results from the light and spectroscopy concept. *Astronomy Education Review*, 6(2), 75-84.

Barnett, M. L., & Mathisen, A. (1997). Tyranny of the p-value: the conflict between statistical significance and common sense. *Journal of dental research*, 76(1), 534–536.

Bloom, B. S. (1956). *Taxonomy of educational objectives*. Longman green and co LTD, England.

Boopathiraj, C., & Chellamani, D.K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Sciences & Interdisciplinary Research*, 2, 189-193.

Coe, R. (2012). It's the Effect Size, Stupid What effect size is and why it is important. Paper Presented at the British Educational Research Association Annual Conference, Exeter, England, UK. 2002

Cohen, A. (2022). Teaching the astronomical visualization used for the explanation of the ancient Ein-Gedi archaeological zodiac and its related inscription. *Journal of Astronomy & Earth Sciences Education*, 9(2), 25–38.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associate, New York.

Galano, S., Colantonio, A., Leccia, S., Marzoli, I., Puddu, E., & Testa, I. (2018). Developing the use of visual representations to explain basic astronomy phenomena. Physical Review Physics Education Research, 14(1), 010145

Hake, R. R. (1998). Interactive-engagement vs traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64-74.

Karaksha, A., Grant, G.D., Nirthanan, S., Davey, A.K., & Anoopkumar-Dukie, S. (2014). A Comparative Study to Evaluate the Educational Impact of E-Learning Tools on Griffith University Pharmacy Students’ Level of Understanding Using Bloom’s and SOLO Taxonomies. *Education Research International*, 2014 (2014): 934854-1-934854-11.

Laphirattanakul, A., Wannawichian, S. & Wattanakasiwich, P. (2024). PARST educational tool for planet positioning. *The Physics Teacher*, 62(4), 272-277.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews of the Cambridge Philosophical Society*, 82(4), 591–605.

Neyman, J. & Pearson, E. S. (1933). *On the problem of the most efficient tests of statistical hypotheses*. Springer Series in Statistics. Springer, New York, NY.

Ole, F. C. B. (2020). Effect of a developed physics laboratory manual on the conceptual understanding of industrial technology student. *European Journal of Education Studies*, 7(6), 113-122.

Peralta, J., Prieto, J. A., Orozco-Sáenz, P., González, J., Trujillo, G., Torres, L., Sánchez, A. & Arnedo, M. (2023). Secondary school students observe Venus with NASA Infrared Telescope Facility (IRTF), *Research Notes of the AAS*, 7(3), 53-57.

Plummer, J.D. (2014). Spatial thinking as the dimension of progress in an astronomy learning progression. *Studies in Science Education*. 50(1), 1-45.

Rakap, S. (2015). Effect sizes as result interpretation aids in single-subject experimental research: description and application of four non overlap method. *British Journal of Special Education*, 42(1), 11-33.

Ruangsuwan, C. & Arayathanitkul, K. (2009). A low-cost celestial globe for hands-on astronomy. *Physics Educucation*, 44(5), 503-508.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. Review of educational research, 77(4), 454-499

Srisawasdi, N. (2018). Transforming chemistry class with technology-enhanced active inquiry learning for the digital native generation. In C. Cox & W. Schatzberg (Eds.) International Perspectives on Chemistry

Education Research and Practice (pp. 221–233). ACS Symposium Series 1142, American Chemical Society: Washington, DC.

Stover, S., & Saunders, G. (2000). Astronomical misconceptions and the effectiveness of science museums in promoting conceptual change. Journal of Elementary Science Education, 12, 41-51.

Sullivan, G. M. & Feinn, R. (2012). Using Effect Size-or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279-282.

Woods, M. & Rosenberg, M. E. (2016). Educational tools: Thinking outside the box. *Clinical Journal of the American Society of Nephrology*, 11(3), 518-526.

Xu, M., Fralick, D., Zheng, J. Z., Wang, B., Tu, X. M. & Feng, C. (2017) The Differences and Similarities Between Two-Sample T-Test and Paired T-Test. *Shanghai Arch Psychiatry*, 29(3), 184-188.

doi: 10.14456/edkkuj.2020.x

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น
https://www.tci-thaijo.org/index.php/edkkuj
ดำเนินการวารสารโดย คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

Appendix

The pre-test and post-test

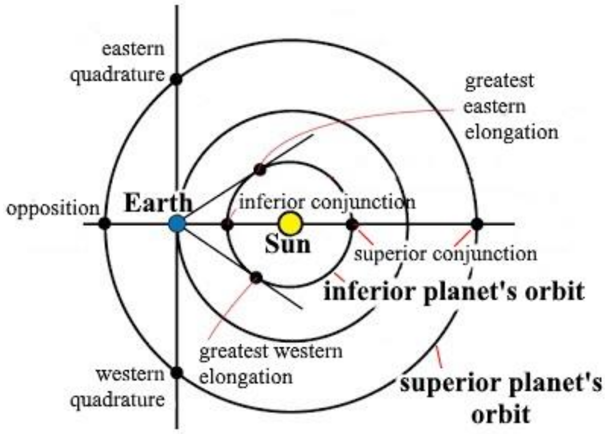
Student name......................................

Topic Effectiveness evaluation for the teaching tool: The Planets Approximating of Rising and Setting Time (*PARST*)

Description This is a multiple-choice test with four choices for each of the 15 questions. The score is 1 point for each question.

Objective To evaluate the student's learning progress about the planet’s motion.

Instruction Choose only one best answer in the answer sheet



Above figure: Important positions of inferior and superior planets

1. Which planet is classified as an inferior planet?

   a. Venus b. Mars c. Pluto d. Eris

2. When is the earliest rising time for Venus?

   a. 1:00 am b. 2:45 am c. 5:30 am d. 7:20 am

3. When is the latest setting time for Venus?

   a. 2:45 pm b. 7:30 pm c. 9:15 pm d. 10:00 pm

*Corresponding author. Tel.:
Email address: arthit.l@cmu.ac.th

4. What is the position of Venus relative to the Earth and the Sun when Venus' latest rising time occurs?
 a. inferior conjunction b. superior conjunction
 c. greatest eastern elongation d. greatest western elongation

5. How many times did Venus rise at the latest time between 2012 and 2019? (Venus' synodic period is 584 days).
 a. 4 b. 5 c. 6 d. 7

6. When is the earliest rising time for Mars?
 a. 12:00 am b. 6:00 am c. 12:00 pm d. 11:00 pm

7. On the day Mars rises at 10:00 pm, what is the setting time of Mars at the same location as the observer?
 a. 4:00 am b. 10:00 am c. 4:00 pm d. 10:00 pm

8. What is the position of Mars relative to the Earth and the Sun when Mars' setting time is 6:00 pm?
 a. superior conjunction b. opposition
 c. eastern quadrature d. western quadrature

9. How many times did Mars rise at 11:00 pm between 2012 and 2019? (Mars' synodic period is 780 days)
 a. 4 b. 5 c. 6 d. 7

10. In another solar system, there are planets, including A, B, C, and D, with properties as follows.

Planets	A	B	C	D
Distances from the host star (AU)	0.48	1.62	4.45	12.39

 An observer is on another planet, R, at a distance of 0.73 AU from the host star. Which planet has the rising time and the setting time varying within a limited duration (shorter than 24 hours), as seen from planet R?
 a. planet A b. planet B c. planet C d. planet D

11. What is the position of the superior planets relative to the Earth and the Sun when they have retrograde motion?
 a. superior conjunction b. opposition
 c. eastern quadrature d. western quadrature

12. Jupiter has a synodic period of about 398 days. If its start position relative to the Earth and Sun is "Opposition", which position will Jupiter be approximately 200 days after the start position?

วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 47(4), 90-111
doi: 10.14456/edkkuj.2020.x

a. superior conjunction b. opposition
c. eastern quadrature d. western quadrature

13. Which planet has the longest synodic period?
a. Mercury b. Venus c. Mars d. Neptune

14. What is the position of Venus relative to the Earth and the Sun when it changes from being a morning star (rises before the sunrise) to an evening star (sets after the sunset)?
a. inferior conjunction b. superior conjunction
c. greatest eastern elongation d. greatest western elongation

15. Which of the following is the most accurate phenomenon observed by an observer on Earth?
a. Venus can be apparent in the sky at 12:00 am.
b. Mars can rise at the same time as Mercury at 9:00 pm.
c. Jupiter and Saturn can be at a very close angular distance ($< 1^{\circ}$).
d. Saturn at the opposition position can set at 12:00 am.