Assessing Skills of Identifying Variables and Formulating Hypotheses Using Scenario-Based Multiple-Choice Questions

Burak Kağan Temiz [iD] [1,*]

[1]Department of Science Education, Faculty of Education, Nigde Omer Halisdemir University, Nigde, Turkey

Abstract: The aim of this study was to investigate the effectiveness of scenario-based multiple-choice questions to assess students' science process skills. To achieve this objective, a test with 32 scenario-based multiple-choice questions evaluating students' skills in formulating hypotheses and identifying variables was prepared and administered to 370 high school freshmen. The questions were involved experiments with two different parts. Both parts of the experiments had the same dependent variable, and in each part the effect of a different manipulated variable on the dependent variable was examined. Therefore, the variables changed roles within the same experiment. In evaluating the test, questions about the first part of the experiments were coded A, and questions about the second part of the experiments were coded B. When the students' scores from the code A and code B items were compared, statistically significant differences were found. Analysis of the data revealed that some students were affected by the different roles played by the variables in the different parts of the experiment.

1. INTRODUCTION

As a result of their natural curiosity, human beings seek to understand the environment in which they live and to acquire new knowledge. The natural sciences that emerged as a result of these efforts embody two main components: the scientific knowledge itself, and the ways in which knowledge can be acquired. The skills that are used for acquiring knowledge in science are called science process skills (SPS). SPS are thus the activities that scientists engage in when they investigate a problem or phenomenon. SPS are mental and physical skills used in collecting, organizing and analyzing data through various methods. These skills are involved in identifying researchable questions, designing investigations, obtaining evidence, interpreting evidence in terms of the question addressed in the research, and communicating the findings of the investigative process. In addition, SPS are needed not only by scientists, but by all citizens in order for them to become scientifically literate people able to function in a society in which science plays a major role and has an impact on everyone's personal, social and global life. In fact, understanding scientific processes is a basic aspect of thinking, used both in science and in other fields to solve problems. For this reason, SPS are also life-long learning skills. In

---

CONTACT: Burak Kağan TEMİZ ✉ bktemiz@ohu.edu.tr ▢ Department of Science Education, Faculty of Education, Nigde Omer Halisdemir University, Nigde, Turkey

elementary and middle school science education, the development of SPS is a major goal of science education.

In the literature different researchers have defined SPS differently (Gabel, 1992; Martin, 2002; Padilla, 1990). In general, SPS are the cognitive skills that we use to process information, to think through problems, and to formulate conclusions. These are the skills that scientists use when they work. By teaching students these important skills, we can enable them to understand their world and learn about it. These skills are fundamental to thinking and to research in science. In the Science-A Process Approach (SAPA), these skills are defined as a set of broadly transferable abilities, appropriate to many science disciplines and reflecting the behavior of scientists. The SAPA has grouped process skills into two categories, basic and integrated. The basic science process skills (BSPS) provide the intellectual ground work of scientific enquiry, such as the ability to order and describe natural objects and events. The BSPS are fundamental to the integrated science process skills (ISPS). The BSPS include observing, classifying, measuring and predicting, while the ISPS are essential skills for solving problems or conducting science experiments. The ISPS include identifying and defining variables, collecting and transferring data, constructing tables of data and graphs, describing relationships between variables, interpreting data, manipulating materials, formulating hypotheses, designing investigations, drawing conclusions and generalizing (Abruscato, 2000; Beaumont-Walters & Soyibo, 2001; Burns, Okey, & Wise, 1985; Carin, 1993; Carin & Bass, 2001; Esler & Esler, 2001; Harlen, 1993, 1999; Hughes & Wade, 1993; Ostlund, 1992; Rezba et al., 1995).

1.1. Formulating Hypotheses and Identifying Variables

When we try to understand things in a scientific way, the complex subject at hand is divided into researchable and understandable elements. These elements of an event or a system are called variables. Variables are the factors, conditions or relations that change or that can be changed in an event or a system. In scientific research, there are three kinds of variables. These are manipulated, responding, and controlled variables (Bailer et al.,1995). A manipulated variable (independent variable) is a factor or a condition which is changed by the researcher on purpose in an experiment. A dependent variable (response variable) is a kind of variable that can be affected by the changes in the factor or the condition. Variables that remain constant through the experiment so as not to interfere with the results are called controlled variables. There can be more than one controlled variable in an experiment.

Formulating a hypothesis is the skill of developing a problem question which can be tested by an experiment about the effect of a manipulated variable on a dependent variable. To formulate a hypothesis means building testable statements based on ideas and experiences which are thought to be true. Hypothesizing means stating a testable solution to a problem. A hypothesis is usually proposed before any experiment or research and is a prediction about the relationships between variables. Being testable is the most important characteristic of a hypothesis.

According to Gabel (1993), a scientist must control all the variables that will affect the outcome of an experiment in order to be able to practice science, that is, to be able to test hypotheses or confirm assumptions. Before controlling variables, the scientist must identify the responding and manipulated variables. Later, a factor is changed on purpose and, as a result, a change occurs in the other variable. The strategy followed in manipulating and controlling variables is to change a variable (the manipulated variable) and examine changes occurring in the other variable (the response variable). At the same time, many other variables (controlled variables) must be defined and kept constant. This is because these variables have the potential to affect the results. If more than one variable is changed at the same time, the result of the experiment is not reliable (Carin & Bass, 2001). Bailer et al. (1995) associated the process of hypothesizing with the process of identifying and controlling variables. On this basis, a hypothesis is a kind of statement that predicts the effect of one variable on another.

1.2. Assessing Science Process Skills

With increased understanding of the importance and value of SPS in science education, the interest of researchers in the subject has also increased. Numerous models have been constructed for the teaching and acquisition of SPS. Additionally, several instruments have been developed to assess achievement in SPS for formative, summative and monitoring purposes. An examination of the literature reveals that numerous tests with various question formats have been developed in order to measure all or some SPS at different levels. Table 1 shows some of these instruments.

As seen in Table 1, most of the SPS assessment instruments were designed using a multiple-choice format, which is relatively easier and less time-consuming to administer. However, several researchers have emphasized the need to develop such instruments using alternative formats. Techniques suggested include systematic observations of students' laboratory work (Lunetta et al.1981), microcomputer simulations (Berger, 1982), technological applications (Kumar, 1996), and open-ended questions (Gabel, 1993). Moreover, Beaumont-Walters and Soyibo (2001) drew attention to the fact that although the commonly used multiple-choice test format has been criticized, only a few researchers have attempted to develop tests for SPS that also involve hands-on tasks. And although considerable attention has been given to assessing the performance of SPS, the development of standardized instruments for participants in a large sample has been difficult. In light of these difficulties, the multiple-choice format may be preferable for large samples (Aydınlı et al., 2011).

Table 1. SPS Assessment Instruments Documented in the Research Literature.

Authors	Title	Year	Test Format
R. S. Tannenbaum	Test of Science Processes	1968	Multiple choice
J. W. Riley	The Test of Science Inquiry Skills	1972	Multiple choice
R. R. Ludeman	The Science Process Test	1974	Multiple choice
L. L. Molitor and K. D. George	The Science Process Test	1975	Multiple choice
F. G. Dillashaw and J. R. Okey	Test of Integrated Process Skills	1980	Multiple choice
K. G. Tobin and W. Capie	Test of Integrated Process Skills	1982	Multiple choice
J. C. Burns, J. R. Okey and K. C. Wise	Test of Integrated Process Skills II	1985	Multiple choice
K. A. Smith and P. W. Welliver	Science Process Assessments for Elementary School Students	1986	Multiple choice
K. A. Smith and P. W. Welliver	Science Process Assessments for Middle School Students	1994	Multiple choice
G. Solano-Flores	The "Bubbles" Task	2000	Hands-on Activity
Y. Beaumont-Walters and K. Soyibo	Test of Integrated Science Process Skills	2001	Multiple Format
Author, M. F. Taşar and M. Tan	Multiple Format Test of Science Process Skills	2006	Multiple Format
Author and M. Tan	Science Process Skills Test	2007	Multiple Format
Shahali E. H. M. and Halim L	Test of Integrated Science Process	2010	Multiple choice
Feyzioglu, B., Demirdag, B., Akyildiz, M., & Altun, E.	Science Process Skills Test	2012	Multiple choice
Aydoğdu B., Tatar N., Yıldız E. and Buldur S.	Science Process Skills Scale	2012	Multiple choice
Aydoğdu, B. and Karakuş, F.	The Scale for Basic Process Skills of Pre-School Students	2017	Multiple choice
Tosun, C	Scientific Process Skills Test	2019	Multiple choice

1.3. Scenario-Based Learning and Assessment

Scenarios are narratives in the form of stories or speeches that emerge from real events or realistic situations. In scenario-based learning the real world is brought into the classroom. Thus, students are given opportunities to think about a problem, to use what they have learned in real or realistic situations, to become aware of their lack of knowledge and to do the necessary work to correct this. Furthermore, scenarios trigger students' higher-order thinking processes such as analysis, synthesis, evaluation and decision-making (Açıkgöz, 2003).

The increasing importance of scenario-based learning in recent years has brought new approaches to the teaching process, and scenarios are now included in many Science and Technology textbooks. With scenario-based learning, students are given the opportunity to discover different problems and situations through scenarios drawn from real life, to use their existing knowledge in these new situations, to offer creative ideas and to implement what they have learned (Erduran Avcı & Bayrak, 2013). Scenarios unique to a specific field can be used in activities involving measurement and evaluation in addition to normal learning activities. According to Thalheimer (2013), scenario-based questions present learners with one or more short paragraphs that describe a situation and include a question that asks learners to make their own decisions. There are many varieties to this basic design. We can use multiple scenes and multiple questions to form a scenario. We can add visual or auditory details to augment or even supplant the text-based scenario. We can also use different types of questions, including multiple-choice, open-ended, and yes-no questions, etc. Scenario-based multiple-choice test items have been used frequently in SPS assessments.

When the multiple-choice tests developed to assess skills in identifying variables and formulating hypotheses are examined, scenario-based questions are frequently encountered. Some of the tests used most frequently in science education research are the Test of Integrated Process Skills (TIPS) (Dillashaw & Okey, 1980; Tobin & Capie, 1982), the Test of Integrated Process Skills II (TIPSII) (Burns et al., 1985), and the Science Process Assessments for Middle School Students (Smith & Welliver, 1995). An examination of items in the multiple-choice format SPS measurement tests used to assess skills in formulating hypotheses and identifying variables shows that question developers generally provide one section from a single-stage experiment and ask the student to identify the hypothesis and the variables in the test. The example given in Figure 1, which is a single-stage experiment, is from an SPS measurement test widely used in Turkey.

Most experiments conducted in science consist of more than one stage. At each stage the effects of a different manipulated variable on a dependent variable is examined. Therefore, the manipulated variable at one stage of the experiment can be a controlled variable at another stage. The idea that the same variable can play different roles in different parts of the experiment should be taken into consideration while developing questions to assess SPS. In the scenarios in the SPS measurement tests widely used in the literature, the idea that an experiment may be made up of more than one stage is not taken into consideration (see Figure 2). Does students' performance change if they are asked questions (see Figure 3) about situations where the effect of a different manipulated variable on a dependent variable at each different stage of the experiment is examined? The aim of this study was to find the answer to this question.

Answer questions 29, 30, 31 and 32 by reading the paragraph given below.

The effects on tomato production of leaves mixed in with the soil are being investigated. In the research an identical quantity and type of soil was placed in four large pots. However, 15 kg of mulched leaves were added to the first pot, 10 kg to the second and 5 kg to the third. No mulched leaves were added to the fourth pot. Tomatoes were then planted in these pots. All the pots were placed in sunlight and watered identically. Tomatoes obtained from each pot were weighed and recorded.

29. What is the hypothesis that was tested in this research?
   a. Plants produce tomatoes in proportion to the sunlight they receive.
   b. The larger the pots are the more mixed leaves are needed.
   c. The more water in the pots, the faster the leaves rot.
   d. The more mulched leaves are in the soil the more tomatoes are produced.

30. What is/are the controlled variable(s) in this research?
   a. The amount of tomatoes obtained from each pot.
   b. The amount of leaves mixed into the pots.
   c. The amount of soil in the pots.
   d. The number of pots with mulched leaves added.

31. What is the dependent variable in this research?
   a. The amount of tomatoes obtained from each pot.
   b. The amount of leaves mixed in the pots.
   c. The amount of soil in the pots.
   d. The number of pots with mulched leaves added.

32. What is the manipulated variable in research?
   a. The amount of tomatoes obtained from each pot.
   b. The amount of leaves mixed in the pots.
   c. The amount of soil in the pots.
   d. The number of pots with mulched leaves added.

Figure 1. Sample item in a scenario from a single-stage experiment

Figure 2. Traditional scenario-based SPS questions about single stage experiments
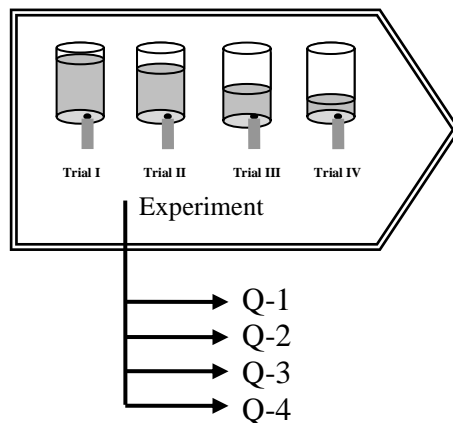
Questions:
Q-1 and 5. What is the manipulated variable in this investigation?
Q-2 and 6. What is the dependent variable in this investigation?
Q-3 and 7. What are the controlled variables in this investigation?
Q-4 and 8. What hypothesis is be ingtested in this investigation?
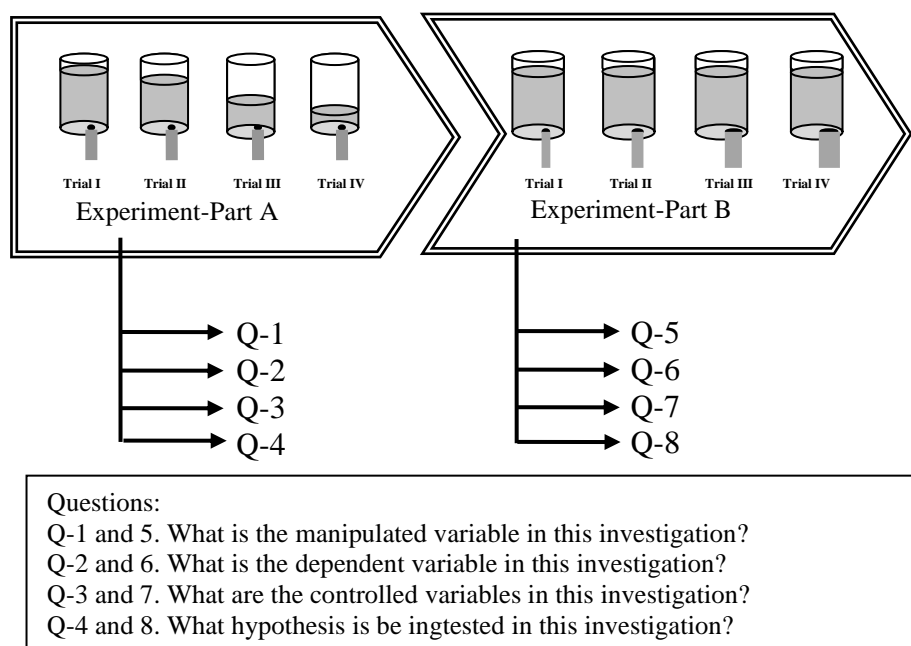
Figure 3. Scenario-based SPS questions about two stage experiments

This study was conducted with the aim of examining whether the students' ability to use SPS (formulating hypotheses and identifying variables) would change with questions about two-part experiments where a different hypothesis was tested in each part and where variables played different roles in different parts. As scenario-based questions are frequently used in the literature to assess the skills of identifying variables and formulating hypotheses, these were the skills that this study examined.

2. METHOD

This study uses a type of descriptive research model with a survey method. Descriptive models that are used commonly aims to describe the situation and find out the factors that are the subjects of the study. The survey type methods contain collecting, classifying, describing, analyzing and inferring results from the data which aim to determine any presence and/or degree of together-change amongst two or more variables (Büyüköztürk et al., 2009; Karasar, 2011).

2.1. Study Group

370 (191 females, 179 males) high school freshmen selected by stratified sampling from five different high schools participated in this study. The majority of students were 15 years old. The participants had just completed their elementary education and had not yet chosen any future field of study.

### 2.2. Data Collection Process and Assessment Tool

In order to measure students' skills in identifying variables and formulating hypotheses, a test with scenario-based multiple-choice items was used. The 40 items in this test were compiled from The Science Process Skills Test (SPST) question pool developed by the author (Temiz, 2007). The SPST was developed for the purpose of assessing skills in identifying variables, formulating hypotheses, controlling variables, recording data (constructing the data table), constructing graphs and interpreting graphs. The SPST is composed of three multiple-choice and three open-ended modules, with a total of six modules. Module 1 assesses the skills of defining variables and formulating hypotheses and has 60 multiple-choice questions; Module 2

assesses the skill of controlling variables (designing experiments) and has five open-ended and 25 multiple-choice questions; Module 3 assesses the skill of constructing a data table and has eight open-ended questions; Module 4 assesses the skill of drawing graphs and has eight open-ended questions; Module 5 assesses the skills of interpreting graphs and has 60 multiple choice questions; and Module 6 assesses the skills of defining the variables and formulating hypotheses and has 10 open-ended questions. The SPST was developed after pilot tests conducted on 1584 Grade 9 students. To collect evidence for the test's validity, content-related, criterion-related and construct-related validity analyses were conducted, and internal-consistency, test-retest and inter-rater agreement analyses were carried out to determine the SPST's reliability. Detailed statistics about test development process can be found in the work "Evaluating students' science process skills in physics teaching" (Temiz, 2007).

The data in this work was collected using 40 multiple-choice questions from among the questions in Module 1 of the SPST. These are related to five experimental scenarios which are individually made of two parts. Each experiment is presented with a paragraph of text and pictures supporting that text, followed by multiple-choice questions come based on what is given. This test was named the Formulating Hypotheses and Identifying Variables Skills (FHIVS) Test.

To examine the test reliability and item indices, the FHIVS test was administered to high school students. A total of 87 students were involved in this pilot test. Students' answers were processed with the Excel software package, and test reliability was investigated by internal consistency analyses. The total scores of the five experimental scenarios test ranged from 4 to 40 (mean=23.9, S.D.=10.6) for the students overall. The total test reliability (KR 20 coefficient) was 0.944. Item difficulty indices ranged from 0.25 to 0.81 with an average of 0.61. Item discrimination indices obtained by using the upper 27% and lower 27% of the sample group showed that 32 of 40 items were above 0.50 with an average of 0.63. Each of these indices fell well within the acceptable range for a reliable test. After the item analysis conducted with the data obtained from the pilot application, one scenario (and eight questions related to this scenario) was taken out of the test.

The revised version of the FHIVS test includes four experimental scenarios; each of which consists of two parts. Each experimental scenario features a single paragraph describing an experiment accompanied by supporting diagrams, and four scenario-based multiple-choice questions about the experiment described. These experimental scenarios are given in Figure 4. The first question related to the experiment was about the manipulated variable, the second was about the dependent variable, the third involved the control variables, and the fourth was about the hypothesis tested in the experiment. The same dependent variable was involved in the first and second parts of all the experimental scenarios given but in each part the effect of a different manipulated variable on the dependent variable was involved while all other variables were kept constant. Consequently, different hypotheses were tested in the first and second parts of the experiments. Additionally, the distractors in the answers to the questions in the first and second parts of the experiments were also identical. One example scenario and eight questions related to this scenario are given in the appendix. In the analysis of students' answers, responses to questions in the first and second parts of experiments were coded as A and B respectively. Items coded A and B were then compared to determine differences in the students' skills in identifying variables and formulating hypotheses in the two parts of the experiment. The contexts of experimental scenarios given in Figure 4 can be summarized as follows:
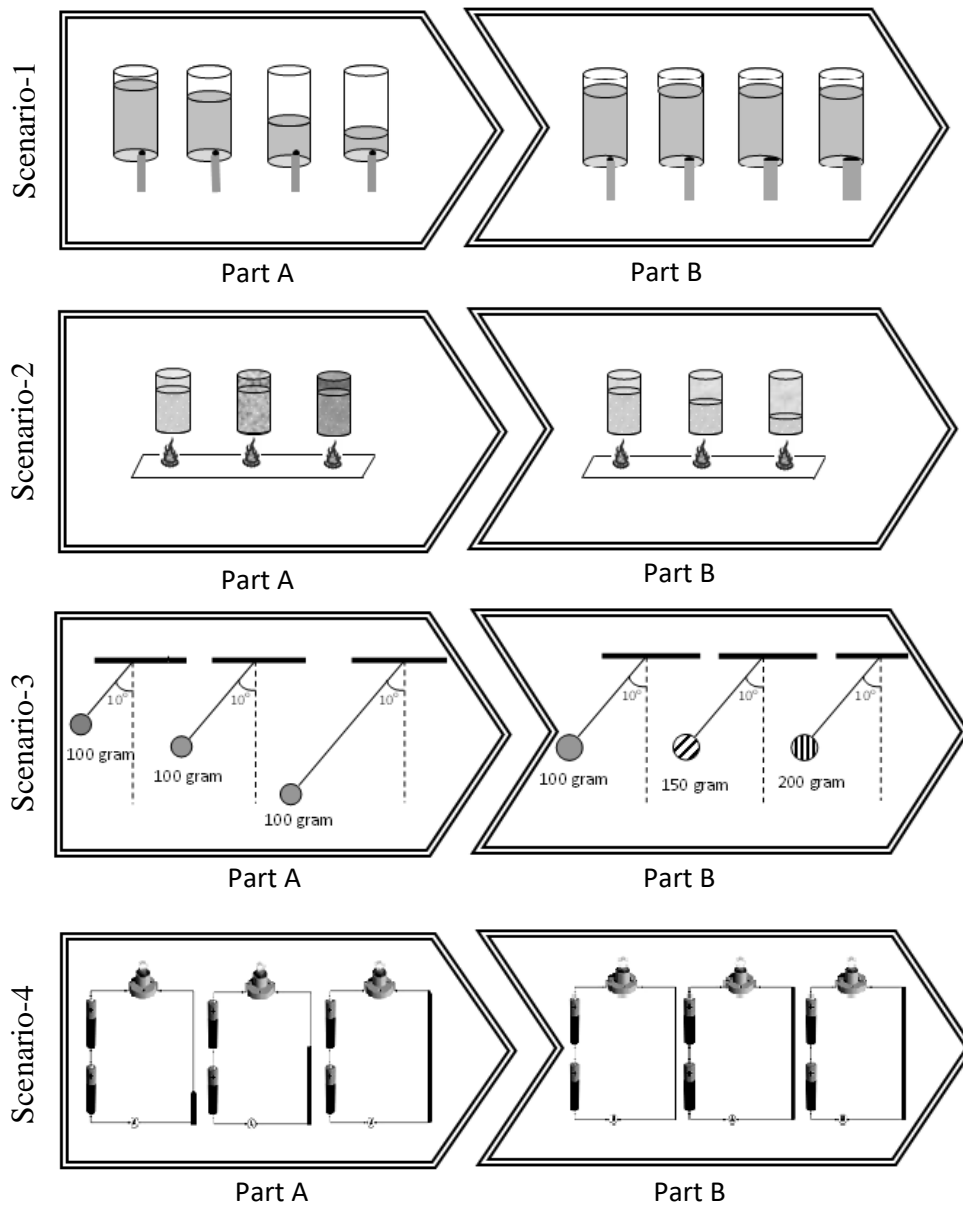
Figure 4. Experimental scenarios used in the FHIVS test

In Scenario 1, two stages of an experiment about the discharge of water from a glass with a hole under it were described. In the first stage of the experiment (Part A), while variables like the size of the hole, the type of liquid and the shape of the container were fixed, the amount of liquid amount was changed and the discharge time was measured. In the second stage of the experiment (part B), while variables like the type of liquid, the amount of liquid and the stage of the container remained the same, the size of the hole changed and the discharge time was measured.

In Scenario 2, two stages of an experiment about boiling water in metal containers were described. In the first stage of the experiment (Part A), while variables like the amount water amount, the amount of heat given to the container and the size of the container were fixed, the metal which the container was made of changed and the boiling times were measured. In the second stage of the experiment (Part B), while variables like the amount of heat given to the container, the size of the container and the metal which it was made of remained the same, the amount of water changed and the boiling times were recorded.

In Scenario 3, two stages of an experiment about a simple pendulum were described. In the first stage of the experiment (Part A), while variables like angle of amplitude, mass of the oscillated object and volume of the object remained the same, the length of the rope and length of oscillation time were measured. In the second stage of the experiment (Part B), while variables like angle of amplitude, length of the rope and volume of the oscillated object remained the same, the mass was changed and the oscillation times were measured.

In Scenario 4, two stages of an experiment about a simple electric circuit were described. In the first stage of the experiment (Part A), while variables like the number of batteries in the circuit, the type of the material the conductive wire is made from and the width of the wire remained the same, the length of the wire changed and the intensity of the current going through the circuit was measured. In the second stage of the experiment (Part B), while variables like the number of batteries, type of the material the conductive wire was made from and length of the wire remained the same, the width of the wire changed and again the intensity of the current going through the circuit was measured.

As shown above, two different stages of an experiment were described in these four scenarios. The number of stages can be increased. In fact, at each stage the effects of a different independent variable on the same controlled variable are investigated and a different hypothesis is tested.

3. RESULTS/FINDINGS

3.1. Consistency of SPS

To examine the consistency of the students' SPS performance, the responses of each student to the questions about the first and second stages, coded as A and B, were compared for accuracy. For this purpose, as shown in Table 2, students' answers were categorized into four groups with different levels of performance consistency.

To describe each group given in Table 2, students' answers to code A and B questions were compared separately for each skill. This comparison was done for all four groups, and the number of students in the groups and percentages in each group were found. The average number of students grouped in terms of skills is given in Table 3. It was found that students falling into Groups 1 and 2 exhibited consistent performances whereas students in Groups 3 and 4 exhibited inconsistent performances.

According to the results presented in Table 3, only about half the students were able to answer both code A and code B items correctly. In questions assessing the skill of identifying controlled variables, this number even dropped to 35%. The percentages of students who answered both code A and code B items incorrectly ranged between 15% and 35%. The percentages of students exhibiting an inconsistent performance by incorrectly answering any one of the code A or B items ranged between 20% and 25%. The skill with the highest level of inconsistent performance was formulating hypotheses (25%). The percentage of students exhibiting consistent performance (Group 1 + Group 2) was in the range 65% - 75%. The skill with the highest level of consistent performance was identifying the dependent variable (75%). All these descriptive statistics demonstrate that some (nearly a fifth) of the students exhibited different performances in the FHIVS test with regard to the two different parts of an experiment.

Table 2. Identification of groups

Groups	A	B	Group description	Performance consistency
Group 1	A Correct	B Correct	Students correctly answered both questions.	Consistent
Group 2	A Wrong	B Wrong	Students incorrectly answered questions.	Consistent
Group 3	A Correct	B Wrong	Students answered code A questions correctly but code B questions incorrectly.	Inconsistent
Group 4	A Wrong	B Correct	Students answered code A questions incorrectly but code B questions correctly.	Inconsistent
Other	?	?	Students left at least one question unanswered in the same experiment.	Undetermined

Table 3. Average Numbers of Students in Groups According to Skills (N = 370)

Skills / Groups	Identifying Manipulated Variable		Identifying Responding Variable		Identifying Controlled Variables		Formulating Hypotheses	
	N	%	N	%	N	%	N	%
Group 1	191	51.62	201	54.19	130	35	186	50.14
Group 2	73	19.73	76	20.41	130	35.2	56	15
Group 3	36	9.8	35	9.32	40	10.88	56	15.2
Group 4	41	11.15	31	8.38	31	8.45	36	9.73
Consistent Performance	264	71.35	276	74.6	260	70.2	241	65.14
Inconsistent Performance	78	20.95	66	17.7	72	19.33	92	24.93
Other	29	7.7	29	7.7	39	10.47	37	9.93
Total	370	100	370	100	370	100	370	100

3.2. Comparison of SPS achievement in different parts of the same experimental scenario

Would the test scores of students be affected when the variables in two different parts of an experiment testing different hypotheses changed roles? To address this question, the test scores for both code A and code B items were compared. For this purpose, a paired samples t-test was conducted for each skill. The results of the paired samples t-test are given in Table 4.

Table 4. Paired Samples t-test Results

Skills	$\bar{X}_A$	$\bar{X}_B$	S_A	S_B	t	p
Identifying Manipulated Variable	2.49	2.58	1.47	1.38	-1,99	0.046
Identifying Dependent variable	2.60	2.51	1.46	1.42	1,99	0.046
Identifying Controlled Variables	1.91	1.76	1.63	1.50	2.96	0.003
Formulating Hypotheses	2.72	2.41	1.31	1.24	6.65	0.000

According to the data in Table 4, there were statistically significant differences in the total scores for the code A and code B items of the test for all the specific skills. These data demonstrate that some students were affected by the variables having different roles in different parts of the experiment. Most differences between the code A and code B questions were found with regard to the skill of formulating hypotheses. When the eta-squared values ($\eta2$ of the manipulated variable=0.01, $\eta2$ of the dependent variable=0.01, $\eta2$ of the controlled variables=0.02, $\eta2$ of formulating hypotheses=0.11) were computed separately for the skills taken into consideration, it could be stated that the two-stage nature of the experiments had a small effect on students' performance scores in terms of identifying variables and a moderate effect on their performance scores for formulating hypotheses.

3.3. Stability of answers in different parts of the same experimental scenario

In a new situation where a different hypothesis is tested, did the students understand the changing role of the variable? To address this question, same responses from each student in both parts of the experiments were compared with one another. The number of students choosing the same response for both code A and code B items for all the experiments and skills were identified. The data obtained are presented in Table 5. The data in Table 5 show that nearly 64% of the students marked the same response in both parts of the experiment while identifying the dependent variable. This can be interpreted as positive since the same dependent variable had been worked on in both parts of all experiments. However, on the other hand, in identifying the manipulated variable 18% of the students marked the same response for the two parts; in identifying the controlled variable 28% of the students marked the same response for the two parts; in formulating hypotheses 14% of the students marked the same response for the two parts. These results are interesting since they demonstrate that some students did not take into consideration the different parts of the experiment while identifying the variables and testing the hypotheses.

Table 5. Percentage of the students who gave the same response for both parts of the experimental scenarios

Skills	Identifying Manipulated Variable		Identifying Dependent variable		Identifying Controlled Variables		Formulating Hypotheses	
Scenarios	N	%	N	%	N	%	N	%
Scenario1	58	15.68	233	62.97	53	14.32	33	8.92
Scenario2	59	15.95	241	65.14	126	34.05	37	10.00
Scenario3	60	16.22	226	61.08	103	27.84	55	14.86
Scenario4	89	24.05	250	67.57	132	35.68	85	22.97
Overall	66.50	17.97	237.50	64.19	103.50	27.97	52.50	14.19

The data collected in the research show that the scores of nearly 22% of the students for formulating hypotheses and identifying variables changed depending on the part of the experiment. In other words, some students' performance changed depending on different parts of the same test. Furthermore, it has been established that a significant portion of the students ignored different parts of the experiment while identifying the variables or hypotheses tested in the experiment. In the second part of the experiment where the hypothesis was tested, these students did not mind putting the same answer they had done in the first part. For example, in the questions given in the Appendix, the effect of the "height of liquid in a glass" variable on the "emptying time" variable was examined in the first part of the experiment. In the second part, the effect of "hole size" variable on the "emptying time" variable was examined. Some students mistakenly selected the "height of liquid in glass" variable as the manipulated variable

in the first and second parts of the experiment. If these questions assessing the skill of identifying variables had only been developed for single stage experiments, this confusion would not have been revealed.

4. DISCUSSION and CONCLUSION

SPS are intellectual and physical skills we use to acquire information, think about problems and formulate conclusions. These skills are an inseparable component of inquiry-based science education. Learning with understanding in science involves using SPS. Thus, the development of SPS is a major goal of science education. Several science education curricula have been developed with the intention of teaching the acquisition of SPS, and measuring and assessing these skills is an important aspect of science education. Over recent years many tools have been developed in various forms with the objective of measuring these important skills (Harlen, 1999; Aydınlı et al., 2011).

The measurement of SPS comes with various difficulties. These difficulties may be discussed from two aspects. The first concerns how SPS should be measured; in other words, it is about the types of question to be used in SPS measurement. Some researchers think that the best way to measure the SPS of students is by using laboratory reports, oral presentations and observations (Lavinghousez, 1973; Gabel, 1992; Ostlund, 1992; Haury & Rillero, 1994; Kazeni, 2005). A more appropriate way of measuring SPS is the use of hands-on activities, but due to their ease of application, simplicity of evaluation, and because they do not require expensive resources, paper and pencil tests are still often currently preferred. According to Rezba et al. (1995), a transition from multiple-choice measurement methods to multi-formatted measurement methods is taking place. However, multiple-choice tests are still frequently preferred because they can be easily applied to large groups of students. According to Burns et al. (1985), assessing students' skills through observation in laboratory situations can be difficult and time-consuming. While an instructor may obtain an intuitive feel for a student's competence in process skills via observation, high-quality tests are needed to achieve accurate measures of students' performance.

The second aspects concern the difficulties in selecting content and contexts when measuring and assessing SPS. Harlen (1999) asserts that SPS have to be used in concert with specific content. Therein lies the difficulty in assessing these skills. Students' performance in any task involving these skills will be influenced by the nature of the content as well as by the students' ability. In the literature is examined various studies have demonstrated that the content of the tasks utilized in SPS measurement tools have an influence on students' performance. Zimmerman and Glaser (2001) conducted a study on this. They investigated whether sixth-grade students were affected by variations in the scenario given while designing an experiment about plants. It was found that student performances were affected when the scenarios were chosen from among topics in the curriculum. These studies also demonstrate that the performance of SPS is affected by whether the content of tests relates to everyday life or to scientific issues. While a question referring school or a laboratory context can point toward a specific idea, a subject from everyday life might not produce a similar association. According to these studies, students demonstrated better SPS when the content was drawn from everyday life, while their application skills were better in scientific contexts (Song & Black, 1991, 1992; Temiz, 2010). In this study, these effects were also taken into consideration when the scenarios were created. Some of the scenarios were created using content from everyday life (scenarios 1 and 2) and some were formulated using scientific contexts (scenarios 3 and 4). The findings obtained in this study add a new dimension to the discussion on content and context selection in SPS measurement. This dimension is the development of multi-stage scenarios.

In this study, two different stages of an experiment used in experimental scenarios were

explained. At each stage, the effects of a different independent variable on the same controlled variable were investigated. In other words, at each stage a different hypothesis was tested. The method of testing a variable's effect on another effect is called "fair testing". According to Hughes and Wade (1993) children have difficulty in controlling variables and see no problem in simultaneously exchanging two or more variables even up to the ages of 13-15. For this reason, the development of the concept of fair testing should commence early in schools. According to Carin and Bass (2001), in controlled variable studies conducted among primary and middle schools, students better understand the experiment when they learn about the fair testing technique. In addition to this, teaching the students that "variables can exchange roles in various parts of an experiment" is a finding which this study contributes to the literature.

Test writers have focused on content validity, reliability, difficulty level and discrimination indices, all of which are important for the development of high-quality tests. Many of the SPS tests widely used in the literature have been developed to meet these requirements. However, due to the nature of SPS, if multiple-choice questions are to be used, the scenarios must be carefully formulated in the question stem. For example, when writing a question, the multi-stage experimental scenario needs to be considered. This study researched the effectiveness of the scenario-based multiple-choice tests widely used in SPS measurement. In multiple-choice SPS tests, item writers generally require the student to determine what hypothesis is being tested in an experiment and to identify the variables in a single-stage experiment. But in science, experiments can have several stages, and a different hypothesis can be tested in each part. Therefore, a manipulated variable in the first part of an experiment can become the controlled variable in the second part. The data collected in this study have demonstrated that multi-stage experiments are effective in ascertaining students' SPS competence. The findings of this research show that students exhibited differing performance in FHIVS questions with regard to differing parts of the same experiment. This variation originates from students' miscomprehension of the reality that variables may play different roles in different parts of an experiment. Nearly one fifth of the students failed to notice that a manipulated variable in the first part of an experiment was the controlled variable in the next part of the experiment. This situation affected their scores for identifying variables in addition to formulating hypotheses.

The results obtained in this study should be considered when assessing the skills of identifying variables and formulating hypotheses, skills which are among the most important SPS. If the students' performance in these areas is to be measured using multiple-choice test items, multi-stage experimental situations where a different hypothesis is tested at each stage should be used instead of single-stage experiments. The ways in which variables can change should be taken into consideration while selecting content to measure SPS. If a student chooses the right answer in a multiple-choice test, this is still not enough to conclude that student's knowledge of the subject is complete and accurate. In addition, a student may choose a distractor as the correct answer due to lack of information and mistakes made during the test. In addition to these factors, not being able to comprehend the changing role of the variables may cause the emergence of Groups 3 and 4 above. If the two-stage scenarios had not been used, this situation would not have been observed. This could have misled the researcher and the researcher may have believed that the student's SPS were more developed (or not as developed) as they were. Some researchers suggest using multiple stages in multiple-choice tests in order to determine misconceptions (Bahar, 2001; Karataş et al., 2003; Aykutlu & Şen, 2012). A similar approach should be followed for measuring SPS. For a student to be considered successful at a skill, she or he should be able to correctly answer the two parts of a related scenario, like the students in Group 1 above.

The advantages of using dual-stage questions while measuring the SPS can be summarized as follows: In reality, scientific experiments consist of multiple stages. Therefore, to use multi-

stage experimental scenarios to measure SPS is more realistic. While a variable can be an "independent variable" at a certain stage of the experiment, the same variable can also be a "controlled variable" at another stage of the experiment. The idea that a variable can play a different role at different stages of the experiment is a part of the "fair testing" strategy. For this reason, while measuring the skills of manipulating variables and formulating hypotheses, using multi-staged scenarios will give more sound results. Data collected from single-stage multiple-choice tests can be misleading. To make more consistent assessments, it is thus better to use multi-stage items.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

ORCID

Burak Kağan TEMIZ 🆔 https://orcid.org/0000-0001-8636-8743

5. REFERENCES

Abruscato, J. (2000). *Teaching children science*, Needham Heights, M.A: Allyn and Bacon.

Açıkgöz Ün, K. (2003). *Aktif öğrenme*, İzmir: Eğitim Dünyası Yayınları.

Temiz, B. K., Taşar, M. F., & Tan, M. (2006) Development and validation of a multiple format test of science process skills. *International Education Journal*, 7(7), 1007-1027.

Temiz, B. K. (2007). *Fizik öğretiminde öğrencilerin bilimsel süreç becerilerinin ölçülmesi [Evaluating students' science process skills at physics teaching]* (Unpublished dissertation). Gazi University, Ankara, Turkey.

Temiz, B. K. (2010). Bilimsel Süreç Becerilerini Ölçmede İçerik Seçiminin Önemi. *e-Journal of New World Sciences Academy,* 5(2), 614-628.

Aykutlu, I., & Şen, A. İ. (2012). Determination of secondary school students' misconceptions about the electric current using a three tier test, concept maps and analogies. *Education and Science*, 37(166), 275-288.

Aydınlı, E., Dökme, I., Ünlü, Z. K., Öztürk, N., Demir, R., & Benli, E. (2011). Turkish elementary school students' performance on integrated science process skills. *Procedia-Social and Behavioral Sciences*, 15, 3469–3475.

Aydoğdu B., Tatar N., Yıldız E., & Buldur S. (2012). The science process skills scale development for elementary school students. *Journal of Theoretical Educational Science*, 5(3), 292–311.

Aydoğdu, B., & Karakuş, F. (2017). Basic Process Skills Scale of towards Pre-School Students: A Scale Development Study, *Journal of Theoretical Educational Science*, 10(1), 49-72.

Bahar, M. (2001). Çoktan Seçmeli Testlere Eleştirel Bir Yaklaşım ve Alternatif Metotlar. *Kuram ve Uygulamada Eğitim Bilimleri*, 1(1), 23-28.

Beaumont-Walters, Y., & Soyibo, K. (2001). An analysis of high school students' performance on five integrated science process skills. *Research in Science and Technological Education*, 19(2). 133-145.

Berger, C. F. (1982). Attainment of skill in using science processes. I. Instrumentation. methodology and analysis. *Journal of Research in Science Teaching*, 19(3), 249-260.

Burns, J. C., Okey, J. R., & Wise, K. C. (1985). Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching*, 22(2), 169-177.

Büyüköztürk Ş, Kılıç Çakmak, E., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2009). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.

Carin, A. A. (1993). *Teaching science through discovery*. Toronto: Macmillan Publishing Company.

Carin, A. A., & Bass, J. E. (2001). *Teaching science as inquiry*. Upper Saddle River, New Jersey: Merrill Prentice Hall.

Dillashaw, F. G., & Okey, J. R. (1980). Test of the integrated science process skills for secondary students. *Science Education*, *64*, 601-608.

Erduran Avcı, D., & Bayrak, E. B. (2013). Investigating Teacher Candidates' Opinions Related to Scenario-Based Learning: An Action Research. *Ilkogretim Online*, *12*(2), 528-549.

Esler, M. K., & Esler, W. K. (2001). *Teaching elementary science. A full-spectrum science instruction approach*. Belmont CA.: Wadsworth Publishing.

Feyzioglu, B., Demirdag, B., Akyildiz, M., & Altun, E. (2012). Developing a science process skills test for secondary students: validity and reliability study. *Educational sciences: Theory and Practice*, *12*(3), 1899-1906.

Gabel, D. L. (1993). *Introductory science skills*, Illinois: Waveland Press, Inc.

Harlen, W. (1993). *Teaching and learning primary science*. London: Corwin Press.

Harlen, W. (1999). Purposes and Procedures for assessing science process skills. *Assessment in Education: Principles, Policy & Practice, 6*(1), 129-146.

Haury, D. L., & Rillero, P. (1994). *Perspectives of hands-on science teaching.* Retrieved from http://www.ncrel.org/sdrs/areas/issues/content/cntareas/science/eric/eric-toc.htm

Hughes, C., & Wade, W. (1993). *Inspirations for investigations in science*. Warwickshire: Scholastic Publication.

Karasar, N. (2000). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.

Karataş, F.Ö., Köse, S., & Coştu, B. (2003). Öğrencilerin Yanılgılarını ve Anlama Düzeylerini Belirlemede Kullanılan İki Aşamalı Testler. *Pamukkale Üniversitesi, Eğitim Fakültesi Dergisi*, *13*(1), 54-69.

Kazeni, M. M. M. (2005). *Development and validation of a test of integrated science process skills for the further education and training learners* (Unpublished master's dissertation). University of Pretoria, South Africa.

Kumar, D. (1999). *Computers and Assessment in Science Education*. Retrieved from ERIC database. (ED395770).

Lavinghousez, W. E. Jr. (1973, February). *The analysis of the biology readiness scale (BRS), as a measure of inquiry skills required*. Paper presented at BSCS Biology, College of Education, University of Central Florida.

Ludeman, R. R. (1975). *Development of the Science Processes Test (TSPT)* (Unpublished doctoral dissertation). Michigan State University. MI.

Lunetta, V. N., Hofstein. A., & Giddings, G. (1981). Evaluating science laboratory skills. *The Science Teacher*, *48*, 22-25.

Martin, D. J. (2002). *Elementary Science Methods a Constructivist Approach*. New York: Delmar Publishers.

Molitor, L. L., & George, K. D. (1976). Development of a test of science process skills. *Journal of Research in Science Teaching*, *13*(5), 405-412.

Ostlund, K. L. (1992). *Science process skills: assessing hands on student performance*. California: Addison Wesley.

Padilla, M. (1990). The science process skills. *Research Matters-to the Science Teacher*. No. 9004. Retrieved from http://www.educ.sfu.ca/narstsite/publications/research/skill.htm

Rezba, R. J., Sprague, C. S., Fiel, R. L., Funk, H. J., Okey, J. R., & Jaus, H. H. (1995) *Learning and assessing science process skills*. Iowa: Kendall/Hunt Publishing Company.

Riley, J. W. (1972). *The Development and Use of a Group Process Test for Selected Processes of the science Curriculum Improvement Study* (Unpublished Doctoral dissertation). Michigan State University. MI.

Shahali E. H. M., & Halim L., (2010), Development and validation of a test of integrated science process skills, *Procedia Social and Behavioral Sciences*, *9*, 142–146.

Smith, K. A., & Welliver, P. W. (1995). *Science process assessments for elementary and middle school students. Smith and Welliver Educational Services.* Retrieved from http://www.scienceprocesstests.com

Solano-Flores, G. (2000). Teaching and assessing science process skills in physics: The "bubbles" task. *Science Activities*, *37*(1), 31-37.

Song, J., & Black, P. J. (1991) The effects of task contexts on pupil's performance in science process skills. *International Journal in Science Education*, *13*, 49-58.

Song, J., & Black, P. J. (1992) The effects of concept requirements of task contexts on pupil's performance in control variables. *International Journal in Science Education*, *14*, 83-93.

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

Thalheimer, W. (2013). The Power of Scenario-Based Questions. Retrieved from http://www.immersivelearninguniversity.com/articlethalheimersep13

Tobin, K. G., & Capie, W. (1982). Development and validation of a group test of integrated science processes. *Journal of Research in Science Teaching, 19*(2), 133-141.

Tosun, C. (2019). Scientific process skills test development within the topic "Matter and its Nature" and the predictive effect of different variables on 7th and 8th grade students' scientific process skill levels. *Chemical Education Researh Practice*, *20*, 160-174.

Zimmerman, C., & Glaser, R. (2001). *Testing positive versus negative claims: A preliminary investigation of the role of cover story in the assessment of experimental design skills* (Tech. Rep. No. 554). Los Angeles, CA: UCLA National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Appendix: Examples of the Items from FHIVS Test

Scenario-A	Scenario-B

Susan has conducted an experiment which is shown below with a glass with a hole under it. Answer the 1st, 2nd, 3rd and 4th questions that follow.


Trial I   Trial II   Trial III   Trial IV
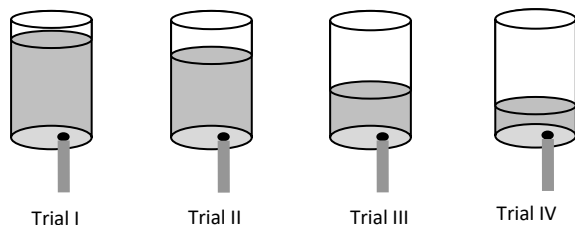
Susan, in her first attempt put liquid into the glass to a height of 15 cm and measured the time as 15 seconds for the glass to be completely emptied. In her second attempt, she put the same liquid into the same glass but this time to a height of 10 cm and measured the time for to empty the glass as 10 seconds. In her third attempt she put same liquid into the same glass to a height of 6 cm and measured the time to empty the glass as 7 seconds. In her fourth and last attempt she put the same liquid into the same glass to a height of 4 cm and measured the time to empty the glass as 5 seconds.

1. What is the manipulated variable in this research?
   a. Height of the liquid in the glass.
   b. Liquid's emptying time.
   c. Number of holes in the bottom of the glass.
   d. The size of the hole in the bottom of the glass.
   e. The type of the liquid in the glass.

2. What is the dependent variable in this research?
   a. Height of the liquid in the glass.
   b. Liquid's emptying time.
   c. Number of holes in the bottom of the glass.
   d. The size of the hole in the bottom of the glass.
   e. The type of the liquid in the glass.

3. What is/are the controlled variable(s) in this research?
   i. Height of the liquid in the glass.
   ii. Liquid's emptying time.
   iii. Number of holes in the bottom of the glass.
   iv. The size of the hole in the bottom of the glass.
   v. The type of the liquid in the glass.

   a. i   b. i and ii   c. ii, iv and v   d. iii, iv and v

   e. ii and iii

4. What is the hypothesis that was tested in this research?
   a. If the size of the hole in the bottom of the glass decreases, then the intensity of the liquid will decrease.
   b. If the height of the liquid in the glass increases, then the emptying time of the liquid will increase.
   c. If the number of the holes' increases, then the emptying time of the liquid will decrease.
   d. If the intensity of the liquid in the glass increases, then the emptying time of the liquid will increase too.
   e. If the size of the hole in the bottom of the glass increases, then the emptying time of the liquid will increase too.

Susan has conducted the new experiment below, with four similar size glasses with different size holes in the bottom. Answer the 5th, 6th, 7th and 8th questions that follow.

Trial I   Trial II   Trial III   Trial IV
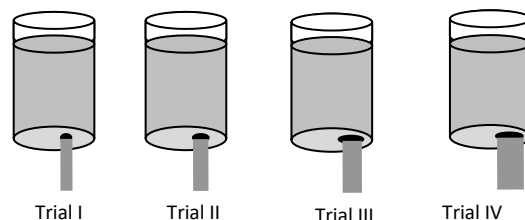
Susan, in her first try put liquid into the glass with 15 cm height and 2 mm hole scale and measured the time as 15 seconds for glass's getting emptied completely. In her second try, she put the same liquid into the same glass but this time with 15 cm height and 3 mm hole scale and measured the emptying time as 10 seconds. In her third try she put same liquid into the same glass with 15 cm height and 4 mm hole scale and measured emptying time as 7 seconds and in her fourth and last try she put same liquid into the same glass with 15 cm height and 5 mm hole scale and measured the emptying time as 7 seconds.

5. What is the manipulated variable in this research?
   a. Height of the liquid in the glass.
   b. Liquid's emptying time.
   c. Number of holes in the bottom of the glass.
   d. The size of the hole in the bottom of the glass.
   e. The type of the liquid in the glass.

6. What is the dependent variable in this research?
   a. Height of the liquid in the glass.
   b. Liquid's emptying time.
   c. Number of holes in the bottom of the glass.
   d. The size of the hole in the bottom of the glass.
   e. The type of the liquid in the glass.

7. What is/are the controlled variable(s) in this research?
   i. Height of the liquid in the glass.
   ii. Liquid's emptying time.
   iii. Number of holes in the bottom of the glass.
   iv. The size of the hole in the bottom of the glass.
   v. The type of the liquid in the glass.

   a. i   b. i and ii   c. ii, iv and v   d. iii, iv and v

   e. ii and iii

8. What is the hypothesis that was tested in this research?
   a. If the size of the hole in the bottom of the glass decreases, then the intensity of the liquid will decrease.
   b. If the height of the liquid in the glass increases, then the emptying time of the liquid will increase.
   c. If the number of the holes' increases, then the emptying time of the liquid will decrease.
   d. If the intensity of the liquid in the glass increases, then the emptying time of the liquid will increase too.
   e. If the size of the hole in the bottom of the glass increases, then the emptying time of the liquid will increase too.