

THE DEVELOPMENT OF SOLO TAXONOMY BASED ASSESSMENT INSTRUMENT ON HUMAN RESPIRATORY SYSTEM TO MEASURE THINKING SKILLS AND MASTERY OF CONCEPTS

Pranutasani Parameswari Putri

Pendidikan Biologi, FMIPA, Universitas Negeri Surabaya
Gedung C3 Lt.1 Jalan Ketintang, Surabaya 60231
e-mail: Pranutasaniputri16030204065@mhs.unesa.ac.id

Raharjo

Jurusan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri
Surabaya Jl. Ketintang Gedung C3 Lt.2 Surabaya 60231, Indonesia
e-mail : Raharjo@unesa.ac.id

Abstract

Appraisal or assessment is the core of education process which create a series of data required for obtaining its level of achievement on the thinking process. The aim of thinking is ultimately developing its individual ability to learn and master the concept which both elements can be acknowledged and measured precisely by using the correct instrument. Those two elements are the essence of this literature to exhibit the ultimate goal of its research based on valid empirical and theoretical data to develop a concept of appraisal solo taxonomy instruments based on human respiratory system for measuring thinking ability and mastering concepts. The human respiratory system is selected due to the most seen clearly and real of its significant positive correlation within group of students daily life. The aim of developing SOLO Taxonomy instruments is measuring and analyzing the group of student ability to learn and master a concept in various point of views to accomplish its level of Piaget cognitive. This research is constructed based upon research and development which the trial phase was done on line and 40 SMAN 12 Surabaya students being sampled randomly, Data analysis method used was a validation method and tests which were then analyzed descriptively. The results of the development research of SOLO taxonomy-based assessment instruments on the Material of the Human Respiration system to measure thinking skills and mastery of theoretically validated concepts with an average value on the type of question A of 100% and the type of question B of 96% which can be interpreted into the category of very valid data. Empirically, the reliability value for type A questions was 0.69 (reliable / high) and type B questions is 0.68 (reliable / high); the value of the level of difficulty of the type A questions is 67% (moderate) and 33% (easy), in the type of question B is 84% (moderate) and 16% (easy); the value of the sensitivity of the questions with 100% sensitive questions on both types of questions.

Keywords : *assessment instruments, SOLO taxonomy, Human Respiratory System*

INTRODUCTION

A good assessment in accordance with the demands required will produce appropriate information so that the objectives of education can be achieved. Information from the assessment includes the assessment process, student thinking outcomes, and thinking progress. Assessment in the 21st Century education has 4C characteristics, namely Communication, Collaboration, Critical Thinking and Problem Solving, Creativity and Innovation (Directorate of High School Development, 2017). These four characteristics will be interconnected and become one of life skills and thinking skills. Thinking skills are divided into two main categories, namely low level thinking skills (LOTS) and high level thinking skills (HOTS). From these two thinking skills, HOTS is one barometer in measuring the

intellectual of a nation. In the process of measuring the intellectuals of a nation, it can be done by using thinking evaluations. According to Subyantoro (2014) the evaluation process in the course must be pursued according to the target and done properly, so that students do not feel burdened but happy due to the feedback of students thinking outcomes must be represented as a motivation in thinking to enable successfully to achieve predetermined standards or values.

One important component in learning evaluation is the assessment instrument. Assessment instruments have meaning as a tool used to measure and find out the success of students in mastering competencies. In the demands of the 2013 Curriculum, students are directed to be able to think highly or HOTS in accordance to the requirement of 21st Century Education. Based on that statement, it

requires an instrument that can classify responses from students to obtain its data of thinking skills and mastery of concepts which are the ultimate goal of learning.

However, by observing the learning process in the field, teachers tend to use LOTS questions so that the information obtained is not in accordance with 21st Century learning concept that emphasizes on HOTS assessment. Rarely do evaluations by looking at student responses in answering questions, especially instruments with the type of problem description. This is not appropriate practices being done because in the problem description there will be many kinds of students responses in solving the problems they encounter and it is important to know that information about thinking skills and mastery of concepts can be appropriately obtained.

To overcome these problems, the teacher can make instruments using HOTS-based questions and analyze student responses base upon a classification that is the taxonomy of SOLO (Structure of Observed Thinking Outcome). In a previous study, Hamdani (2009) explained that by the SOLO taxonomic model, teachers can assess the quality of students' responses to a given problem.

Developed in 1982 by Bigg and Collis, the SOLO taxonomy has a classification for students' responses to the structure of observed learning outcomes into five levels, namely structural, unistructural, multistructural, relational, and expanded abstracts. The five levels show the ability of students to think using several alternatives and as comprehensive as the multistructural level, students are directed to be able to think several alternative answers. At the relational level, students are directed to enable thinking comprehensively. Finally, at an expanded abstract level, students are directed to be able to think comprehensively and make generalizations for solutions to given problems.

The level in SOLO taxonomy is similar to the level in Bloom's taxonomy which is generally used in making assessment instruments namely C1 through C6. However, there are the differences between SOLO taxonomy and Bloom's taxonomy. One of them is Bloom's taxonomy used to determine students' abilities based on students' cognitive process abilities in understanding a problem, while SOLO's taxonomy is used to measure students' ability to respond to a problem through the complexity of understanding the given problem (Ardiani, 2013). Based on these similarities and differences, SOLO taxonomy can be

used as a new perspective in measuring and analyzing students' abilities and as a complement in Piaget's cognitive level.

This difference is in line with previous research that explains that the development of instruments using the SOLO taxonomy can stimulate students to think more critically and creatively in solving problems and teachers are able to know the knowledge inside the brain of students precisely. By knowing the knowledge available to students, it can also be determined how the students' understanding and mastery of concepts of a material.

Respiratory system material was selected because it has direct visible phenomena such as the effect of cigarette smoke, air pollutants on the body, as well as conditions that are often experienced related to the respiratory system. Students with different backgrounds and thinking abilities will also provide diverse responses, therefore these responses must be analyzed appropriately in order to produce the required informations.

RESEARCH METHODS

This research is a development study using the Research and Development (R&D) model. The time and place of research until the development stage was carried out in October 2019 - March 2020, while the limited trial phase was conducted online in March 2020 at the homes of students of SMAN 12 Surabaya.

Theoretically, data validation that have been obtained are analyzed and interpreted according to the following table.

Tabel 1. Validity Interpretation

Percentage of Validity (%)	Validity Interpretation
$P \geq 81,50$	Very valid
$81,49 \geq P \geq 62,75$	Valid
$62,74 \geq P \geq 44,00$	Moderate valid
$43,99 \geq P \geq 25,00$	Invalid
$24,99 \geq P \geq 00,00$	Invalid

(Source : Retnawati, 2016)

Whereas in theoretical validation data which are obtained, its consist of reliability value, level of difficulty, and item sensitivity. The reliability value and success rate data are analyzed and interpreted in the following table.

Tabel 2. Interpretation of Test Reliability Values

Interval	Interpretation
$1,00 \geq r_{11} \geq 0,80$	Very high
$0,79 \geq r_{11} \geq 0,60$	high
$0,59 \geq r_{11} \geq 0,40$	Moderate
$0,39 \geq r_{11} \geq 0,20$	Low
$0,19 \geq r_{11} \geq 0,00$	Very Low

(Source : Arikunto, 2010)

Tabel 3. Interpretation of Difficulty Tests

Difficulty Index	Interpretation
0,00 – 0,30	Difficult
0,31 – 0,70	Moderate
0,71 – 1,00	Easy

(Source : Arikunto, 2010)

Through carefully calculate those datas obtained by employing theoretical and empirical validity, then the further step taken is to analyze the students' answers of each item. The process is carried out to determine the extent to which students' abilities in thinking and mastering concepts based on SOLO taxonomy.

RESULTS AND DISCUSSION

Instrument Development

After carrying out the development process, an indicator is produced which will then be developed into a SOLO Taxonomy based assessment instrument.

Tabel 4. These basic competencies and indicators are chosen as the basis for developing instruments based on the 2013 curriculum

Basic competencies	Indicator
Analyzing the correlation between the inner structure of organ building blocks in the respiratory system correspondence to bioprocess and impaired function that can occur in the human	3.8.1 Analyze the correlation between the structure of the organ building blocks and its function..
	3.8.1 Analyze the cause of a disease to explain the function of an organ.
	3.8.1 Analyzing the correlation between activities undertaken and the condition of the lungs.

Basic competencies	Indicator
respiration system.	3.8.1 Identifying the respiratory mechanism based on the exhale gas exchange takes place and the respiratory muscles mechanism being used.
	3.8.1 Evaluate the correlation of gas exchange process with air pressure.

Based on the basic competencies and indicators in table 4, assessment instruments can be developed by employing both basic competencies and indicators as one parameter also SOLO Taxonomy mapping simultaneously, the outcome of implementing both parameters can be explained by the following table.

SOLO taxonomic levels listed are expected to be the abilities level of students achievement.

Table 5. Correlation between indicators and SOLO Taxonomy

Indicator of Achievement	SOLO Level
analyze the correlation between the structure of the organ's constituents and their functions.	Relational
analyze the causes of a disease to find out the function of an organ.	Relational
Capability of analyzing the correlation between activities undertaken with the condition of the lungs.	Relational
	Relational
Capability to explain various respiratory mechanisms based on the gas exhale exchange takes place and respiratory mechanism based on the muscles used.	Multistructural
Capability to explain the correlation between the gas exchange processes with air pressure.	Extended Abstract

Based on table 5, we can develop a SOLO taxonomy-based valuation instrument which will be explained in the following table.

Table 6. correlation between items with SOLO Taxonomy

Question	SOLO Level
1.A. By viewed from its function and location, why do human tracheal parts consist of cartilage rings and rigid walls?	Relational
1.B. By viewed from its function, why must the alveoli be composed of a layer of epithelial tissue?	
2.A. one of the phenomena of asfliction is carbon monoxide poisoning. What is the effect of carbon monoxide with a high quantity on the human body?	Relational
2.B. Asthma sufferers are not recommended to do strenuous exercise because the disease can recur. How is the relationship between strenuous exercise against asthma?	
3.A. the respiratory frequency of people who rarely exercise will be different from athletes. Even people who rarely exercise will get tired more quickly than athletes in sports. From this phenomenon, how can this happen?	Relational
3.B. Exercising at night in a lush area with trees has a greater risk of having difficulty breathing, why does that happen?	
4.A. The Bajau are a tribe that lives on the coast, so the majority of the Bajau tribe occupations are diving and fishing. In fact this tribe is able to dive for 13 minutes with only one breath. From this case, how is the influence of these habits on the body condition of the Bajau people?	Extended Abstract
4.B. When you are on a bus with poor maintenance (a door that remains open, a thick smoky exhaust / exhaust system) for quite a long time, what is the likelihood of that happening if it is associated with the human respiratory system?	
5.A. What is the correlation between external breathing (external) with deep breathing (internal) in the process of human breathing?	Multistructural

Question	SOLO Level
5.B. How is the interoperability of the muscles between the ribs and diaphragm to the process of human breathing?	Extended Abstract
6.A. If the air pressure in the lungs is lower than the condition of the surrounding environment, what is the likelihood that will occur?	
6.B. If the air pressure in the lungs is higher than the conditions in the surrounding environment, what is the likelihood that will occur?	

Theoretical Validity

Based on the results of the assessment of experts / experts in the field of material and education obtained results in the form of values in each of the listed criteria namely material, construction, and language. The assessment is done by filling out the validation sheet that the researcher has provided. In the material criteria and for the type questions A and B get an average value of 100, while the language criteria get a value of 88.



Figure 1. Diagram of the recapitulation results of the theoretical validation of the test instrument to analyze the mastery of the concept of SOLO taxonomy on Respiratory System Material.

Empirical Validation

Empirical validity consists of reliability, item difficulty level, and sensitivity. The reliability value is obtained from previous data which is then processed using the Cronbach Alfa method in the SPSS 20 program.

Table 7. Reliability Table of Types A and B with the Cronbach Alpha Method.

Types	Consist of Reliability
A	0,69
B	0,68

The calculation of the level of difficulty in the type of questions A and B obtained results in the following diagram

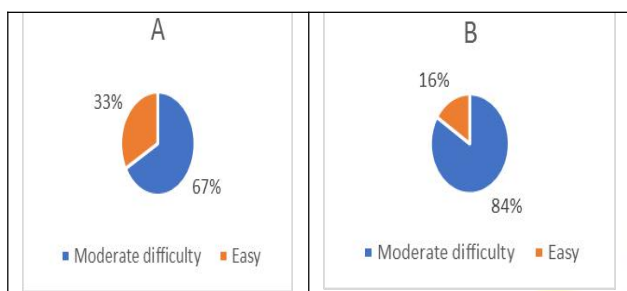


figure 2. Difficulty Level Diagram for Each Item of Types A and B in the SOLO Taxonomy-Based Assessment Instrument on Human Respiration System Material to Measure Thinking Skills and Concept Mastery.

In sensitivity, each item tested has the following results.

Tabel 8. Results of Calculation of Sensitivity of Problem Items

No.	Sensitivity	
	Type A	Type B
1.	0,8	0,7
2.	0,7	0,6
3.	0,7	0,7
4.	0,7	0,7
5.	0,8	0,7
6.	0,7	0,8

Theoretical Validity

In the type of question A, the resulting validation is 100% and the type of question B is 96%, the average consists of material, construction, and linguistic criteria. In the material criteria, the types of questions A and B get a percentage of 100% which can be interpreted very valid. This shows that the assessment instruments have met the material validity criteria which include the compatibility between the questions with basic competencies and achievement indicators, the compatibility of the items with the truth of existing concepts or theories, the compatibility of the domains in the SOLO taxonomy,

the limits of questions with clear answers and in accordance with the level in the SOLO taxonomy, the content of the material in question is in accordance with the level of education in schools (Retnawati, 2016).

In the construction criteria, the types of questions A and B have a result an average percentage of 100% which can be considered very valid (Retnawati, 2016). Based on this interpretation it can be concluded that the test instrument is in accordance with aspects related to the validity of the construction which includes scoring instructions in accordance with the keywords, instructions for working on questions that are easy to understand, questions that do not depend on the answers to previous questions, and for parts of tables, graphics, or pictures (if any) cannot be given a value because of its problem impossibilities.

In the linguistic criteria, type of question A has a percentage result value of 100% and type of question B 88% which can be interpreted very valid (Retnawati, 2016). Aspects of linguistic criteria consist of two things namely the sentence that does not lead to double interpretation and the use of grammar and spelling in accordance with Indonesian rules. In the type B, there are problems that are not in accordance to the linguistic aspects listed, among others, in questions number 3 and 5 do not meet the linguistic aspects of grammar and spelling in accordance to Indonesian language norms, in question number 4 there are sentences that cause double interpretations. But all three of these problems have been fixed in accordance to its input from the experts / experts before conducting a limited trial to students.

Empirical Validity

The test item that have been developed and theoretically validated were then tested on 40 students online. The use of online methods is due to conditions (the Corona virus) which makes it is impossible to conduct face-to-face meetings at school. There are several aspects of empirical validity, namely the value of reliability, the level of difficulty items, and sensitivity.

In finding the reliability value for this instrument, the Alpha Cronbach method is used in the SPSS 20 software. The use of this software is explained by Widoyoko (2014) who said that besides using the manual method of calculating the reliability value there

are several softwares that can be used to facilitate the calculation between Other SPSS for Windows, Lisner Excel, and so on.

The results of the reliability calculation for both types of questions produce a value of 0.69 for questions A and 0.68 for questions B which the averaged of both, become 0.68. Both values can be interpreted is high according to Arikunto (2010) which means that both test instruments are reliable. Reliable definition for an instrument is the result of its measurement which is fixed and can be trusted. Measurable measurement means that if this test is given to a student and repeated at a different time, the student will still be in the same rank or group.

After knowing the value of reliability, it must also be known the level of difficulty of the two types of questions from the instrument. Based on figure 2 it can be seen that in type A questions it is known that 67% of questions have moderate difficulty level and 33% are easy, while in type B questions 84% of questions have moderate difficulty level and 16% are easy. These two types of questions were not found difficult levels because almost all students answered the questions well. This can occur due to two things namely extrinsic factors (from outside the student) and intrinsic factors (from within the student) (Ristiyani & Bahriah, 2016). One of the extrinsic factors is the environment, in working on this instrument students work online so that it can be done anywhere which in a conducive learning environment can make students more concentrated in working on the problems. One of the intrinsic factors is that psychics, students feel learning is fun and able to receive lessons well which in this case found no difficult problems in both types.

In the calculation of sensitivity in both types of questions, the results show a value of more than 0.3 which can be interpreted all the problems in both are sensitive types (Syaifulloh, Rizal, B., & Jatmiko, B., 2014). This shows that in many students are able to master the indicators and basic abilities that are interrelated.

CONCLUSION

The development of SOLO-based taxonomy assessment instruments on the Respiratory System Material can be concluded theoretically and empirically valid. The results of theoretical validation on the types of questions A and B are 100% and 96%. The results of empirical validation are the reliability of both with an average of 0.68; difficulty level for type

A questions (67% moderate and 33% easy) and type B questions (86% moderate and 14% easy); and sensitivity with an average yield of 0.7 for both types of questions.

SUGGESTION

Further development research is required on SOLO's taxonomy-based assessment instruments on different material, with more samples on several schools to enable to show valid and reliable data obtained and assessment instruments are developed more flexible.

ACKNOWLEDGMENT

I as a researcher would like to thank Dra. Nur Kuswanti, M.Sc.St and Prof. Dr. Muslimin Ibrahim, M.Pd as the examiner and validator. Thank you also to Surabaya 12 Public High School for their availability during this study.

REFERENCES

- Ardiani, N. F. 2013. *Taksonomi Bloom VS Taksonomi (SOLO, Fink, Marzano) dalam Pembelajaran*. Universitas Kristen Satya Wacana.
- Arikunto, S. 2010. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT. Bumi Aksara.
- Bigg, J. B., & Collis. 1982. *Evaluating The Quality of Thinking : The SOLO Taxonomy*. New York: Akademik Press Inc.
- Brown, B. 2015. Twenty First Century Skills: A Bermuda College. *Twenty First Century Skill*, 58-64.
- Direktorat Pembinaan SMA . 2017. *Implementasi Pengembangan Kecapakan Abad 21 dalam Perencanaan Pelaksanaan Pembelajaran (RPP)*. Jakarta: Kementrian Pendidikan dan Kebudayaan.pdf.
- Ennis, R. 2013. *The Nature of Critical Thinking*. Diambil kembali dari Outlines of General Critical Thinking Dispositions and Abilities: <http://www.criticalthinking.net/longdefinition.html> (Diakses pada 19 November 2019)
- Griffin, P., & Care, E. 2015. *Assessment and Teaching of 21st Century Skills : Methods and Approach*. Dordrecht: Springer Business Media.
- Guo, Z. 2016. The Cultivation of 4C's in China Critical Thinking, Communication. *International Conference on Education, Management, and*

Applied Social Science, 1-4.

Hamdani, A. S. 2009. Pengembangan Sistem Evaluasi Pembelajaran Pendidikan Agama Islam Berbasis Taksonomi SOLO. *Jurnal Pendidikan Islam Vol.01, No.01 ISSN 2085-3033*.

Kivunja, C. 2014. Innovative Pedagogies in Higher Education to Become Effective Teachers of 21st Century Skills : Unpacking The Thinking and Innovations Skills Domain of The New Thinking Paradigm. *International Journal of Higher Education*, 3(4), 37.

Osman, K., Hiong, L. C., & Vebrianto, R. 2013. 21st Century Biology : An Interdisciplinary Approach of Biology, Technology, Engineering, and Mathematics Education. *Procedia-Social and Behavioral Sciences*, 102, 188-194.

Rahayu, P. 2019. *Pengembangan Instrumen Penilaian Berbasis Taksonomi The Structure Of Observed Thinking Outcome pada Materi Konsep Larutan Penyanga dan Hidrolisis*. Semarang: Universitas Negeri Semarang.

Retnawati, H. 2016. *Validitas, Reliabilitas, dan Karakteristik Butir (Panduan untuk Peneliti, Mahapeserta didik, dan Psikometrian) Berbasis Software*. Yogyakarta: Parama Publishing.

Ristiyani, E., & Bahriah, E. S. 2016. Analisis Kesulitan Belajar Kimia Siswa di SMAN X Kota Tangerang Selatan. *Jurnal Penelitian dan Pembelajaran IPA*, 18-29.

Subyantoro. 2014. Pengembangan Evaluasi Pembelajaran Bahasa Indonesia Berbasis Taksonomi SOLO. *LITERA VOL 13, NO 1*, 67-78.

Syaifulloh, Rizal, B., & Jatmiko, B. 2014. Penerapan Pembelajaran dengan Modul Guided Discovery dengan Lab Virtual PhET untuk Meningkatkan Hasil Belajar Peserta Didik Kelas XI di SMAN 1 Tuban pada Pokok Bahasan Teori Kinetik Gas. *Jurnal Inovasi Pendidikan Fisika*.

Widoyoko, S. E. 2014. *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Belajar.