

Development of Higher Order Thinking Skills Assessment Refer the Theory of Bloom in Laboratory Work

Soka Hadiati¹, Eti Sukadi², Adi Pramuda³

DOI: 10.35445/alishlah.v13i3.1075

Article Info**Abstract**

Keywords:
Assessment; *Bloom*,
HOTS; *Laboratory Work*

Higher-order thinking skills (HOTS) are one of the important components that students must master to be successful in the 21st century. The HOTS of students in several regions in Indonesia is low because HOTS-based learning and the instruments are not appropriate. This study aims to determine the validity and reliability of the HOTS assessment content, empirical, and constructs referring to bloom theory in laboratory work. The development model used is the Plomp development model, which consists of five stages: initial investigation, design, realization/construction, test, evaluation and revision, and implementation. The HOTS assessment was proven relevant by experts and then distributed to 70 Physics Education students. The HOTS assessment was analyzed by rash model theory analysis with the Winstep program to obtain empirical validity. SEM (Structural Equation Model) was used to obtain construct validity. The results showed that the HOTS assessment was valid and reliable. The results show that the assessment has been able to measure every aspect of HOTS well in laboratory work so that it can be used.

Kata kunci:
Asesmen; *Bloom*; *HOTS*
Kerja Laboratorium

Abstrak

Higher order thinking skills (HOTS) merupakan salah satu komponen penting yang harus dikuasai siswa untuk sukses pada abad 21. Penelitian ini bertujuan untuk mengetahui validitas dan reliabilitas isi, empiris, dan konstruk dari assesmen HOTS mengacu teori bloom dalam kerja laboratorium. Model pengembangan yang digunakan yaitu model pengembangan Plomp yang terdiri atas lima tahap yaitu investigasi awal, desain, realisasi/konstruksi, tes, evaluasi dan revisi, implementasi. Asesmen HOTS di buktikan relevansinya oleh pakar kemudian di sebarakan ke 70 mahasiswa Pendidikan Fisika. Asesmen HOTS dianalisis dengan analisis *rash model theory* dengan program *Winstep* untuk mendapatkan validitas empiris. SEM (*Structural Equation Model*) digunakan untuk mendapatkan validitas konstruk. Hasil penelitian menunjukkan bahwa asesmen HOTS valid dan reliabel. Hasil penelitian menunjukkan bahwa asesmen telah mampu mengukur setiap aspek dari HOTS dengan baik dalam kerja laboratoium sehingga dapat digunakan.

¹ IKIP PGRI Pontianak, Pontianak Kalimantan Barat
Email: sokahadiati@gmail.com

¹ IKIP PGRI Pontianak, Pontianak Kalimantan Barat
Email: etisukadi_sukadi@yahoo.com

³ IKIP PGRI Pontianak, Pontianak Kalimantan Barat
Email: adipramuda@gmail.com

INTRODUCTION

Higher Order Thinking Skills are essential to solve problems in everyday life as a prerequisite for success in the 21st century (Saavedra & Opfer, 2012). Big countries such as Singapore, the United States, Taiwan, and China have inserted HOTS in the curriculum (Hwang, Lai, Liang, Chu, & Tsai, 2018; Yeung, 2015). HOTS needs special attention because of their role in making the right decision and logically solving problems in everyday life, so it needs special attention.

Several studies have examined the importance of developing HOTS. HOTS is needed to solve problems efficiently, communicate and collaborate with others effectively, and understand creative ideas (Kosturko, McQuiggan, & Sabourin, 2015). HOTS involves making conclusions from concrete facts and arrive at logical conclusions by making clear connections (Scott, Barbarin, & Brown, 2013). Learning that trains HOTS will lead students to think logically by using reasoning and associating with the knowledge they have to solve problems.

HOTS of students in Malang was low namely 59%, 58%, and 57%, respectively for independent thinking, critical thinking, and creative thinking (Husamah, 2018). 60% of students in Padang have HOTS in the low category (Ahmad, Prahmana, Kenedi, Helsa, & Zainil, 2017). The low HOTS can be caused because the HOTS assessment instrument has not been able to measure HOTS accurately and accurately, so it is necessary to develop a HOTS assessment based on the right theory so that in this study the development of a HOTS assessment based on Bloom's theory was carried out.

The HOTS assessment is specifically designed for laboratory work activities. Laboratory work activities expose students to equipment and methodologies, where students receive a set of instructions on how to conduct certain experiments and the proper procedures needed to analyze data (Sterna, Echeverria, & Porta). In laboratory work, there are many problem-solving activities and process skills, so that it has the potential to develop student HOTS. This study aimed to determine: (1) the content validity of the higher-order thinking skills assessment referring to bloom theory in laboratory work, (2) the empirical validity and reliability of the online higher-order thinking skills assessment referring to bloom theory in laboratory work, (3) the validity and reliability construct online assessment of higher-order thinking skills referring to bloom theory in laboratory work.

METHOD

The development model used refers to the Plomp development model (Buchory & Setiawati, 2015). The development model consists of five stages, namely: (1) the initial investigation stage; (2) Design; (3) Realization/construction stage; (4) tests, evaluations and revisions; (5) complete implementation in Figure 1.

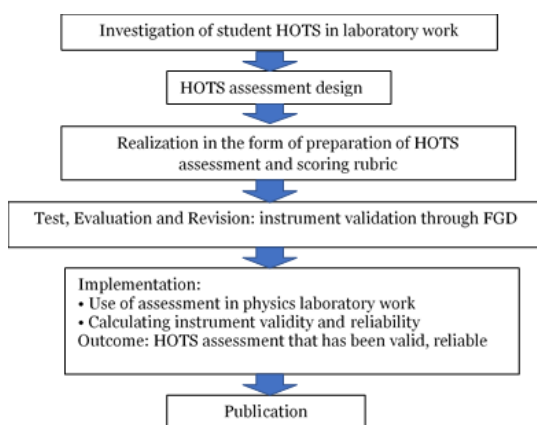


Figure 1. Research Method

The HOTS assessment, which is based on theory, will be proven relevant (content validation) by experts to test the data, find the meaning of the data and position the data in the right concept/theory frame, which is carried out in a small forum consisting of competent people who represent the field group. Practitioners, observers in the field of study, academics in the field of study, bureaucrats who can contribute to data stability (Satori & Komariah, 2014). The HOTS assessment was distributed to 70 Physics Education students in West Kalimantan to get the HOTS score. The HOTS score was analyzed for empirical validity of the instrument using rash model theory analysis with the winsteps program. Reliability using Cronbach's Alpha. Validity is used to test the consistency of the developed assessment. Then to test the construct validity, SEM (Structural Equation Model) was used.

FINDINGS AND DISCUSSION

The research instrument was validated by seven validators: physics learning media experts, material experts, and lecturers in Basic Physics and Electrical Measuring Instruments courses. The results of the assessment instrument's content validity and inter-rater reliability are presented in Table 1. Table 1 presents the results of the validator's assessment of the test items. Table 1 shows that all test questions are valid and reliable.

Table 1. Results of Content Validity and Inter-rater Reliability

Item	V	Kategori validitas	Reliability
1	0.75	Moderate	67% Reliabel
2	0.75	Moderate	
3	0.75	Moderate	
4	0.75	Moderate	
5	0.93	High	
6	0.75	Moderate	
7	0.75	Moderate	
8	0.96	High	
9	0.96	High	
10	0.96	High	
11	0.75	Moderate	
12	0.75	Moderate	
Mean	0.82	High	

Empirical validity was calculated using the Winsteps program, which refers to the Rasch Model Theory. The output of the Winsteps program shows the validity, reliability, and item discrimination. The results of the calculation of the validity and reliability of the HOTS are presented in Tables 2 and 3.

Table 2. Empirical Validity of Test Instruments

Item	Outfit MNSQ	Category
1	Maximum	Valid
2	0,65	Valid
3	0,86	Valid
4	1,05	Valid
5	1,36	Valid
6	1,05	Valid
7	1,00	Valid
8	0,99	Valid
9	1,11	Valid
10	0,93	Valid
11	1,17	Valid
12	0,70	Valid
Mean	0,99	Valid

Table 2 shows that all HOTS items are valid because they have MNSQ OUTFIT values in the range of $0.5 < \text{MNSQ} < 1.5$. This shows that all items can measure accurately. Table 3 shows that the HOTS instrument has item reliability 0.89 and person reliability 0.48 with good and moderate categories. This shows that the consistency of the answers from the subject is still weak, but the quality of the items in the instrument's reliability aspect is quite good.

Table 3. Empirical Reliability of Test Instruments

Instrument	Type	R	Reliability category
HOTS	Item Reliability	0,89	Good
	Person Reliability	0,48	Moderate

Item discrimination refers to the ability of an item to differentiate among students. Item discrimination can be determined referring to the Pt Measure Corr value in Table 4 with the following criteria: very good (> 0.4), good (0.3-0.39), sufficient (0.2-0.29), not able to discriminate (0.00-0.19), requires examination of items (< 0). Item number 1 needs to be considered because it has poor.

Table 4. Item discrimination

Item	Pt Measure Corr.	Category
1	0,00	Poor
2	0,31	Good
3	0,42	Sangat baik
4	0,34	Good
5	0,49	Very good
6	0,37	Good
7	0,35	Good
8	0,47	Very good
9	0,30	Good
10	0,38	Good
11	0,54	Very good
12	0,50	Very good

The results of the analysis of validity, reliability and item discrimination identified that the HOTS instrument question number 1 was not used. The results of the empirical validation become a reference in eliminating several questions on the HOTS test instrument. In the HOTS test item number 1 (B1) is eliminated. The next stage is the analysis of construct validity and reliability for HOTS questions with first-order CFA as presented in Figure 2 and Figure 3.

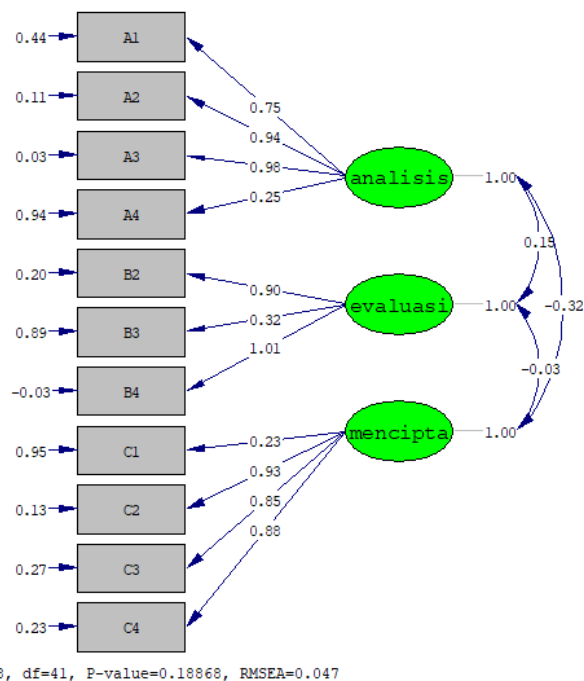


Figure 2. Basic Model Standardized Solution HOTS

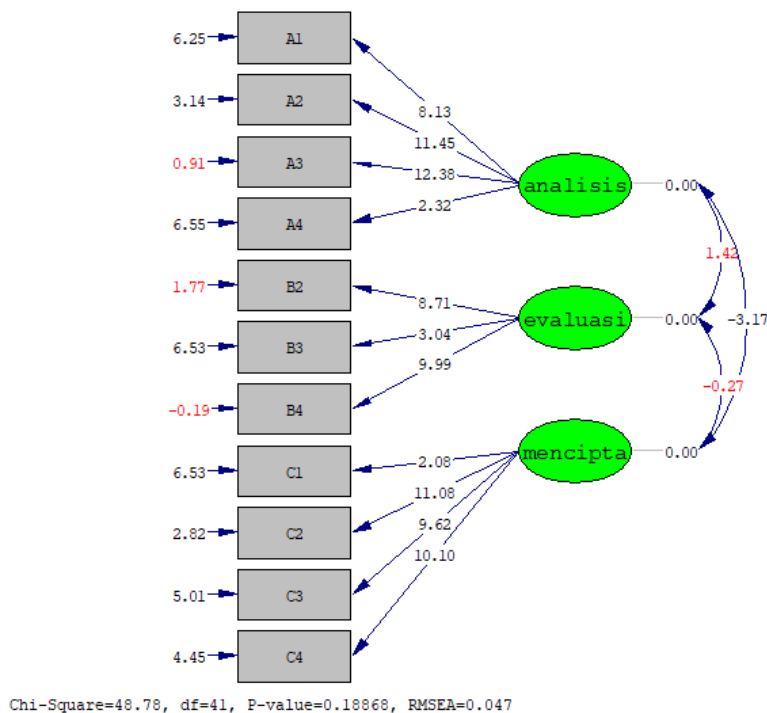


Figure 3. Basic Model t-values HOTS

Figure 2 shows the first-order CFA on 11 HOTS items. The results of the first-order CFA analysis on the HOTS instrument obtained a P-value > 0.05, and RSMEA was close to 0, so that it was concluded that there was a model fit. Figure 2 shows the estimated value of the factor load. Figure 3 shows the estimated significance to determine the effect of the observable variable to the latent variable. Figure 2 and Figure 3 serve as a reference for determining the instrument's construct validity. The summary of the validity and reliability of the HOTS construct from the results of the CFA analysis can be seen in Table 5.

Table 5. Construct Validity and Reliability of HOTS Instruments

Item	Loading factors	Standart Errors	Validity construct	t-value	Significance	Reliability
A1	0,75	0,44	Good	8,13	Yes	Construct Reliability = 0,93
A2	0,94	0,11	Good	11,45	Yes	
A3	0,98	0,03	Good	12,38	Yes	
A4	0,25	0,94	Poor	2,32	Yes	
B1	0,90	0,20	Good	8,71	Yes	Maximal Reliability = 0,91
B2	0,32	0,89	Poor	3,04	Yes	
B3	1,01	0,03	Good	9,99	Yes	
B4	0,23	0,95	Poor	2,08	Yes	
C1	0,23	0,95	Poor	2,08	Yes	
C2	0,93	0,13	Good	11,08	Yes	
C3	0,85	0,27	Good	9,62	Yes	
C4	0,88	0,23	Good	10,10	Yes	

Table 5 shows that all items influence the latent variable. Items A4, B2, C1 have less significant construct validity for latent variables, while items A1, A2, A3, B1, B3, B4, C2, C3, C4 have good construct validity. The results of this analysis indicate that most of the items have been able to measure every aspect of the HOTS so that they can be used entirely. Laboratory work is able to present measurement tools that can stimulate students' reasoning so that it can lead to an increase in HOTS. Students' reasoning was stimulated by presenting models, modelling and teaching (Heijnes & Van Joolingen, 2018). Modelling using measurement tools helps students process scientific information better (Jong, Chiu, & Chung, 2015). Scientific modelling in learning contributes to improving cognitive, metacognitive and social abilities (Louca & Zacharia, 2011; Malone, Schunn, & Schuchardt, 2018). Experiments in laboratory work will stimulate reasoning so that students can develop HOTS (Richland & Simms, 2015; Zohar & Alboher, 2018). Each step of the laboratory work in designing experiments, conducting experiments, and analyzing results can explore students' HOTS. The instrument that has been made refers to the laboratory work objectives, laboratory work steps, and the right theory HOTS be able to measure students' abilities accurately. This is in line with the research that found that reasoning will be formed through collaboration, collaboration, and communication activities in laboratory work encourage an increase in HOTS (Hwang, Lai, Liang, Chu, & Tsai, 2018; Manoney & Harris, 2017). In addition, laboratory work helps students develop understanding and skills empirically (Girault, d'Ham, Ney, Sanchez, & Wajeman, 2012). This suggests that the laboratory work can be applied easily in learning and being able to help learners achieve good competence. Laboratory work activities can stimulate students' reasoning because practical tools require logic and skills in conducting experiments.

CONCLUSION

The HOTS assessment was valid and reliable. The results show based on content validity. It is known that the HOTS assessment has a validity value of 0.82 with a high category and 67% reliability. Empirical validity shows that the HOTS assessment is valid (0.99) and has item reliability 0.89 and person reliability 0.48 with good and moderate categories. Construct validity shows that most of the items have been able to measure every aspect of the HOTS to be used entirely. The assessment has been able to measure every aspect of HOTS well in laboratory work so that it can be used in laboratory work. The instrument trial is limited to the West Kalimantan region, so further testing is needed in other areas. It is necessary to develop HOTS instruments referring to different theories as comparisons.

REFERENCES

- Ahmad, S., Prahmana, R., Kenedi, A., Helsa, Y., & Zainil, M. (2017). The instruments of higher order thinking skills. *Journal of Physics: Conference Series*, 943, 012053. doi:<https://doi.org/10.1088/1742-6596/943/1/012053>
- Girault, I., d'Ham, C., Ney, M., Sanchez, E., & Wajeman, C. (2012). Characterizing the Experimental Procedure in Science Laboratories: A preliminary step towards students experimental design. *International Journal of Science Education*, 36(4), 825–854.
- Heijnes, D., & Van Joolingen, W. (2018). Stimulating Scientific Reasoning with Drawing-Based Modeling. *Journal of Science Education and Technology*, 27(1), 45-56.
- Husamah, F. D. (2018). OIDDE Learning Model: Improving Higher Order Thinking Skills of Biology Teacher Candidates. *International Journal of Instructio*, 11(2), 249-264. doi:<https://doi.org/10.12973/iji.2018.11217a>
- Hwang, G., Lai, C., Liang, J., Chu, H., & Tsai, C. (2018). A long-term experiment to investigate the relationships between high school students' perceptions of mobile learning and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, 66(1), 75–93. doi:<https://doi.org/10.1007/s11423-017-9540-3>.
- Hwang, G., Lai, C., Liang, J., Chu, H., & Tsai, C. (2018). A long-term experiment to investigate the relationships between high school students' perceptions of mobile learning and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, 66(1), 75-93. doi:<https://doi.org/10.1007/s11423-017-9540-3>.
- Jong, J., Chiu, M., & Chung, S. (2015). The use of modeling-based text of ideal gas law to improve students' modeling competencies. *Science Education*, 99(5), 986–1018.
- Kosturko, L., McQuiggan, J., & Sabourin, J. (2015). *Mobile Learning: A Handbook for Developers, Educators, and Learners*. USA: John Wiley & Sons, Inc.
- Louca, L., & Zacharia, Z. (2011). Modeling-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64, 471-492. doi:[10.1080/00131911.2011.628748](https://doi.org/10.1080/00131911.2011.628748).
- Malone, K., Schunn, C., & Schuchardt, A. (2018). Improving conceptual understanding and representation skills through Excel-based modeling. *Journal of Science Education and Technologi*, 27(1), 30-44.
- Manoney, J., & Harris, B. R. (2017). The effects of collaborative testing on higher order thinking: Do the bright get brighter?. *Active Learning in Higher Education*, 1-13.
- Richland, L., & Simms, N. (2015). Analogy, higher order thinking, and education: Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(2), 177–192. doi:<https://doi.org/10.1002/wcs.1336>
- Saavedra, A., & Opfer, V. (2012). Learning 21st-Century Skills Requires 21st-Century Teaching. *Phi Delta Kappan*, 94(2), 8–13. doi:<https://doi.org/10.1177/003172171209400203>
- Satori, D., & Komariah, A. (2014). *Metodologi Penelitian Kualitatif*. Bandung: Alfabeta.
- Scott, K., Barbarin, O., & Brown, J. (2013). From higher order thinking to higher order behavior: Exploring the relationship between early cognitive skills and social competence in Black boy. *American Journal of Orthopsychiatry*, 83(2-3), 185–193.
- Sterna, C., Echeverria, C., & Porta, D. (n.d.). Teaching Physics through Experimental Project. *Procedia IUTAM*, 20, 189–194. doi:<https://doi.org/10.1016/j.piutam.2017.03.026>
- Yeung, S. S. (2015). Conception of teaching higher order thinking: perspectives of Chinese teachers in Hong Kong. *The Curriculum Journal*, 26(4), 553–578. doi:<https://doi.org/10.1080/09585176.2015.1053818>
- Zohar, A., & Alboher, V. A. (2018). Raising test scores vs. teaching higher order thinking (HOT): senior science teachers' views on how several concurrent policies affect classroom practices. *Research in Science & Technological Education*, 36(2), 243–260. Retrieved from <https://doi.org/10.1080/02635143.2017.1395332>