High Order Thinking Skills and Digital Literacy Skills Instrument Test

Dzulfiqar Satria Waliyuddin[1], Dwi Sulisworo[2]
Ahmad Dahlan University, Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia[1,2]
dzulfiqar2008050027@webmail.uad.ac.id[1], dwi.sulisworo@uad.ac.id[2]

Abstract: Developing evaluation instruments in the student learning process in interactive, fun, challenging, motivating, and developing students 'creativity and independence are needed to improve students' High Order Thinking Skills (HOTS) and digital literacy skills (DLS). HOTS and digital literacy are part of the 21st-century skills that are important for every individual. This study developed a test instrument integrated with Augmented Reality (AR) and Google Scholar (GS) to encourage HOTS and students' digital literacy skills. This type of research was development research. The subject matter was the scale and comparison. Product trials were applied to 11th-grade students in Sleman, Indonesia. Product validity was based on expert judgment, item validity, distinguishing power, difficulty level and reliability to reach a good category in general. AR was created using the ARLOOPA application operated with android. The research instrument was designed to increase HOTS and DLS.
Keywords: ARLOOPA; Augmented Reality; High Order Thinking Skills; Digital Literacy Skills; mathematics.

Tes Instrumen Keterampilan Berpikir Tingkat Tinggi dan Keterampilan Literasi Digital

Abstrak: Pengembangan instrumen Evaluasi dalam proses pembelajaran siswa secara interaktif, menyenangkan, menantang, memotivasi, dan mengembangkan kreativitas dan kemandirian siswa diperlukan untuk meningkatkan High Order Thinking Skills (HOTS) dan digital literacy siswa (DLS). HOTS dan literasi digital adalah bagian dari keterampilan abad ke-21 yang penting bagi setiap individu. Penelitian ini mengembangkan instrumen tes yang terintegrasi dengan Augmented Reality (AR) dan Google Scholar (GS) untuk mendorong HOTS dan keterampilan literasi digital siswa. Jenis penelitian ini adalah penelitian pengembangan. Pokok bahasannya adalah skala dan perbandingan. Uji coba produk diterapkan pada siswa kelas 11 di Sleman, Indonesia. Validitas produk didasarkan pada penilaian ahli, validitas item, daya pembeda, tingkat kesulitan dan reliabilitas mencapai kategori baik secara umum. AR dibuat menggunakan aplikasi ARLOOPA yang dioperasikan dengan android. Instrumen penelitian dirancang untuk meningkatkan HOTS dan DLS.
Kata kunci: ARLOOPA; Augmented Reality; Kemampuan Berpikir Tingkat Tinggi; Kemampuan Literasi Digital; matematika.

1. Introduction

The 21st-century learning framework must incorporate technology into the learning process that focuses on creativity, critical thinking, communication, and collaboration. Six components must be achieved in 21st-century learning, namely: creativity, collaboration, information fluency, critical thinking, problem-solving, decision making, citizenship and concepts and technology. First, students can demonstrate the ability to think creatively, build knowledge, and develop innovative products. Second, learners can use digital media to communicate and collaborate remotely to support the learning process for themselves and together. Third, students can apply digital tools to collect information, evaluate and process information. Fourth, students can do problem-solving and decision making. Then students have critical thinking skills to plan and carry out problem-solving and decision making based on information using digital technology. Fifth, students can understand human, cultural and social problems related to technology legally and ethically. Sixth, students can show a good understanding of technology concepts (Rochelle, 2000). In other words, the current role of education in the learning process in schools is more focused on developing student resources, including Digital Literacy Skills (DLS) and High Order Thinking Skills (HOTS).

DLS is an individual's ability to understand and use information in various contexts using his ability to think critically and his skills in using information and communication technology. Today it can be seen together that smartphones have become the primary medium for communication. Therefore, individuals are required to have the skill to understand the text, symbols and graphics digital. In general, it can be said that human life today has a new type of knowledge closely related to the development of the digital world. Digital literacy is knowledge and skills in using digital media, communication tools, or networks to find, evaluate, use, and create information. It also related to utilizing it in a healthy, wise, intelligent, accurate, precise, and law-abiding manner in the context of fostering communication in everyday life (Gee, 2003). While HOTS is an individual's ability to connect, manipulate, and change the knowledge and experience already possessed critically and creatively in making decisions to solve the problems at hand (Brookhart, 2010).

Technologies that are considered capable of fulfilling the needs of DLS and HOTS students is the Augmented Reality (AR) and Google Scholar (GS). AR is a smartphone-based media that can display computer graphics to the real world. Thus, AR allows users to see the natural world and the virtual world simultaneously (Billinghurst, 2002). One of the current AR development companies is ARLOOPA inc. The company created an application called ARLOOPA that can be used on both Android and IOS software. Of course, this is a distinct advantage for students and teachers. Meanwhile, GS is a Google service that allows users to search for scientific materials in text in published manuscripts. However, there are growing challenges among school teachers, especially in Madrasah Aliyah Miftahunnajah Sleman, regarding whether AR and GS can be used as effective technologies in developing HOTS and DLS -based test instruments.

Instrument test based on High Order Thinking Skills (HOTS) is a test instrument that serves to measure higher cognitive thinking skills, namely thinking skills that are not merely remembering (recall), restating, or referring without processing (reading). HOTS-based assessment in the context of the evaluation measures the ability to: transfer one concept to another, process and apply information, seek information from several different sources, use the information to solve problems and analyze ideas or information critically. Lots of research has been done on developing a test instrument to measure students' HOTS, but it has not met the characteristics of the digital era.

This research was conducted to develop a test instrument based on HOTS and DLS. The instrument test in question is an instrument that contains items designed based on the four HOTS components. It covers reasoning ability, argumentation ability, critical thinking ability, and metacognition ability (Schraw, 2011). In addition to being developed based on the HOTS component, the items were also developed based on the four DLS components, namely operational skills, formal skills, information skills and strategic skills (Lordache, 2017). In addition to measuring students' HOTS and DLS, it is hoped that the instrument test results from this development can be a reference for researchers to increase research on the development of assessment instruments that encourage students' HOTS and DLS (Djawad, 2018).

2. Method

The research method used is research and development. The developed product is based on the following development steps: information gathering, planning, initial product development, limited trial, initial product revision, field trial, and final product revision (Brog & Gall, 1983). The instrument to measure the validity is a validation sheet which will then be analyzed qualitatively. Validation of instrument carried out aimed to evaluate the validity of the instrument that is being developed. The instrument validation process was carried out at the initial development stage and was carried out by five mathematics education experts. After the instrument was deemed valid, the instrument was tested on a limited basis to 5 non-experimental students. At the same time, field trials were conducted on 11th-grade students of Madrasah Aliyah Miftahunnajah Sleman, totalling 17 students. Trial data analysis was carried out using classical test theory parameters to determine the quality of the instrument empirically as the basis for revising the questions. The study was carried out on the results of the expert validation sheet to see the uniformity of the value of the instrument validation level. Data in the form of the value of each item of expert assessment results were analyzed using the Aiken's V formula to calculate the content validity coefficient using the following formula (Aiken, 1985).

$$V = \frac{\sum S}{n(c-1)}$$

where:
$\sum S$ = Total difference between the score given by the expert and the minimum score.
n = Number of experts involved.
c = Number of score options that can be given.

The data from the analysis is then compared with the validity categorization that follows the class range according to Table 1.

Table 1. Expert Validity Category
(Koestoro, 2006)

Score Range	Category Validity	Remark
$0,8 - 1,00$	Very high	Proper to use
$0,6 - 0,799$	High	Proper to use
$0,4 - 0,599$	Enough	Proper to use
$0,2 - 0,399$	Low	Not worth using
$0 - 0,199$	Very low	Not worth using

Meanwhile, the data from the test results to students were analyzed using SPSS and Microsoft Excel. The analysis was conducted to determine the validity, reliability, level of difficulty and discriminating power of questions. First, the validity analysis process was carried out using the product-moment correlation of the person with a significance level. In contrast, the reliability analysis used the Alfa Cronbach statistical test with the reliability criterion (Yusup, 2018). Then the difficulty level analysis was carried out with the help of Microsoft Excel using the formula.

Using obtained the value of the difficulty level, then compared based on the categorization in Table 2.

Table 2. Difficulty Level Category

Difficulty Level	Category
Less than 0.3	Very Difficult
0.3-0.7	Medium
More than 0.7	Easy

With the help of the same application, then the discriminatory power analysis is carried out using the following formula.

$$Discriminatory\ power = \frac{BA \times JB - BB \times JA}{JA \times JB}$$

where:
BA = The Number of participants in the upper group who answered correctly.
BB = The Number of participants in the lower group who answered correctly.
JA = The Number of participants in the upper group.

JB = The Number of participants in the lower group.

After obtaining the value of distinguishing power, we then compared it based on the categorization in Table 3.

Table 3. Distinguishing Power Category
(Arikunto, 2013)

Distinguishing Power Level	Category
0.71-1.00	Very well
0.41-70	Good
0.21-0.40	Enough
0.00-0.20	Not good

Data from the test instrument to students in the form lists the value obtained by the student based on the scoring rubric Brookhart (2010) as Table 4.

Table 4. Instrument Scoring Rubric

Score	Information
2	The answers given are following the question's intent, and the calculation process/source of information is correct.
1	The answers given are following the question's intent, but there are still errors in the calculation process/source of information.
0	The answers given do not match the intent of the question or do not answer at all.

3. Result and Discussion

The initial process carried out in this study was the preparation of a test instrument product that contained six HOTS and DLS questions. Expert validation was carried out by SZN as mathematics teacher at Muhammadiyah Junior High School Ayah, Kebumen, ST as mathematics teacher at Muhammadiyah SMK Mungkid, Magelang, and TDD as mathematics teacher at SD Muhammadiyah Pakem, Sleman. Then to improve the quality of the test, the researchers involved two students of Master of Mathematics Education at Ahmad Dahlan University, Yogyakarta, namely NR and SY. Next, validation activities are carried out by giving the instrument test script to the validators. Then analyzed using Aiken's V formula and obtained a coefficient of 1.00. In other words, in general, the test instrument has question items that are very feasible to use.

Based on the results of the limited trial to 5 non-experimental students, it was found that the average time required by the four students to fill out the test instrument was 90 minutes. Then from the results of the assessment of student work, data obtained by rounding to three decimal places the results of the validity analysis using the SPSS application. Based on the Pearson correlation analysis results, the conclusion of the item validity test results is as Table 5.

Table 5. Data from the Validity Analysis of Limited Test Items

Question Items	Information
1a	Valid
1b	Invalid
2a	Valid
2b	Valid
3a	Valid
3b	Valid

Meanwhile, in the analysis of the reliability, the statistical test results were obtained Cronbach's Alpha is .810.

Based on the results of Cronbach's alpha analysis above, it was concluded that the test instrument was categorized as reliable. After a limited trial was carried out, then field trials were carried out. However, before the trial was carried out, item number 1b was corrected based on the suggestions of the limited trial participants. The advice given is to clarify the work order regarding building construction in AR, which will be calculated. After the repair of the item questions is completed, then a field test is carried out.

Field trials were conducted to determine the validity of the item items, the instrument's reliability, level of difficulty and the discriminating power of the questions. Field trials were carried out on 11th-grade students of Madrasah Aliyah Miftahunnajah Sleman, totalling 17 students. At the stage of working on the test instrument, students are given 90 minutes of processing time according to the average length of time for students in a limited trial. After the field trials, data on the scores achieved were obtained on each item of the question. An analysis of the validity of the item items was carried out using the Pearson correlation statistical test with the help of the SPSS application.

Based on the Pearson correlation analysis results above, the conclusion of the item validity test results is as Table 6.

Table 6. Data Item Validity Results Field Test Results

Question Items	Information
1a	Valid
1b	Valid
2a	Valid
2b	Valid
3a	Valid
3b	Valid

Meanwhile, at the stage of reliability analysis using the Cronbach Alpha statistical test, the results were .840. Based on the results of Cronbach's alpha analysis above, it can be concluded that the test instrument is categorized as reliable.

The next stage is to test the level of difficulty and discriminating power of the questions. The analysis process was carried out using the help of Microsoft Excel application. Determination of students in the upper and lower groups based on students' average daily test scores compared to the KKM (Criteria for Complete Learning) Compulsory Mathematics. From the results of the analysis obtained the following Table 7.

Table 7. Data on the Level of Difficulty and Distinguishing Power of Questions

Question Items	Difficulty Level	Category	Distinguishing Power	Category
1a	0.47	Medium	0.67	Good
1b	0.65	Medium	0.63	Good
2a	0.53	Medium	0.47	Good
2b	0.53	Medium	0.75	Good
3a	0.41	Medium	0.58	Good
3b	0.65	Medium	0.63	Good

It can be seen from the table above obtained 100% (all) of the items on the test instrument are in the medium category. This is not a problem in developing HOTS questions considering that HOTS questions do not mean difficult questions. Dr Abdur Rahman Asári explained that the items designed to encourage students to use HOTS in solving problems did not always have a high level of difficulty. Students do not only use recall, restate or recite throughout the problem-solving process, but must analyze, evaluate or create, then such question items are questions that encourage students' HOTS. On the other hand, regardless of the difficulty level of the question, if it is only solved by using recall, restate or recite, then the question is not a HOTS question (As'ari, 2019). Meanwhile, from the data, it was obtained

that 100% (all) of the items on the test instrument had good discriminating power.

Overall, the statistical instrument test results based on the data obtained from the field test scores can be seen in the following Table 8.

Table 8. Overall Statistical Test Results Table

Statistical Scale	Amount	Interpretation
Average sig.(2-tailed) correlation test	.002	Valid
Cronbach's Alpha Value	.840	Reliable
Average difficulty level	.74	Medium
Average discriminating power	.62	Good

Based on the data above, the average sig. (2-tailed) correlation test and Cronbach's alpha value were obtained, respectively, equal to 0.002 and 0.840. That is, the test instrument that has been made contains useful items. In addition, at the level of reliability, the instrument test created is categorized as reliable. Then, in analyzing the level of difficulty and discriminating power of the item items, the average level of difficulty and discriminating power was 0.74 and 0.62, respectively. This means that the items made on the test instrument generally have a moderate level of difficulty and good discriminating power. At the field trial stage, the researcher did not receive any suggestions for improvement regarding the test instrument's quality. Therefore, the test instrument used at the field trial stage was used as the final product of this development research.

On the other hand, from the results of the research that has been shown in the previous section, it also explicitly explains that AR can be used effectively and efficiently as a learning media to improve students' HOTS. The development of augmented reality-based learning media is very possible to do in order to improve students' critical thinking skills (Faridi, 2021). The use of GS as a media to improve students' DLS also needs to be tried by teachers in the learning process. DLS must begin to be integrated in every learning process in schools to improve the abilities of 21st century students (Perdana, 2019). Students are expected to be not only skilled in using technological tools, but also need to build the ability to capture information properly and have mutual respect in the interaction process directly or indirectly (Santos, 2017).

4. Conclusion

Based on the results of the research and discussion, the following conclusions were obtained. The final product of this research is a test instrument that aims to measure the HOTS and DLS of 11th-grade high school students. The test instrument obtained consisted of 6 items in the form of a description. Based on the instrument analysis, the average sig. (2-tailed) correlation test and Cronbach's alpha value were obtained, respectively, equal to 0.002 and 0.840. That is, the test instrument that has been made contains valid items. In addition, at the level of reliability, the instrument test created is categorized as reliable. Then, in analyzing the level of difficulty and discriminating power of the item items, the average level of difficulty and discriminating power was 0.74 and 0.62, respectively. This finding means that the items made on the test instrument generally have a moderate level of difficulty and good discriminating power.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement*, *45*(1), 131-142.

As'ari, A.R. Ali, Muhammad. Basri, Hasan., Kurniati, Dian. Maharani, Swasti. (2019). *Mengembangkan HOTS (High Order Thinking Skills) melalui Matematika*. Malang: Penerbit Universitas Negeri Malang.

Arikunto, Suharsimi. (2013). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Billinghurst, Mark. (2002). Augmented Reality in Education. *New horizons for learning 12*(5). 1-5.

Brookhart, S. M. (2010). *How to Assess High-er Order Thinking Skills in Your Class-room*. Alexandria: ASCD

Borg, W. R. & Gall, M.D. (1983). *Educational researcher: An introduction*, (7th ed.). United States: Pearson education, Inc

Chan, B. S., Churchill, D., & Chiu, T. K. (2017). Digital literacy learning in higher education through digital storytelling approach. *Journal of International Education Research (JIER), 13*(1), 1-16.

Djawad, Y., Suhaeb, S., & Jaya, H. (2018). *Innovation In Learning Through Digital Literacy at Vocational School of Health*. In International Conference on Indonesian

Technical Vocational Education and Association (APTEKINDO 2018). Atlantis Press. 239-245.

Faridi, H., Tuli, N., Mantri, A., Singh, G., & Gargrish, S. (2021). A framework utilizing augmented reality to improve critical thinking ability and learning gain of the students in Physics. *Computer Applications in Engineering Education, 29*(1), 258-273.

Gee, J. P. (2003). *What video games have to teach as about learning and literacy?* (1st ed.). New York: Palgrave Macmillan.

Schraw, G. (2011). *Assessment of Higher Order Thinking Skills*. Austin: The University of Texas.

Koestoro, B., & Basrowi, H. M. (2006). *Strategi Penelitian Sosial dan Pendidikan*. Surabaya: Yayasan Kampusina.

Perdana, R., Yani, R., Jumadi, J., & Rosana, D. (2019). Assessing students' digital literacy skill in senior high school Yogyakarta. *JPI (Jurnal Pendidikan Indonesia), 8*(2), 169-177.

Rochelle, Jeremy M., et al. (2000). *Changing how and what children learn in school with computer-based technologies.* The Future of Children. 76-101.

Santos, A. I., & Serpa, S. (2017). The importance of promoting digital literacy in higher education. Int'l J. Soc. Sci. Stud., 5, 90.

Yusup, F. (2018). Uji validitas dan reliabilitas instrumen penelitian kuantitatif. *Tarbiyah: Jurnal Ilmiah Kependidikan, 7*(1). 17-23.