*JPMIPA, 22 (1), 2021, 45-51*

**Jurnal Pendidikan MIPA**

e-ISSN: 2550-1313 | p-ISSN: 2087-9849

http://jurnal.fkip.unila.ac.id/index.php/jpp/

# The Development of Computer Based Test Instruments using Wondershare Quiz Creator to Measure HOTS in Temperature and Heat

Isna Istikomah*, Agus Suyatna , Viyanti, Doni Andra

Department of Physics Education, Universitas Lampung, Indonesia.

**Abstract:** This study aims to develop a computer-based test instrument using the valid Wondershare Quiz Creator (WQC) to measure HOTS in temperature and heat material. This study uses research and development research methods and the research design used refers to the research design by Borg & Gall. Data analysis was performed by means of internal validation tests by a team of experts. This test instrument has been tested by experts consisting of the material validation test with a score of 3.10 (valid), construction with a score of 2.88 (valid), and language with a score of 3.63 (very valid). The CBT setting on this product is also very valid with an average score of 3.68. The sig value of the difference in the level of validity and complexity in the types of multiple choice, multiple response, matching, and sequence items obtained was> 0.05, which indicates that there was no difference in the average validity and complexity between the four types of questions at each cognitive level.

**Keywords:** Test instruments, Wondershare Quiz Creator (WQC), HOTS.

***Abstrak:*** *Penelitian ini bertujuan mengembangkan instrumen tes berbasis komputer menggunakan Wondershare Quiz Creator (WQC) yang valid untuk mengukur HOTS pada materi suhu dan kalor. Penelitian ini menggunakan metode penelitian research and development dan desain penelitian yang digunakan mengacu pada desain penelitian oleh Borg & Gall. Analisis data dilakukan dengan uji validasi internal oleh tim ahli. Instrumen tes ini telah diuji ahli yang terdiri dari uji validasi materi dengan skor 3,10 (valid), konstruksi dengan skor 2,88 (valid), dan bahasa dengan skor 3,63 (sangat valid). Setting CBT pada produk ini juga sudah sangat valid dengan perolehan rata-rata skor 3,68. Nilai sig perbedaan tingkat validitas dan kompleksitas pada tipe soal multiple choice, multiple response, matching, dan sequence yang diperoleh >0,05 yang menandakan bahwa tidak terdapat perbedaaan rata-rata validitas dan kompleksitas antar keempat tipe soal tersebut pada tiap level kognitif.*

***Kata kunci:*** *Instrumen tes, Wondershare Quiz Creator (WQC), HOTS.*

## ▪ INTRODUCTION

21st century education is formed due to the strong influence of the development of Techno-Science, which is triggered by advances in the world of computers, information and communication technology, and the internet (BSNP, 2010: 32). Technology allows individuals to gain access to information (real-time data), provides simulations about an object as it is (real world), and opportunities to connect with various learning objects according to interests (Anggraeni and Sole, 2018). Meeting the demands of today's world that requires a change in the assessment strategy to measure skills that are now valued in a complex global environment (Partnership, 2009). In accordance with the demands of the 21st century, students are required to be able to go through several stages of complex thinking or high-order thinking skills in dealing with various problems both in school and

Isna Istikomah et al.
Corresponding Email:
isnaistikomah13@gmail.com

DOI: http://dx.doi.org/10.23960/jpmipa/v22i1.pp45-51
Received: 12 March 2021
Accepted: 22 April 2021

social life, as well as information and communication technology, so that students are able to compete with the outside world and all the demands of the era to come.

According to Heong, et al (2011) higher order thinking is defined as the expansion of the use of the mind to meet new challenges. High-order thinking requires a person to apply new information or knowledge that he has and manipulate information to reach possible answers in new situations (Kusuma, et al. 2017). Ariyana, et al (2018: 5) state that this skill is also used to underline various high-level processes according to Bloom's taxonomic ladder. According to Bloom, skills are divided into two parts. The first is low-level skills that are important in the learning process, namely remembering, understanding, and applying, and the second is classified into higher-order thinking skills in the form of analyzing, evaluating, and analyzing skills. , and create (create).

At the SMA / MA level, physics is considered important to be taught as a separate subject with several considerations. First, in addition to providing knowledge to students, the subject of Physics is intended as a vehicle for fostering thinking skills that are useful for solving problems in everyday life (BSNP, 2006). One of the physics concepts that is considered to be quite closely related to student life and students often experience misconceptions, namely the concept of temperature and heat (Silung, et al., 2016). Not only students but scientists also have difficulty applying their scientific knowledge related to heat and temperature to everyday situations (Alwan, 2011).

The 2013 curriculum currently used strongly emphasizes educators to have skills in developing HOTS (Higher Order Thinking Skill) assessment instruments, which is an evaluation tool capable of training students' creative and critical thinking processes (Khaldun, et al., 2019). According to the Directorate of High School Development (2017), the assessment for SMA should more assess higher-order thinking skills or Higher Order Thinking Skills (HOTS), which is a form of questions that has a level of thinking, analyzing, evaluating, and creating. Along with the development of sophisticated technology, assessments can be carried out using computer-based assessments (Haq, 2019), however, currently the practice questions given to students still use paper and they rarely use computer facilities to work on practice questions or evaluations. Khaldun, et al., 2019). Meeting the demands of the times, a computer-based evaluation or assessment tool or CBT (Computer Based Test) is needed, one of which is using the Wondershare Quiz Creator (WQC) application.

WQC is a software application that can be used to create various types and difficulty levels of questions in flash format (Dafitri, 2017). In this program (WQC) various kinds of stimuli can be inserted in the form of images, graphics, and appropriate videos so that it can increase students' understanding and interest in working on questions (Khaldun et al., 2019). Based on the problems described above, one effort that can be made to overcome these problems is to develop a computer-based test instrument using (WQC) to measure HOTS on temperature and heat material.

## ▪ METHOD

This research is a research development (Research and Development). The research design in product development uses a development model according to Borg & Gall (1983) which consists of 10 development steps. In this study, researchers only used 6 steps, which consisted of: (1) research and information gathering, (2) planning, (3) product development, (4) limited trials, (5) product revision, and (6) the final product. The development carried out is the creation of HOTS questions on temperature and heat parameters with 4 different types of questions, namely multiple chice, multiple responses,

matching, and sequences. In this development an expert test was applied. The expert validation test was carried out by a team of experts, namely three lecturers and two physics teachers. Expert testing is carried out to determine the feasibility level of the product that has been developed, namely in the form of material aspects, construction, language, the complexity of the stages of thinking, and the CBT setting.

The data obtained in this study are quantitative data in the form of a questionnaire instrument given during the validation and practicality tests. The validation test questionnaire instrument used has several aspects that need to be assessed, namely regarding the material, construction, language, types of questions, and CBT settings. The data collection technique was carried out by using a questionnaire technique. The questionnaire used was an expert validation test instrument. This questionnaire test instrument is used to collect and assess the appropriateness of the product developed as a learning evaluation.

The validation test is carried out by an expert test assessment using a questionnaire. The assessment questionnaire used has five rating scales based on the question indicators, starting from the scale or the highest score, namely “Very Fit”, “Appropriate”, “Sufficiently Suitable”, “Not Suitable”, and “Not Suitable”. Each score represents a different answer based on the validator's assessment or opinion on the suitability of the material, construct, language, type of question, and CBT setting.

Obtaining the results of the validation of the test instrument from the validator's assessment, then categorized according to the evaluation result criteria, namely the acquisition of an average value of 1.01-1.75 with the validity criteria "invalid", the average value of 1.76-2.50 with the validity criteria "quite valid", average value -the average 2.51-3.25 with the validity criteria "valid", and the acquisition of an average value of 3.26-4.00 with the validity criteria that is "very valid".

## ▪ RESULT AND DISCUSSION

Computer-Based Test Instruments Using WQC to Measure HOTS in Material Temperature and Heat has been completed by researchers. Research on the development of this test instrument was carried out by collecting information, product planning, product development, and validation testing. The objectives to be achieved in this product development research are to develop a computer-based test instrument using a valid WQC and to describe the setting of a computer-based test instrument using the appropriate WQC to measure HOTS on temperature and heat material. The development of a test instrument that was developed was in the form of computer-based HOTS questions which were presented with a more attractive appearance due to stimuli related to the material and questions, so that they could attract the attention of students or students in working on the questions. The product developed by the researcher is expected to help students to be more motivated in thinking HOTS and more interested in working on the questions with the various stimuli used in WQC, as well as helping students in making assessments or correcting students' test answers more efficiently and effectively .

After testing the material validation, construction, and language of the results of the instrument development, this test was declared valid with the average score on the material aspect was 3.10 and categorized as valid. The valid question material for measuring HOTS on temperature and heat material is that it has a level of conformity to the questions with indicators, measures cognitive levels C4, C5, and C6, uses questions that are accompanied by stimuli (images, graphics, videos, discourses, and tables) which function to stimulate HOTS, has a dimension of knowledge with appropriate indicators

and questions, and has a homogeneous and logical choice of answers. This agrees with (Silaholo, Rosidin, and Suyatna, 2020) which states that the HOTS assessment instrument model is able to enrich the level of thinking (C1, C2, C3, C4, C5, C6) which will be measured, enriching the competency achievement indicators (GPA) measured. , and reduce verbalism (too long questions).

The average validation score in the construction aspect is 2.88 with the valid category. The construction of a valid question for measuring HOTS in this temperature and heat material is to use a short, clear, and firm stimulus. Stimulus formulations and answer choices are statements that are needed only, the stimulus used does not provide clues to the answer key, is free from multiple statements, and the information provided by the stimulus must be clear and function properly.

**Table 2.** Validation results of CBT settings for HOTS questions on temperature and heat materials

| Arrangement Aspect | Average Score | Quality Statement |
|---|---|---|
| Time Limit | 3.68 | Very Valid |
| Random | 3.68 | Very Valid |
| Answer Submition | 3.68 | Very Valid |
| Feed Back | 3.68 | Very Valid |
| Scoring | 3.68 | Very Valid |
| **Amount Average** | **3.68** | **Very Valid** |

Validation of the CBT setting has five points of assessment aspects, including time setting, random setting (randomization) of questions, setting answer submissions, setting feed back (feedback), and setting scoring. The average validation score on time setting is 3.68 which indicates that the time setting given to the questions is in accordance with the very valid category. This is in accordance with the opinion (Pranata, Suyatna, and Rosidin, 2020) which states that the appropriate setting for measuring the CBT-based HOTS assessment is the time setting that displays the remaining time for the test.

The average validation score on the random setting (randomization) of questions and answer submissions has the same average score of 3.68 with the very valid category. Questions made with a random system aim to increase the effectiveness of the questions, reduce the level of student cheating with each other when working on questions, so that they can help relieve the exam supervisor team during the supervision process. This agrees with (Gunawan & Prabowo, 2017) that randomizing the questions carried out will make the questions that students work on are not the same as others (random) so that the test-taking process is more conducive and able to prevent cheating. The results of the validation to determine the differences in the level of validity, complexity, and cognitive thinking levels on this test instrument are as follows.

**Table 3.** Data on differences in the levels of validity, complexity, and levels of cognitive thinking

| No. | Type of Question | Thinking Level | Average Validity | Average Complexity |
|---|---|---|---|---|
| 1 | *Multiple Choice* | C4 | 4.06 | 4.40 |
| 2 | | C5 | 4.07 | 4.40 |
| 3 | | C6 | 4.07 | 4.40 |

| | | | | |
|---|---|---|---|---|
| 4 | *Multiple Response* | C4 | 4.05 | 4.20 |
| 5 | | C5 | 4.19 | 4.60 |
| 6 | | C6 | 4.21 | 4.60 |
| 7 | *Matching* | C4 | 4.20 | 4.20 |
| 8 | | C5 | 4.19 | 4.60 |
| 9 | | C6 | 3.99 | 4.60 |
| 10 | *Sequence* | C4 | 4.01 | 4.40 |
| 11 | | C5 | 3.98 | 4.20 |
| 12 | | C6 | 3.95 | 4.20 |

Based on Table 3, to look for differences in the level of validity and complexity between types of questions at each level of cognitive thinking tested using SPSS using One Way Anova. The results of One Way Anova using SPSS to measure differences in the level of validity and complexity can be seen in Table 4 below.

**Table 4.** Results of homogenity of variences

| Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|
| 1.925 | 2 | 9 | .201 |

Based on Table 4, it is known that the output results from the Test of Homogenity of Variences obtained a significance value of 0.201. The significance value obtained is more than 0.05, so that H0 is accepted, which means that the sample or test result has a homogeneous variance.

**Table 5.** One Way Anova Results

| **Score** | **Sum of Squares** | **Df** | **Mean Square** | **F** | **Sig.** |
|---|---|---|---|---|---|
| Between Groups | .025 | 2 | .012 | 1.549 | .264 |
| Within Groups | .072 | 9 | .008 | | |
| Total | .097 | 11 | | | |

Based on Table 5, it is known that the output of One Way Anova obtained a significance value of 0.264. The significance value obtained is more than 0.05, then H0 is accepted, which means H1 is rejected. These results indicate that there is no difference in the average validity and complexity between the multiple choice, multiple item types response, matching, and sequence at each cognitive level.

The highest mean value of validity and complexity at the C4 thinking level is the multiple response and matching question types. This indicates that the four types of questions at the C4 cognitive level are equally valid, but multiple responses and matching are more valid for use than multiple choices and sequences at the C4 cognitive level. The multiple response question type had the highest mean value of validity and complexity at the cognitive level C5. This means that the four types of questions at the cognitive level C5 are equally valid, but the multiple responses are more valid to use than the other three types of questions. At the cognitive level C6, the four types of questions were both valid and complex, but the level of validity was higher using the multiple choice, multiple response, and matching type, while the sequences had lower levels of validity and complexity to measure HOTS at the C6 cognitive thinking level.

## ▪ CONCLUSION

Has produced a computer-based test instrument using WQC to measure HOTS on temperature and heat material with aspects of the material, construction, and language tested through a validation questionnaire and has been declared valid. CBT settings for computer-based test instruments using WQC to measure HOTS on the material temperature and heat developed have been tested through a validation questionnaire and have been declared valid. The appropriate CBT settings for this question instrument are by setting the time setting, activating random (randomization) questions, providing feedback on each question, and giving a score according to the weight of the questions.

There is no difference in the average validity and complexity between multiple choice, multiple response, matching, and sequence types at each level of cognitive thinking and these four types can be used on computer-based test instruments using WQC to measure HOTS on temperature and heat material.

## ▪ REFERENCES

Alwan, A. A. (2011). Misconception of heat and temperature among physics students. *Procedia-Social and Behavioral Sciences*, *12*, 600-614.

Anggraeni, D. M., & Sole, F. B. (2018). E-Learning Moodle, Media Pembelajaran Fisika Abad 21. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, *1*(2), 57-65.

Arikunto. (2012). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.

Ariyana, Y., Pudjiastuti, A., Bestary, R., & Zamroni. (2018). *Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berpikir Tingkat Tinggi.* Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.

BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. Jakarta: BSNP Depdiknas.

BSNP. (2010). *Laporan BSNP*. Jakarta: BSNP Depdiknas.

Dafitri, H. (2017). Pemanfaatan Wondershare Quiz Creator Dalam Tes Berbasis Komputer. *QUERY: Jurnal Sistem Informasi.* 2579-5341.

Gunawan, G., & Prabowo, D. A. (2017). Sistem ujian online seleksi penerimaan mahasiswa baru dengan pengacakan soal menggunakan linear congruent method(Studi kasus di Universitas Muhammadiyah Bengkulu). Jurnal Informatika Upgris, 3(2).

Haq, M. A. (2019). *Pengembangan Instrumen Penilaian Ulangan Harian Menggunakan Wondershare Quiz Creator pada Materi Statistika.* Surabaya: Universitas Islam Negeri Sunan Ampel Surabaya.

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. B., & Mohamad, M. M. B. (2011). The level of marzano higher order thinking skills among technical education students. *International Journal of Social Science and Humanity*, *1*(2), 121.

Kadir. (2015). Statistika Terapan. Jakarta: PT Raja Grafindo. 364 hlm.

Khaldun, I., Hanum, L., & Utami, S. D. (2019). Pengembangan soal kimia higher order thinking skills berbasis komputer dengan wondershare quiz creator materi hidrolisis garam dan larutan penyangga. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education), 7*(2), 132-142.

Kusuma, M. D., Rosidin, U., Abdurrahman, A., & Suyatna, A. (2017). The development of Higher Order Thinking Skill (Hots) instrument assessment in physics

study. *IOSR Journal of Research & Method in Education (IOSR-JRME)*, *7*(1), 26-32.

Partnership for 21st Century Skills. (2009). *P21 framework definitions. Washington, DC.*

Pranata, B., Suyatna, A., & Rosidin, U. (2020). Pengembangan Asesmen Higher Order Thinking Skills (HOTS) Berbasis Computer Based Test (CBT) pada Materi Induksi Elektromagnetik. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, *9*(2), 83-98.

Pratiwi, D. E., Suyatna, A., & Viyanti, V. (2020, February). Design of Computer-Based Testing for Higher-order Thinking Skills on Static Fluid Material. In *Journal of Physics: Conference Series*, 1467(1). IOP Publishing.

Sihaloho, Y. E. M., Rosidin, U., & Suyatna, A. (2020). CBT HOTS assessment instrument model: teacher perception analysis of a role of illustration as well as quiz and question setting according to HOTS. *Jurnal Pembelajaran Fisika*, *8*(1), 9-17.

Kusairi, S., & Zulaikah, S. (2017). Diagnosis miskonsepsi siswa SMA di Kota Malang pada konsep suhu dan kalor menggunakan three tier test. *Jurnal Pendidikan Fisika dan Teknologi*, *2*(3), 95-105.