JTK: Jurnal Tadris Kimiya 7, 1 (June 2022): 1-13
Website: http://journal.uinsgd.ac.id/index.php/tadris-kimiya/index
ISSN 2527-9637 **(online) ISSN** 2527-6816 **(print)**

The Making of Evaluation Instrument Based on HOTS with Wondershare Quiz Creator on Ion Balance and Buffer Solution pH

Asmadi M. Noer[1], Puji Pebrianti[1], Betty Holiwarni[1], and Sunarti[2]
[1]Chemistry Education Department, FKIP, Universitas Riau, JL.H. R. Soebrantas KM 12,5 Panam, Pekanbaru, 28293, Indonesia
[2]SMA Negeri 5 Pekanbaru, Jl. Bawal No. 43, Kel. Wonorejo-Marpoyan damai, Pekanbaru
**E- mail : asmadi.m@lecturer.unri.ac.id*

Received: 15 May 2022; Accepted: 13 June 2022; Published: 30 June 2022

Abstract

The ability of students to solve problems is still low; therefore, a Higher Order Thinking Skills (HOTS)-based evaluation instrument through media is needed. This study aims to produce an evaluation instrument based on HOTS using Wondershare Quiz Creator (WQC) implemented to a valid ionic balance and pH buffer solution subject according to material and media experts. This research uses research and development (R&D) with the Plomp model. Besides, the research subjects are three teachers and 30 students in 11th grade from SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru. Literature studies and field studies were carried out as the data collection techniques. The research data is obtained from the validation results of two material validators, two media validators, and user trials. The validation result to material validators has obtained the average based on the material aspect 96.00%, construction aspect 93.10%, and language aspect 99.00%. According to media validators, based on the content aspect 90.00%, learning design 84.00%, display (visual communication) 85.00%, and software utilization 95.71%. The analysis of the questions showed that ten questions were valid, had very high reliability, a medium difficulty level of five questions, and a difficult category of five questions and had good distinguishing power. User response scores are 88.00% by teachers and 95.23% by students with very good criteria.

Keywords: HOTS, ionic balance and pH buffer solution, wondershare quiz creator

DOI: https://doi.org/10.15575/jtk.v7i1.14392

1. Introduction

Learning is an interactive relationship between teachers and students in an educational institution that aims to achieve an expected result (Pane & Darwis, 2017). In the learning process, the teacher plays an essential role in facilitating students in learning activities, one of which is providing an evaluation instrument to students.

Besides, students will get various sciences in senior high school (*Sekolah Menengah Atas/SMA*), one of them is chemistry. Learning chemistry is not only calculating the content of substances stoichiometry but also requires skills and a thought process to understand abstract concepts such as the material on ion balance and the pH of buffer solutions. Therefore, to make students' HOTS skills in accepting the ion balance and pH of the buffer solution material, they need the design and presentation/product of the Wondershare media with additional concept illustrations, sketches, and animation related to the material. Hence, HOTS abilities have become the demands of 21st century learning. HOTS is the ability to connect and change one's knowledge critically and creatively when solving a problem in new situations (Faruq & Huda 2020). Dinni (2018) states that the HOTS indicators are divided into three indicators: critical thinking skills, problem-solving, and creative thinking skills. On the other hand, HOTS is essential to be trained by educators in a learning activity (Ahmad & Sukiman, 2019).

In addition, Kempirmase et al. (2019) stated that students' HOTS could be improved through HOTS-based evaluation instrument exercises. Thus, in a classroom learning process, such as during daily tests, mid-semester examinations, and semester final examinations, a teacher needs to be accustomed to giving HOTS-level questions to students. Awareness of the importance of making and taking HOTS-level evaluation questions is also getting higher after seeing the test results of Indonesian students in several international standard test trials such as the Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS). Meanwhile, Indonesia has participated in PISA from 2001 to 2018, and the result was that science skills in Indonesia are still relatively low and Indonesia only occupies position 70 of the 78 countries evaluated (Schleicher, 2019). These low-skill require the world of education in Indonesia to prepare for the rapid development of 21st century knowledge and skills, such as equipping students with HOTS-based evaluation instruments. Meantime, Indonesian students are only good at working on questions whose level is only rote, but the concept of analyzing is still low (Laili et al, 2020).

One of the supporting factors is that Indonesian students' scientific is still relatively low, namely the lack of students' skills in HOTS. In addition, from the interview with three chemistry teachers in September 2020, the teachers admitted that the implementation of the evaluation on the material of ion balance and pH of the buffer solution was still taking questions from textbooks that are still categorized as understanding (C2), applying (C3), and analyzing (C4). As a result, the teachers are not yet skilled in making HOTS-based evaluations. Besides, the teacher should stimulate HOTS in teaching chemistry and alternative assessment offerings so that students are expected to trigger an understanding of cognitive processes and get used to high-level thinking skills. Moreover, the evaluation instrument product is based on Wondershare Quiz Creator (WQC); it is hoped that teachers will be helped in their duties while optimizing the HOTS assessment of students.

According to cognitive learning theory, a person learns through information/knowledge processing which involves mental processes of thinking, solving problems, and making decisions. Therefore, it is necessary to make a HOTS-based evaluation instrument that could be used to train students' HOTS skills.

Furthermore, the Covid-19 has impacted several aspects of life, especially in education. The current student evaluation process is carried out at home using the online method. The online learning process is an alternative solution to overcome the current limitations of face to face learning (Herliandry et al., 2020). In addition, the development of information and communications technology (ICT) has had a vital influence on learning evaluation. The implementation of the current learning evaluation is no longer conducted with printed sheets because a paper has several weaknesses: students are less interested and motivated in working on questions, some students cheat, the class atmosphere becomes crowded, students do not immediately collect their answers, and the assessment process takes a long time (Febriani et al., 2021).

This problem shows that implementing learning evaluation with printed sheets is less effective and efficient. However, the application of learning evaluation that utilizes technology can make it easier for teachers to conduct online evaluations. The evaluation could be arranged at a specific time. Also, students will get the description of information regarding the learning evaluation. Eventually, the evaluation results could be seen immediately, and it will be helpful for teachers to make corrections (Meryansumayeka et al., 2018).

Some online platforms can be used for evaluation, including Edmodo, Socrative, Classmaker, Kahoot, WQC, and others. The Kahoot application limits the number of characters in making questions and responses, and an Overhead Projector must be available during the learning process, which is a drawback (Plump & LaRosa, 2017). On the other side, the Classmaker application requires an English-language programming language to operate, which will make it difficult for teachers to provide online evaluations to students (Nardjosoeripto et al., 2017). In the end, the teacher could use a WQC to solve that problem because it is easier to operate.

A WQC is software for making online questions without the need for programming language skills to operate. WQC will make it easier for teachers to change the rules for randomizing questions. Of course, this is important to reduce the cheating of students who exchange answers and is helpful for teachers in assessing students (Meryansumayeka et al., 2018). Besides, Satriana (2019) states that the advantage of the WQC software is that educators could enter pictures and videos related to the material. Furthermore, it could help in online learning today, where the current learning process has limitations to doing a practicum.

The HOTS-based evaluation instrument with WQC as a display media presents an evaluation instrument that could improve students' ability in HOTS. The first was conducted by Sa’adah et al. (2019) with the title "Development of HOTS Questions with WQC as a Display Media on Stoichiometry Material for Class X SMA/MA Equivalent." The result showed that the WQC application on stoichiometry material is feasible to use with validation by the material validator based on the material aspect is 95.34% (very valid), construction is 95.20% (very valid), and language is 93.14% (very valid). The result showed that the substance of the material is 84.00% (valid), learning design 86.00% (very valid), visual communication 84.00% (valid), software utilization 88.57% (very valid), and user response trials that conducted by teachers and students showed a percentage of 90.67% and 91.20% with excellent criteria. Moreover, Sa'adah et al. (2019) used a media validator to validate the data. The second is Iqbal et al. (2018), entitled "Development of an Evaluation Tool Based on WQC on Colloidal Material Class XI Science at SMA Koperasi Pontianak" the results obtained are the validity of the assessment of material experts with a percentage of 83.98% and an assessment of media experts with a percentage of 90,69%. Based on the assessment of material and media experts, it can be concluded that the evaluation tool belongs to the perfect category.

Different from those previous researches, this study aims to develop a HOTS-based evaluation instrument based on the WQC as a display media on ion balance material and pH buffer solution for 11th grade and equivalent levels valid and reliable.

2. Research Method

This research was conducted at the Chemical Education Study Program, FKIP Riau University, Pekanbaru, with trials at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru, and the research time starting from September to May 2021. The development of HOTS-based evaluation instruments with WQC as a display medium for the ionic balance material and the pH of the buffer solution for 11th grade from *SMA/MA* equivalent, designed using a research and development design based on Research & Development (R&D) by the Plomp development model which consists of five phases, which are the preliminary investigation phase, the design phase (design), realization/construction phase, validation phase, trial and revision (evaluation, test, and revision), and implementation phase (Arianti & Hardiyanto, 2018). The type of data is needed in the form of numbers from validation results and results of limited trials that are described qualitatively. The HOTS-based evaluation instrument with the WQC as a display media developed by the researcher was assessed by four validators to test its validity, consisting of two material validators and two media validators. The validation process incorporates many times, with a validator assessment at the end of the validation process. This way aims to obtain valid results based on material and media aspects.

Data analysis is the validity analysis of the material, construction, and language aspects. In contrast, media validity is based on content substance, learning design, display (visual communication), and the use of software obtained from the assessment by the validator. The type of Likert scale is used the Likert scale with a score of 1 to 5, then calculate the percentage score with the equation 1 (Eq.1).

$$P = \frac{n}{N} \times 100\% \quad (1)$$

Description:
P = Percentage score (%)
n = Total score obtained
N = Total maximum score

The percentage score is converted into a qualitative value that links to the measurement scale assessment criteria, see Table 1.

Table 1. Validity Criteria

Percentage	Description
80.00 – 100	Very Good/ Very Valid/ Very Proper
60.00 – 79.99	Good/ Valid/ Proper
40.00 – 59.99	Pretty Good/ Pretty Valid/ Pretty Proper
20.00 – 39.99	Not Good/ Not Valid/ Not Proper
0.00 – 19.99	Poor/ Invalid/ Improper

(Budiastuti & Bandur, 2018)

Four validators declared that the HOTS-based evaluation instrument with the WQC on the ion balance material and the pH of the buffer solution is valid. Further, a one-at-a-time trial was conducted on three students with high, medium, and low abilities at SMAN 2 Pekanbaru. Furthermore, a limited trial was followed by 30 students at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru to get the value of validity, reliability, level of difficulty, and distinguishing power.

A validity test is used to see a measuring instrument's accuracy in carrying out its measuring function. Arikunto (2018) clarifies that the validity test formula shown by equation 2 (Eq. 2).

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2}} \quad (2)$$

Description:
r_{xy} = Correlation coefficient sought
N = Number of respondents
X = Value of variable X (item score)
Y = Value of variable Y (item score)
$\sum X$ = sum of x scores
$\sum Y$ = sum of y scores
$\sum X^2$ = sum of squared x scores
$\sum Y^2$ = sum of squared y scores
$\sum XY$ = sum of the product paired source

The reliability test was carried out by using the Spearman-Brown formula as follows (Eq.3).

$$ri = \frac{2rb}{1+rb} \quad (3)$$

Description:
ri = Internal reliability of all instruments
rb = Product moment correlation between the first and second halves

Tabel 2. Reliability Criteria

Reliability	Criteria
0.81– 1.00	Very high
0.61– 0.80	High
0.41– 0.60	Enough
0.21– 0.40	Low
0.01– 0.20	Very low

(Arikunto, 2018)

An analysis of the difficulty level is used to determine the level of difficulty of the items. The formula used is as follows (Eq. 4).

$$P = \frac{B}{JS} \quad (4)$$

Description:
P = difficulty index
B = the number of students who answer correctly
JS = total students (tested)

Tabel 3. **The Characteristic of The Difficulty Index**

Difficulty Index	Criteria
0.00 – 0.30	Difficult
0.31 – 0.70	Enough
0.71 – 1.00	Easy

(Arikunto, 2018)

Distinguishing power is defined as the ability of an item to distinguish between high-ability and low-ability students. The distinguishing power of the questions was analyzed by using the formula (Eq. 5).

$$DP = \frac{BA}{JA} - \frac{BB}{JB} \quad (5)$$

Description:
DP = distinguishing power
J = total of participants
JA = total of upper group
JB = total of lower group
BA = total of upper group who answer correctly
BB = total of lower group who answer correctly

The analysis of validity, reliability, level of difficulty, and distinguishing power were also tested by using the Anates V4 software to analyze items and prove the criteria for testing validity, reliability, level of difficulty, and distinguishing power.

The trial was limited to teachers, and students were given a user response questionnaire arranged based on a Likert scale with the criteria as shown in Table 4.

Tabel 4. Classification of teachers and Student's Response

Percentage (%)	Classification of Students' Achievement
80.00 -100	Very good
60.00 – 79.99	Good
40.00 – 59.99	Enough
20.00 – 39.99	Poor
0.00 -19.99	Fail

(Arikunto, 2018)

3. Result and Discussion

This development research results in a HOST-based design evaluation with a WQC as a display medium for the ion balance material and the pH of the buffer solution for 11th grade from *SMA/MA* of the same level.

3.1. Prelimenary Investigation

In the initial investigation phase, various analyses have been carried out: preliminary research analysis, student needs, competencies, and materials by analyzing related information. The front-end analysis has been structured through interviews at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru through three chemistry teachers and the distribution of student questionnaire data.

An analysis of students has been carried out by digging up information on relevant literature studies. Students who are used as practitioners have studied the material for ion balance and the pH of buffer solutions. The competency analysis was implemented through a study of attitudes, knowledge, and skills competence on the senior high school chemistry subjects syllabus. Additionally, the information was obtained that the material for ion balance and pH of buffer solutions was in the essential competencies of 3.12 and 4.12 (Noviyanti et al., 2020). Material analysis is information related to the material selected for the development of the HOTS-based evaluation instrument, which is the material for ion balance and the pH of the buffer solution.

Table 5. Pre-Research Findings at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru and MAN 1 Pekanbaru

Activity	Result
An interview with three chemistry teachers.	1. The teachers have participated in training on making HOTS questions, but they gave the daily tests, midterm exams, and final exams at a few HOTS levels. 2. Unfinished learning indicators on the solution material and the pH of the buffer solution, which is the indicator for calculating the pH of a buffer solution 3. Applications in providing evaluations to students that teachers use during online learning have not fully utilized technology-based applications that can increase students' motivation to do evaluations.
Spread the questionnaire to SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru	1. Questionnaire data from three schools shows that 70% of students strongly agree that the material for ion balance and pH of a buffer solution requires students' HOTS skills. 2. As many as 85% of students have not known the HOTS skills problem well.

3.2. Design Phase

The initial design of the HOTS question instrument was based on the ion balance material and the pH of the buffer solution. A comparison of questions made by the teacher with HOTS questions designed by researchers in detail can be seen in Table 6.

Table 6. The Comparison of Questions Made by Chemistry Teachers at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru with HOTS Questions Designed by Researchers

Learning Indicators	The Questions Made by Chemistry Teachers		The HOTS Questions Desgined by Researchers	
	C2	C3	C4	C5
Observe the pH of the buffer solution after it is diluted, and add a small acid or base	√	-	√	-
Designing probation to make buffer solutions with a specific pH		A buffer solution with pH 5 - log 6 can be made with a mix of 100 mL of solution C_6H_5COOH 0.2 M and 100 mL of solution NaOH 0.1 M, then the value of K_a C_6H_5COOH is....	A student mixed sodium solution nicotinic with nicotinic acid. Nicotinic acid (vitamin B3) can improve brain performance and protect skin from sunlight by setting acid ionization (K_a = 1.4 x 10^{-5}). If they want to make a solution that supports pH= 4, how can they make the solution?	
Determining pH buffer solution		200 mL of 0.4 M NH_4OH solution (K_b = 10^{-5}) and 200 mL of 0.2 M HCl are mixed, then pH has a big as?	If the strong acid is reacted with Aniline (C6H5NH2) will shape its acid conjugate and anilinium ion with K_b= 4.0 x 10^{-10}. Then 5.58 grams of aniline dissolved in water up to a volume of 600 mL solution, so ionized as much as 2%. If into 300 mL the solution added with 200 mL the HCl solution pH is 2, then pH mixed solution is... (A_r C = 12, H = 1, N = 14)	
Discuss the role of buffer solutions in the body of living things and industry.	√	-	√	-

In the next section, the researchers compose the answer key. The answer key is made after the questions are completed to make it easier for researchers or teachers to assess the results of the HOTS-based evaluation on ion balance material and the pH of the buffer solution. And then, the researchers compile validation sheets and material validation sheets that have been arranged based on material, construction, and language aspects. In contrast, media validation sheets are arranged based on content substance, learning design, software display and utilization, material validation sheet rubrics, and media validation sheet rubrics. Next, the researchers make the initial design of the HOTS-based evaluation instrument display media with the WQC, which includes filling in the identity, a complete cover page, the WQC user guide page, the content page of the HOTS question, and the results page.

3.3. Realization and Construction

The realization/construction phase is carried out by realizing the design of the HOTS-based evaluation instrument and the prototype. The realization was produced: (1) the HOTS-based evaluation instrument includes the initial setup of the HOTS-based evaluation instrument grid and the HOTS-based evaluation instrument on the ionic balance material and pH of the buffer solution, and (2) the result of the realization of the prototype design is a HOTS-based evaluation instrument with a WQC as a display media on the ion balance material and the pH of the buffer solution, which follows the analysis results based on the characteristics of the students.

As a result, this phase obtained the difficulty of the HOTS questions.

3.4 Evaluation, Test and Revision Phase

This phase results in validation, testing, and revision. The validation activities were carried out by four validators consisting of two lecturers of the Chemical Education Study Program at the Riau University as material validators. On the other hand, two other validators came from a lecturer in the Department of Informatics, Syarif Kasim State Islamic University, Riau, as media validators. The material validator focuses on the material, construction, and language aspects. Meanwhile, media validators have a role in assessing aspects of content substance, learning design, display (visual communication), and software utilization. Therefore, the assessment produces quantitative and qualitative data. Quantitative data contains numbers for which the average value will be calculated. Furthermore, qualitative data is a validator's way of assessing improvement.

On the other hand, suggestions and improvements from the material validator which are: (1) adding an explanation of the benefits of buffer solutions in everyday life; (2) increase question indicators; (3) improve question editors, (4) improve numbers according to questions, and (5) make some questions to improve students' HOTS in analyzing. Validation of material, construction, and language aspects can be seen in Figure 1.

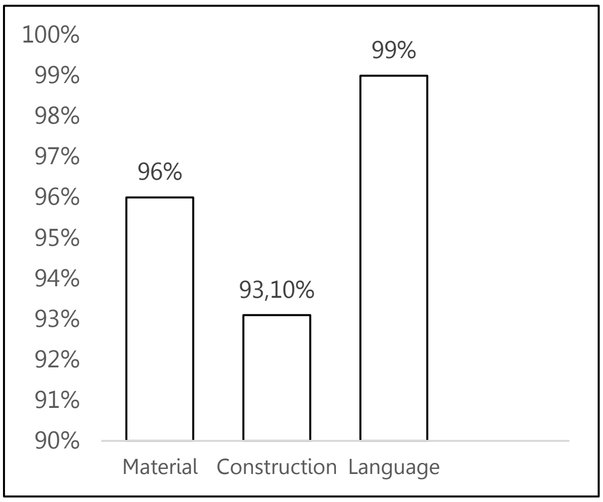

Figure 1. The Final Result of Two Validators Assesment on the HOTS Questions.

The HOTS-based evaluation instrument on the ionic equilibrium materials and the pH of the buffer solutions was initially developed as many as 15 items, but after the close of the validation stage; then, it was suggested that only ten items were valid. Last but not least, those ten items could be categorized as HOTS questions, which are C4 and C5. Besides, learning indicators analyzing buffer solution pH have been made in questions number 1,3,6 and 8. Meanwhile, learning indicators conducting experiments to make buffer solutions with a specific pH have been made in questions number 2,5,7, and 9. Furthermore, the learning indicator investigating the components and pH buffer solution has been made in question 4. In addition, the learning indicator analyzes the pH of the buffer solution when it is diluted, added a little acid, or added a little base had been made to the questions.

A recap of the assessments of the HOTS-based evaluation instruments on the ionic equilibrium materials and the pH of the buffer solutions assessed by the two material validators is provided in Table 7.

Table 7. Final Result of Validation by Material Validator

Number	Aspects		
	Material	Construction	Language
1	100	97.20	100
2	100	93.60	100
3	100	93.60	100
4	100	98.00	100
5	100	95.40	100
6	100	94.40	100
7	100	92.60	100
8	90.00	91.80	100
9	80.00	82.60	90.00
10	90.00	91.80	100
Averages(%)	96.00	93. 10	99.00
(%) The total of validation	96.03		
Very Valid Category			

Therefore, the results of media validation with WQC based on aspects of content substance, learning design, display (visual communication), and the use of software get suggestions and improvements.

The suggestions and improvements from the media validator include: (1) having an attractive cover; (2) synchronizing the terms in the user manual with those in the application; (3) having feedback; and (4) having a set time and displaying more detailed identity information for the compiler on the front page.

Next, the researchers corrected according to the suggestions the media validator gave and obtained valid HOTS questions. After the HOTS questions are more valid than before; that suggestion is packaged in question-making in WQC.

The diagram of the WQC media assessment's final results of the two media validators based on aspects of content substance, learning design, display (visual communication), and software utilization can be seen in Figure 2. Furthermore, a recap of the WQC media assessed by two media validators is provided in Table 8.

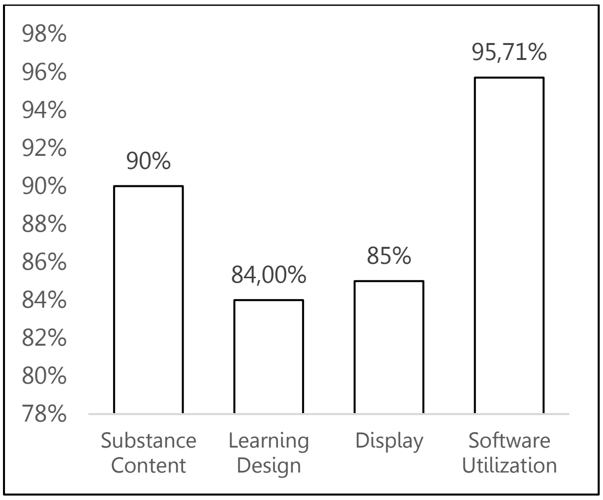

Figure 2. Final Result of Media Assessment

Table 8. Final Result of Validation by Material Validator

Assessment Aspect	Rating Percentage (%)	Category
Substance Content	90.00 %	Very Valid
Learning Design	84.00 %	Very Valid
Display (Visual Communication)	85.00 %	Very Valid
Software Utilization	95.71 %	Very Valid
Average total validation percentage	88.67	Very Valid

The results of media validation show that the average percentage score of the media validator is 88.67%. This data shows that the developed HOTS-based evaluation instrument with the WQC has fulfilled all components in content substance, learning design, display (visual communication), and software utilization with the result that trials can be implemented.

After the WQC Media Validator and the HOTS-based evaluation instrument with WTQ on ion balance material and pH buffer solution are valid; the trial stage will be conducted. Certainly, one-on-one and limited trials are the types of trials that will be realized in this study. First, the one-on-one trial begins at SMAN 2 Pekanbaru for three students with high, medium, and low abilities. In addition, the researcher briefly explained the WQC media as a display and how to use it in the one-on-one trial. Furthermore, the students were asked to work on ten multiple-choice questions, and they had 90 minutes available in the WQC software.

Thus, students with high ability could answer seven questions; students with medium ability could also answer four questions which can be seen in Table 9. Meanwhile, students with low ability correctly answer two questions. Finally, the students took 60 minutes as an average time. Meantime, the students are directly supervised while working on the questions. The results of field observations show that students still have difficulty analyzing the calculation questions and determining the pH of mixed solutions; it could be concluded that students still need more analytical skills to solve problems.

The one-on-one trial was fulfilled to obtain responses to the HOTS-based evaluation instrument with the WQC as a display medium for the ion balance material and the pH of the buffer solution for improvement during limited trials. One of the students' responses was the improvement of the editorial questions on question number 6. Additionally, the researchers made editorial improvements to the questions so that students quickly understood them.

A limited trial was conducted on 30 students at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru and MAN 1 Pekanbaru. The selection of students as respondents were chosen directly by the chemistry teacher concerned. The limited trial at SMAN 8 Pekanbaru and MAN 1 Pekanbaru was carried out boldly through the WhatsApp group and the Zoom application; the researcher sent the publication results through the WhatsApp group and asked students to work on it, then supervised the students in working through Zoom. Meanwhile, the limited trial at SMAN 2 Pekanbaru was carried out directly by observing health protocols. All students are required to wear masks. The time given is 90 minutes. The limited trials were implemented at different times: April 26 at SMAN 8 Pekanbaru, April 27 at SMAN 2 Pekanbaru, and May 3, 2021, at MAN 1 Pekanbaru.

The results of the limited trial were obtained at SMAN 8 Pekanbaru and MAN 1 Pekanbaru, which were conducted through an online test. During the test, students get ten questions, and they can answer five answers. In other words, students only get a score of 50. During the limited trial at SMAN 2 Pekanbaru, the test was carried out directly by giving ten questions to students. From the test results, the average student only answered two questions. It means that students at SMAN 2 Pekanbaru only got an average score of 20 in the limited trial. From the test results, students showed significant differences in two trial test sessions: online and offline. In addition, the results of working on the HOTS-based evaluation instrument with the online wondershare quiz maker display media are higher than doing it in person.

Table 9. The Students' Results of One-on-one Testing Evaluation based on HOTS with WQC.

Number	Name	Ability	Processing Time (Minutes)	Correct Answer
1	Student 1	High	60	70
2	Student 2	Medium	75	60
3	Student 3	Low	85	40

Students represented the one-on-one test instrument questions with varying abilities (high, medium, and low) shown in Table 10, and took longer and resulted in fewer correct numbers.

Table 10. Level of Questions Difficulty

Question Items	Difficulty Level	Criteria
1	0.23	Difficult
2	0.66	Medium
3	0.66	Medium
4	0.30	Difficult
5	0.70	Medium
6	0.30	Difficult
7	0.30	Difficult
8	0.26	Difficult
9	0.66	Medium
10	0.66	Medium

Based on data processing, the reliability value calculated by the Spearman-Brown formula and the Anates V4 software obtained a number of 0.85, so the score is very high.

Although the score obtained by students is low, the reliability value is very high. These data indicate that the HOTS-based evaluation instrument developed by the researcher consistently provides assessment results. Based on the data processing, the difficulty level value, five questions were obtained in the difficult category and five in the medium category. The index of the difficulty level of small questions states that the questions are getting more complex and vice versa. In other words, if the index is extensive, the questions will be more accessible to students.

Furthermore, based on the analysis of the difficulty level, the evaluation instrument that the researcher developed has a good difficulty level: medium and difficult. Indeed, this phase follows Alfajri et al. (2019), which states that good questions are questions that are not too easy and not too difficult. On the other side, based on the index of difficulty level, information was obtained that items 3 and 6 had a minor difficulty index because the number of students who answered correctly was minimal, causing a small difficulty index.

Table 11. Results of Calculation of Distinguishing Power Item Question (formula)

Question Items	Distinguishing Power	Criteria
1	0, 62	Question accepted
2	0, 37	Question accepted
3	0, 37	Question accepted
4	0, 37	Question accepted
5	0, 62	Question accepted
6	0, 37	Question accepted
7	0, 87	Question accepted
8	0, 37	Question accepted
9	0,75	Question accepted
10	0,50	Question accepted

Moreover, all categories of questions were accepted on distinguishing power data processing which can be seen in Table 11. This problem can be interpreted that the ten items being acceptable and suitable for use. Basuki & Hariyanto (2016) states that a good question is an item with very good criteria, then the question is accepted. Question number 7 has a high distinguishing value because many students answered correctly in the upper group, and there were no students who answered correctly in the lower group.

Hence, distinguishing power's value is greater than other items' value.

Based on the tabulation of student response data from the three schools, the total average percentage is 92.53%. This value gives a score in the range of 80.00%-100% with very good criteria. Meanwhile, students who get an impression after working on the HOTS-based evaluation instrument with the WQC system have very good criteria: students are able to equip students with straightforward and exciting media. Finally, a HOTS-based evaluation instrument with a WQC as a display medium for the ion balance material and the pH of the final buffer solution.

On the other hand, according to the data tabulation of the responses of three chemistry teachers at SMAN 8 Pekanbaru, SMAN 2 Pekanbaru, and MAN 1 Pekanbaru. Researchers also accept that the average score is 92.53%, so the average total score of the criteria is very good. Furthermore, the researchers received positive responses from the three chemistry teachers who stated that the evaluation instrument developed on the ion balance material and the buffer solution pH met the HOTS criteria and followed the basic competencies. Finally, the use of an attractive WQC application is expected to increase students' motivation in working on HOTS-based evaluation instruments.

4. Conclusion

The results showed five phases of the Plomp development model on HOTS-based evaluation with WQC display media in ion balance material and buffer solution pH for 11^{th} grade from senior high school. First, the initial investigation phase is a front-end analysis based on interviews with three chemistry teachers, an analysis of the questions used in schools, and an analysis of the needs of students through the distribution of online questionnaires. Second, the design phase is to design an initial HOTS-based evaluation design, create an answer key and design a media WQC. The third is the realization phase. Fourth is the validation phase. The validation phase is a trial and revision which produces 10 HOTS questions with valid WQC use, and the fifth is the implementation phase. The process of testing user responses by teachers with a percentage of 88.00% and students with a percentage of 95.23% with very good criteria. HOTS-based evaluation instrument with WQC as a display medium on ion balance material and pH of the buffer solution for 11^{th} grade from senior high school level equivalent valid according to material validators based on material aspects 96.00%, construction aspects 93.10%, and language aspects 99.00%. Besides, the media validator based on the aspect of substance, content is 90.00%, learning design is 84.00%, display (visual communication) is 85.00%, and software utilization is 95.71%.

References

Ahmad, I. F., & Sukiman, S. (2019). Analisis higher order thinking skills (HOTS) pada soal ujian akhir siswa kelas 6 KMI dalam kelompok mata pelajaran Dirosah Islamiyah di Pondok Modern Tazakka Batang. *Jurnal Pendidikan Agama Islam, 16(*2), 137-164. https://doi.org/10.14421/jpai.2019.162-02

Alfajri, A. R., Maizora, S., & Agustinsa, R. (2019). Kepraktisan soal-soal higher order thinking untuk menghasilkan soal yang praktis untuk siswa kelas XI MAN 1 Kota Bengkulu. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS), 3*(2), 188-195. https://doi.org/10.33369/jp2ms.3.2.205-217

Arianti, A., & Hardiyanto, H. (2018). Pengembangan Perangkat Pembelajaran Mandiri Berbasis Soal Terbuka dalam Pembelajaran Kalkulus pada Prodi Pendidikan Matematika Universitas Negeri Makassar. *Matematika dan Pembelajaran, 6*(1), 35-47. ttp://dx.doi.org/10.33477/mp.v6i1.441

Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan.* Jakarta: Bumi Aksara.

Basuki, I., & Hariyanto. (2016). *Asesmen Pembelajaran*. Bandung: Rosda Karya

Budiastuti, D., & Bandur, A. (2018). Validitas dan reliabilitas penelitian. Jakarta: Mitra Wacana Media.

Dinni, H. N. (2018). HOTS (High Order Thinking Skills) dan kaitannya dengan kemampuan literasi matematika. *PRISMA, Prosiding Seminar Nasional Matematika*, *1,* 170-176. Retrieved from https://journal.unnes.ac.id/sju/index.php/prisma/article/download/19597/9507

Faruq, U., & Huda, M. M. (2020). Bahasa Arab Berbasis Peningkatan Pembelajaran HOTS (Higher Order Thinking Skills) (Kajian Pembelajaran Bahasa Arab Di Madrasah Aliyah Unggulan Darul 'Ulum Step 2 Kemenag RI). *Al-Hikmah: Jurnal Pendidikan dan Studi Islam, 8*(1), 1-20. Retrieved from https://jurnal.staiba.ac.id/index.php/Al-Hikmah/article/view/135/0

Febriani, A., Elvia, R., & Handayani, D. (2021). Pengembangan Alat Evaluasi Pembelajaran Kimia Berbasis Komputer Menggunakan Wondershare Quiz Creator Pada Materi Larutan Penyangga. *ALOTROP, 5*(2), 191-197. Retrieved from https://ejournal.unib.ac.id/index.php/alotropjurnal/article/view/17141

Herliandry, L. D., Nurhasanah, N., Suban, M. E., & Kuswanto, H. (2020). Pembelajaran pada masa pandemi Covid-19. *JTP-Jurnal Teknologi Pendidikan, 22*(1), 65-70. https://doi.org/10.21009/jtp.v22i1.15286

Iqbal, W. M. G., Fadhilah, R., & Hadiarti, D. (2018). Pengembangan Alat Evaluasi Berbasis Wondershare Quiz Creator Pada Materi Koloid Kelas XI di SMA Koperasi Pontianak. *Ar-Razi Jurnal Ilmiah*, *6*(1), 11-19. Retrieved from https://123dok.com/document/download/yeokxxeq

Kempirmase, F., Ayal, C. S., & Ngilawajan, D. A. (2019). Kemampuan Berpikir Kritis Siswa dalam Menyelesaikan Soal-Soal Higher Order Thinking Skill (HOTS) pada Materi Barisan dan Deret Aritmatika Di Kelas XI SMA Negeri 10 Ambon. *Prosiding Seminar Nasional Pendidikan Matematika Universitas Pattimura*, 21-24. Retrieved from https://ojs3.unpatti.ac.id/index.php/matedu/article/view/1610

Laili, M., Aini, N., & Christanti, A. (2020). High Order Thinking Skills (HOTS) dalam Penilaian Bahasa Inggris Siswa SMA. Lintang Songo: *Jurnal Pendidikan, 3*(1), 18-25. Retrieved from https://journal.unusida.ac.id/index.php/jls/issue/view/28

Meryansumayeka, M. V. S. M., Virgiawan, M. D., & Marlini, S. (2018). Pengembangan kuis interaktif berbasis e-learning dengan menggunakan aplikasi wondershare quiz creator pada mata kuliah belajar dan pembelajaran matematika. *Jurnal Pendidikan Matematika, 12*(1), 29-42. Retrieved from https://ejournal.unsri.ac.id/index.php/jpm/article/view/5094

Nardjosoeripto, P., Sari, N., & Purbosari, P. M. (2017). Analisis Ahli dalam Pengembangan Evaluasi Pembelajaran IPS Online Berbasis Classmaker. Edudikara: *Jurnal Pendidikan dan Pembelajaran, 2*(3), 268-278. Retrieved from https://ojs.iptpisurakarta.org/index.php/Edudikara/article/view/58

Noviyanti, N., Haryati, S., & Herdini, H. (2020). Pengembangan lembar kegiatan peserta didik (LKPD) berbasis search, solve, create and share (SSCS) pada pokok bahasan kesetimbangan ion dan pH larutan penyangga. *J-PEK (Jurnal Pembelajaran Kimia), 5*(1), 8-16. http://dx.doi.org/10.17977/um026v5i12020p008

Pane, A., & Darwis D. M. (2017). Belajar Dan Pembelajaran. *FITRAH: Jurnal Kajian Ilmu-Ilmu Keislaman, 3*(2), 333. https://doi.org/10.24952/fitrah.v3i2.945

Plump, C. M., & LaRosa, J. (2017). Using Kahoot! in the classroom to create engagement and active learning: A game-based technology solution for eLearning novices. *Management Teaching Review, 2*(2), 151-158. https://doi.org/10.1177/2379298116689783

Sa'adah, S. I., Rasmiwetti, R., & Linda, R. (2019). Pengembangan soal HOTS dengan wondershare quiz creator sebagai media display pada materi stoikiometri kelas X. *JTK (Jurnal Tadris Kimiya), 4*(2), 177-188. https://doi.org/10.15575/jtk.v4i2.5469

Satriana, D. (2019). Pengembangan Perangkat Lunak Wondershare Quiz Creator Berbasis Blog untuk Evaluasi Pembelajaran Bahasa Indonesia Kelas X. *Jurnal Siliwangi: Seri Pendidikan P-ISSN, 5*(1). Retrieved from https://jurnal.unsil.ac.id/index.php/jspendidikan/article/view/739

Schleicher, A. (2019). *PISA 2018: Insights and Interpretations*. Paris: OECD Publishing.