CC Thinker: Mobile-based Assessment to Train Creative and Critical Thinking in Students

https://doi.org/10.3991/ijim.v16i15.29991

Wahyu Ridhoni[1,2], Punaji Setyosari[1](✉), Dedi Kuswandi[1], Saida Ulfa[1], Dahlia Janan[3]
[1]Universitas Negeri Malang, Malang, Indonesia
[2]Politeknik Hasnur, Barito Kuala, Indonesia
[3]Sultan Idris University of Education, Perak, Malaysia
punaji.setyosari.fip@um.ac.id

Abstract—This research aims to combine assessment of creative and critical thinking in a mobile device context, to see its effect on the creative and critical thinking of students. Design-based research as the method was used to develop the assessment as learning. It is begun with the assessment model and android application development, also experiments in the end. Three assessment sessions are participated by 46 students. The analysis using Kruskal Wallis gains the result that there are significant differences in the creative thinking (Chi-square 18.245, Asymp. Sig 0.000) and the critical thinking (Chi-square 7.620. Asymp. Sig 0.022). A significant difference in creative thinking is seen since the second session, while the difference in critical thinking needs the time up to the third session.

Keywords—creative thinking, critical thinking, mobile-based assessment, assessment as learning

1 Introduction

The challenge of the students in the 21st century, which is also a skill deemed to be very important by many researchers and teachers is how the students can think creatively and critically [1]. Both skills are also important in preparing the students to enter the professional world as well as for the future [2]. Creative thinking is a skill of using the imagination to explore new ways to generate innovation [3], while critical thinking is a skill of proving a point using a strong analysis to solve a problem [4]. Although it seems to be the opposite of each other, the two skills are intersecting and complete one another [3]. If one only thinks critically, then the alternative they create may not be varied. Also, if one only thinks creatively, then the solution created may not answer the problem.

The assessment is necessary to find out whether a student has been on a certain level in terms of creative and critical thinking. Previously, if one wants to measure creative and critical thinking, then two partial instruments are needed. For example, measuring creative thinking, it needs the Instrument Torrance Test of Creative Thinking (TTCT) [5][6], and for critical thinking, it uses the California Critical Thinking Disposition

Inventory (CCTDI) [7]. The combination of the two assessments here is very reasonable since the two skills are completing each other. Creative and critical thinking has a medium correlation [8] and is also positively correlated [9].

A mobile application can be used for the educational environment [10] and growing fast [11]. This research uses mobile application with the Android operating system to deliver the assessment considering the majority of the subjects own an android smartphone (91.5%) so that "bring your own device" can be applied [12] and teachers do not need to provide any device to facilitate this assessment. The assessment approach is assessment as learning, so it is used more as a self-reflection [13]. More specifically, the following research question was identified:

RQ1: Does the developed mobile-based assessment affect students' Creative Thinking?
RQ2: Does the developed mobile-based assessment affect students' Critical Thinking?

2 Literature review

2.1 Creative thinking

The research on the assessment of creative thinking that is specialized on the students has been done by Karpova et al [14] Involves students from various majors by using the existing instrument, which is the Torrance Test of Creative Thinking (TTCT) figural form. On the TTCT, Torrance uses Flexibility, Originality, Fluency, and Elaboration as the sub-test [15].

Meanwhile, Moffat et al [16] involve the design students in the "serious game design" subject by developing the framework themselves where they can gain the score of their creativity. Besides, the partners can also gain knowledge together while scoring each other. This type of scoring is classified as an "assessment as learning". The instrument that is also developed by oneself is the Widening-Connecting-Reorganizing (WCR) by Pizzingrilli et al [17]. Meanwhile, Lipman [18] mentions Imaginative, Holistic, Inventive, and Generative as the sub-variable of creative thinking.

2.2 Critical thinking

The research on the assessment of critical thinking that is done upon the students as the subjects that are developed by Reynders et al [19] created the ELIPSS Rubrics that is aimed at the students of STEM. The rubric is the scoring in points of 1, 3, or 5 according to the condition that is described on the rubric in every of the following factor: Evaluating, Analyzing, Synthesizing, and Arguments (Structure & Validity). Besides, with the subjects of the students from various backgrounds, Liu et al [20] Test the validity and reliability of the instruments of HEIghten™, where the result is adequate whether in the level of group or individual. Meanwhile, Rickles et al [21] use Structure of the Observed Learning Outcome (SOLO) Taxonomy as the scoring rubric for the students of Sociology to score the critical thinking in writing or even when in-class discussion.

On critical thinking, Watson & Glaser [22] mentions recognizing assumption, evaluating arguments, and drawing conclusions, also known as RED as sub-variables.

Davies [23] Divides critical thinking into cognitive elements (argumentation, inference making, and reflective judgment) and prosperity (disposition, abilities, and attitudes). Then, Facione [24] mentions that the following sub-variables are critical thinking elements: analysis, evaluation, inference, and explanation.

2.3 Mobile-based assessment

The mobile phone is one of the devices that belongs to almost every student in Indonesia because of the affordable price for all layers of society [25]. The experiment of the mobile-based assessment has been done on students of law major where the students show a positive attitude in interacting with the mobile application [26]. Besides, there is also research done on computer major in the form of the development of Mobile Response System (MRS) that facilitate the scoring of the problem-solving activities interactively [27].

Mobile-based assessment has the strength since the size of the compact mobile device allows students to apply the "bring your own device" [12]. Similar to the other online-based assessment, the mobile-based assessment can be designed with rich visuals [28] and the most important thing is that it can simplify the management of the test [29]. Various types of assessment are now able to be done with the help of the mobile device, the students can even cross evaluate with the other students [30]. Most of the respondents think positively of this mobile-based assessment because it is easy to use. However, being disconnected from the internet in the middle of the assessment becomes a note in itself [31].

3 Method

This research uses Design-Based Research (DBR) as the method. The research procedure consists of 3 phases that are adapted from the research of Ulfa et al [32] also uses DBR, the phases are illustrated in Figure 1.

The first phase focuses on the development of the assessment model where the output is a question bank that is valid and reliable, the questions also provide a timer and score shown based on the level of difficulties. In this phase determining the subtest variable based on the literature review. Generating the questions items are limited to objective test questions, it can be in the form of multiple choices, grouping, matching, sorting, and another type of questions. Validating the content from the experts consists of 3 lecturers with research interest in creative and critical thinking also the minimum qualification of a doctoral degree. Every question item will be classified into three categories of invalid, revision required, and valid. On instrument testing to the limited subject, try out is conducted to about 30 students through Google Forms and it will be conducted in several sessions. The score will be used to analyze the validity using Pearson Correlation (<5%), reliability with Cronbach's Alpha (>0.6), and level of difficulty by identifying the Mean value (0–0.33: hard, 0.34–0.66: medium, 0.67–1: easy). Only valid and reliable items are used for the next phase.

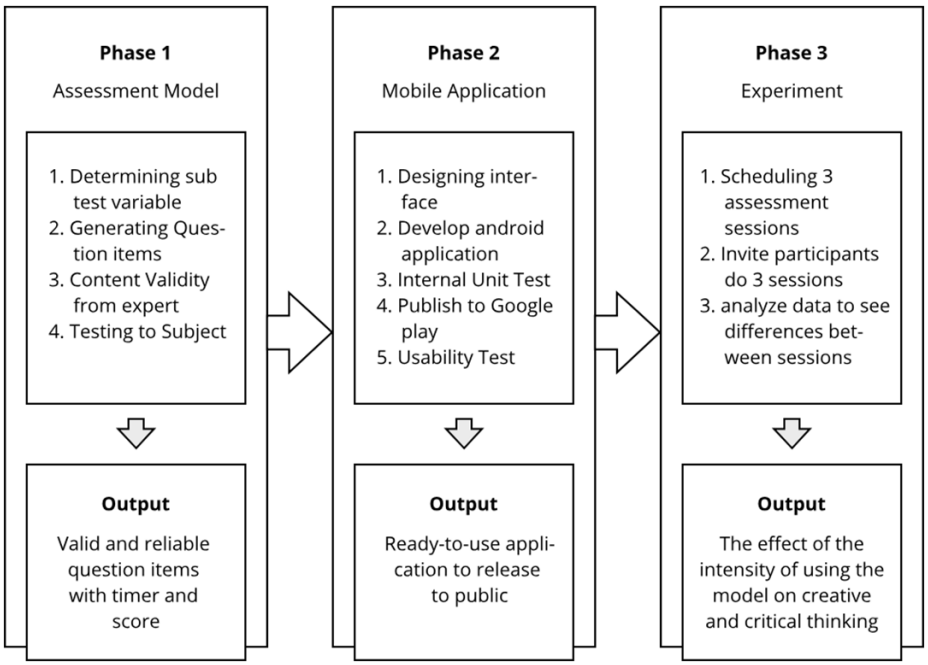


Fig. 1. Research procedure

The second phase is the development of the mobile application by referring to the model developed in the previous phase. In the final part of this phase, scoring by the users is done to see whether it needs some revision or not, so that the final application is ready to use publicly. After respondents used the application, they answered indicators with 5-points Likert scales on Google Forms (very less to very good). The mean of all indicators is used to categorize the validity of the mobile applications. Less than 20 Very Invalid and 20.1–40 Invalid (cannot be used), 40.1–60 Less Valid (can be used but needs major revision), 60.1–80 Valid (can be used but needs minor revision), and 80.1–100 Very Valid (can be used without revision).

The final phase is the experiment using the mobile application developed that is conducted in three sessions of assessment so that one can see whether there is an intensity influence on the score of creative and critical thinking. It was started by inviting students from various universities to participate as respondents. Invitations are sent via WhatsApp groups and Social Media. Participating students are limited to those who are active students. Verification is needed by matching the data in the national higher education database so that the respondents who take part in the experiment are truly valid. Experiments were carried out in 3 sessions to see how the progress of creative and critical thinking and whether there were significant differences from session to session. Manova can be used to see the effect on several variables. An alternative to Manova if the data is not normally distributed and the variation is not homogeneous, then the test is carried out non-parametrically with Kruskal Wallis. After answering the hypothesis, it can be explored further by a post hoc test.

4 Result and discussion

4.1 Development of assessment model

The determination of creative thinking refers to three main sources, those are Lipman, Torrance in Rad et al and Pizzingrilli & Antonietti. The three sub-tests that resulted are the Parse (C1), Deviate (C2), and Modify (C3). The interrelation with the references is illustrated in Figure 2.

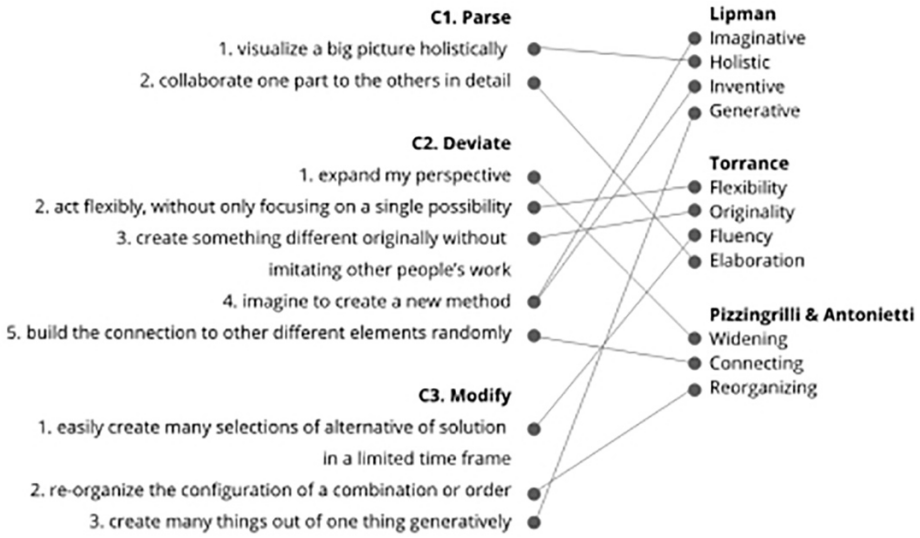


Fig. 2. Construct of creative thinking

While for critical thinking, it refers to the three main sources, are Facione, Watson & Glaser, dan Davies. The three sub-tests results are Verify (C1), Compare (C5), and Conclude (C6). The interrelation with the references is illustrated in Figure 3.

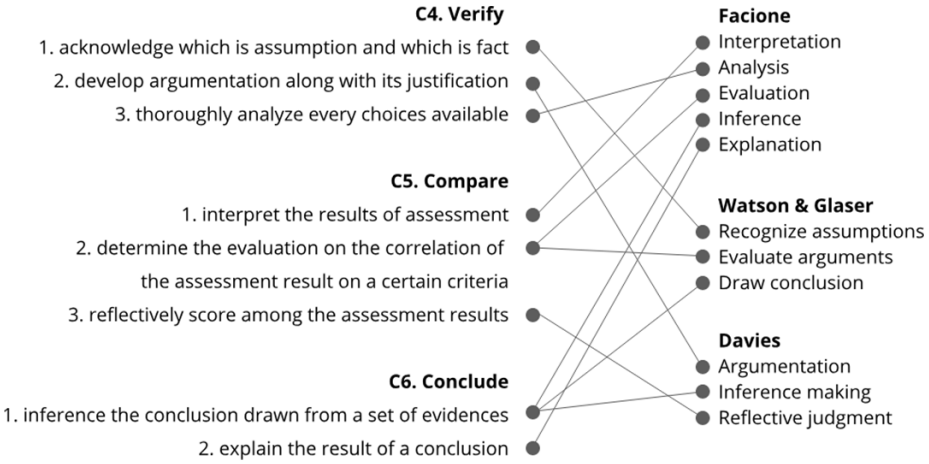
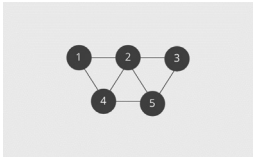


Fig. 3. Construct of critical thinking

There are 8 objective-type questions, which are taken from the question bank randomly for each sub-test so that there are 240 questions prepared that are then evaluated in terms of content validity by 3 experts in the topic of creative and critical thinking. Every sub-test has a different question that is adjusted to the characteristics of the variable that will be built. Parse (C1) uses selection type of question, Deviate (C2) uses multiple-choice, Modify (C3) uses matching, Verify (C4) uses true-false, Compare (C5) uses arranging sequence, and Conclude (C6) uses classifying. Based on the scoring from the experts, 198 items (82.5%) are valid, yet 42 items (17.5%) require some correction before it is processed to the next phase of experimenting with the subjects. Table 1 contains the questions for each sub-test.

Table 1. Question samples

Item	Question	Answer Key
C1-2	Choose the best answers according to the result of the summation of numbers that are directly connected by the line. You may choose more than one answer among the options a, b, c, d, or e.  ☐ a. 12 = 3+5+4 ☐ b. 12 = 1+4+2+5 ☐ c. 12 = 5+2+6 ☐ d. 12 = 3+1+3+5 ☐ e. 12 = 2+1+5+4	**a, b**

(Continued)

Table 1. Question samples *(Continued)*

Item	Question	Answer Key
C2-4	Choose the best answer that consists of an uncommon pair of words. Choose only one most appropriate answer among the options a, b, c, d, or e. a. Brush – Shoes b. Balloon – Air c. Male – Female d. Tent – Fly e. Table – Wood	**d**
C3-2	Connect the choice of answer on the left to ones on the right which consist of the same numbers in a different order. a. 2 4 / 6 8 b. 2 7 / 8 6 c. 2 9 / 5 6 1. 6 4 / 2 8 2. 2 7 / 8 6 3. 2 6 / 8 7 4. 8 2 / 6 4 5. 7 2 / 6 8	**a1, a4, b2, b3, b5**
C4-20	Today, the evacuees are happy. Urgent support has been distributed in the neighboring village. **True** or **False**?	**False**
C5-2	Compare and arrange the times below from the highest number of minutes to the lowest 1. 2 hours, 20 minutes, 10 seconds 2. 1 hour, 10 minutes, 15 seconds 3. 3 hours, 30 minutes, 50 seconds 4. 2 hours, 15 minutes, 40 seconds	**3, 1, 4, 2**
C6-4	A survey on beverage flavor was conducted by a producer for 100 people. There are only two flavors that received the most positive feedback are being produced. Determine which flavors will be produced and those that are not. 1. Lemon, 23 people approve 2. Strawberry, 56 people approve 3. Chocolate, 75 people approve 4. Manggo, 45 people approves A. Being Produced B. Not Being Produced	**A (2, 3)** **B (1, 4)**

The questions bank, which consists of 240 questions that have been corrected based on the expert evaluation, is then tested on the limited subjects through Google Forms. The participants are the students of Information Engineering at the Polytechnic of Hasnur, Southern Kalimantan, Indonesia. From 240 questions, then the test is divided into 4 sessions with 60 questions each with a 1-week gap. There are 31 participants

consisting of 16 male students (51.61%) and 15 female students (48.39%) who followed all four sessions and can be used in the data analysis. After the validity analysis using Pearson Correlation Sig. (2-tailed) scores less than 5% and using Cronbach's Alpha test that scores higher than 0.6, then 169 question items gained are the remaining valid and reliable items (Table 2). The questions also have a timer and score that is calculated from the degree of difficulty, hard (mean 0 to 0.33), medium (mean 0.34 to 0.66), and easy (mean 0.67 to 1). The category of hard is given a timer of 1 minute and 30 seconds and it also has a score of 3, the medium category is given the timer of 1 minute and has a score of 2, while the category of easy is given the timer of 45 seconds and has the score of 1 if the answer is correct. There is an exception on Verify (C4) where there are only 6 items left, therefore on the implementation of the questions from this sub-test, all the items are going to be used without random choosing like the other sub-tests.

Table 2. Final questions

Sub Test	Hard	Medium	Easy	N
C1. Parse	21	7	–	28
C2. Deviate	–	21	8	29
C3. Modify	4	25	10	39
C4. Verify	1	4	1	6
C5. Compare	9	24	1	34
C6. Conclude	3	25	5	33
Total				169

4.2 Development of mobile application

Need analysis is adjusted with the assessment model that has been developed in the previous phase. The list of needs is arranged as follows:

Assessment Participants:

1. The Verified identity of the participants
2. Assessment can be done repeatedly by the same participants/class
3. The tests consist of 6 types where each question are limited with a different timer and different score shown
4. The score of the creative and critical thinking as well as the detailed score of the sub-tests
5. The top 3 scores of the participants of assessments are visible
6. The discussion of the answers which one is correct and false

Evaluator/Assessment Owner:

1. Log in to see the assessment result that has been done by a group
2. Assessment can be made for a particular date
3. Choosing the question package of creative and critical thinking in Bahasa Indonesia or English
4. Adding the essay question out of the question package of creative and critical thinking

Based on the needs mentioned above, then the android application is designed and developed and is downloadable through Google Play under the name of CC Thinker. Figure 4 here are several displays of the interface.

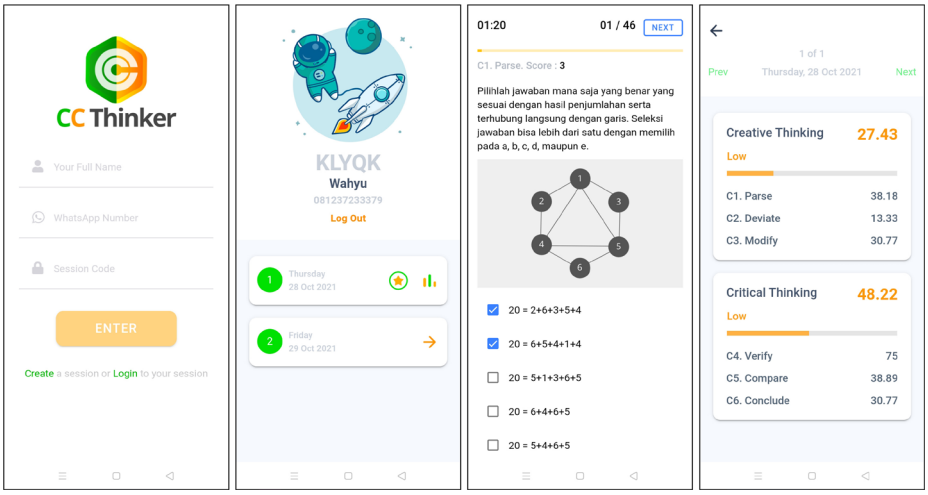


Fig. 4. Application interface

Usage testing by the students of various universities is conducted, involving the respondents that are not only from the students of the Information Engineering study program. The respondents are 47 students, 24 of them are males (51.06%) and 23 of them are female (48.94%), 31 of them are from polytechnic (65.96%) and 16 of them are from universities (34.04%). They try and explore the features provided in the application and then they are asked to fill in the scoring through Google Forms with 11 indicators: (1). Ease of installation, (2). Ease of use, (3). Ease of search, (4). The processing of the application, (5). Navigation to switch page, (6). Layout, (7). Texts readability, (8). Picture quality, (9). Clarity of data graph, (10). Display aesthetics, and (11). Handling the wrong input.

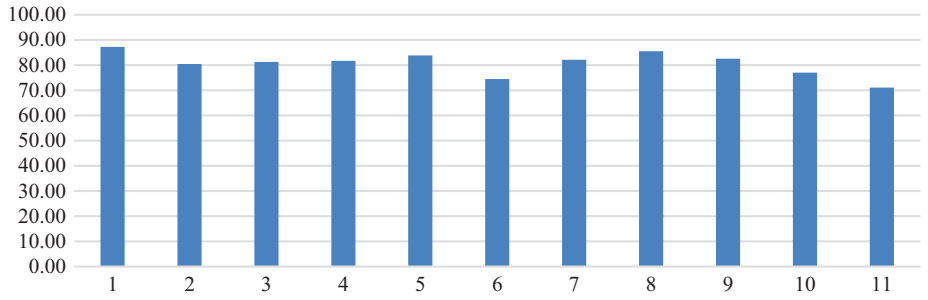


Fig. 5. Scoring on usage testing

The accumulated average score of the indicator from all respondents gives the score of 80.66 which means that it falls into the category of Very Valid (80–100) so that the application can be used without the need for any revision. The highest scoring is given to the ease of installation with 87.23 and the lowest score, but still in the category of Good, is in the indicator of Handling on wrong input with the score of 71.06.

4.3 Experiment with three assessment sessions

The experiment is done using one group repeated measures design, where every respondent works on the assessment three times. The questions on every assessment are shown randomly by the system. 46 students follow all three assessment sessions, 24 students are from Southern Kalimantan (52.17%), 12 students are from East Java (26.09%), and there rest 10 students are from various locations (21.74%). Meanwhile, 15 students of which are pursuing associate degrees (32.6%), and 31 students are pursuing bachelor's degrees (67.4%). The gender of the respondents is 24 students of them are male (52.2%) and 22 students are female (47.8%).

The score of creative thinking is calculated from the average score of C1, C2, and C3. Meanwhile, the critical thinking score is calculated from the average score of C4, C5, and C6. The result gained from the first session until the third session is shown in Table 3. There are improvements from each session to the next ones, either in creative thinking or critical thinking. The average score the critical thinking is higher than creative thinking. Deviate (C2) is a sub-test of creative thinking with the highest score, and Conclude (C6) is a sub-test of critical thinking with the highest score. This result is consistent from the first session to the third session. The average score of the creative thinking from sessions 1 to 3 falls into the category of medium (50–75). Meanwhile, the average score of critical thinking from sessions 1 to 2 falls into the category of medium and the third session falls into the category of high (>75).

Table 3. The average score of the creative and critical thinking

Session	C1	C2	C3	C4	C5	C6	Creative	Critical
1	67.37	70.54	41.12	62.14	68.96	78.74	59.68	69.95
2	70.48	77.89	57.09	67.93	71.75	83.06	68.48	74.25
3	74.23	83.00	63.93	75.18	71.18	84.30	73.72	76.89

The normality test shows the data for critical thinking with Sig. 0.026 is lower than 0.05, so it is not normally distributed. Kruskal Wallis is used to analyze data that is not normally distributed. Table 4 shows the presence of significant differences in creative thinking (Asymp Sig. 0.000) as well as in critical thinking (Asymp Sig. 0.022).

Table 4. Kruskal wallis statistics

Variable	Chi-Square	df	Asymp. Sig.
Creative	18.245	2	0.000
Critical	7.620	2	0.022

RQ1: Does the developed mobile-based assessment affect students' Creative Thinking?

The post hoc was done to see deeper the comparison among sessions so that the significant difference was identified to happen in which session. In creative thinking, the significant difference happened since the second session (Figure 6). This shows that the influence of mobile-based assessment on creative thinking is stronger than on critical thinking.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
session 1-session 2	-23.043	8.337	-2.764	.006	.017
session 1-session 3	-35.033	8.337	-4.202	.000	.000
session 2-session 3	-11.989	8.337	-1.438	.150	.451

Fig. 6. Post hoc kruskal wallis for creative thinking

RQ2: Does the developed mobile-based assessment affect students' Creative Thinking?

However, in critical thinking, the significant difference occurred in a longer period, specifically needing time up to the third session (Figure 7). This shows that the influence of mobile-based assessment on critical thinking is lower than on critical thinking.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
session 1-session 2	-15.109	8.337	-1.812	.070	.210
session 1-session 3	-22.587	8.337	-2.709	.007	.020
session 2-session 3	-7.478	8.337	-.897	.370	1.000

Fig. 7. Post hoc kruskal wallis for critical thinking

5 Limitations of the study

In the assessment model development stage, the Verify (C4) items remaining after the validity and reliability test are only 6. Ideally, the number should be more than 8 so that the questions that appear during the assessment can be randomized by the system. This research also requires comparison with other creative and critical thinking assessment instruments so that it can be seen whether the results tend to be the same in a subject. The measurement results for creative and critical thinking using mobile technology between men and women, or between science and art majors are also interesting to study.

6 Conclusion

The development of an assessment model in the form of an objective test with 169 questions that have been tested on the validity and reliability. The questions also provide a timer and score shown based on the difficulty levels. The android application is then developed to deliver the questions. After the testing is done, the score gained is very valid, so that the application can be concluded to be used widely without the need for revision. Meanwhile, in the experiment phase, it gained the score that there are significant differences in creative thinking and critical thinking so it can be concluded that the intensity of the application use affects the creative and critical thinking of the users. Significant differences in creative thinking is seen since the second session. On critical thinking, it needs the time up to the third session to show the significant differences.

7 Acknowledgment

This research has been supported by The PNBP Grant Commission, under the State University of Malang Research Grant 2021.

8 References

[1] Battelleforkids.org, "P21 Partnership for 21st Century Learning," 2020. [Online]. Available: https://www.battelleforkids.org/networks/p21. [Accessed: 03-Dec-2020].

[2] D. P. Dewi, S. Soekopitojo, A. Larasati, M. F. Kurniawan, and E. R. S. Hartanti, "Developing Instrument to Measure Student's Capability for Future Work in Industry 4.0 at Vocational Education Culinary Program," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 12, p. 110, Jul. 2020. https://doi.org/10.3991/ijim.v14i12.15589

[3] R. DiYanni, *Critical and Creative Thinking: A Brief Guide for Teachers*. Wiley-Blackwell, 2016.

[4] H. D. Kurniawan and H. Kuswanto, "Improving Students' Mathematical Representation of Physics and Critical Thinking Abilities Using the CAKA Mobile Media Based on Local Wisdom," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 02, p. 72, Jan. 2021. https://doi.org/10.3991/ijim.v15i02.11355

[5] G. Krumm, V. Arán Filipppetti, V. Lemos, J. Koval, and C. Balabanian, "Construct validity and factorial invariance across sex of the Torrance Test of Creative Thinking – Figural Form A in Spanish-speaking children," *Think. Ski. Creat.*, vol. 22, pp. 180–189, 2016. https://doi.org/10.1016/j.tsc.2016.10.003

[6] G. Krumm, M. Aranguren, V. Arán Filippetti, and V. Lemos, "Factor Structure of the Torrance Tests of Creative Thinking Verbal Form B in a Spanish-speaking Population," *J. Creat. Behav.*, vol. 50, no. 2, pp. 150–164, 2016. https://doi.org/10.1002/jocb.76

[7] A. S. Mahmoud and H. A. Mohamed, "Critical Thinking Disposition among Nurses Working in Puplic Hospitals at Port-Said Governorate," *Int. J. Nurs. Sci.*, vol. 4, no. 2, pp. 128–134, 2017. https://doi.org/10.1016/j.ijnss.2017.02.006

[8] S. M. Wechsler *et al.*, "Creative and critical thinking: Independent or overlapping components?," *Think. Ski. Creat.*, vol. 27, no. January 2017, pp. 114–122, 2018. https://doi.org/10.1016/j.tsc.2017.12.003

[9] K. C. Tsai, "Investigating the Empirical Links between Creative and Critical Thinking," *Psychol. Soc. Educ.*, vol. 11, 3, pp. 267–280, 2019. https://doi.org/10.25115/psye.v11i3.1064

[10] M. Kalogiannakis and S. Papadakis, "Combining Mobile Technologies In Environmental Education: A Greek Case Study," *Int. J. Mob. Learn. Organ.*, vol. 11, no. 2, pp. 108–130, 2017. https://doi.org/10.1504/IJMLO.2017.10005249

[11] S. Papadakis and M. Kalogiannakis, "Mobile Educational Applications For Children: What Educators And Parents Need To Know," *Int. J. Mob. Learn. Organ.*, vol. 11, no. 3, pp. 256–277, 2017. https://doi.org/10.1504/IJMLO.2017.085338

[12] P. N. Chou, C. C. Chang, and C. H. Lin, "BYOD or not: A Comparison Of Two Assessment Strategies For Student Learning," *Comput. Human Behav.*, vol. 74, pp. 63–71, 2017. https://doi.org/10.1016/j.chb.2017.04.024

[13] X. Li, "Self-Assessment As 'Assessment As Learning' In Translator And Interpreter Education: Validity And Washback," *Interpret. Transl. Train.*, vol. 12, no. 1, pp. 48–67, 2018. https://doi.org/10.1080/1750399X.2017.1418581

[14] E. Karpova, S. B. Marcketti, and J. Barker, "The Efficacy Of Teaching Creativity: Assessment Of Student Creative Thinking Before And After Exercises," *Cloth. Text. Res. J.*, vol. 29, no. 1, pp. 52–66, 2011. https://doi.org/10.1177/0887302X11400065

[15] I. S. Rad, L. Karimi, V. Ramezani, M. Ahmadi, R. Heshmati, and E. Jafar, "Psychometric Properties of Torrance test (Persian version) of Creative Thinking (A form)," *Procedia - Soc. Behav. Sci.*, vol. 5, pp. 1429–1433, 2010. https://doi.org/10.1016/j.sbspro.2010.07.301

[16] D. C. Moffat and O. A. Shabalina, "Structures, Frameworks And Assessments For Student Exercises For Creative Thinking In Design," *Commun. Comput. Inf. Sci.*, vol. 754, pp. 711–722, 2017. https://doi.org/10.1007/978-3-319-65551-2_51

[17] P. Pizzingrilli, C. Valenti, L. Cerioli, and A. Antonietti, "Creative Thinking Skills from 6 to 17 Years as Assessed Through the WCR Test," *Procedia – Soc. Behav. Sci.*, vol. 191, pp. 584–590, 2015. https://doi.org/10.1016/j.sbspro.2015.04.498

[18] M. Lipman, *Thinking in Education*. Cambridge, UK: Cambridge University Press, 2003.

[19] G. Reynders, J. Lantz, S. M. Ruder, C. L. Stanford, and R. S. Cole, "Rubrics To Assess Critical Thinking And Information Processing In Undergraduate Stem Courses," *Int. J. STEM Educ.*, vol. 7, no. 1, 2020. https://doi.org/10.1186/s40594-020-00208-5

[20] O. L. Liu, L. Mao, L. Frankel, and J. Xu, "Assessing Critical Thinking In Higher Education: The Heightentm Approach And Preliminary Validity Evidence," *Assess. Eval. High. Educ.*, vol. 41, no. 5, pp. 677–694, 2016. https://doi.org/10.1080/02602938.2016.1168358

[21] M. L. Rickles, R. Z. Schneider, S. R. Slusser, D. M. Williams, and J. F. Zipp, "Assessing Change in Student Critical Thinking for Introduction to Sociology Classes," *Teach. Sociol.*, vol. 41, no. 3, pp. 271–281, 2013. https://doi.org/10.1177/0092055X13479128

[22] G. Watson and E. M. Glaser, *Watson–Glaser II Critical Thinking Appraisal: Technical manual and user's guide*. San Antonio, TX: NCS Pearson, 2010.

[23] M. Davies, "A Model of Critical Thinking in Higher Education," in *Higher Education: Handbook of Theory and Research*, vol. 30, no. 1, M. B. Paulsen, Ed. Cham: Springer International Publishing, pp. 75–81, 2015

[24] P. A. Facione, *The California Critical Thinking Skills Test: College Level Experimental Validation and Content Validity*. Millbrae, CA: California Academic Press LLP, 1990.

[25] S. Ulfa, "Mobile Technology Integration into Teaching and Learning," *IEESE Int. J. Sci. Technol.*, vol. 2, no. 1, pp. 1–7, 2013.

[26] M. Barreiro-Gen, "Evaluating The Effects Of Mobile Applications On Course Assessment: A Quasi-Experiment On A Macroeconomics Course," *Int. Rev. Econ. Educ.*, vol. 34, May, 2020. https://doi.org/10.1016/j.iree.2020.100184

[27] M. Fuad, D. Deb, J. Etim, and C. Gloster, "Mobile Response System: A Novel Approach To Interactive And Hands-On Activity In The Classroom," *Educ. Technol. Res. Dev.*, vol. 66, no. 2, pp. 493–514, 2018. https://doi.org/10.1007/s11423-018-9570-5

[28] K. Andrews, M. Zimoch, M. Reichert, M. Tallon, U. Frick, and R. Pryss, "A Smart Mobile Assessment Tool for Collecting Data in Large-Scale Educational Studies," *Procedia Comput. Sci.*, vol. 134, pp. 67–74, 2018. https://doi.org/10.1016/j.procs.2018.07.145

[29] G. Meletiou, I. Voyiatzis, V. Stavroulaki, and C. Sgouropoulou, "Design And Implementation Of An E-Exam System Based On The Android Platform," *Proc. 2012 16th Panhellenic Conf. Informatics, PCI 2012*, pp. 375–380, 2012. https://doi.org/10.1109/PCi.2012.76

[30] C. L. Lai and G. J. Hwang, "An Interactive Peer-Assessment Criteria Development Approach To Improving Students' Art Design Performance Using Handheld Devices," *Comput. Educ.*, vol. 85, pp. 149–159, 2015. https://doi.org/10.1016/j.compedu.2015.02.011

[31] J. R. Stowell, "Use Of Clickers Vs. Mobile Devices For Classroom Polling," *Comput. Educ.*, vol. 82, pp. 329–334, 2015. https://doi.org/10.1016/j.compedu.2014.12.008

[32] S. U. Ulfa, D. Barnabas Lasfeto, and I. Fatawi, "Applying Fuzzy Logic to Customize Learning Materials in e-Learning Systems," *Ubiquitous Learn. An Int. J.*, vol. 14, no. 2, pp. 49–61, 2021. https://doi.org/10.18848/1835-9795/CGP/v14i02/49-61

9 Authors

Wahyu Ridhoni received the scholarship for Doctoral of Instructional Technology, State University of Malang, Indonesia. He is currently working as a lecturer in Informatics Engineering, Hasnur Polytechnic, Indonesia. His research interest is applications of information system technology for education, especially in mobile learning. He is also an IT project manager on several private and government projects. (email: wahyu@polihasnur.ac.id)

Punaji Setyosari is a professor in Instructional Technology, Department of Educational Technology, Faculty of Education, State University of Malang, Indonesia. His research interests are teaching assessment, instructional technology and media, and learning strategies. He is also the author of research and development methodology and instructional design books. (email: punaji.setyosari.fip@um.ac.id)

Dedi Kuswandi is an Associate Professor in Instructional Technology, Department of Educational Technology, Faculty of Education, State University of Malang, Indonesia. His research interests are learning resources, curriculum development, learning model, teacher training, and program development. (email: dedi.kuswandi.fip@um.ac.id)

Saida Ulfa is currently working as a Lecturer in Instructional Technology, Department of Educational Technology, Faculty of Education, State University of Malang, Indonesia. She obtained her master's and a doctoral degree from Saga University, Japan. Her research interests are soft computing in education, adaptive learning, and personalization in online learning systems. (email: saida.ulfa.fip@um.ac.id)

Dahlia Janan is a Professor and Head of the Department of Malay Language and Literature at the Sultan Idris University of Education, Malaysia. Her research interests are Readability, Reading Process, Reading Pedagogies, Literacy, and Assessing text difficulty. She has won the best Ph.D. thesis prize throughout the United Kingdom for the field of literacy awarded by the United Kingdom Literacy Association. (email: dahlia@fbk.upsi.edu.my)

Article submitted 2022-02-05. Resubmitted 2022-05-31. Final acceptance 2022-06-14. Final version published as submitted by the authors.