

*Badrin Kartowagiran, Samsul Hadi, Nur
Wahyumiani, Fitri Alfarisa,
Widowati Pusporini*

Indonesia

Effectiveness of the AA “4C” Authentic Assessment Model: A Single-Case-Research (SCR)

DOI: 10.15804/tner.2019.57.3.16

Abstract

The purpose of this study is to show the effectiveness of the AA “4C” model in increasing teachers’ abilities. The study is a Single Case Research using the ABABA design. The subject of the study was a natural-science teacher. Results of the observation conducted continuously by three observers show that each of the teacher’s ability items improves from the first condition, to the second condition, and to the third condition. This finding is supported by the results of the Manova technique showing that the AA “4C” model is shown to be effective in improving teachers’ abilities in conducting authentic assessment.

Key words: authentic assessment; effectiveness of the AA “4C”

Introduction

In this 4.0 industrial revolution era, the demands for qualified teachers are inevitable. Teachers are required to teach students higher-order thinking skills such as analysis, evaluation, and creation. The higher the teachers’ quality, the higher the quality of the instructional processes, and, in its turn, the higher the quality of education. This is in line with Marzano, Frontier, & Livinnhston (2011) who state that the more positive activities in the classroom, the higher the learners’ achievements. Meanwhile, Barber & Mourshed (2012) state that students’ achievement begins with effective teachers and the school principal. The importance of teachers’ quality in improving the quality of graduates also applies to teachers of natural

science (NS) as a subject matter. However, it is a fact that this is not always true; there are teachers who are not capable of carrying out their task. Kartowagiran & Jaedun (2016) found that the level of understanding of authentic assessment by Junior High School NS teachers is low. One reason for this are numerous definitions of authentic assessment.

The results of the study of theories and/or ideas of Wiggins (1998), Tombari & Borich (1999), Earl, Hargreaves, & Schmidt (2002), Svinicki (2004), Frey, Schmitt, & Allen (2012), Gulikers, Kester, Kirschner, & Bastiaens (2008), Nitko & Brookhart (2011), Raymond, Homer, Smith, & Gray (2012), Vu & Alba (2014), Brown, Irving, & Keegan (2014) are summarized and used in this research. Within this terminological framework, authentic assessment is one that is real and factually assesses from inputs, processes, to outputs. It is integrated with teaching and it includes the assessment of skill competence, the assessment of knowledge competence, especially at the higher order (HOT), the assessment of attitude, especially the aspects of attitudes attached to the basic competence being studied.

Using the results of a 2016 study, Kartowagiran & Jaedun (2017) developed an authentic assessment model called an AA "4C". This model covers 4C, i.e. four competences, including assessing: (1) knowledge competence, (2) skill competence, (3) attitudinal competence, (4) the consistency in implementing the model. The consistency of the model use is characterized by four principles: (a) setting the knowledge competencies at higher-order thinking skills (HOTs), (b) assessing attitude competencies, especially those of the basic competencies, (c) focusing on four students at a session, and (d) running the assessment holistically, realistically, and as integrated into the subject matter. This coverage of the four competencies is obtained from various theories and research findings. This model has been validated theoretically, but has not been verified in field practice. The present study, therefore, is intended to show the effectiveness of the AA "4C" model in improving teachers' abilities while implementing authentic assessment.

Methodology of Research

The study was a Single-Case-Research (SCR) used the ABABA design standing for the preliminary situation (A1), first intervention (B1), second condition (A2), second intervention (B2), and third condition (A3) (Kratochwillv & Levin, 2014). The SCR model was selected in order to describe in detail the improvement in each item of the teacher ability while conducting authentic assessment. This purpose is not attainable when an experiment, action research, or a case study are used.

The study was conducted in a Junior High School in Yogyakarta City during September and October 2017, involving one natural science teacher. Data were obtained by three observers who performed continual observations on three instructional sessions, namely the preliminary condition, the second condition, and third condition. The instructional subject matters were Earth Structure, Solar System, and Solar and Lunar Rotation and Revolution. In addition to the observations, the teacher's ability in conducting authentic assessment was also assessed by 31 students. The development of the teacher's ability in conducting the authentic assessment from preliminary condition to the second condition and the third condition was observed by three observers and assessed by 31 students.

'Intervention' in this context is the training given to the teacher who is the subject of this study. She was trained and given examples of how to make lesson plans, to prepare teaching materials, and to make a test. The training guide and examples of the way how to make lesson plans, to prepare teaching materials, and to write a test are included in the AA "4C" model.

The research data were taken by the use of: (a) observation sheets equipped with an interview guide and (b) assessments sheets. The observation sheet was validated through expert-judgment and estimated using the Aiken Formula (Aiken, 1985). The results showed that, out of 24 items, four were deleted since their V value was lower than 0.88. So, the observation sheet ended up with 20 items. The observation sheet reliability was measured using the inter-rater reliability technique and estimated using the interclass coefficient correlation. It showed $r_{ii} = 0.79$ meaning the observation sheet was reliable (Feldt & Brennan, 1989).

The assessments sheet consisted of 20 items concerning teacher's abilities in implementing authentic assessment. The assessments sheet was subjected to 96 Junior High School students where the subject teacher taught. Using the exploratory factor analysis technique (EFA), items that had a factor loading lower than 0.3 were discarded (Hair, et.al., 2014). The EFA results showed that the assessments sheets consisted of four components of teacher's ability conducting authentic assessment, assessing: (1) knowledge competence, (2) skill competence, (3) attitudinal competence, (4) the consistency in implementing the AA "4C" model. The EFA results also showed that each component had five items and its reliability was $\alpha = 0.73$ meaning that the assessment sheet was reliable (Feldt & Brennan, 1989).

The observation sheet and assessments sheet each consisted of 20 items. Both the observation and assessments sheets each had four options producing a "very low" criterion for a score of 1.00–1.33, "low" for 1.67–2.33, "high" for 2.67–3.33, and "very high" for 3.67–4.00. The observation sheet was answered by three observers, and an assessments sheet was to be completed by the 31 students. Data of the

results of observations equipped by the interviews were subjected to a qualitative analysis technique to describe the development of each item of teacher’s ability from the preliminary condition (A_1), second condition (A_2), to third condition (A_3). Meanwhile, data of the results of students’ assessment were analysed by using the Manova technique, univariate Anova, and pair-*t*.

Results of Research

At the beginning of the study, the teacher conducted a preparation stage that was followed by a teaching stage. This was observed by three observers and assessed by 31 students. Results of this activity were called Condition 1, or the preliminary situation, and the results of the students’ assessment became the pre-condition of the study. Results of the on-going observation of the teacher’s abilities in doing authentic assessment can be seen in Figure 1.

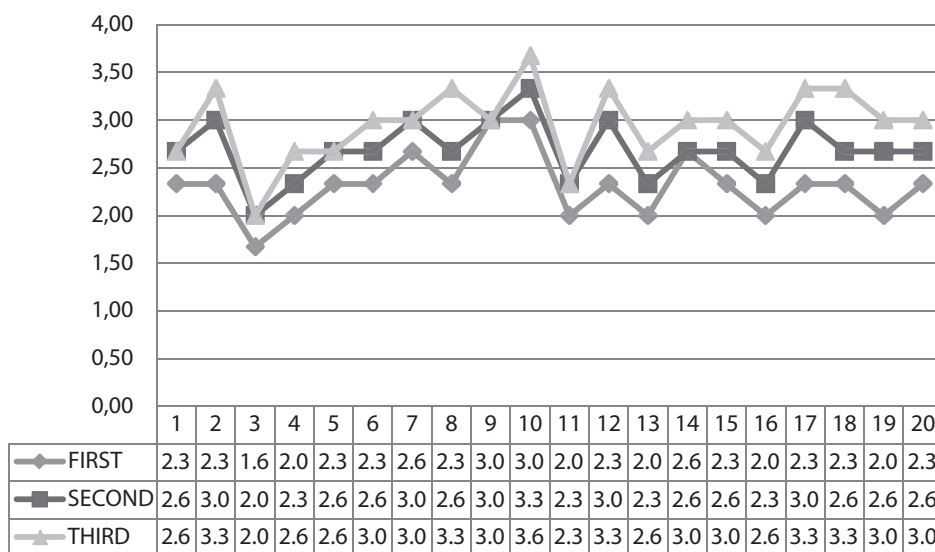


Figure 1. Teacher’s development in doing authentic assessment by competency items

Figure 1 shows that each item of the teacher’s abilities in doing authentic assessment improves from Condition 1, Condition 2, to Condition 3. None of the ability item stays in the same state. Using the foregoing criteria, 16 ability items still have a low score and four have a high score. Based on the results of Observation 1

(Preliminary condition), the teacher was intervened by Intervention 1, specifically on ability items with low scores. Following Intervention 1, the teacher carried out Teaching 2, the observers Observation 2, and the students Assessment 2. Results of Observation 2 show that the teacher's ability items with a high score elevated drastically from 4 to 16, leaving four items of abilities with low scores, these are: Item 3 (Teacher's technique in answering students' questions), Item 4 (Teacher doing skill assessment during instructional interaction), Item 11 (Cognitive level teacher delivers to students), and Item 16 (Teacher prepares notes for doing assessment).

Results of observation 2 were used as an input for Intervention 2. The teacher conducted Teaching 3, the observers Observation 3, and the students Assessment 3, called the post condition. Results of Observation 3 show that the number of ability items with high scores increased to 18 and only two ability items had low scores: Item 3 (Teacher's technique in answering students' questions) and Item 11 (Cognitive level teacher delivers to students). Figure 1 also reveals that Item 3, "Teacher's technique in answering students' questions", had the lowest score in the initial situation. After Intervention 1, i.e. 'teacher given training in answering students', improvement of the 'teacher's ability in answering students' was not significant. Even up to the third condition, this item stays low. The other item that stays low up to the third condition is Item 11, "a cognitive level of the questions asked to students".

Data from the students' assessment are analysed using the Manova technique followed by an Anova and *t*-test. Results the Manova procedure give a Wilks' Lamda value of $F = 23.030$ with $p = 0.00$. This indicates that, according to students, there is a significant difference in the teacher's abilities in assessing attitudes, skills, knowledge, and the consistency in implementing the AA "4C" model, in the periods before and after using the model. The teacher's abilities in these four aspects are higher in the post condition than they are in the pre-condition.

Discussion

Observation results from the three observers in the study indicate that out of 20 items in the natural science teacher's abilities in doing authentic assessment, 16 items have a low score. Of the 16 items, six have a "very low" category: Items 3, 4, 11, 13, 16, and 19. Item 3 is related to techniques used by the teacher to respond to students' questions. It is observed that the teacher seldom answers students' questions well; his answers are inappropriate such as: "How on earth did you ask this question?", "That belongs to Year 1 material. Why did you ask this question?"

etc. It seems that the teacher forgets that attitude skills, that are written in the lesson plan and taught and modelled by the teacher in the class, have an impact on students' attitudes. Kartowagiran & Maddini (2015) have the same conclusion in their study.

Another item with a very low score is Item 4: "Teacher does not assess attitude during the classroom process". This is in line with Kartowagiran & Jaedun (2016) who found that many teachers do not carry out attitude assessment for the following reasons: (1) not enough time; should attitude assessment be done after every lesson; (2) too many students in a class and in the whole school, and (3) the teachers themselves do not know well how to plan and conduct attitude assessment.

Ability Item 11 deals with the level of the teacher's cognitive questions. In the early situation, the teacher's test items are of the multiple-choice type and represent mostly the knowledge level (C1), and the comprehension level (C2). Meanwhile, for the junior high school, teachers need to set up their assessment of application (C3), analysis (C4), evaluation (C5), and even creation (C6). Questions of the higher-order thinking will help students be able to think critically. The study by Abrami, *et al.* (2008) demonstrates that giving students the opportunity to have a dialogue helps students to face problems, gives students examples of critical thinking, and offers students some sort of guidance that will have a positive impact on their skills in critical thinking.

Ability Item Number 13 is related to the teacher's abilities in choosing test types for the daily quiz. In this case, teachers only use the multiple-choice test items and hardly ever use essay types or mixtures of test types or testlets. This is understandable as, according to Edward (Hamdi & Kartowagiran 2018), the testlet aims to combine the strengths of the essay and multiple-choice tests, and they complement the weaknesses of these tests. Item 16 is the fifth that has a very low score, i.e. preparing notes for assessment. In the initial situation, the teacher only used random pieces of paper to take notes in doing the assessment. However, the teacher should make use of a specific book or notebook as it is described in the principles of authentic assessment. Natalia, *et al.* (2018) found that her English-teacher subjects used rubrics or journals for doing attitude assessment.

The last item that has a very low score is Item 19 concerning consistency of the teacher in implementing authentic assessment of the AA "4C" model. The main principle for the teacher is to focus on four students at each time. However, this is not conducted appropriately. Actually, if the teacher is consistent in taking four students at a time, the teacher will be able to obtain quite accurate assessment of the students at the end of the semester. Accuracy of assessment can be achieved

because of an appropriate ratio between students and the teacher in the class. For this, many schools apply team-teaching techniques to maintain small ratios. Good team teaching will be able to improve the students' achievement. It is in line with Ronfeldt, M., *et al.*'s (2015) research. This research involved 9,000 teachers from 336 schools in the Miami-Dade County in two years, and it found that there is a significant positive correlation between the quality of team teaching and the achievement of students.

Besides increasing the accuracy of observation, a focus on only four students at a time makes it easy for the teacher to handle the class. This can be learned from the interview with the teacher who said that, "after applying the AA '4C' assessment model, I no longer feel frustrated in conducting authentic assessment, including assessing students' attitudes since I only have to deal with four items of attitude and four students at each session."

Findings of the study also reveal that two ability items do not show any significant improvement even up to the third condition. Up to the end of the study, these two items stay in the "medium" category. These are Item 3 (techniques used to answer students' questions) and Item 11 (the cognitive level of teacher's questions). Item 3 is unique. The student's question is concerned with knowledge, and the teacher's response is related to knowledge. However, the teacher's answer is related to attitudinal matters. For example, the teacher responded, "That's of last week's class, isn't it?" with a slight tone of blaming the student. This is an attitude that is not too easy to change. Attitudinal abilities are almost identical to personal characters that include the dimensions of moral reasoning, moral feeling, and moral action. It is understood that it often takes long to change from moral reasoning to moral action (Lickona, 2009).

Another element that does not increase during the study is Item 11, concerning the level of the teacher's cognitive questions. Before the research intervention, the teacher's questions range around C₁ and C₂. After receiving the training, the teacher improves to C₂ and sometimes C₃. However, this is still low, and the teacher's abilities in producing higher-order thinking (HOT) questions are still low, too. This is identical with Apino & Retnawati (2017) who found that teachers' understanding of HOT items is low in that case.

Findings that are derived from students' assessment are consistent with those derived from observers' results. According to the students, there is an improvement in the teacher's abilities in assessing attitudes, skills, and knowledge and in the consistency of using the AA "4C" model from the pre- to the post- conditions stage. The fact that the teacher's abilities improve from the first condition to the third condition can be seen as a natural phenomenon. This is because the prac-

tices are done and monitored repeatedly and continuously during the classroom interaction.

The discussion above underlines that, after using the AA “4C” assessment model, the teacher’s abilities in implementing authentic assessment are of the high category. The use of good authentic assessment in learning is bound to produce good learning effects. This agrees with Hargreaves, Earl, & Schmidt (2002) who stated that the use of authentic assessment tends to motivate students to be more responsible towards their own learning, make assessment part of the integral process of interaction, and support students to be more creative in applying their knowledge.

Results of the continuous observation by the three observers show that the teacher’s abilities in implementing authentic assessment improve from Condition 1, to Condition 2, and Condition 3. Neither of the items stays the same throughout Condition 1 to Condition 3. In the initial stage, out of 20 items, sixteen have a low score and only four have a high score. During the second condition, teacher’s ability items drastically increase from 4 to 16. For the third condition, high teacher’s abilities increase to 18 and only two items remain low. This means that there is an increase in the teacher’s abilities in implementing authentic assessment. Results of this observation stay in agreement with the results of the students’ assessment. There is a difference in the teacher’s abilities in doing authentic assessment before and after using the AA “4C” model. Teacher’s competencies in doing authentic assessment after using the model are higher than before using the model.

Conclusion and Research Implication

First, the AA “4C” assessment model is effective in improving the teacher’s abilities in doing authentic assessment; however, the teacher is not fully able to develop test items that have higher-order thinking skills and to select the techniques for responding to students’ questions. **Second**, the AA “4C” assessment model is efficient and effective. It is efficient in that, in each learning session, the teacher is to assess only four students on four character items. It is also effective in the use of the AA “4C” model that can improve teacher’s abilities of doing authentic assessment.

In order to improve teachers’ abilities in doing authentic assessment, they need to be given training in using the AA “4C” model within an adequate time allocation. The model has a comprehensive coverage of assessment and needs to focus only on four students and four items of attitude in each learning session.

Recommendations

First, it is necessary to provide training in the implementation of the AA “4C” model that has been found to be an efficient and affective model. **Second**, development of authentic assessment packages need to be realized for other school subject matters.

Reference

- Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M.A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. DOI:10.3102/0034654308326084
- Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement, 45*(1), 131–142. DOI:10.1177/0013164485451012
- Apino, E., & Retnawati, H. (2017). Developing instructional design to improve mathematical higher order thinking skills of students. *Journal of Physics: Conference Series, 812*, 012100). DOI:10.1088/1742-6596/812/1/012100
- Barber, M., & Mourshed, M. (2012). *Professional development international*. New York: Pearson
- Brown, G.T.L., Irving, S.E., & Keegan, P.J. (2014). *An introduction to educational assessment, measurement & evaluation*. Auckland, New Zealand: Dunmore Publishing Ltd.
- Feldt, L.S., & Brennan, R.L. (1989). “Reliability,” *Educational measurement*, edited by Robert L Linn. New York: Macmillan.
- Frey, B.B., Schmitt, V.L., & Allen, J.P. (2012). Defining authentic classroom assessment. *Practical assessment, research & evaluation, 17*(2).
- Gulikers, J.T.M., Kester, L., Kirschner, P.A., & Bastiaens, T.J., (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction, 18*(2), 172–186. DOI: 10.3102/0162373715576075
- Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). *Multivariate data analysis: Pearson new international edition*. Harlow: Pearson Education Limited.
- Hamdi, S., & Kartowagiran, B. (2018). Developing a Testlet Model for Mathematics at Elementary Level. *International Journal of Instruction, 11*(3), 375–390. DOI:10.12973/iji.2018.11326a
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39*(1), 69–95.
- Kartowagiran, B & Maddini, H. (2015). Evaluation Model for Islamic Education Learning in Junior High School and Its Significance to Students’ Behaviours. *American Journal of Educational Research, 3*, 8, 990–995 DOI:10.12691/education-3-8-7
- Kartowagiran, B., & Jaedun, A. (2016). Model assessment autentik untuk menilai hasil belajar siswa Sekolah menengah Pertama (SMP): Implementasi asesmen autentik di

- SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(2), 131–141 DOI:10.21831/pep.v20i2.10063
- Kartowagiran, B., & Jaedun, A. (2017). Model Asesmen Autentik untuk Menilai Hasil Belajar Siswa SMP. Laporan Penelitian Hibah Penelitian Tim Pascasarjana Tahun ke III (tidak diterbitkan)
- Lickona, T. (2009). *Educating for character: How our schools can teach respect and responsibility*. Bantam.
- Marzano, R.J. Frontier, T., & Livinnhston, D. (2011). *Effective supervision*. Alexandria: ASCD
- Natalia, E.N., Asib, A., Kristina, D. (2018). The Application of authentic assessment for students writing skill. *Journal of Education and Human Development*, 7, 2, 49–53. DOI: 10.15640/jehd.v7n2a5
- Nitko, A. J & Brookhart, S.M. (2011). *Educational assessment of the student*. Boston, MA: Pearson
- Raymond, J.E., Homer, C.S.E., Smith, R., & Gray, J.E. (2012). Learning through authentic assessment: An evaluation of new development in the undergraduate midwifery curriculum. *Nurse Education in Practice*, 30, 1–6. DOI:10.1016/j.nepr.2012.10.006
- Ronfeldt, M., Farmer, S.O., McQueen, K., & Grissom, J.A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, 52(3), 475–514. DOI:10.3102/0002831215585562
- Svinicki, M.D. (2004). Authentic assessment: Testing in reality. *New Directions for Teaching and Learning*, 100, 23–29 DOI:10.1002/tl.167
- Tombari, M.L., & Borich, G.D. (1999). *Authentic assessment in the classroom: Applications and practice*. Upper Saddle River, NJ: Prentice-Hall.
- Vu, T.T., & Alba, G.D. (2014). Authentic assessment for student learning: An ontological Conceptualisation. *Educational Philosophy and Theory*, 46(7), 778–791. DOI:10.1080/0131857.2013.795110
- Wiggins, G. (1998). *Educative assessment. designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.