International Journal of Instruction *October 2019 ● Vol.12, No.4*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp.49-64

Received: 01/01/2019
Revision: 22/05/2019
Accepted: 27/05/2019
OnlineFirst:26/07/2019

Development of Assessment for the Learning of the Humanistic Model to Improve Evaluation of Elementary School Mathematics

Winarno
Assoc. Prof., Tadris Mathematics Program, State Institute for Islamic Studies (IAIN) Salatiga, Indonesia, *winarno@iainsalatiga.ac.id*

Muh Zuhri
Prof., Islamic State Law Faculty, IAIN Salatiga, Indonesia, *muh_zuhri@iainsalatiga.ac.id*

Mansur
Prof., Education Faculty, IAIN Salatiga, Indonesia, *mansur@iainsalatiga.ac.id*

Imam Sutomo
Dr., Lecturer, IAIN Salatiga, Indonesia, *imam_sutomo@iainsalatiga.ac.id*

Khusna Widhyahrini
Dr., Postgraduate Program, IAIN Salatiga, Indonesia, *khusna@iainsalatiga.ac.id*

This research aims at finding the way to develop the assessment for learning of the humanistic (AfL-H) model in Mathematics at elementary schools. This is developmental research of the Hopkins & Clark model. The experimental subjects are 67 mathematics teachers and 375 elementary students in Grade. The data were analyzed in two stages using qualitative and quantitative methods. The experiment of the AfL-H model uses repeated measures analysis. The research results indicate that the AfL-H model developed here has the following main characteristics: 1. the learning provides greater fairness and respects all (learners), and 2. The assessment respects each student's varied abilities based on their own abilities. Several findings in this research are: (1) students receive greater respects regarding the abilities they have in the assessment; (2) the information obtained using the AfL-H model is accurate and as actually needed by students, (3) the application of the AfL-H model in mathematics learning improves student's motivation, confidence, self-awareness, behavior during learning, and ability in Mathematics, and (4) students' learning progress is shown by individual and class profiles.

Keywords: assessment, formative assessment, assessment for learning, humanistic, mathematics

Citation: Winarno, Zuhri, M., Mansur, Sutomo, I., & Widhyahrini, K. (2019). Development of Assessment for the Learning of the Humanistic Model to Improve Evaluation of Elementary School Mathematics. *International Journal of Instruction*, 12(4), 49-64. https://doi.org/10.29333/iji.2019.1244a

INTRODUCTION

Evaluation serves as information for consideration in making decisions in assessing achievement (Kusmaryono et al., 2019; Mcintosh, 1997). Meanwhile, the evaluation can be defined as a process of planning, obtaining, and providing information which is highly needed to make alternative decisions for teacher (Amri, 2017; Sawant, 2016). Evaluation can be used to check to what extent a program has succeeded in relation to its environment and it is also a judgement, whether the program shall be continued, delayed, upgraded, institutionalized, accepted, or rejected (Uttl et al, 2017; Yildizh, 2018).

At the time, and perhaps even now, such a prescription might seem simple, students do not learn what they are taught. Even when instruction was planned with great care, delivered effectively, and in a way that engages students, the learning outcomes often bear little or no relation to what was intended. If what a student learns as a result of a particular sequence of instructional activities is impossible to predict, even in the unlikely event that all the learners in an instructional group are at the same place when the instruction starts, within minutes, students will have reached different understandings. That is why assessment is a, perhaps the, central process in effective instruction. It was only through assessment that we could find out whether a particular sequence of instructional activities has resulted in the intended learning outcomes (Wiliam, 2011; Hursen, 2011).

Nevertheless, what happens in the field, at schools, in both international and domestic contexts, is that formative assessment has not functioned well. In the international context, the finding of the Fair Test Examiner (1999) research shows that formative assessment is relatively infrequently performed in classrooms and most teachers have no idea of how to use this assessment. A similar finding is suggested by Black and William (1998) who find that most tests in classrooms lead to superficial learning and memorization. Teachers failed to help their fellow teachers to be good assessors and they frequently put more emphasis on quantity rather than on quality of works. Teachers usually replicate standardized tests in their own assessment practice; thus, they provide less information ~~on~~ about their students.

Furthermore, in the Indonesian context, formative assessment in mathematics learning is not that different from the international phenomenon. The performing mathematics learning assessment, teachers still use the weak assessment format Zulkardi (2002). Although, formative assessment could be used to increase student and teacher quality by using some indicator properly (Bennett & Bennett, 2011; Rakoczy et al., 2018). The assessment instrument which is poorly designed and emphasizes more on results rather than on the process still dominates their assessment (Bailey et al., 2010; Melguizo et al., 2017). Moreover, Kusnanto (2006) reported results of the survey about distributed to high schoolers in Semarang indicates that students tend to have an incorrect attitude when learning mathematics. It was caused of the assessment system poority.

Recently, an observation has surfaced among education experts in relation to the education concept, education as a humanistic process or it was called as the process of making humans human. The process of making humans human of course goes beyond merely the physical realm, rather it should deal with all dimensions and potentials existing within oneself and the surrounding reality. Education is essentially the process of making mankind human, i.e. being aware of the independent human. An independent human is the creative one as manifested in their culture (Boyd & Richerson, 2009; Tilaar, 2005). In a humanistic education, students are viewed as unique creatures with numerous varied potentials and intelligences. Thus, it will create a democratic learning, which acknowledges the children's rights to do a learning activity which matches their characteristics. What needs to be present in children's learning environment is reality. Children have weaknesses in addition to strengths, courage besides fear, they can be mad, disappointed and happy. Children will be viewed as unique personalities who are capable of developing the potentials, they own optimal. The created learning situation will be more relaxed, fun and gives no burden to students.

Rooted in constructivist social perspective, humanistic education tries to engage students in interactional practices. In this regard, the educator should be able to create social relations together with a positive atmosphere in the classroom, and organize cooperative language work, by enhancing the students emotions and inner self (Cheung et al., 2015). Finding a motivating force in learners but also in herself/himself should be a primary goal to the teacher in order to promote humanistic language teaching, which can influence the personal development of each participant in the classroom (Davis, 2001; Indrayati, 2017). Finally, teachers should also find some time to learn to think about their role and to reflect on what they are going to do in the classroom as well as on what is happening around them. Only if they are willing to explore their own emotional reactions to students along with their potential and power in the classroom, can they grow and develop both personally and professionally and, as a result, can foster the growth and development of the learners‟ knowledge as well (Khatib, 2013)

According to the characteristics of mathematics learning at elementary schools involve: students learning to move from a stage of concrete nature to a more abstract one, and students can use symbols and formal representations and naturally develop from a more concrete stage, shaping logical, critical, creative, careful and disciplined attitudes (McGrath & van Bergen, 2015). Hence, the assessment process of elementary school mathematics learning needs to use a mix of assessment frames and consider the humanistic nature of students

The results from the Research and Development model previously developed by Hopkins & Clark (Havelock, 1976) aimed at testing although the AfL-H model can improve student's mathematics ability. This research aims at (1) finding the way to develop the AfL-H model, (2) discovering what information may ~~be~~ emerge~~d~~ when the AfL-H model is used, (3) finding out the utilization of information resulting from the assessment using the AfL-H model, (4) figuring out whether the AfL-H model could improve students' comprehension, behavior, and ability, and (5) finding the procedure for reporting assessment results using the AfL-H model in mathematics learning. To deal with the weakness of formative assessment, the practice as explained above and to improve the quality of mathematics ability, the researcher offers an assessment model which is integrated with learning, i.e., a combination between of the Assessment for learning (AfL) model and humanistic education that the researcher calls as the assessment for Learning-Humanistic (AfL-H) model.

As an application of the definition, goal, and principles of the assessment for learning of the humanistic in the performance of assessment in the classroom, the syntax of the AfL-H model was prepared. This syntax was the practical guideline and manual for education practitioners, teachers and lecturers in its implementation in the classroom. As a guideline in implementing the model an empirical test, the hypothesis of this research is then formulated as "The application of the AfL-H model can improve the Quality and Evaluation of Elementary School Mathematics".

LITERATURE REVIEW

The assessment for learning concept is basically nothing new in educational assessment. However, the form of its implementation in the context of learning quality improvement and refinement, this assessment for learning is better, well-planned, more directed and focused. These are at least reflected by the definition of assessment for learning stated in the Assessment Reform Group (2002) which suggests that assessment for learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there. The emphasis of this definition of assessment for learning is on the process of obtaining and utilizing information.

Information is obtained through cooperation between teachers and students and this information is used by them (teachers and students) for improving and refining the next teaching and learning quality. For teachers, this information is used to improve and revise the teaching strategy as is actually needed by their students. Meanwhile, for students, it can be used as the basis in modifying their learning strategy which suits them better. Similar definition is also proposed in Pearson Education (2006), i.e., assessment for learning is a collaborative process between teacher and students, and with students engaging with each other in structuring their own learning. It is built on a foundation of shared learning objectives and shared criteria for success. Students are given the criteria for success and the support they need to achieve that success. Feedback, either during or on completion of the task, is essential if students are to know what else must be done to ensure further learning. Students are provided with opportunities to participate in self- or peer-assessment as this develops an understanding of personal responsibility in learning (Alnasser & Alyousef, 2015). It can be seen in the last definition that the emphasis is on the collaboration between teachers and students and among students themselves (Andrian, 2018). Their collaboration is related to teaching and learning activities in the effort of making all students successful. To achieve this success, teacher's task such as sharing the teaching and learning goals and success criteria at the beginning of the lesson has also been an emphasis of this definition.

Having assigned problems in a structured way for students to work on, then teachers should give feedback to their works, hence information will be obtained regarding students' strengths and weaknesses. The AfL model in this research adopts the two definitions of assessment for learning suggested by the Assessment Reform Group and Pearson Education. Meanwhile, the term humanistic means placing human in the potential they have and giving emphasis on the collaboration between teachers and students as well as utilizing information resulting from assessment to modify the teaching and learning strategy and technique (Indrayati, 2017). The operationalization of definition of assessment for learning within the framework of purpose is stated in CEA (2003) which says that the purposes of assessment for learning are to: 1). provide teaching and learning insight to teachers and students in an effort to improve success for all, 2). help the process of setting goals, 3). enable continuous reflection on what students know now and what they need to know next, 4). measure what is assessed, 5). set the right and immediate intervention needed to achieve the teaching and learning goals, and 6). improve the standard obtained by students. In relation to these purposes, the success in teaching and learning is intended for both teachers and students. Teachers are demanded to have the insight and professional competence in teaching and learning, such as mastering the contents, planning the lessons, setting learning goals, and making the right decisions, so that students are motivated to revise and improve their learning.

To realize the purposes explained above, the Assessment Reform Group (2002) gives the ten main principles, that are assessment for learning: 1). should be part of effective planning of teaching and learning, 2). should focus on how students learn, 3). should be recognized as central to classroom practice, 4). should be recognized as a key professional skill for teachers, 5). should be sensitive and constructive because any assessment has an emotional impact, 6). should take account of the importance of learner motivation, 7). should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed, 8). learners receive constructive guidance about how to improve, 9). develops learners capacity for self-assessment so that they can become reflective and self-managing, and 10). Should recognize the full range of achievements of all learners

Teachers became key participants in the learning process. As Palmer explains, teachers need to resume their traditional roles as mentors (1997). Education in a democracy was geared toward and powered by a particularly precious and fragile ideal. This ideal is founded on the belief that every person possesses an "infinite and incalculable value". Based on this premise, the role of the educator becomes much more complex than that of simple transmitter of information. So, eloquently explains, the pedagogy of questioning is part of a concerted effort to liberate and humanize education (Ayers, 2009)

Meanwhile, according to the humanism theory, the learning process should begin with and be addressed to the purposes of making humans human. "The nature of humanistic learning theory is more abstract and closer to the field of philosophy, personality theory, and psychotherapy, rather than the study of learning psychology" (Leonard, 2002). The assessment for learning of the humanistic theory prioritizes heavily the learning process to achieve the learning goals. The humanistic learning theory talks more about the concepts of education to shape the idealized human being, humanistic mankind, as well as on the learning process in the most ideal form. In other words, this theory is interested more in the learning process in its most ideal form**.**

According to Mangunwijaya (2001) the main concept of humanistic educational thought is to respect human dignity. Brown and Knight (2004) suggested the fundamental matter in humanistic education was the desire to realize an educational environment which sets students free from tight competition, high discipline, and fear of failure. Humanistic education as follows: Humanistic education was came from the teaching of humanism assumptions (Friedman & MacDonald, 2006). The education model was more of humanity education rather than education on specific knowledge for certain professions. Education is a general education, rather than a specialist one. The interpretation of unique strengths of humans can basically result in the same form as non-specialist education referred to as humanistic.

Humanistic education views the human as a subject who was free and independent to determine where their life is heading. Humans are fully responsible for their own life and also for the life of others. Therefore, the assessment for learning of the humanistic education can never force its wish upon any child. Educators should help students develop themselves, i.e. help each individual recognize~~s~~ themselves as a unique human and realize the potentials existing within. Any goal which does not match the kid's potential cannot be the goal of humanistic education.

The assessment for the learning-humanistic model developed has such main characteristics as, 1). The learning provides greater fairness and respects all (learners), and 2). The assessment made respects each student's varied abilities based on their own abilities (Friedman & Mcdonald, 2006). Students receive greater respect regarding the ability they have in the assessment. The information obtained through the use of the AfL-H model is accurate and matches what students actually need and the application of the AfL-H model in mathematics learning improves students' motivation, confidence, self-awareness, behavior during teaching and learning, and students' ability at mathematics; and, finally, students' learning progress is presented via individual and class profiles.

METHOD

This is developmental research by adopting the model developed by Hopkins & Clark, i.e. the R, D & D model" (Havelock, 1976). During the research stage, ~~4~~ four activities are done, namely preliminary research, study of research results, curriculum analysis, and model prototype preparation. During the development stage, five ~~5~~ activities are done, namely expert validation, readability test, teacher training, limited trial, and expanded trial. During the diffusion stage, three activities are done, namely dissemination, training, and demonstration. The design of this developmental research was shown in Figure 1.

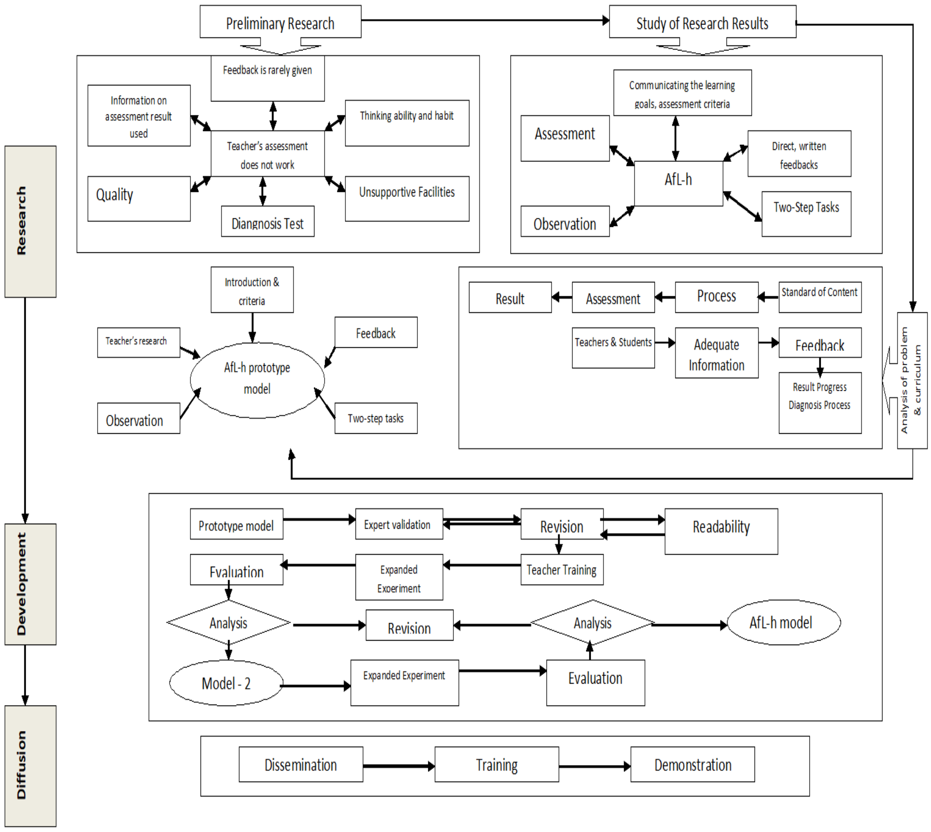

Figure 1
Design of AfL-H Model Development

Model refers to the design presented above. Operationally, the development procedure performed include: 1) Preparing the model set, i.e., practical guidelines for using the model, instrument for assessing the model effectiveness, instrument for assessing students, two-step task instrument, scoring criteria and rubric, and lesson plan; 2). Expert validation; 3). Revision; 4). Readability of model set to elementary school mathematics teachers; 5). Expert validation, readability, and revision of repeated process; 6). Teacher training; 7). Limited trial; 8). Model revision; 9). Expanded trial; and 10). The AfL-H model which fits the theoretical framework and the empirical data is found.

The empiric verification of the model effectiveness within the development research framework uses a quasi-experimental approach with single-group interrupted time-series design (Creswell, 1994). The Single-Group Interrupted Time Series was shown at Table 1.

Table 1
Single-Group Interrupted Time-Series Design for Model Experiment

Meeting	1	2	3	4	5
Class A	X/O1	X/O2	X/O3	X/O4	X/O5

Note:
X = treatment
Oi = students measurement Mathematics ability, i = 1, 2, 3, 4, 5.

The limited and expanded trials were performed for two months (equal to 16 meetings), starting from November 2017 to February 2018, with the measurement being performed five times. The treatment and observation are done consecutively in one meeting. Prior to the trial, the researcher first controls the internal and external validities. The research respondents consist of teachers and students. The involvement of these respondents depends on the need at the development phase. For the limited trial respondents, one class is chosen as the treatment class. Four respondents from teachers are chosen who teach mathematics in the school where the limited trial (State Elementary Schools 1 and 2 Salatiga) is held. One of these teachers is assigned the task to teach, and two others are to be the observers. For the trial subject in the expanded trial, two schools are chosen, each consisting of two classes. Each class was taught by a teacher and observed by an observer (from teachers). The total number teachers serving as the research respondents is 20 and prior to their teaching in the classroom they are trained by the researcher on how to use the AfL-H model. The research subjects involved at the model development phase are presented in Table 2.

Table 2
Distribution of Research Subject

Subject from	Number of Trial Subjects			Total
	Readability	Limited	Expanded	
Students	53	99	223	375
Teachers	17	20	30	67

The method to collect data and the instrument to be used in this development consist of various techniques whose uses are adjusted to the need and the type of data collected according to the development activity stages. Basically, there are two instrument groups, i.e. data collecting instrument and treatment instrument. The treatment instrument consists of students' self-assessment, and two-step task. This treatment instrument serves as data collecting instrument at the same time. So, the data collecting instrument consists of a behavior observation sheet, two-step task (to measure the student's ability and misconceptions ~~to~~ in the learning content given), and a students' self-assessment sheet. In regard to the validity and reliability of the measurement results of the instruments used, they will be subjected to validation (Nitko & Brookhard, 2007). The two-step task instrument will be seen for its test gauge requirements including difficulty level, reliability, validity, and item discrimination power (Cohen & Swerdlik, 2005). Furthermore, to measure the inter-rater reliability of such instruments as self-assessment sheet, behavior observation sheet, model effectiveness questionnaire, model applicability in classroom, and model assessment sheet of the validation result, the

Cohen's Kappa coefficients and percentages of agreements (Cohen, 2001; Grinnell, 1988; Tosto et al., 2016) are used. To calculate the Cohen's Kappa (κ) coefficient, the formula suggested as below:

$$K = \frac{\sum f_o - \sum f_e}{N - \sum f_e}$$

Where:

K = inter-rater reliability level (inter-rater reliability coefficient)

f_o = observation result frequency

f_e = expected frequency

N = number of question items rated (clarified)

Moreover, to calculate the level of percentages of agreements between both raters which contain only yes or no responses, the formula suggested as below:

Percentages of agreements = $\frac{agreements}{dissagreements + agreements}$ X 100%

(Grinell, 1988)

The lower limit of reliability coefficient used for a good test was ammount 0.70 (Linn, 1989). The data from preliminary research were analyzed using a descriptive approach equipped with a narrative which serves research purpose. The data during the model development are analyzed using both qualitative and quantitative approaches. The qualitative analysis is performed to analyze the data from model validation by experts who give feedback for the purpose of improving the model. The qualitative analysis is also used to analyze the data resulting from students' self-assessment against the open-ended questions that they were given. The analysis to figure out student's ability development, self-assessment, and behavior during or after the AfL-H model trial uses repeated measures analysis (Field, 2000). The data analysis design was presented in Table 3.

Table 3
Repeated Measures Design for Data Analysis

Meeting						
1	2	3	4	5		
Subject	1	Y11	Y12	Y13	Y14	Y15
2	Y21	Y22	Y23	Y24	Y25	
3	Y31	Y32	Y33	Y34	Y35	
N	Yn1	Yn2	Yn3	Yn4	Yn5	

FINDINGS AND DISCUSSION

In this study to obtain the AfL-H model make Focus Group Discussion (FGD) was conducted with 3 measurement experts, 3 mathematicians and 3 education experts. The results of model validation by experts indicate that the model set acts as if it is a guideline and the model instruments have met the validity and reliability requirements. All results of expert validation confirm that the model is feasible for use without any revision. The results of empirical tests of model effectiveness constitute~~s~~ the answers to the research hypotheses. The empirical test of the AfL-H model effectiveness is conducted using repeated measures analysis. In this analysis, the variable to which the attention is paid to is students' ability in repeated measures. In this case the measurement is done for ~~5~~ five times in two schools, each for two classes. To analyze using repeated measures, an assumption test was performed, i.e. sphericity assumption (Li et al., 2014; Field, 2000, Stevens, 1996). For the purpose of testing this assumption, Mauchly's Test of Sphericity is used. The tested assumption is:

Ho : The covariance matrix of errors resulting from orthonormal transformations of the variable under consider = opponent identity matrix

H1 : The covariance matrix of errors resulting from orthonormal transformations of the variables observed ≠ identity matrix

Test criteria: accept the hypothesis Ho if the significance value from the calculation of $p \geq \alpha$, otherwise Ho is rejected (Miller and Miller, 2004).

Results of this calculation of assumption test are summarized in Table 4. From the information presented in Table 4, it is found that the statistical value for Mauchly's test was 0.235 and the significance from the calculation of *p* is 0.000001. If the significance value from the calculation is compared to the significance $\alpha = 0.05$, then $\alpha > p$. Thus, the sphericity assumption is unfulfilled. Field (2000: 333) says that if the sphericity assumption is unfulfilled, then a correction can be made by seeing one of the Epsilons (Greenhous e-Geisser, Huynh-Feldt, or Lower-bound). In this analysis, the researcher chose the Greenhous e-Geisser Epsilon whose significance value from the calculation of ε is 0.863. If this ε value is rated against the criteria, i.e. $\leq \varepsilon \leq 1$ (k being times of measurement, in this case k = 5), thus the obtained ε is within this interval. Hence, it can be concluded that the sphericity assumption is fulfilled, therefore the repeated measures test can proceed. Summary of the Result of Calculation of Mauchly's Test of Sphericity was shown at Table 4.

Table 4
Mauchly's Test of Sphericity Calculation

Measure: MEASURE 1

Within Subject Effect	Mauchly's W	Approx Chi-Square	df	Sig.	Epsilon		
					Greenhous e-Geisser	Huynh-Feldt	Lower-bound
Meeting	.345	210.89	9	.000	.875	.765	0.25

Tests the null hypothesis that the error covariance matrix of the Orto-normalized transformed dependent variable is proportional to an identity matrix

The requirement of repeated measures analysis is fulfilled, therefore, the next analysis can be done. Results of calculation of repeated measures analysis of the multivariate test are summarized in Table 5. Treatment (application of the AfL-H model) for each meeting and interaction between meetings with the classes show significant influence.

At least, this is shown by the significance value from the calculation of (p) of each meeting and interaction of meeting with the classes being less than the chosen significance value $\alpha = 0.05$. Therefore, it can be concluded that the application of the AfL-H model to mathematics learning gives significant influence on students' increased mathematics ability. A summary of the Results of Calculation of the Multivariate Test in the analysis are presented at Table 5.

Table 5
Multivariate Test in the Analysis Calculation

Measure: MEASURE 1						
Effect		Value	F	Hypothesis df	Error df	Sig.
Meeting	Pillai's Trace	.521	24.086[a]	4.000	144.000	.000
	Wilks' Lambda	.602	24.086[a]	4.000	144.000	.000
	Hotellng's Trace	.689	24.086[a]	4.000	144.000	.000
	Roy's Largest Root	.689	24.086[a]	4.000	144.000	.000
Meeting* Class	Pillai's Trace	.325	4.471	12.000	438.000	.000
	Wilks' Lambda	.700	4.579	12.000	381.280	.000
	Hotellng's Trace	.389	4.628	12.000	428.000	.000
	Roy's Largest Root	.255	9.323[b]	4.000	146.000	.000

a. Exact statistic
b. The statistic is un upper bound on F that yields a lower bound on the significance level.

This multivariate test result is confirmed by the result of tests of within-subjects effects as presented in Table 6.

Table 6
Test of Within-Subjects Effects in Repeat Measures Analysis

Measure: MEASURE 1						
Source		Type III Sum o Squares	df	Mean Square	F	Sig.
Meeting	Sphericity Assumed	478.179	4	119.520	74.611	.000
	Greenhouse-Geisser	478.088	2.503	191.025	74.611	.000
	Huynh-Feldt	478.088	2.601	183.772	74.611	.000
	Lower Bound	478.088	1.000	478.079	74.611	.000
Meeting*Class	Sphericity Assumed	80.265	12	6.688	4.175	.000
	Greenhouse-Geisser	80.265	7.508	10.690	4.175	.000
	Huynh-Feldt	80.265	7.804	10.284	4.175	.000
	Lower Bound	80.265	3.000	26.754	4.175	.007
Error (Meeting)	Sphericity Assumed	941.835	588	1.602		
	Greenhouse-Geisser	941.835	367.896	2.560		
	Huynh-Feldt	941.835	382.418	2.463		
	Lower Bound	941.835	147.000	6.408		

Computed using $\alpha = 0.05$

Observing Table 6, it was clear that treatment (application of the AfL-H model) for each meeting and interaction between meetings with classes show significant influence. This is shown by the significance value from the calculation of each (p) in the meeting row (p = 0.0001) and meetings* classes (p = 0.0001) being less than the chosen significance value α= 0.05. Therefore, it can be concluded that the application of the AfL-H model to mathematics learning gives significant influence on students' increased mathematics ability. Furthermore, test was done to discover whether the influence is linear or not. For this purpose, the tests of within-subjects contrast was used and its results are summarized in Table 7. It can be seen in Table 7 at the meeting row that the appropriate influence is linear. This is shown by the significance value from the calculation of p at 0.00001 being less than the significance α= 0.05 for being linear. Therefore, it can be said that an increase to the number of meetings in the application of the AfL-H model to mathematics learning will improve students' ability in mathematics. A summary of Results of the Calculation of Tests of Within-Subjects Contrast in Repeated Measures Analysis as presented in Table 7.

Table 7
Test of Within-Subjects Contras in Repeated Measure Analysis

Measure : MEASURE 1

Meeting		Type III Sum of Squares	df	Mean Square	F	Sig
Meeting	Linear	479.01	1	469.01	138.258	.000
	Quadratic	7.156	1	7.066	5.464	.021
	Cubic	1.75	1	1.74	2.055	.154
	Order 4	.265	1	.262	.300	.585
Meeting* Class	Linear	44.827	3	14.941	4.404	.005
	Quadratic	23.573	3	7.845	6.066	.001
	Cubic	5.14	3	1.707	2.016	.114
	Order 4	6.775	3	2.261	2.582	.056
Error (Meeting)	Linear	498.688	147	3.392		
	Quadratic	190.117	147	1.293		
	Cubic	124.425	147	.846		
	Order 4	128.714	147	.876		

Next step is to see which meeting pairs give different means. For this purpose, results of the post hoc test (in this case using the Bonferroni method) was observed, whose summary of results is presented in Table 8. It can be seen in this Table 8 that the significance values from the calculation of p for all pairs observed are the same at 0.00001. If the significance value from the calculation is compared to the significance α = 0.05, then $\alpha > p$. Therefore, it can be concluded that each pair of meetings observed gives significant difference of student's ability. A summary of Results of the Calculation of the Post Hoc Test in Repeated Measures Analysis of Meeting are presented in Table 8.

Table 8
Post Hoc Test in Repeated Measures Analysis of Meeting Calculation

Measure : MEASURE 1				
(I) Metting	(J) Meeting	Mean Difference (I-J)	Stdr error	sig
1	2	-.808*	-159	.000
3		-1.405*	.168	.000
4		-1.792*	.165	.000
5		-2.322*	.191	.000
2	1	.819*	.155	.000
3		-.599	.137	.000
4		-.965*	.095	.000
5		-1.490*	.115	.000
3	1	1.407*	.165	.000
2		.598*	.139	.000
4		-.384*	..090	.000
5		-.895	.116	.000
4	1	1.790*	.162	.000
2		.985*	.145	.000
3		365*	.092	.000
5		-.521*	.072	.000
5	1	2.301*	.191	.000
2		1,484*	.164	.000
3		.897*	.11	.000
4		.523*	.0.073	.000

Based on estimated Marginal means
* The mean difference is significant at .05 level
a. adjustment for multiple comparison: Bonferroni

After observing the differences in students' ability for each meeting, the next step is to see whether the difference in classes gives different students' ability in mathematics. To figure it out, observe Table 9. From the information presented in Table 9, it can be seen that the significance value from the calculation of p for the class pair VII-5 and VII-A is 0.021 and the class pair VII-5 and VII-B is 0.005. If this significance value from the calculation is compared to the significance α= 0.05, then α> p. Therefore, it can be concluded that the two class pairs have different abilities in mathematics. The other four pairs have no differences or have similar ability in mathematics. A summary of the Results of Calculation of the Post Hoc Test in Repeated Measures of Class Case as presented in Table 9.

Table 9
Calculation of the Post Hoc Test in Repeated Measures of Class Case

Class VII (I)	Class VII (J)	Mean Difference (I-J)	Std. Error	Sig.
Class VII-1	Class VII-5	-.2885	.34505	1.000
	Class VII-A	.7246	.33111	.181
	Class VII-B	.8878	.33318	.052
Class VII-5	Class VII-1	.2874	.34505	1.00
	Class VII-A	1.0121*	.34096	.021
	Class VII-B	1.1749*	.34296	.005
Class VII-A	Class VII-1	-.7247	.33111	.181
	Class VII-5	-1.0121*	.34096	.021
	Class VII-B	.1629	.32893	1.000
Class VII-B	Class VII-1	-.8875	.33318	.052
	Class VII-5	-1.1749*	.34296	.005
	Class VII-A	-.1629	.32893	1.000

Based on Observed means
* The mean difference is significant at -05 level

Based on the results of repeated measures analysis above, it is found that the application of the AfL-H model to mathematics learning gives significant influence on students' increased mathematics ability. This students' increased ability follows a linear trend, i.e., each increase in numbers of meeting (of course it is equipped with the assignment of two-step tasks to measure students' ability) will give students better understanding on the contents of mathematics. This fact indicates that the AfL-H model applied to the teaching and learning is effective in improving the quality of mathematics learning. In addition to influence increased ability, the research results also show that it has some influence on the students' improved awareness, motivation, responsibility, and behavior in teaching and learning. Teacher can improve their teaching and learning system towards to humanistic assessment, they use humanistic approach to support the student based on constructivism, integration of thinking and human empower (Gurses et al., 2015).

CONCLUSION

Based on the research results and discussion above in the previous part, the conclusions which are also the findings of this research are: 1) the AfL-H model in mathematics learning at elementary schools is developed using research and a developmental method. The research stages include pre-survey, problem analysis, curriculum analysis, study of research results, expert consultation, and model prototype preparation. Afterwards, the development stage includes 1) expert validation, readability test, teacher training, limited trial, and expanded trial so that the model fits both theoretically and empirically; 2) the information emerging when using this AfL-H model in learning is the accurate information which matches students' actual need in terms of their comprehension on the lesson content, students' behavior during teaching and learning, and their mathematics ability; 3) the use of information resulting from the assessment in the AfL-H model is made through feedback and reflection, 4) the teaching and learning give greater fairness and respect all (students), 5) the assessment made appreciates the varied abilities of each students based on their own abilities, 6) students receive greater appreciation for the abilities they own in the assessment, and 7) students' learning progress is shown through individual and class profiles. Both profiles indicate a trend of developing individual (or class) comprehension of the lesson content, students' behavior during teaching and learning, and students' ability in mathematics for each meeting. The application of the AfL-H model to the assessment of elementary school mathematics learning improves student's comprehension, behavior, and mathematics ability.

REFERENCES

Andrian, D. (2018). The instrument development to evaluate local curriculum in Indonesia. *International Journal of Instruction*, *11*(4), 921-934.

Bailey, T., Jeong, D. W., & Cho, S. (2010). Economics of education review referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255-270.

Bennett, R. E., & Bennett, R. E. (2011). Assessment in education: Principles, policy & practice formative assessment: A critical review formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions Royal Society B*, *364*, 3281-3288.

Brown, S., & Knight, P. (2004). *Assessing learners in higher education*. London: Kogan Page.

Cheung, K. C., Taylor, R., Cheung, K. C., & Taylor, R. (2006). Towards a humanistic constructivist model of science learning: changing perspectives and research implications. *Journal of Curriculum Studies*, *23*(1), 37-41.

Davis, H. A. (2001). The Quality and impact of relationships between elementary school students and teachers. *Contemporary Educational Psychology*, *26*(4), 431-453.

Gurses, A., Demiray, S., & Do, C. (2015). A Design practice for interactive-direct teaching based on constructivist learning (IDTBCL): Dissolution and solutions. *Procedia-Social and Behavioral Sciences*, *191*, 44-49.

Harris L. F., & Douglas A. M. (2006). Humanistic testing and assessment. *Journal of Humanistic Psycology*, *46*(4), 510-529.

Hursen, C. (2011). Psychological counselling and guidance departments students assesment on learning to learn perception of competence. *Procedia-Social and Behavioral Sciences*, *30*, 2452-2456.

Indrayati. (2017). The implementation of PAIKEM based on project-based learning and cooperative learning. *IOSR Journal of Humanities and Social Science*, *22*(7), 37-43.

Kusmaryono, I., Suyitno, H., Dwijanto, D., & Dwidayati, N. (2019). The effect of mathematical disposition on mathematical power formation: Review of dispositional mental functions. *International Journal of Instruction*, *12*(1), 343-356.

Li, C., Xue, Y., Hoyningen-huene, W. Von, & Zhang, J. (2014). Quantitative evaluation of uncertainties in satellite retrieval of dust-like aerosols induced by spherical assumption. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *136*, 45-57.

McGrath, K. F., & Van Bergen, P. (2015). Who, when, why and to what end? Students at risk of negative student-teacher relationships and their outcomes. *Educational Research Review,* 14, 1-17. https://doi.org/10.1016/j.edurev.2014.12.001.

Mcintosh, M. E. (1997). Formative assessment in mathematics. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *71*(2), 92-96. https://doi.org/10.1080/00098659709599333.

Melguizo, T., Kosiewicz, H., Prather, G., Bos, J., Melguizo, T., Kosiewicz, H., … Melguizo, T. (2014). How are community college students assessed and placed in developmental math? Grounding our understanding in reality. *The Journal of Higher*, *85*(5), 691-722.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2016). Formative assessment in mathematics: Mediated by feedback’s perceived usefulness and students’ self-efficacy. *Learning and Instruction*, *25*(2), 160-182.

Sawant, D. G. (2016). Role of IQAC in maintaining quality standards in teaching, learning and evaluation. *Pacific Science Review A: Natural Science and Engineering*, *2*(2), 66-69.

Suliman, M. A., & Hesham S. A. (2015). Improving the effectiveness of the peer feedback technique: The impact of focusing EFL student-writers on macro level features. *International Journal of English Language Education*, *3*(1), 92-112.

Tuba, A. E., & Hulya. Y. (2018). Classroom assessment practices of teachers in Turkey. *International Journal of Instruction*, *11*(3), 587-602.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty’s teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22-42.