Available online https://ejournal.ressi.id/index.php/arisen

Assessment and Research on Education, 1 (2), 2019, 93-105

Authentic assessments effectiveness on students' motivation and interactions in English speaking

Angesti Palupiningsih

SMP Budi Mulia Dua. Jalan Panjen Wedomartani Bimo Martani Ngemplak Sleman, 55281, Indonesia
Corresponding Author. Email: palupiangesti@gmail.com

Abstract

This research aims at (1) revealing the differences of motivation and interaction between the students who are assessed using authentic assessment and those who are not in the English speaking practices and 2) trying out whether or not authentic assessment is more effective than non-authentic assessment on the students' motivation and interaction to deal with English speaking practices of SMPN 2. This study was a quasi-experimental study using non-equivalent with one control group and one experimental group design. The data collecting instruments were a questionnaire for students' motivation measurement and a questionnaire for students' interaction measurement. The validity of the instruments was gained using expert judgment and item validity. Bivariate Pearson was employed to check the item validity, whereas the reliability employed Alpha Cronbach. To see the difference in the effect of authentic assessment to the students' motivation and the students' interaction, the data were analyzed using the MANOVA at the significance level of 5%. The results are as follows. (1) There is a different effect of the authentic assessments on the students' motivation and interaction. (2) Authentic assessment is more effective than non-authentic assessment on the students' motivation and interaction to deal with English speaking practices of SMPN 2.

Keywords: authentic assessments, students' motivation, students' interaction, English speaking practices

INTRODUCTION

The first indicator of mastering a language is able to communicate orally using that language (Nunan, 1999, p. 225). Moreover, (Harmer, 2003) states that the ability to speak fluently indicates the ability to process information and language 'on the spot'. Therefore, speaking needs to be trained as early as possible.

One of the reasons why speaking is important, in Indonesian context, is because Indonesia welcomes AFTA in 2015. Suyanto, quoted by (Nurhadi, 2013) said that human resources in Indonesia should have competency to communicate verbally. That is why it is important that Indonesia students also have to be accustomed to speak English.

However, most students failed in speaking. It can be seen from their score in doing the English test. Most of them get high scores but they still cannot speak English. They thought that speaking is not important since there is no speaking test. As the consequence, they are not motivated. In fact, one of the key successes of learning is motivation (Gardner, 2012; Harmer, 2003; Lennon, 1993). Motivation in speaking still becomes a problem. As what (Hidayanti, 2011) mentions, students' motivation in speaking is low because they think that it is difficult since English is not their own language. That was a problem which is faced by teachers. It can be seen from their behavior, such as students do not want to talk or say anythings using English, they just end up in chatting in their own language, and finally they make too much noise and lose control.

What has been mentioned above can be as a backwash effect of the assessments. From the students' viewpoint, the curriculum is defined by assessment (Conti-Ramsden & Durkin, 2012; Surgenor, 2010). It means that, students do not pay attention on what is curriculum prevail but

in what is the assessment. As a result, students learn what they think they will be assessed on – this term has been called backwash (Elton, 2000; Surgenor, 2010). They know that their effort will be assessed by a standard test called National Examination which emphasizes on reading. Although in the classroom assessment of speaking, they still meet assessments in a traditional model of assessment. They are simply asked to do multiple choice tests or memorize a dialogue and then perform it in front of the class. In the real life, speaking is not memorizing. As it was stated before, the ability to speak fluently indicates the ability to process information and language 'on the spot'. In other words, those traditional assessments for speaking are not authentic. Meanwhile, the authenticity of an assessment is influential on students' motivation. As what (Harmer, 2003) says the need to pass an exam is one of the outside factors that influence extrinsic motivation. Therefore, the students' motivation in speaking is low. It is because they think that there is no use to speak actively because it will be useless in their life or there is no relation between speaking in the classroom and speaking in the real life. As a result, the interaction in the speaking class is also low.

According to (Harmer, 2003, p. 51), it is accepted for most fields of learning that motivation is important to success. Without motivation, learners will almost certainly fail to make the necessary efforts. At its most basic level, motivation is some kind of internal drive to push someone to do things in order to achieve something (Harmer, 2003, p. 51). Motivation is influenced by a lot of factors, whether it comes from the students or something beyond them. In a speaking class, the students are different. Some of them are motivated but some are not. Therefore, teachers must work for this. In some speaking classes, students are still assessed in traditional assessment. They are asked to memorize a text, and then perform it and they are also assessed in multiple choice/matching tests. In the real life, it does not work.

Besides, it will not facilitate the students to interact with others. In a speaking class, interaction is essential. Interaction is mediating and supporting the practice of learning (Kumpulainen & Wray, 2003). Authentic tests are expected to give the students a clear condition for speaking. It is because communication does not always happen in monologue but also in dialogue, both interpersonal and transactional dialogue. How can teachers assess the students' interaction when the tasks are only memorizing or selecting A, B, C options?

There are a lot of strategies to encourage students' motivation and interaction in language teaching and learning. For example: using media, designing interesting tasks and materials, managing the settings, and using varieties of methods. However, sometimes some teachers forget the backwash effects of an assessment are really influential for students. It raises their motivation in learning. Therefore, teachers should turn their assessment model to a model of assessment that encourages students to really 'speak' like in the real situation. It is assumed that it would affect their motivation to speak because their 'real speaking' would be assessed. In other words, the assessment should assess their authentic learning. Therefore, the appropriate assessment model is authentic assessments. As what Biggs and Tang state (Surgenor, 2010), the assessment will ensure that students are also learning and demonstrating the intended outcomes.

Authentic assessment is a term used to describe real tasks that require students to perform and/or produce knowledge rather than reproduce information others have discovered (Vyortkina, 2003). It means that this assessment emphasizes on student's abilities in demonstrating their knowledge in meaningful activities or real-world tasks. This assessment not only asks students' knowledge but show his/her performances of the mastered knowledge and skills. There are many kinds of authentic assessments, such as performances, projects, writings, demonstrations, debates, simulations, presentations, or other sorts of open-ended tasks. Those can be applied in all subjects, one of them is English.

There are some theories that link authentic assessments and students' motivation and interaction. Increasing the authenticity of an assessment is expected to have a positive influence on student learning and motivation (Herrington & Herrington, 2007; Komsi, Hambali, & Ramli, 2018; Schuyten & Ferla, 2007). In line with this, (Harmer, 2003, p. 116) says genuinely

communicative speaking activities facilitate the students a real desire to speak and a communicative purpose for doing so. It is also supported by the cognitive theory, students should be viewed as active learners who discover and construct meaning and to help them set goals, plan, and reach goals; associate and link new information with existing knowledge in meaningful ways; think reflectively, critically, and creatively; develop self-monitoring skills; have positive expectations for learning and confidence in their skills; be enthusiastically and internally motivated to learn; apply what they learn to real-world situation; and communicate effectively.

Therefore, the appropriate model of assessments to improve motivation and interaction is authentic assessments. Authentic assessment meets the above requirements. Assessment will encourage students to become more meaningfully involved in the subject matter and more intrinsically motivated to study the topic. Assessment must be challenging but fair. It will increase students' enthusiasm for learning (Santrock, 2010, p. 551). If it is too difficult will lower students' self-esteem and self-efficacy, as well as raise their anxiety. On the other hand, if it is too easy will bore them and not motivate them to study hard enough. In each assessment event, a teacher communicates with the students (Brookhart & Nitko, 2008; Santrock, 2010, p. 552). That one of the ways of how authentic assessment can improve students' interaction. She also suggests that teacher should evaluate students using variety of performances, especially performances that are meaningful to students (Santrock, 2010, p. 552). Many classroom assessment experts emphasize that if teachers think that motivated and active learning are important goals of instruction, they should create alternative assessments that are different from traditional assessments which do not evaluate how students construct knowledge and understanding, set and reach goals, think critically and creatively. Moreover, there is a theory uttered by (Harlen & Crick, 2003) which corroborate the above theories state that assessments give effect on students, curriculum, and teachers. Two of the students' effects are students' motivation and interaction (Brown & Abeywickrama, 2004; Harmer, 2003; Rivers, 1987; Rosenthal & Zimmerman, 2014; Stiggins, Arter, Chappuis, & Chappuis, 2004).

Based on the theories and stated problems above, some hypotheses rise. (1) There are differences between students' motivation and interaction in speaking who are assessed using authentic assessment and those who are not; (2) Authentic is more effective than non authentic assessment on the students' motivation and interaction to deal with English speaking practices.

METHOD

To answer the hypotheses above a quasi-experimental research design was employed. Population of this research was all year-8 students of SMPN 2 Kalasan in the academic year of 2013/2014, consisting of 192 students. The sample of this research consisted of two groups of students that were selected randomly from the population. It used cluster random sampling. It is because the population has already grouped and every group has the same opportunity. It is done by taking two lotteries consisted of the class' name. One was as the experimental group and the other was as the control group. All year-8 classes were drawn to determine experimental group and control group.

There are three variables in this research. The first is the authentic assessment which belongs to the independent variable. The second is students' motivation which belongs to the dependent variable. Then, the third is students' interaction which also belongs to the dependent variable.

The data was directly gotten by the researcher by giving the treatment to the experimental class. Therefore, the data of this research was primary data. This research used questionnaires for data collection technique. It evaluated students' motivation and interaction in the speaking class which are related to the authentic assessment. The procedure was as follows: (1) writing the instruments, (2) trying it out, (3) validating the research instrument, (4) estimating reliability of the research instrument, (5) revising the research instrument, (6) conducting the pre-test to

the sample to know the subject condition related the dependent variable. It is used to know the initial condition of both of the group, (7) giving the treatment, that was authentic assessment in the form of improvisation and role play to the experiment group and without authentic assessment to the control group, (8) conducting the post-test to the samples in the form of questionnaire to the both groups to compare the result.

There were two questionnaires. This research adapted model of motivation assessments proposed by (Gardner, 2012). It is called The Attitude/Motivation Test Battery (AMTB), whereas the interaction questionnaire was adapted from a definition proposed by (Wals, 2011).

The researcher used two kinds of validity; they were expert judgments and item validity. The expert was Ari Purnawan, M.A. as the expert of Language Assessments, whereas for the item validity, Bivariate Pearson was employed. The experts said that some items need to be revised.

After the questionnaires were revised, those were tried out to get the item validity and the reliability of the instruments. Bivariate Pearson was employed because the instruments were in the form of questionnaires which have 1 – 4 score for each response. After conducting the validity test, the reliability test was done. This research employed Alpha Cronbach for the questionnaire. It is because the instruments used Likert scale in the scale of four.

Then, to see the improvement, MANOVA tests (Multivariate Analysis of Variance) were used because this research consisted of one independent variable and two dependent variables. However, before MANOVA tests were employed, there were preliminary analyses. They were test of normality and test of homogeneity to make sure that the data have normal distribution and homogenous.

RESULT AND DISCUSSION

The data consists of the initial students' motivation and interaction score and the last students' motivation and interaction score. The initial scores were gathered from the pre-test whereas the last scores were gathered from the post test. The results of the measurements of the students' motivation and interaction from the both experiment and control group are presented below.

The Statistical Description of the Data

The Statistical Description of the Pre-Tes Result of the Students' Motivation and Interaction of the Experimental Group

The experimental group is a group of students that was treated with authentic assessment. Before the treatments were given to this group, they had got the pre-test. Those were tests of motivation and interaction in the form of questionnaires. There were 35 items for the motivation test and 24 items for interaction in the scale of 4 and 3 respectively for each number. For the motivation test, the highest score was 87 and the lowest score was 38 (the maximum score is 140). For the interaction test, the highest score was 61 and the lowest score was 25 (the maximum score is 72).

From the SPSS 20 computation of the students' motivation, it shows that the result of the pretest mean of the experimental group was 57.13; the mode was 51; the median was 53.00; and the standard of deviation was 12.265. It means that the students' motivation score were far from the maximum score, 140, even from the half of the maximum score, 70.

Based on the statistical computation, the categorization of students' motivation are presented in the Table 1.

Table 1. The Categorization of the Students' Motivation Pretest Score of the Experimental Group

Categories	Interval	Percentage	Cumulative Percentage
High	>105	0	0
Average	71 – 105	18.8	100.0

Low	<71	81.2	81.2
Total		100.0	

Table 1 shows that most of the students occupied the *low* category. Meanwhile, for the interaction measurement, the SPSS 20 computation shows that the result of the pretest mean of the experimental group was 35.44; the mode was 25.00; the median was 36.00; and the standard of deviation was 8.74. It means that the students' interaction score were far from the maximum score, 72. It has not achieved the half of the maximum score, 36. Based on the statistic computation, the categorization of the students' interaction are presented in the table 2.

Table 2. The Categorization of the Students' Interaction Pretest Score of the Experimental Group

Categories	Interval	Percentage	Cumulative Percentage
High	57 – 72	3.1	100.0
Average	41 – 56	25.0	96.9
Low	24 – 40	71.9	71.9
Total		100.0	

The Statistical Description of the Pre-Test Result of the Students' Motivation and Interaction of the Control Group

The Control group is a group that was treated without authentic assessment (with traditional assessments). Before the treatments were given to this group, they had got the pre-test. Those were tests of motivation and interaction in the form of questionnaires. There were 35 items for the motivation test and 24 items for the interaction test in the scale of 4 and 3 respectively for each number. For the motivation test, the highest score was 81 and the lowest score was 38 (the maximum score is 140). For the interaction test, the highest score was 55 and the lowest score was 25 (the maximum score is 72).

From the SPSS 20 computation of students' motivation, it shows that the result of pretest mean of the control group was 56.16; the mode was 51.00; the median was 52.00; and the deviation standard was 10.51. It means that the students' motivation score were far from the maximum score, 140, even from the half of the maximum score, 70. Based on the statistic computation, the categorization of the students' motivation are presented in the table 3.

Table 3. The Categorization of Motivation Pretest Result of the Control Group

Categories	Interval	Percentage	Cumulative Percentage
High	>105	0	0
Average	71 – 105	15.6	100.0
Low	<71	84.4	84.4
Total		100.0	

Meanwhile, for the interaction measurement, the SPSS 20 computation shows that the result of pretest mean of the control group was 34.69; the mode was 25.00; the median was 33.50; and the deviation standard was 8.67. It means that the students' interaction score were far from the maximum score, 72. It has not achieved the half of the maximum score, 36. Based on the statistic computation, the inclination categorization of students' interaction are presented in the Table 4.

Table 4. The Categorization of the Students' Interaction Pretest Result of the Control Group

Categories	Interval	Percentage	Cumulative Percentage
High	>56	0	0
Average	41 – 56	25.0	100.0
Low	<41	75.0	75.0
Total		100.0	

The Table 4 shows that most of the students still belong to the *low* category.

The Statistical Description of the Post Test Result of the Students' Motivation and Interaction of the Experimental Group

The post tests to the experimental group were conducted to see the students' motivation and interaction treated with authentic assessments. The result shows that the highest score was 99 and the lowest score was 55 (the maximum score is 140). For the interaction test, the highest score was 65 and the lowest score was 28 (the maximum score is 72).

From the SPSS 20 computation of the post test of students' motivation, it shows that the result of the post test mean of the experimental group was 78.69; the mode was 74.00; the median was 74.50; and the standard of deviation was 10.21. It means that the students' motivation increase. Although it still has not achieved the maximum score, but it has achieved the half of the maximum score. Based on the statistical computation, the categorization of the students' motivation are presented in the Table 5.

Table 5. The Categorization of the Students' Motivation Post Test Result of the Experimental Group

Categories	Interval	Percentage	Cumulative Percentage
High	> 105	0	0
Average	71 – 105	90.6	100.0
Low	< 71	9.4	9.4
Total		100.0	

The Table 5 shows that there are a lot of students move from the *low* category to the *average* category although there is still no students belong to the *high* category.

Meanwhile, for the interaction measurement, the SPSS 20 computation shows that the result of the post test mean of the experimental group was 39.93; the mode was 40.00; the median was 38.00; and the standard of deviation was 8.75. It means that the students' interaction increase. Although it still has not achieved the maximum score, but it has achieved the half of the maximum score. The complete computation was attached for further information. The frequency distribution of the post test score is presented as follows. Based on the statistical computation, the categorization of the students' interaction are presented in the Table 6.

Table 6. The Categorization of the Interaction Post Test Result of the Experimental Group

Categories	Interval	Percentage	Cumulative Percentage
High	>56	9.4	100.0
Average	41 – 56	34.4	90.6
Low	<41	56.2	56.2
Total		100.0	

The Table 6 shows that although the most students still belong to the *low* category, there is movements of students interaction categorization from the *low* category to the *average* category, even to the *high* category.

The Statistical Description of the Post Test Result of the Students' Motivation and Interaction of the Control Group

The post tests to the control group were conducted to see the students' motivation and interaction effects without authentic assessments treatment. The result shows that the highest score was 81 and the lowest score was 38 (the maximum score is 140). For the interaction test, the highest score was 42 and the lowest score was 25 (the maximum score is 72).

From the SPSS 20 computation of the post test of students' motivation, it shows that the result of the post test mean of the control group was 57.06; the mode was 51.00; the median was 52.00; and the standard of deviation was 11.73. It means that the students' motivation was not effected significantly. Based on the statistical computation, the categorization of the students' motivation are presented in the Table 7.

Table 7. The Categorization of the Students' Motivation Post Test Result of the Control Group

Categories	Interval	Percentage	Cumulative Percentage
High	>105	0	9.4
Average	71 – 105	21.9	100.0
Low	<71	78.1	78.1
Total		100.0	

The Table 7 shows that there was increasion on the average category, from 15.6% to 21.9%. Meanwhile, for the interaction measurement, the SPSS 20 computation shows that the result of the post test mean of the control group was 31.84; the mode was 25.00; the median was 33.00; and the standard of deviation was 5.73. It means that the students' interaction still has not achieved the maximum score, even the half of the maximum score. Based on the statistical computation, the categorization of the students' interaction are presented in the Table 8.

Table 8. The Categorization of the Interaction Post Test Result of the Control Group

Categories	Interval	Percentage	Cumulative Percentage
High	>56	0	0
Average	41 – 56	9.4	100.0
Low	<41	90.6	90.6
Total		100.0	

The Table 8 shows that most students was still in the *low* category.

The Prerequisite Tests

The Normality Test

The normality tests were employed to the motivation and the interaction pre-test and post-test of the control group and the experimental group. SPSS 20 computation shows that the data have a normal distribution. The data are said to have a normal distribution if the p is higher than the alpha coefficient 0.05 (5%). The following table is the summary of the normality tests. The table shows that the p value is higher than 0.05. In other word, the data have a normal distribution.

Table 9. The Summary of The Normality Tests of Motivation

Data	P	Explanation
Pretest of Ex G.	0.363	p > 0,05 = normal
Post-test of Ex G.	0.059	p > 0,05 = normal
Pretest of Con G.	0.224	p > 0,05 = normal
Post-test of Con G.	0.162	p > 0,05 = normal

Table 10. The Summary of The Normality Tests of Interaction

Data	p	Explanation
Pretest of Ex G.	0.581	p > 0,05 = normal
Post-test of Ex G.	0.101	p > 0,05 = normal
Pretest of Con G.	0.634	p > 0,05 = normal
Post-test of Con G.	0.275	p > 0,05 = normal

The Homogeneity Tests

The homogeneity tests of the data were included in MANOVA computation. The variance is said to be homogeneous if the significance value is higher than the significance level of 5 % (0,05). From the table below, it can be seen that the p value is more than 0.05. This shows the homogenity of the data. Here is the summary of the results of homogeneity test of prettes and postest of the experimental group and the control group.

Table 11. The Summary of Homogeneity Tests of Students' Motivation and Interaction

Data	P	Explanation
Motivation and Interaction Pretest of the Experimental and the Control Group	0.870	p. 0.870 > 0,05 = homogeneous
Motivation and Interaction Pretest-Post test of the Experimental Group	0.587	p. 0.587 > 0,05 = homogeneous
Motivation and Interaction Post test of the Experimental and the Control Group	0.087	p. 0.087 > 0,05 = homogeneous
Motivation and Interaction Pretest-Post test of the Control Group	0.053	p. 0.053 > 0,05 = homogeneous

The Data Analyses

The data analyses aim to test the research hypotheses. Those are (1) to find out whether there are different effects on students' motivation and interaction in speaking who are assessed using authentic assessments and those who are not and (2) to find out whether the use of authentic assessment at improving students' motivation and interaction in speaking is more effective than improving students' motivation and interaction without authentic assessment. In this case, MANOVA tests were employed. This technique was used to see whether there are significance differences of students' motivation score and students' interaction score of the experimental group from those of the control group. It is said to have significant differences if the sig. value is lower than 0.05. The computations of MANOVA tests were done using SPSS 20. The computations show that authentic assessments are more effective.

MANOVA Test to the Motivation Pretest of the Experimental Group and the Control Group

The data of the pretest of the experimental group and the control group were analysed using MANOVA test to know the students' motivation and interaction in the begining (before the treatments). The null hypothesis, there is no significant difference of motivation and interaction in speaking between the students of the experimental and the control group before the treatments, is accepted if the F value of *Pillae Trace, Wilk Lambda, Hotelling Trace, Roy's Largest Root* have sig > 0.05. The following table is the result of SPSS.20 computation.

Table 12. The SPSS 20 Computation Result of MANOVA Test of Students' Motivation and Interaction Pretest Score of the Experimental and the Control Group

Effect		Value	F	Hypothesis df	Sig.
Intercept	Pillai's Trace	.977	1269.558[b]	2.000	0.000
	Wilks' Lambda	.023	1269.558[b]	2.000	0.000
	Hotelling's Trace	41.625	1269.558[b]	2.000	0.000
	Roy's Largest Root	41.625	1269.558[b]	2.000	0.000
Assessments	Pillai's Trace	.004	.114[b]	2.000	0.893
	Wilks' Lambda	.996	.114[b]	2.000	0.893
	Hotelling's Trace	.004	.114[b]	2.000	0.893
	Roy's Largest Root	.004	.114[b]	2.000	0.893

The Table 12 shows that the F value for *Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root* have higher sig than 0.05. Those mean that there is no difference between experimental group and control group. They have same condition in the students' motivation and interaction before the treatment.

MANOVA Test to the Motivation-Interaction Pretest and Post Test of the Experimental Group

The data of the pretest and the post test of the experimental group were analysed using MANOVA test to compare the students' motivation and interaction after the treatment (using authentic assessments).

The alternative hypothesis, which states that there is a significant difference of motivation and interaction in speaking before the students were assessed using authentic assessments and after the students were assessed using authentic assessments, is accepted if the F value of *Pillae Trace, Wilk Lambda, Hotelling Trace, Roy's Largest Root* have sig < 0.05 (lower than). The following table is the result of SPSS.20 computation.

Table 13. The SPSS 20 Computation Result of MANOVA Test of the Students' Motivation and Interaction Pretest – Post Test Score of the Experimental Group

Effect		Value	F	Hypothesis df	Sig.
Intercept	Pillai's Trace	.981	1550.247[b]	2.000	0.000
	Wilks' Lambda	.019	1550.247[b]	2.000	0.000
	Hotelling's Trace	50.828	1550.247[b]	2.000	0.000
	Roy's Largest Root	50.828	1550.247[b]	2.000	0.000
Assessments	Pillai's Trace	.491	29.386[b]	2.000	0.000
	Wilks' Lambda	.509	29.386[b]	2.000	0.000
	Hotelling's Trace	.963	29.386[b]	2.000	0.000
	Roy's Largest Root	.963	29.386[b]	2.000	0.000

The Table 13 shows that the F value for *Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root* have lower sig than 0.05. Those mean that there is a significant difference between the pretest - the posttest of the experimental group. Their motivation and interaction increase after they got the treatments.

MANOVA Test to the Motivation-Interaction Post Test between the Experimental Group and the Control Group

The data of the post test of experimental group and the control group were analysed using the MANOVA test to compare the students' motivation and interaction after each treatment. The alternative hypothesis which *states* that there is a significant difference of motivation and interaction in speaking between the students' who were assessed using authentic assessments in speaking and those who were not, is accepted if the F value of *Pillae Trace, Wilk Lambda, Hotelling Trace, Roy's Largest Root* have sig < 0.05 (lower than). The following table is the result of SPSS.20 computation.

Table 14. The SPSS 20 Computation Result of MANOVA Test of the Students' Motivation and Interaction Post Tests Score of the Experimental Group and the Control Group

Effect		Value	F	Hypothesis df	Sig.
Intercept	Pillai's Trace	.981	1544.624[b]	2.000	0.000
	Wilks' Lambda	.019	1544.624[b]	2.000	0.000
	Hotelling's Trace	50.643	1544.624[b]	2.000	0.000
	Roy's Largest Root	50.643	1544.624[b]	2.000	0.000
Assessments	Pillai's Trace	.521	33.139[b]	2.000	0.000
	Wilks' Lambda	.479	33.139[b]	2.000	0.000
	Hotelling's Trace	1.087	33.139[b]	2.000	0.000
	Roy's Largest Root	1.087	33.139[b]	2.000	0.000

Table 14 shows that the F value for *Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root* have lower sig than 0.05. This means that there is significant difference of post test between the experimental group and the control group. The students' motivation and

interaction scores of the experimental group are higher than students' scores of the control group.

Based on the data above, it can be concluded that: (1) there is no significant difference of students' motivation and interaction pretest score between the experimental group and the control group in the begining; (2) there is significant difference of students' motivation and interaction in speaking of the experimental group before the treatment and after the treatment using authentic assessments; and (3) there is a significant difference of students' motivation and interaction post test score between the experimental group and the control group in speaking. Based on the statements above, it can be inferred that authentic assessments have effects on students' motivation and interaction. They are more effective than non-authentic assessments on the students' motivation and interaction to deal with English speaking practices of SMPN 2 Kalasan.

The Hypotheses Results

The Result of the First Hypothesis

The different effects on motivation and interaction between the group which is assessed using authentic assessments and that which is not in speaking can be checked by finding the difference of the post test score between the experimental group and the control group. The summary of the MANOVA test result can be seen in Table 14.

The analysis of SPSS 20 shows that the *p* value is 0.00 (at the significant level of 5%). The *p* value is lower than 0.05 (0.00<0.05). From the data, the result of the first hypothesis testing is as follows. The Ha, which states that there are different effects on students' motivation and interaction in speaking between who are assessed using authentic assessment and those who are not, **is accepted**.

The Result of the Second Hypothesis

The effectiveness of authentic assessments on the students' motivation and interaction in speaking can be checked by finding the difference of the pretest- postest score of the experimental group. The summary of the MANOVA test result can be seen in Table 13.

The analysis of SPSS 20 shows that the significance value is 0.00 (in the significant level of 5%). The significance value is lower than 0.05 (0.00<0.05). From the data, the result of the first hypothesis testing is as follow. The Ha, authentic assessment is more effective than non-authentic assessment on the students' motivation and interaction to deal with English speaking practices of SMPN 2, **is accepted**.

Discussion

The improvement of the mean score of pretest to post test of the experimental group, from 57.13 to 78.69 (for motivation) and 35.44 to 39.94 (for interaction), shows that there is a significant difference. On the other hand, the mean scores of the control group have no significant improvement, 56.16 to 57.60 (for motivation) and 34.69 to 31.84 (for interaction). The improvement of experimental group is 21.56 (for motivation) and 4.5 (for interaction), whereas the improvement of the control group is 1.44 (for motivation), decreased for 2.85 (for interaction). However, when the proportion was checked, it was increased. Before the treatment, the mean of interaction in control group was 34.69 in 10 meetings, whereas after the treatment, it was 31.84 in 8 meetings. The proportion calculation is as follow.

$P_1 = x_1/n_1 = 34.69/10 = 3.46$
$P_2 = x_2/n_2 = 31.84/8 = 3.98$
P_1 = proportion before the trearment
P_2 = proportion after the treatment
x_1 = the interaction mean before the trearment
x_2 = the interaction mean after the trearment
n_1 = the total meeting before the trearment

n_2 = the total meeting after the trearment

From the calculation above, it can be seen that P_2 (3.98) > P_1 (3.46). In other words, the interaction improvement also happens to the control group after the treatment.

Based on the MANOVA test to the post test of the experimental and the control group, the significance value is 0.00 (<0.05) to the both of the variables. It shows that the use of authentic assessments is more effective at improving students' motivation and interaction than without authentic assessments. Besides, it shows that there is significant different of post test score between the experimental group and the control group.

The research finding also shows the categorization inclination of the pretest and post test from both the experimental group and the control group. For the motivation of the experimental group, there are 81.25% students belong to the Low category, 18.75% students belong to the Average category, and no student belong to the High category at the begining. After the treatment, there are only 9.38% students belong to the Low category and the Average category increase to 90.62% although there is still no student belong to the High category. Similarly, for the interaction of the experimental group, there are 71.88% students belong to the Low category, 25.00% students belong to the Average category, and there are only 3.12% student belong to the High category at the begining. After the treatment, there are 56.25% students belong to the Low category and the Average category increase to 34.38% as well as the High category increases to 9.38%.

The condition above is caused by the following things, (1) The students could know directly their achievement, (2) the students felt new atmosphere, (3) They directly could feel the benefit, (4) the students think that it is not too difficult, and (5) the students had to show their performance of speaking when they were assessed.

For the motivation of the control group, there are 84.38% students belong to the Low category, 15.62% students belong to the Average category, and no student belong to the High category at the begining. After the treatment, there are still 78.12% students belong to the Low category and the Average category increase only to 21.88%, and there is still no student belong to the High category. For the interaction of the control group, there are 75.00% students belong to the Low category, 25.00% students belong to the Average category, and no student belong to the High category at the begining. Suprisingly, after the treatment, there are 90.62% students belong to the Low category and the Average category decrease to 9.38%, and there is still no student belong to the High category.

The condition above is caused by the following problems. (1) The students are bored. (2) Some students think that the assessments are not important. (3) It is easy for the students to only copy the answer from their friends since they think that the important thing is the score not their performance and competencies. (4) They think that it is difficult.

The result of this finding verifys the objectives of giving authentic assessments, that is authentic assessments foster students intrinsic motivation (Brown & Abeywickrama, 2004, p. 12). What (Harmer, 2003, p. 116) says is also proven by this, genuinely communicative speaking activities facilitate the students a real desire to speak and a communicative purpose for doing so. The ways to present the authenticity can facilitate the students to feel that they are assessed in meaningful (relevant, interesting) way and closely approximate that those are useful because they are real-world tasks. Moreover, it also proves the theory of learning and motivation stated by (Rosenthal & Zimmerman, 2014; Schunk, Pintrich, & Meece, 2008) that people will not demonstrate skills until they are motivated to display them. Authentic assessments need the students to display or perform the skills they have learnt because in this model of assessments, they are demanded to perform the skills as in the real world. Besides, recent thinking reconfigures ways in which assessment can motivate students to want to learn (Stiggins et al., 2004, p. 38). Moreover, theories have said that students' motivation will rise when the quality of the learning experience is improved. Although the students from the experimental group and the control group have a same motivation, finally the students of the experimental group have

higher motivation after they got different assessments. Experimental group had authentic assessments treatment, whereas control group had traditional assessments treatment.

Besides, the finding of this research proves what (Rivers, 1987; Tuan & Nhu, 2010, p. 20) says that in interaction, students can use all they possess of the language – all they have learned or casually absorbed–in real life exchanges. While authentic assessment is a model of assessment which needs students to perform in real-world context. Therefore, it can be said that interaction happens in authentic assessments. As what is showed by this finding that the students who got authentic assessments (the experimental group) have more interaction than those who did not get authentic assessments (the control group). That can be seen from the mean of the both of the groups after the treatment. The experimental group was 39.94and the control group was 31.84. It happens because the learning activities in the authentic assessments facilitate the interaction happens, such as group discussion, giving opinion spontaneously, information gap, improvisation, etc.

However, in its application, authentic assessment has some difficulties, such as: (1) since each student needs to be assessed one by one, it takes more time than traditional assessments. (2) Some students still be noisy (impatient while waiting for their turn), so that it will obstacle for the teacher to listen how was the performance. (3) it is quite difficult to control all students' performance since team teaching in Indonesia still have not been applied and the students number in a class still more than 25 students.

The finding of the research cunducted in the year-VIII students of SMPN 2 Kalasan shows that there is higher improvements of students' motivation and interaction scores in the experimental group compare to those scores in the control group. The means improvement in experimental group show that authentic assessments help the students in improving their motivation and interaction. Besides, it can be said that authentic assessments also help the students to achive the learning goals. However, the application of authentic assessments need to be adjusted with students' condition and the environment condition. The use of authentic assessment is effective for teachers to deal with students motivation and interaction in speaking.

CONCLUSION

Based on the analyses of the research finding and the discussion, it can be concluded that, first, there is a significant different of motivation and interaction between the year-VIII students of SMPN 2 Kalasan who were assessed using authentic assessments and those who were assessed without authentic assessments. The second, authentic assessment is more effective than non authentic assessment on the students' motivation and interaction to deal with English speaking practices of SMPN 2. Those were proven using the result of MANOVA test on the posttest score between the experimental group and the control group. It was drawn from the p value 0.00, it is lower than 0.05 (significant if $p<0.05$). The conclusion above is also based on the improvement of the mean of the posttest score of the experimental group, it was 21.56 (78.69-57.13) for motivation and 4.50 (39.94-35.44) for interaction. The improvement of mean score of the experimental group is higher than the control groups' which is only 0,90 (57.06-56.16) for motivation and for the interaction as well. All of these signify that authentic assessemts are more effective than traditional assessments.

REFERENCES

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Prentice Hall.

Brown, H. D., & Abeywickrama, P. (2004). Language assessment. In *Principles and Classroom Practices.* White Plains, NY: Pearson Education Inc.

Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology Review*, *22*(4), 384–401. https://doi.org/10.1007/s11065-012-9208-z

Elton, L. (2000). The UK research assessment exercise: Unintended consequences. *Higher*

Education Quarterly, *54*(3), 274–283. https://doi.org/10.1111/1468-2273.00160

Gardner, J. (2012). *Assessment and learning*. Sage.

Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 169–207. https://doi.org/10.1080/0969594032000121270

Harmer, J. (2003). *The practice of English language teaching. Completely revised and updated*. Edinburgh Gate: Pearson Education Limited.

Herrington, A., & Herrington, J. (2007). Authentic mobile learning in higher education. *AARE 2007 International Educational Research Conference*. Association for Research in Education.

Hidayanti, F. (2011). *Peningkatan keterampilan berbicara Bahasa Inggris melalui lagu dan permainan pada siswa kelas IV di SDN Madiredo 02 Kecamatan Pujon Kabupaten Malang*. Universitas Negeri Malang.

Komsi, D. N., Hambali, I., & Ramli, M. (2018). Kontribusi pola asuh orang tua demokratis, kontrol diri, konsep diri terhadap motivasi belajar siswa. *Psychology, Evaluation, and Technology in Educational Research*, *1*(1), 55. https://doi.org/10.33292/petier.v1i1.21

Kumpulainen, K., & Wray, D. (2003). *Classroom interactions and social learning: From theory to practice*. Routledge.

Lennon, P. (1993). The advanced learner: affective, social and motivational factors. *The Language Learning Journal*, *8*(1), 39–43. https://doi.org/10.1080/09571739385200351

Nunan, D. (1999). *Second language teaching & learning*. Boston, MA: Heinle & Heinle publishers.

Nurhadi, N. (2013). *Kesiapan SDM Indonesia menghadapi AFTA 2015*. Yogyakarta.

Rivers, W. M. (1987). *Interactive language teaching*. Cambridge University Press.

Rosenthal, T. L., & Zimmerman, B. J. (2014). *Social learning and cognition*. Academic Press.

Santrock, J. (2010). Educational psychology. In *Handbook of Educational Psychology*. https://doi.org/10.1146/annurev.ps.05.020154.002041

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications*. New Jersey: Pearson/Merrill Prentice Hall.

Schuyten, G., & Ferla, J. (2007). Can authentic assessment help in delivering competent consumers of statistics for non-academic professions? *Proceedings IASE Satellite Conference on Assessing Student Learning in Statistics*. International Statistical Institute.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Assessment Training Institute.

Surgenor, P. (2010). *Teaching toolkit: Effect of assessment on learning*. Dublin: University College Dublin.

Tuan, L. T., & Nhu, N. T. K. (2010). Theoretical review on oral interaction in EFL classrooms. *Studies in Literature and Language*, *1*(4), 29.

Vyortkina, D. (2003). *Portfolio assessment in educational leadership programs at Master's level*. Florida State University.

Wals, A. E. J. (2011). Learning our way to sustainability. *Journal of Education for Sustainable Development*, *5*(2), 177–186. https://doi.org/10.1177/097340821100500208