
Keefektifan penilaian autentik untuk menilai keterampilan berbicara siswa sekolah menengah atas

Ningsih Kusumastuti

SD Muhammadiyah 28 Jakarta. Jalan Panjang Cipulir, Kota Jakarta Selatan, 12230, Indonesia
Email: ningsih_kusumastuti@yahoo.co.id

Abstrak

Penelitian ini bertujuan untuk mengetahui keefektifan penilaian autentik untuk menilai keterampilan berbicara siswa kelas XI SMA Negeri 11 Yogyakarta. Model penilaian autentik yang digunakan adalah penilaian guru, penilaian diri dan penilaian kawan. Validitas instrumen penelitian diuji dengan menggunakan *expert judgement*, sedangkan reliabilitas instrumen diuji dengan menggunakan reliabilitas *Alpha Cronbach*. Normalitas data diuji dengan menggunakan metode *Kolmogorov Smirnov Z*. Keefektifan penilaian diri, penilaian kawan dan penilaian guru sebelum dan setelah perlakuan diuji dengan analisis *covariance* atau ANACOVA. Hasil analisis menunjukkan bahwa terdapat perbedaan kemampuan berbicara yang signifikan dengan penggunaan penilaian diri, penilaian kawan dan penilaian guru. Ketiga model penilaian tersebut memberi pengaruh yang signifikan terhadap kemampuan berbicara siswa. Penilaian kawan merupakan penilaian yang paling efektif digunakan dalam pembelajaran berbicara siswa dibandingkan dengan penilaian diri dan penilaian guru. Penilaian diri yang dikombinasikan dengan penilaian kawan paling efektif digunakan daripada penilaian tradisional dan penilaian diri yang dikombinasikan dengan penilaian guru dalam pembelajaran berbicara siswa.

Kata kunci: penilaian, penilaian berbicara, penilaian autentik

The effectiveness of authentic assessment in learning speaking skill for students of high school

Abstract

The objective of this research was to know the effectiveness of authentic assessment in learning speaking skill for the XI graders of SMA Negeri 11 Yogyakarta. The model of authentic assesment taken in this research was teacher assessment, self assessment and peer assessment. Expert judgement was used to analyze the instrument validity, while the instrument reliability was analyzed with Alpha Cronbach test. The data normality was analyzed using Kolmogorov Smirnov Z test. ANACOVA test was used to see the effectiveness of self assessment, peer assessment and teacher assessment before and after the treatment. The result showed that there were some significant differences to the students' skill by the use of self assessment, peer assessment, and teacher assessment. The three assessments gave significant effect to the students' speaking skill. Peer assessment was proven the most effective in learning speaking skill compared to self assesment and teacher assessment. Self assessment which was combined with peer assessment was proven the most effective one in learing speaking skill compared to traditional assessment and self assessment which was combined with teacher assessment.

Keywords: assessment, oral assessment, authentic assessment

PENDAHULUAN

Dalam instansi pendidikan, ada 3 hal yang menjadi pilar pokok yang harus dimiliki dan dilakukan yaitu kurikulum, proses pembelajaran, dan penilaian. Kurikulum merupakan penjabaran dari pokok-pokok bahasan yang menjadi target pembelajaran di sekolah. Pembelajaran merupakan upaya yang dilakukan guru untuk bisa mencapai target dan standar yang telah ditentukan dalam kurikulum. Sedangkan penilaian merupakan bentuk pengukuran dan penilaian yang ditujukan untuk mengetahui tingkat ketercapaian materi dan keberhasilan



proses belajar mengajar. Dari penilaian tersebut dapat diambil tindakan untuk mempertahankan proses pembelajaran dan menetapkan kriteria ataukah perlu merevisi kedua aspek sebelumnya.

Proses pembelajaran dikatakan berhasil jika yang dilakukan berdasar pada kurikulum yang ada dan hasil yang berupa nilai yang didapat juga memenuhi standar nilai yang ditetapkan. Begitu juga dengan aspek penilaian, suatu proses penilaian akan dikatakan berhasil jika proses tersebut mencakup aspek yang tersebut dalam kurikulum dan merefleksikan hasil dari proses pembelajaran. Jadi bisa disimpulkan bahwa ketiga aspek tersebut saling terkait satu sama lain.

Para guru telah berusaha melakukan usaha yang ditujukan bagi tercapainya target yang ditetapkan dalam kurikulum, salah satunya dengan menciptakan suasana pembelajaran yang nyaman dan efektif, yaitu melalui pembelajaran komunikatif. Akan tetapi pada kenyataannya, meskipun para guru telah menggunakan metode komunikatif yang dianggap merupakan metode yang paling efektif dalam pengajaran bahasa, *output* yang dihasilkan siswa belum mencapai hasil yang dikatakan memuaskan. Seperti yang dinyatakan (Wright, 2007, p. 249) *“Such activities may be enjoyable to all involved but will result in poor performance on the mandated high-stakes assessments.”* Suatu aktivitas pembelajaran mungkin saja menyenangkan bagi semua pihak yang terlibat akan tetapi menghasilkan performan yang buruk pada saat penilaian. Pada saat proses pembelajaran siswa secara aktif terlibat dalam komunikasi pembelajaran, akan tetapi saat mereka dihadapkan pada penilaian kemampuan berbicara mereka akan merasa kesulitan dalam mengekspresikan kembali ujaran yang telah dipelajari dan dipraktikkan sebelumnya.

Penilaian atau juga dikenal dengan istilah asesmen merupakan proses akhir dalam suatu pembelajaran. Penilaian juga disebutkan sebagai suatu proses sistematis dalam pengumpulan, analisis, dan penafsiran informasi untuk menentukan seberapa jauh seorang peserta didik dapat mencapai tujuan pendidikan (Nurgiyantoro, 2010: 7). Lebih jauh lagi Schwartz & Webb (2014: 2) menjelaskan bahwa *“Assessment is interpreted in terms of the routine tasks that students undertake in order to receive feedback on their learning and a mark or grade signifying their achievement.”* Penilaian diartikan sebagai tugas rutin yang dilakukan siswa dengan tujuan untuk mendapatkan *feedback* terhadap pembelajaran dan nilai yang merepresentasikan pencapaian mereka. *Feedback* yang diberikan biasanya berupa *descriptive feedback* atau berupa penjelasan. *Descriptive feedback* tersebut bisa berupa komentar, masukan, informasi yang berasal dari berbagai sumber yang memberikan informasi untuk memungkinkan siswa secara bertahap melakukan perubahan ke arah yang lebih baik.

Penilaian berbicara merupakan penilaian yang mengacu kepada segala bentuk penilaian pembelajaran siswa yang dilakukan secara lisan (Joughin, 2010, p. 3). Penilaian berbicara bertujuan untuk melihat kemampuan siswa dalam berkomunikasi baik secara komunikasi dasar maupun komunikasi akademik. Dalam penilaian berbicara bukan hanya siswa saja yang dituntut untuk berbicara, akan tetapi asesor (dalam hal ini guru) bisa juga terlibat dalam proses komunikasi tersebut. Selain itu penilaian proses penilaian berbicara juga harus melibatkan siswa untuk bisa aktif mengaplikasikan dan mengembangkan pengetahuan. Ada empat kategori yang ada dalam penilaian berbicara, yaitu: *Grammar* dan *Vocabulary*, *Discourse Management*, *Pronunciation*, dan *Interactive Communication* (Aslanishvili, 2011, p. 105).

Penilaian dilakukan oleh seorang guru dalam menentukan tingkat ketercapaian siswa dalam menyerap informasi yang diberikan. Pada saat melakukan penilaian, guru bertugas untuk mengumpulkan informasi tentang pemahaman siswa yang digunakan untuk mengetahui hasil proses pengajaran dan membantu siswa untuk bisa belajar lebih banyak lagi (Davies, 2011, p. 1). Sebelum melakukan penilaian, guru perlu melakukan berbagai persiapan yang matang dan terencana. Hal tersebut dimaksudkan agar nantinya penilaian dapat berjalan dengan lancar dan sesuai dengan target yang telah ditentukan.

Dari hasil observasi awal yang dilakukan di lokasi penelitian, guru memberikan penugasan berupa presentasi kepada siswa. Dalam penilaian terhadap tugas tersebut guru menggunakan

penilaian tradisional dimana guru tidak menggunakan instrumen penilaian tertentu, akan tetapi memberikan nilai berupa angka secara langsung tanpa memberikan deskripsi khusus terhadap penilaian yang diberikan pada tugas yang dilakukan oleh siswa. Selain itu guru juga tidak memberikan umpan balik yang berupa saran terhadap kesalahan aspek bahasa dan materi yang digunakan oleh siswa selama proses penilaian. Sebagai dampak dari penilaian tradisional tersebut, peneliti menemukan beberapa kesalahan bahasa yang dilakukan siswa secara berulang-ulang. Bahkan peneliti menemukan beberapa kesalahan yang sama juga dilakukan oleh siswa yang lain.

Terkait dengan alasan dan kondisi tersebut, peneliti menawarkan penilaian yang lebih memberikan kesempatan kepada siswa untuk bisa merefleksi diri dan mengembangkan pemahaman dalam berbicara. Model penilaian seperti ini yang disebut sebagai penilaian autentik. Model penilaian autentik yang ditawarkan berupa penilaian guru (*teacher assessment*), penilaian diri (*self-assessment*), dan penilaian kawan (*peer-assessment*). Ketiga penilaian tersebut diaplikasikan pada penugasan berbicara yang berupa kegiatan presentasi. Melalui kegiatan tersebut siswa menampilkan kemampuan dan menunjukkan pemahaman materi dari hasil pembelajaran. Kegiatan penugasan tersebut selanjutnya dinilai dengan menggunakan model penilaian guru, penilaian diri, dan penilaian kawan.

Pemilihan siswa SMA Negeri 11 sebagai obyek penelitian dikarenakan sampai saat ini ketiga model penilaian tersebut belum digunakan dalam penilaian berbicara siswa. Model penilaian yang digunakan guru adalah penilaian tradisional.

Dalam penelitian tersebut, peneliti membatasi pada keefektifan penggunaan assesmen autentik berupa penilaian formatif yang berbentuk penilaian diri, penilaian kawan, dan penilaian guru dalam pembelajaran berbicara siswa. Peneliti berpendapat bahwa melalui ketiga penilaian tersebut siswa akan mendapatkan kesempatan untuk mendapatkan pemahaman yang baik terhadap materi yang dipelajari dan mengembangkan kemampuan berbicara mereka.

Tujuan dari penelitian tersebut adalah untuk mengetahui apakah terdapat perbedaan kemampuan berbicara siswa setelah digunakan penilaian autentik yang berupa penilaian diri, penilaian kawan, dan penilaian guru dalam pembelajaran berbicara siswa. Peneliti juga ingin mengetahui keefektifan masing-masing model penilaian autentik dan kombinasi penilaian autentik dibandingkan dengan penilaian tradisional dalam pembelajaran berbicara siswa.

Penelitian tersebut diharapkan dapat membantu siswa dalam memaksimalkan kemampuan berbicara, khususnya setelah dilakukannya proses penilaian. Dengan menggunakan model penilaian autentik tersebut, diharapkan siswa bisa secara langsung terlibat dalam proses penilaian dan sekaligus mengembangkan kemampuan secara mandiri. Selain itu diharapkan juga penelitian tersebut dapat membantu guru dalam mengembangkan model penilaian berbicara dan menentukan jenis penilaian berbicara yang paling tepat bagi siswa.

Penilaian autentik atau yang juga dikenal dengan asesmen autentik merupakan penilaian yang melibatkan siswa dalam tindakan yang nyata. Model penilaian tersebut bertujuan untuk meningkatkan kualitas hasil proses pembelajaran melalui sistem penilaian atau assesmen bermakna. Melalui penggunaan penilaian tersebut diharapkan siswa dapat mengetahui kekurangan dan kesalahan yang dilakukan, serta dapat memperbaiki kesalahan tersebut yang pada akhirnya siswa dapat lebih memahami dan mengembangkan materi yang dipelajari. Secara simpel (Nurgiyantoro, 2010) mendeskripsikan asesmen autentik sebagai bentuk penilaian yang mementingkan penilaian proses dan hasil sekaligus. Lebih lanjut dalam bukunya, (Nurgiantoro, 2018, p. 23) menjelaskan bahwa penilaian autentik menekankan kemampuan siswa untuk mendemonstrasikan pengetahuan yang dimiliki secara nyata dan bermakna.

Asesmen autentik juga disebut sebagai penilaian performan (*performance assessment*) karena dalam penilaian tersebut siswa diminta untuk melakukan suatu tugas yang bermakna. Asesmen autentik, seperti yang dideskripsikan Ataç (2012: 10), juga disebut sebagai penilaian performan, penilaian alternatif, dan juga penilaian langsung. Sitting dalam Ataç (2012: 10)

memberikan deskripsi singkat tentang asesmen autentik sebagai berikut “*Performance assessments call upon the examinee to demonstrate specific skills and competence, that is, to apply the skills and knowledge they have mastered*”. Penilaian performan menuntut siswa untuk mendemonstrasikan keahlian dan kecakapan tertentu yang bertujuan untuk mengaplikasikan keahlian dan pengetahuan yang telah mereka kuasai. Weigle dalam (McKay, 2006, p. 98) juga menjelaskan bahwa “*Performance assessment refers to assessment that involves either the observation of behaviour in the real world or a simulation of a real-life activity*”. Penilaian performan mengacu pada penilaian yang melibatkan baik observasi terhadap tingkah laku dalam dunia nyata ataupun simulasi terhadap aktivitas yang nyata. Tingkah laku yang dimaksudkan adalah bagaimana siswa menggunakan kemampuan bahasa dalam dunia nyata. Sejalan dengan pendapat tersebut, (Palm, 2001, p. 4) juga mendefinisikan performan asesmen sebagai penilaian yang berdasarkan pada pengamatan dan *judgement*. (McKay, 2006) juga menambahkan “*Performance assessments give children opportunities to use the language for real purposes, and in real or realistic situations, and assess their attempts to do so successfully*”. Penilaian performan memberikan kesempatan kepada siswa untuk menggunakan bahasa untuk tujuan yang nyata dan menilai usaha yang sukses dilakukan. Melalui penilaian performan tersebut siswa dapat menggunakan bahasa untuk mengungkapkan maksud dan tujuan mereka. Selain itu siswa juga bisa menilai apakah mereka bisa mengatasi kesulitan dalam menggunakan bahasa selama berkomunikasi.

Penilaian autentik memiliki beberapa manfaat baik bagi siswa maupun bagi guru. Melalui penggunaan penilaian autentik siswa dapat secara langsung terlibat dalam proses penilaian, sehingga mereka mengetahui kesalahan maupun kekurangan yang mereka lakukan dan aspek yang perlu diperbaiki maupun dikembangkan. Bagi guru penilaian tersebut dapat membantu meningkatkan keaktifan siswa dalam proses pembelajaran. Lebih jauh lagi, Muller dalam (Nurgiyantoro, 2010) menjelaskan bahwa ada beberapa manfaat yang bisa diambil dalam penggunaan model penilaian autentik. Manfaat dari penggunaan model penilaian tersebut adalah sebagai berikut: *Pertama*, penggunaan penilaian autentik memungkinkan dilakukannya pengukuran secara langsung terhadap kinerja siswa sebagai indikator capaian kompetensi yang diajarkan. Asesmen autentik menuntut siswa untuk menggunakan ekspresi dalam situasi yang konkret dan sekaligus bermakna yang secara otomatis juga mencerminkan penguasaan dan ketrampilan keilmuannya. *Kedua*, penilaian autentik memberikan kesempatan kepada siswa untuk mengkonstruksikan hasil belajarnya. Hal ini berarti bahwa dalam mengeksposisikan kalimat siswa dituntut untuk bisa menyeleksi ungkapan berdasarkan pengetahuan yang dimiliki dan menganalisa situasi agar nantinya ungkapan yang digunakan akan relevan dan bermakna. *Ketiga*, asesmen autentik memungkinkan terintegrasikannya kegiatan pengajaran, belajar, dan penilaian menjadi satu kegiatan yang terpadu. *Keempat*, asesmen autentik memberi kesempatan pada siswa untuk menampilkan hasil belajarnya, untuk kerjanya, dengan cara yang dianggap paling baik dan efektif.

Terdapat sejumlah langkah yang perlu ditempuh seorang guru dalam pengembangan asesmen autentik, yang meliputi: mengidentifikasi tujuan, merencanakan kegiatan penilaian, membuat rubrik dan/atau prosedur penilaian, menentukan standar, melibatkan siswa dalam penilaian, memilih kegiatan penilaian, mencatat informasi (Pierce and O'Malley, 1996: 63). Dari definisi asesmen autentik secara jelas dapat dilihat bahwa jenis penilaian tersebut memberikan dampak positif terhadap siswa. Melalui model tersebut, siswa diberi kesempatan untuk merefleksi penampilan mereka sendiri, sehingga siswa tahu kesalahan dan kekurangan dari hasil kerja mereka. Penilaian autentik juga diharapkan mampu menstimulasi siswa untuk mengembangkan kemampuan dan kompetensi yang sejalan dengan dunia kerja mereka di masa yang akan datang.

Penilaian guru atau *teacher assessment* merupakan bentuk penilaian kinerja siswa yang berupa angka maupun komentar ataupun *feedback*. Penilaian tersebut dilakukan oleh guru yang bersangkutan. Informasi yang didapat dalam penilaian tersebut bisa berupa wawancara langsung dengan siswa, pengamatan terhadap kinerja siswa ataupun berupa penilaian langsung

terhadap performan siswa (Brookhart & Nitko, 2008).

Penilaian diri atau *self assessment* merupakan kegiatan siswa dalam menilai dan mengevaluasi kinerja yang telah dilakukannya. Lebih jauh lagi (Orsmond, Maw, Wilson, & Sears, 2004, p. 8) mendefinisikan *self-assessment* sebagai keterlibatan siswa dalam mengidentifikasi standar atau kriteria yang diaplikasikan pada kerja mereka dan menilai apakah mereka telah memenuhi standar dan kriteria tersebut. Ketika siswa melakukan penilaian dan mengevaluasi kinerjanya sendiri, mereka dituntut untuk berpikir tentang apa yang telah dipelajarinya dan sekaligus menambah pemahaman terhadap subyek yang dinilai. Untuk memudahkan siswa mendapatkan ranah yang jelas tentang aspek diri yang dinilai, dibutuhkan format dan batasan yang jelas yang berupa rubrik penilaian diri yang akan menjabarkan aspek kinerja yang dinilai. Refleksi terhadap kinerja dan pemahaman yang didapat dari penilaian diri tersebut diharapkan dapat dijadikan sebagai bahan dan tolak ukur untuk pemahaman dan peningkatan pada performan selanjutnya.

Penilaian kawan atau *peer assessment* merupakan kegiatan menilai dan mengevaluasi kinerja yang telah dilakukan oleh temannya. Topping et al dalam Orsmond, Maw, Wilson, & Sears (2004: 8) mendefinisikan *peer-assessment* sebagai rancangan bagi pasangan dalam kinerja untuk melihat tingkat, nilai, bobot, kualitas atau kesuksesan hasil pembelajaran dari siswa lain pada status yang sama. Ketika siswa terlibat aktif dalam memberikan penilaian terhadap kinerja kawan, maka sebenarnya saat itu siswa juga belajar dan melakukan evaluasi terhadap apa yang akan dilakukannya kemudian. *Feedback* dan masukan yang diberikan kepada kawan diharapkan akan menjadi bahan pembelajaran bagi kawan maupun bagi dirinya. Seperti pada penilaian diri, dalam proses penilaian kawan tersebut dibutuhkan ranah dan batasan yang jelas tentang aspek kinerja yang dinilai.

METODE

Jenis penelitian ini adalah penelitian eksperimen semu dimana tidak semua variabel yang berhubungan dengan penilaian berbicara bisa diukur dalam penelitian. Penulis membatasi penelitian pada penggunaan penilaian autentik yang berupa penilaian diri, penilaian kawan, dan penilaian guru dalam pembelajaran berbicara. Penelitian dilakukan di SMA Negeri 11 Yogyakarta. Observasi awal dilakukan pada awal bulan Oktober 2012 untuk mengetahui kondisi siswa. Penelitian dilakukan pada bulan Maret-Mei 2013 pada semester genap tahun ajaran 2012/2013.

Populasi penelitian ini adalah siswa kelas XI IPA SMA Negeri 11 Yogyakarta tahun pelajaran 2012/2013 sejumlah 192 siswa. Dalam penelitian tersebut peneliti mengambil sampel dengan *teknik cluster random* sampling melalui dua tahap, yaitu tahap pertama menentukan sampel kelas dengan cara acak mengambil tiga kelas yang berjumlah 95 siswa. Pengambilan sampel tersebut berdasar pertimbangan bahwa ketiga kelas tersebut mendapatkan materi berdasarkan kurikulum yang sama, kelas yang diajar oleh guru yang sama, kelas yang mempunyai rata-rata kemampuan berbahasa yang setingkat, kelas yang berdasarkan kondisi kelas sesuai dengan tujuan penelitian. Pada tahap selanjutnya peneliti menentukan satu kelas yang dijadikan sebagai kontrol kelas dan dua kelas yang dijadikan sebagai eksperimen kelas secara acak. Tabel 1 merupakan deskripsi sampel penelitian secara detail:

Tabel 1. Jumlah Sampel Penelitian

Kelas	Keterangan	Jumlah
XI IPA 3	Kelas eksperimen (Xa)	32
XI IPA 4	Kelas eksperimen (Xb)	31
XI IPA 5	Kelas kontrol	32
	Jumlah	95

Desain penelitian yang digunakan adalah *randomized control-group pretest-posttest design*. Desain penelitian dapat ditunjukkan dalam tabel 2.

Tabel 2. Desain Penelitian Randomized Control-Group Pretest-Posttest

Group	Pretest	Treatment	Posttest
First Experimental Group (R)	T ₁	X _a (Method a)	T ₂
Second Experimental Group (R)	T ₁	X _b (Method b)	T ₂
Control Group (R)	T ₁	--	T ₂

(Isaac & Michael, 1995, p. 66)

Penelitian ini melibatkan dua variabel bebas, yaitu penilaian berbicara dengan menggunakan penilaian guru dan penilaian diri (X_a) dan penilaian berbicara dengan menggunakan penilaian kawan dan penilaian diri (X_b). Variabel terikat dalam penelitian ini adalah kemampuan berbicara siswa kelas XI SMA Negeri 11 Yogyakarta (Y).

Instrumen yang digunakan dalam penelitian tersebut adalah rubrik penilaian berbicara. Model rubrik penilaian yang digunakan diadopsi dari rubrik penilaian berbicara pada *English Oral Test for Cambridge International Examination for IGCSE Level period October – November 2012*. *English Oral Test for IGCSE Level* adalah tes kemampuan berbicara bahasa Inggris sebagai bahasa kedua bagi siswa sekolah lanjutan usia 15-16 tahun yang diselenggarakan oleh Cambridge University, yang dalam hal ini juga dianggap sesuai untuk siswa kelas XI. Selanjutnya rubrik tersebut diadaptasi sesuai dengan kebutuhan dan kemampuan objek penelitian. Sebelum instrumen digunakan dalam penelitian, terlebih dahulu instrumen diuji dengan menggunakan pendapat dari ahli (*judgement expert*). Reliabilitas instrumen diuji dengan menggunakan reliabilitas *Alpha Cronbach*.

Data penelitian diperoleh melalui observasi, penilaian selama pelaksanaan pembelajaran, dan tes. Data observasi berupa daftar nilai akhir semester siswa pada semester gasal untuk mengetahui kondisi siswa. Penilaian selama pelaksanaan pembelajaran digunakan sebagai bentuk hasil keterlibatan siswa dalam penggunaan model penilaian autentik. Tes dilakukan sebanyak dua kali, yaitu pretest dan *posttest*. Model tes yang digunakan dalam penelitian tersebut berupa penugasan kepada siswa.

Untuk menguji apakah data penelitian bersifat homogen, digunakan metode *Alpha Cronbach*. Normalitas data diujikan dengan menggunakan metode *Kolmogorov Smirnov Z*. Penilaian autentik yang berupa penilaian diri, penilaian kawan, penilaian guru, penilaian diri yang dikombinasikan dengan penilaian guru, dan penilaian diri yang dikombinasikan dengan penilaian kawan masing-masing dibandingkan dengan penilaian tradisional sebagai penilaian non autentik. Keefektifan penilaian autentik diuji dengan menggunakan analisis *covariance* atau ANACOVA.

HASIL DAN PEMBAHASAN

Hasil Uji Normalitas Pretest dan Posttest

Uji normalitas bertujuan untuk mengetahui distribusi data dalam variabel yang digunakan dalam penelitian. Data yang baik dan layak digunakan dalam penelitian adalah data yang memiliki distribusi normal. Apabila data berdistribusi normal maka data hasil penelitian tersebut dianalisis dengan metode parametrik, sedangkan jika data tidak berdistribusi normal, maka data hasil penelitian dianalisis dengan metode non parametrik. Suatu data dikatakan mempunyai sebaran normal jika $p > 0,05$. Hasil analisis menunjukkan bahwa nilai signifikan yang didapat dari hasil uji normalitas seluruh variabel lebih dari 0,05, sehingga dapat disimpulkan bahwa data pada seluruh variabel penelitian berdistribusi normal, pengujian hipotesis dapat dilakukan dengan metode parametrik. Hasil uji normalitas nilai pretest dan *posttest* ditunjukkan pada Tabel 3.

Tabel 3. Hasil Uji Normalitas Nilai *Pretest* dan *Posttest*

	N	Normal Parameters ^{a,b}		Most Extreme Differences			Kolmogorov-Smirnov Z	Asymp.Sig (2-tailed)
		Mean	Std.Dev	Absolute	Positive	Negative		
Pretest_3	32	74.6812	2.55891	.165	.165	.083	.931	.351

<i>Posttest_3</i>	32	77.3125	2.28244	.211	.211	.155	1.192	.117
<i>Pretest_4</i>	31	76.4290	2.34068	.184	.184	.109	1.1025	.244
<i>Posttest_4</i>	31	79.8258	2.35244	.185	.185	.121	1.029	.241
<i>Pretest_5</i>	31	74.8161	2.51715	.213	.213	.139	1.185	.121
<i>Posttest_5</i>	32	75.0500	2.77988	.163	.163	.084	.924	.360

a. Test distribution is Normal.

b. Calculated from data.

Hasil Uji Homogenitas Nilai Pretest dan *Posttest*

Homogenitas data pretest dan *posttest* diuji dengan menggunakan Lavene Tes. Kriteria pengambilan keputusan adalah sebagai berikut:

Bila sig. atau signifikansi atau nilai probabilitas > 0,05, data yang berasal dari populasi memiliki varians sama (homogen). Bila sig. atau signifikansi atau nilai probabilitas < 0,05, data yang berasal dari populasi memiliki varians tidak sama (tidak homogen)

Interpretasi dilakukan dengan memilih salah satu statistik, yaitu statistik yang didasarkan pada rata-rata (Based on Mean). Hasil analisis menunjukkan bahwa sig. Based on Mean untuk pretest sebesar 0,918. Dan nilai sig. Based on Mean untuk *posttest* adalah 0,598. Bila nilai-nilai tersebut dibandingkan dengan nilai alpha 0,05, maka 0,918 > 0,05 dan 0,598 > 0,05. Oleh karena itu data pretest dan *posttest* memiliki varians yang sama atau homogen. Hasil uji homogenitas nilai pretest dan *posttest* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Uji Homogenitas Nilai Pretest dan *Posttest*

		Lavene Statistic	df1	df2	Sig.
PRETEST	Based on Mean	.085	2	92	.918
	Based on Median	.046	2	92	.955
	Based on Median and with adjusted df	.046	2	89.764	.955
	Based on trimmed mean	.088	2	92	.915
POSTTEST	Based on Mean	.518	2	92	.481
	Based on Median	.738	2	92	.481
	Based on Median and with adjusted df	.738	2	91.491	.481
	Based on trimmed mean	.655	2	92	.522

Hasil Pengujian Hipotesis

Pengujian hipotesis dilakukan dengan menggunakan analisis ANACOVA Faktorial. Dalam pengujian dengan ANACOVA Faktorial, skor pretest dijadikan sebagai covariate karena pretest merupakan variabel berskala kuantitatif, sedangkan skor *posttest* dijadikan sebagai dependent variabel. Nilai ini menunjukkan berapa besar pengaruh covariate terhadap variabel dependen. Signifikan < 0,05 berarti pengaruh signifikan. Model penilaian merupakan variabel berskala kualitatif, maka ia menjadi peubah bebas atau disebut juga fixed factor. Nilai ini menunjukkan berapa besar pengaruh peubah bebas terhadap variabel dependen. Signifikan < 0,05 berarti pengaruh signifikan.

Hasil analisis menunjukkan bahwa angka signifikansi untuk peubah pretest adalah 0,000. karena nilai Sig.<0,05 maka Ho ditolak. Hal ini berarti bahwa pada tingkat kepercayaan 95% dapat dikatakan ada hubungan linier antara pretest dengan nilai *posttest*. Pernyataan ini mengindikasikan bahwa asumsi ANACOVA telah terpenuhi. Pengujian ini dilakukan dengan menghilangkan pengaruh perbedaan model penilaian dari siswa terlebih dahulu. Selanjutnya dilakukan pengujian untuk mengetahui pengaruh model penilaian terhadap nilai yang diperoleh siswa. Pengujian ini dilakukan dengan menghilangkan pengaruh nilai pretest dari siswa. Dari hasil pengolahan terlihat bahwa angka signifikansi untuk variabel model penilaian adalah 0,000. Karena nilainya jauh di bawah 0,05 maka Ho ditolak. Sehingga dapat disimpulkan bahwa tanpa pengaruh nilai pretest, pada tingkat kepercayaan 95% ada pengaruh perbedaan model penilaian terhadap nilai yang diperoleh siswa.

Untuk mengetahui pengaruh nilai pretest siswa dan perbedaaan model penilaian terhadap nilai yang diperoleh siswa secara simultan dapat dilihat dari angka signifikansi pada bagian *Corrected Model*. Terlihat bahwa angka signifikansinya adalah sebesar 0,000. Karena nilai signifikansi jauh di bawah 0,05 maka H_0 ditolak. Sehingga pada tingkat kepercayaan 95% dapat disimpulkan bahwa secara simultan nilai pretest dan model penilaian berpengaruh terhadap nilai yang diperoleh siswa. Hasil analisis juga menunjukkan nilai signifikan pada variabel model penilaian adalah $0,000 < 0,05$, sehingga H_0 ditolak dan dapat disimpulkan bahwa tanpa pengaruh nilai pretest, pada tingkat kepercayaan 95% ada pengaruh perbedaan model penilaian terhadap nilai *posttest*. Hasil analisis pengaruh model penilaian terhadap nilai yang diperoleh siswa ditunjukkan pada Tabel 5.

Tabel 5. Hasil Uji Hipotesis

Dependent Variable: *POSTTEST*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	683,642 ²	3	227,881	85,431	.000
Intercept	39,295	1	39,295	14,731	.000
Model Penilaian	193,632	2	96,816	36,296	.000
Pretest	324,339	1	324,339	121,593	.000
Error	242,735	91	2,667		
Total	569615,220	95			
Corrected Total	926,377	94			

a. R Squared = ,738 (Adjusted R Squared = ,729)

Uji ANACOVA juga dilakukan untuk menguji interaksi pretests dan model penilaian terhadap nilai *posttest*. Hasil analisis menunjukkan pengaruh perbedaan model penilaian terhadap nilai *posttest*. Nilai signifikan yang didapat dari variabel grup adalah 0,043, nilai ini lebih kecil dari 0,05 yang berarti bahwa model penilaian berpengaruh terhadap nilai *posttest* siswa. Tabel tersebut juga menunjukkan perbedaan model penilaian terhadap nilai *posttest*. Nilai signifikan yang didapat dari variabel pretest adalah 0,000, nilai ini lebih kecil dari 0,05 yang berarti bahwa nilai pretest berpengaruh signifikan terhadap nilai *posttest*. Untuk mengetahui pengaruh interaksi model penilaian dan nilai pretest siswa terhadap nilai *posttest*, dapat dilihat nilai R square. Nilai signifikan yang didapat dari variabel grup* pretest adalah 0,034, nilai ini lebih kecil dari 0,05 yang berarti bahwa interaksi nilai pretest dan model penilaian berpengaruh signifikan terhadap nilai *posttest*. Hasil analisis interaksi nilai pretest dan model penilaian terhadap nilai *posttest* ditunjukkan pada Tabel 6.

Tabel 6. Hasil Uji Interaksi Nilai Pretest dan Model Penilaian terhadap *Posttest*

Dependent Variable: *POSTTEST*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	694,380 ²	6	138,876	53,276	.000	.750
Intercept	38,605	1	38,605	14,810	.000	.443
Grup	13,318	2	6,659	2,555	.043	.754
Pretest	322,877	1	322,877	123,863	.000	.582
Grup*pretest	10,737	2	5,369	2,060	.034	.744
Error	231,998	89	2,607			
Total	569615,220	96				
Corrected Total	926,377	94				

a. R Squared = ,738 (Adjusted R Squared = ,729)

Pembahasan

Penilaian Tradisional

Penilaian tradisional merupakan penilaian non autentik yang diberikan pada kelas kontrol. Dalam penilaian tersebut, guru memberikan penilaian terhadap performan siswa tanpa menggunakan rubrik penilaian, dan hasil penilaian tidak diinformasikan kepada siswa. Komunikasi yang terjadi dalam proses ini bersifat satu arah saja, karena tidak ada interaksi antara guru dan siswa untuk membahas dan mendiskusikan hasil performan siswa dan masukan dari guru. Dalam penelitian tersebut penilaian tradisional dibandingkan dengan ketiga model penilaian autentik, yaitu penilaian diri, penilaian kawan, dan penilaian guru untuk mengetahui keefektifan dari masing-masing model penilaian.

Penilaian Diri

Penilaian diri sebagai model penilaian autentik merupakan bentuk penilaian inovatif yang mendukung kegiatan pembelajaran siswa. Penilaian tersebut merupakan proses dimana siswa terlibat dan bertanggung jawab dalam menilai hasil kerjanya sendiri, yaitu menilai kemampuan berbicara diri sendiri. Melalui refleksi diri tersebut diharapkan siswa dapat mengetahui kesalahan yang dilakukannya dan memahami aspek yang harus diperbaiki dan dikembangkan pada performan selanjutnya. Penilaian diri dapat mendorong siswa untuk bisa mandiri dan meningkatkan motivasi mereka. Selain itu, penilaian diri juga dapat digunakan dalam membantu mengembangkan kemampuan siswa untuk memeriksa dan berpikir kritis mengenai proses pembelajaran yang mereka jalani.

Hasil analisis menunjukkan bahwa penilaian diri berpengaruh terhadap nilai *posttest* siswa, yang berarti bahwa penilaian diri memberikan pengaruh pada peningkatan kemampuan berbicara siswa. Pengaruh simultan antara penilaian diri dan nilai *pretest* terhadap nilai *posttest* adalah sebesar 0,718. Selanjutnya penilaian diri dibandingkan dengan penilaian tradisional dengan melihat nilai rata-rata *posttest*. Kelompok yang menggunakan penilaian diri memiliki rata-rata *posttest* sebesar 78,5492, sedangkan kelompok yang menggunakan penilaian tradisional memiliki rata-rata *posttest* sebesar 75,05. Kelompok yang menggunakan penilaian diri memiliki rata-rata *posttest* yang lebih besar dibandingkan dengan kelompok yang menggunakan penilaian tradisional dengan selisih angka rata-rata sebesar 3,4992. Ini berarti bahwa penilaian diri lebih efektif digunakan dalam menilai berbicara siswa dibandingkan dengan penilaian tradisional.

Tabel 7. Keefektifan Penilaian Tradisional dan Penilaian Diri

Dependent Variable: *POSTTEST*

Grup	Mean	Std. Dev	N
Tradisional	75.0500	2.77988	32
Diri	78.5492	2.62426	32
Total	77.3705	3.13928	64

Penilaian Kawan

Penilaian kawan atau lebih dikenal dengan penilaian antar teman sebagai model penilaian autentik juga merupakan bentuk penilaian inovatif yang mendukung kegiatan pembelajaran siswa. Penilaian kawan dalam pembahasan ini adalah proses di mana siswa terlibat dan bertanggung jawab dalam penilaian kerja siswa lain, yaitu menilai kemampuan berbicara siswa. Secara umum penilaian kawan digunakan untuk memberikan nilai atau memberi umpan balik pada teman mereka mengenai hasil kerja atau kemampuan mereka berdasarkan suatu kriteria yang telah dibuat bersama mereka. Beberapa keuntungan penilaian kawan antara lain: 1) Dapat meningkatkan hasil belajar, 2) Dapat meningkatkan kolaborasi belajar melalui umpan balik dari teman sejawat, 3) Siswa dapat membantu temanya dalam pemahaman dan belajar mereka dan merasa lebih nyaman dalam proses belajar, dan 4) Siswa dapat memberi komentar pada kinerja temannya.

Hasil analisis menunjukkan bahwa penilaian kawan berpengaruh terhadap nilai *posttest* siswa, yang berarti bahwa penilaian kawan memberikan pengaruh pada peningkatan

kemampuan berbicara siswa. Pengaruh simultan antara penilaian diri dan nilai pretest terhadap nilai *posttest* adalah sebesar 0,801. Selanjutnya penilaian kawan dibandingkan dengan penilaian tradisional dengan melihat nilai rata-rata *posttest*. Kelompok yang menggunakan penilaian kawan memiliki rata-rata *posttest* sebesar 79,8258, sedangkan kelompok yang menggunakan penilaian tradisional memiliki rata-rata *posttest* sebesar 75,05. Kelompok yang menggunakan penilaian kawan memiliki rata-rata *posttest* yang lebih besar dibandingkan dengan kelompok yang menggunakan penilaian tradisional dengan selisih angka rata-rata sebesar 4,7758. Ini berarti bahwa penilaian kawan lebih efektif digunakan dalam menilai berbicara siswa dibandingkan dengan penilaian tradisional.

Tabel 8. Keefektifan Penilaian Tradisional dan Penilaian Kawan

Dependent Variable: *POSTTEST*

Grup	Mean	Std. Dev	N
Tradisional	75.0500	2.77988	32
Kawan	79.8258	2.35244	31
Total	77.4000	3.51201	63

Penilaian Guru

Penilaian guru sebagai bentuk model penilaian autentik dalam penelitian tersebut merupakan bentuk penilaian kinerja siswa yang berupa angka yang dilakukan oleh guru dan digunakan sebagai umpan balik terhadap siswa. Informasi yang didapat guru dalam penilaian tersebut berdasarkan dari pengamatan terhadap performan siswa yang selanjutnya digunakan sebagai bahan masukan yang disampaikan kepada siswa. Melalui cara tersebut siswa akan mengetahui kesalahan yang dilakukan, dan selanjutnya akan dijadikan sebagai bahan evaluasi dan peningkatan pada performan selanjutnya.

Hasil analisis menunjukkan bahwa penilaian guru berpengaruh terhadap nilai *posttest* siswa, yang berarti bahwa penilaian guru memberikan pengaruh pada peningkatan kemampuan berbicara siswa. Pengaruh simultan antara penilaian guru dan nilai pretest terhadap nilai *posttest* adalah sebesar 0,678. Selanjutnya penilaian guru dibandingkan dengan penilaian tradisional dengan melihat nilai rata-rata *posttest*. Kelompok yang menggunakan penilaian guru memiliki rata-rata *posttest* sebesar 77,3125, sedangkan kelompok yang menggunakan penilaian tradisional memiliki rata-rata *posttest* sebesar 75,05. Kelompok yang menggunakan penilaian guru memiliki rata-rata *posttest* yang lebih besar dibandingkan dengan kelompok yang menggunakan penilaian tradisional dengan selisih angka rata-rata sebesar 2,2625. Ini berarti bahwa penilaian guru lebih efektif digunakan dalam menilai berbicara siswa dibandingkan dengan penilaian tradisional.

Tabel 9. Keefektifan Penilaian Tradisional dan Penilaian Guru

Dependent Variable: *POSTTEST*

Grup	Mean	Std. Dev	N
Tradisional	75.0500	2.77988	32
Guru	77.3125	2.28244	33
Total	76.1813	3.76875	64

Penilaian Diri yang Dikombinasikan dengan Penilaian Kawan

Penilaian diri dan penilaian kawan dalam proses pembelajaran memiliki peran yang sangat besar dalam meningkatkan kemampuan siswa, karena melalui kedua model penilaian tersebut siswa dapat merefleksikan hasil kerjanya dan mengetahui kesalahan yang dilakukannya secara lebih detail. Penilaian tersebut secara otomatis memberi pengaruh terhadap performan selanjutnya. Kedua model penilaian autentik tersebut tentu saja memiliki keefektifan masing-masing. Dalam penelitian tersebut, peneliti melihat keefektifan jika penilaian diri dikombinasikan dengan penilaian kawan selama proses pembelajaran.

Hasil analisis menunjukkan bahwa penilaian diri yang dikombinasikan dengan penilaian kawan berpengaruh terhadap nilai *posttest* siswa, yang berarti bahwa kombinasi kedua penilaian autentik tersebut memberikan pengaruh pada peningkatan kemampuan berbicara siswa. Pengaruh simultan antara kombinasi dua penilaian autentik dan nilai pretest terhadap nilai *posttest* adalah sebesar 0,75. Selanjutnya kombinasi dua penilaian autentik dibandingkan dengan penilaian tradisional dengan melihat nilai rata-rata *posttest*. Kelompok yang menggunakan kombinasi dua penilaian autentik memiliki rata-rata *posttest* sebesar 79,8258, sedangkan kelompok yang menggunakan penilaian tradisional memiliki rata-rata *posttest* sebesar 75,05. Kelompok yang menggunakan kombinasi dua penilaian autentik memiliki rata-rata *posttest* yang lebih besar dibandingkan dengan kelompok yang menggunakan penilaian tradisional dengan selisih angka rata-rata sebesar 4,7758. Ini berarti bahwa penilaian diri yang dikombinasikan dengan penilaian kawan lebih efektif digunakan dalam menilai berbicara siswa dibandingkan dengan penilaian tradisional. Keefektifan penilaian diri yang dikombinasikan dengan penilaian kawan dijelaskan secara singkat pada Tabel 10.

Penilaian Diri yang Dikombinasikan dengan Penilaian Guru

Seperti halnya penilaian diri dan penilaian kawan, penilaian guru sebagai salah satu model penilaian autentik juga memiliki peran yang besar dalam meningkatkan kemampuan siswa. Melalui penilaian guru siswa mendapatkan informasi dan masukan terhadap kesalahan yang dilakukannya dari orang yang benar-benar berkompeten dan menguasai materi dengan baik, dalam hal ini adalah guru, sehingga siswa memiliki masukan yang lebih sebagai bahan untuk refleksi dan perbaikan pada performan selanjutnya. Dalam penelitian tersebut, peneliti melihat keefektifan jika penilaian diri dikombinasikan dengan penilaian guru selama proses pembelajaran.

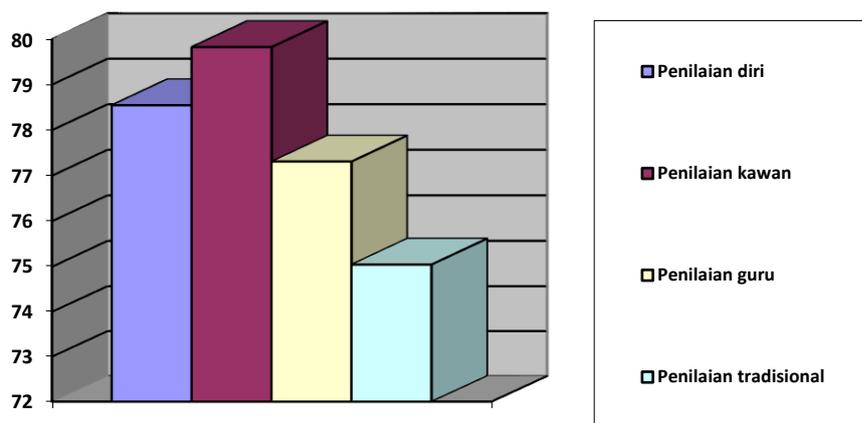
Hasil analisis menunjukkan bahwa penilaian diri yang dikombinasikan dengan penilaian guru berpengaruh terhadap nilai *posttest* siswa, yang berarti bahwa kombinasi kedua penilaian autentik tersebut memberikan pengaruh pada peningkatan kemampuan berbicara siswa. Pengaruh simultan antara kombinasi dua penilaian autentik dan nilai pretest terhadap nilai *posttest* adalah sebesar 0,75. Selanjutnya kombinasi dua penilaian autentik dibandingkan dengan penilaian tradisional dengan melihat nilai rata-rata *posttest*. Kelompok yang menggunakan kombinasi dua penilaian autentik memiliki rata-rata *posttest* sebesar 77,3125, sedangkan kelompok yang menggunakan penilaian tradisional memiliki rata-rata *posttest* sebesar 75,05. Kelompok yang menggunakan kombinasi dua penilaian autentik memiliki rata-rata *posttest* yang lebih besar dibandingkan dengan kelompok yang menggunakan penilaian tradisional dengan selisih angka rata-rata sebesar 2,2625. Ini berarti bahwa penilaian diri yang dikombinasikan dengan penilaian guru lebih efektif digunakan dalam menilai berbicara siswa dibandingkan dengan penilaian tradisional. Keefektifan penilaian diri yang dikombinasikan dengan penilaian guru dijelaskan secara singkat pada Tabel 10.

Tabel 10. Keefektifan Kombinasi Penilaian Autentik dan Penilaian Tradisional

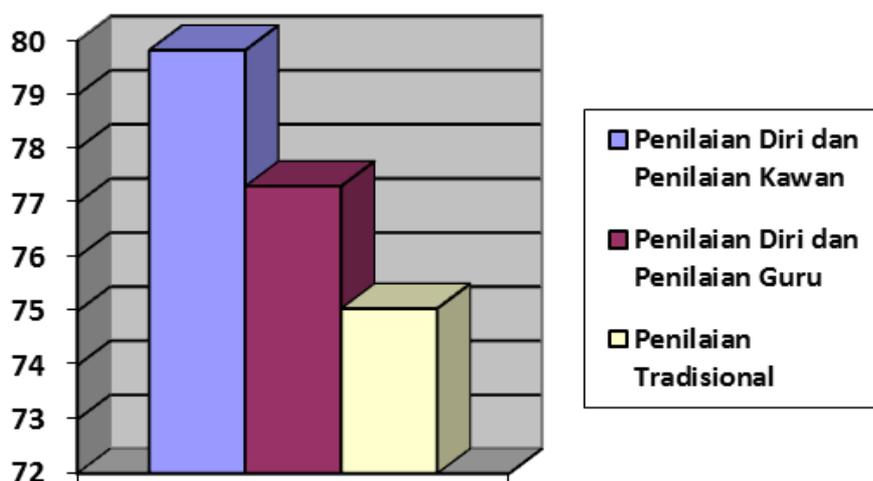
Dependent Variable: *POSTTEST*

Grup Penilaian	Mean	Std. Dev	N
Guru dan diri	77.3125	2.28244	32
Kawan dan diri	79.8258	2.35244	31
tradisional	75.0500	2.77988	32
Total	77.3705	3.13928	95

Berikut ilustrasi keefektifan model penilaian yang digunakan dalam penelitian.



Gambar 1. Keefektifan Penilaian Autentik dan Penilaian Tradisional



Gambar 2. Keefektifan Kombinasi Penilaian Autentik dan Penilaian Tradisional

Gambar 1 menunjukkan keefektifan ketiga penilaian autentik, yaitu penilaian diri, penilaian kawan, dan penilaian guru serta penilaian tradisional sebagai penilaian non autentik. Gambar tersebut menjelaskan bahwa penilaian kawan merupakan model penilaian yang paling efektif digunakan dalam menilai kemampuan berbicara siswa dibandingkan dengan penilaian diri, penilaian guru, dan penilaian tradisional.

Gambar 2 menunjukkan keefektifan kombinasi penilaian autentik, yaitu penilaian diri yang dikombinasikan dengan penilaian kawan dan penilaian diri yang dikombinasikan dengan penilaian guru, dan penilaian tradisional. Gambar tersebut menjelaskan bahwa penilaian diri yang dikombinasikan dengan penilaian kawan paling efektif digunakan dalam menilai kemampuan berbicara siswa dibandingkan penilaian diri yang digabungkan dengan penilaian guru dan penilaian tradisional.

SIMPULAN

Hasil penelitian dan analisis data menghasilkan beberapa kesimpulan. Berikut kesimpulan yang dapat penulis susun: pertama bahwa terdapat perbedaan kemampuan berbicara siswa yang signifikan pada kelas eksperimen setelah perlakuan setelah digunakannya penilaian diri,

penilaian kawan, penilaian guru, dan kombinasi penilaian autentik, kedua bahwa terdapat perbedaan kemampuan berbicara siswa yang diberi perlakuan dengan menggunakan penilaian autentik dibandingkan dengan siswa yang menggunakan penilaian tradisional sebagai penilaian non autentik, dan kesimpulan yang ketiga bahwa penilaian diri yang dikombinasikan dengan penilaian kawan merupakan model penilaian autentik yang paling efektif digunakan dalam menilai kemampuan berbicara siswa dibandingkan dengan penilaian autentik yang lain yaitu penilaian diri yang dikombinasikan dengan penilaian guru dan penilaian tradisional sebagai penilaian non autentik.

DAFTAR PUSTAKA

- Aslanishvili, T. (2011). How should be speaking assessed. *IBSU Scientific Journal*, 5(1), 101-109.
- Ataç, B. A. (2012). Foreign Language Teachers' Attitude toward Authentic Assessment in Language. *Dil ve Dilbilimi Çalışmaları Dergisi*, 8(2).
- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Prentice Hall.
- Davies, A. (2011). *Making classroom assessment work*. ERIC.
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences*. Edits publishers.
- Joughin, G. (2010). *A short guide to oral assessment*. Leeds Met Press in association with University of Wollongong.
- McKay, P. (2006). *Assessing young language learners*. Ernst Klett Sprachen.
- Nurgiantoro, B. (2018). *Penilaian otentik dalam pembelajaran bahasa*. UGM PRESS.
- Nurgiyantoro, B. (2010). *Penilaian pembelajaran bahasa berbasis kompetensi*. Yogyakarta: BPFE-Yogyakarta.
- Orsmond, P., Maw, S., Wilson, J., & Sears, H. (2004). *Self-and Peer-Assessment: Guidance on Practice in the Biosciences: Guidance on Practice in the Biosciences*. Centre for Bioscience.
- Palm, T. (2001). *Performance and authentic assessment, realistic and real life tasks: A conceptual analysis of the literature*. Umeå universitet: Department of educational measurement,.
- Schwartz, P., & Webb, G. (2014). *Assessment: Case studies, experience and practice*. Routledge.
- Wright, R. J. (2007). *Educational assessment: Tests and measurements in the age of accountability*. Sage Publications.