

**ANALISIS KUALITAS BUTIR SOAL ULANGAN AKHIR SEMESTER GASAL MATA PELAJARAN BAHASA JEPANG KELAS X BAHASA SMA NEGERI 1 DRIYOREJO TAHUN AJARAN 2018/2019**

**Marendra Yuda Pratama**

Jurusan Bahasa dan Sastra Jepang, Fakultas Bahasa dan Seni, Universitas Negeri Surabaya, email: marendrapratama@mhs.unesa.ac.id

**Amira Agustin Kocimaheni, S.Pd., M.Pd.**

Dosen Jurusan Bahasa dan Sastra Jepang, Fakultas Bahasa dan Seni, Universitas Negeri Surabaya, email: amiraagustin@unesa.ac.id

**Abstract**

A teaching and learning activity in schools is important to be followed by a learning evaluation. In general, a teacher conducts a learning evaluation with the aim to measuring the achievement of a student, but the purpose of learning evaluation is not only limited to that, evaluation is also used to measure how much success the teacher has in the learning process. When researchers interviewed teachers of Driyorejo Public High School 1, the teacher never conducted activities to analyze the quality of the items. Based on that, this research was carried out so the teacher could find out the quality of the school test items based on the validity, reliability, difficulty level of items, difference power, effectiveness of distractor.

In this study, researcher used a descriptive quantitative approach. Which has the aim to analyzing and describing the quality of the observed test items. For calculations, the researcher used IBM SPSS Statistics 26 and Microsoft Excel 2020 application. The researcher analyzing data according to the rules of each item and described according to the results.

The result of validity is the items with multiple choice forms 8 items (80%) showed valid, then for essay form 13 items (83%) were showed valid. Based on that it is known that the validity of the test items is very good. The reliability of the items shows 0.66 for the multiple choice form, so it is known to be unreliable because the score is less than 0.70. The essay forms are reliable because the score is 0.92. The results of the analysis of the difficulty level of the items showed that out of 10 multiple choice items, 7 items were classified as easy (70%), and 3 items were classified as moderate (30%). For essay form of 15 essay items, 9 items can be classified as easy (60%) and 6 items are classified as medium (40%). The results of the analysis of the difference power, for multiple choice forms got the results 8 items with good difference power (80%) and 2 items with sufficient difference (20%). The essay form got the results 9 items with very good differences (60%), 4 items with good difference (27%), and one items with sufficient difference (13%). The results of effectiveness of distractor, 2 items had very good effectiveness (10%), 3 items had a good effectiveness (20%), 4 items had sufficient effectiveness (40%) and one items had poor effectiveness (10 %).

**Keywords:** learning evaluation, quality items, items test

**要旨**

学校の教育の中で学習評価のことが重要である。一般的に、教師は生徒の成績を測定するために学習評価を行うが、評価の目的はそれだけでなく、学習評価が学習プロセスでどれだけ成功しているかを測定することでもある。研究者がドリヨレジョ公立高校 1 の教師にインタビューしたとき、教師はアイテムの品質を分析するための活動を行ったことはなかった。それに基づいて、本研究は、妥当性、信頼性、項目の難易度、差異の力、ディストラクタの有効性に基づいて学校のテスト項目の品質を見つけることができるように実施された。

この研究では、研究者は記述的な定量的アプローチを使用した。これは、観察されたテスト項目の品質を分析し、説明することを目的としている。計算には、研究者は IBM SPSS Statistics26 と MicrosoftExcel2020 アプリケーションを使用した。研究者は、各項目のルールに従ってデータを分析し、結果に従って説明する。

## Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Bahasa Jepang Kelas X Bahasa SMA Negeri 1 Driyorejo Tahun Ajaran 2018/2019

妥当性の結果は、複数の選択形式で 8 項目 (80%) が有効であることが示され、次にエッセイ形式で 13 項目 (83%) が有効であることが示された。項目の信頼性は多肢選択形式で 0.66 を示しており、スコアが 0.70 未満であるため、信頼性が低いことが分かっている。スコアが 0.92 であるため、エッセイフォームは信頼できる。項目の難易度を分析した結果、多肢選択式の 10 項目のうち、7 項目が簡単 (70%)、3 項目が中程度 (30%) に分類された。15 のエッセイ項目のエッセイ形式では、9 つの項目が簡単 (60%) に分類され、6 つの項目が中程度 (40%) に分類される。差異の力の結果、複数の選択形式について、結果は、良好な差異パワー (80%) の 8 項目と、十分な差異 (20%) の 2 項目だった。エッセイフォームの結果は、非常に良い違いがある 9 項目 (60%)、良い違いがある 4 項目 (27%)、十分な違いがある 1 項目 (13%) だった。ディストラクタの有効性の結果、2 項目は非常に有効性が高く (10%)、3 項目は有効性が良好 (20%)、4 項目は十分な有効性 (40%)、1 項目は有効性が低かった (10%)。

キーワード：学習評価、品質項目、項目テスト

### PENDAHULUAN

Sebuah kegiatan belajar mengajar di sekolah pada akhirnya diikuti dengan evaluasi pembelajaran. Kegiatan tersebut setidaknya dilakukan oleh seorang pengajar pada kurun periode tertentu. Arikunto (2016:4) memaparkan dalam sebuah kegiatan pembelajaran yang terjadi di kelas, guru adalah pihak yang paling memiliki tanggung jawab. Dengan demikian, guru sudah seharusnya memiliki keterampilan untuk mengevaluasi hasil belajar siswa, selain itu guru juga harus memiliki ketrampilan dalam megolah materi agar tujuan pembelajaran yang diharapkan dapat tercapai.

Nurgiyantoro (2012:30) juga menjelaskan bahwa kegiatan evaluasi hasil belajar memiliki tujuan untuk mengetahui sejauh mana kompetensi yang telah dicapai dalam sebuah pembelajaran yang dilakukan. Oleh karena itu, evaluasi hasil belajar dapat digunakan oleh seorang pengajar sebagai acuan untuk mengetahui pencapaian sebuah proses belajar mengajar yang berkesinambungan dengan kompetensi-kompetensi yang ada pada sebuah kurikulum. Pada evaluasi hasil belajar pada dasarnya mengikutsertakan banyak kegiatan struktural untuk menentukan metode-metode yang dilakukan untuk memperoleh informasi yang dibutuhkan untuk menentukan keputusan dalam kegiatan belajar mengajar.

Dalam mengidentifikasi sebuah penilaian seorang penilai memerlukan sebuah keterampilan identifikasi, selain itu penilai juga harus memahami jenis-jenis penilaian baik kontekstual proses dan juga penilaian hasil. Arikunto (2016:58) memaparkan bahwa instrumen penilaian meliputi validitas, reliabilitas, objektivitas, dan ekonomis.

Dalam sebuah pembelajaran penilaian sebagai pusat untuk kontrol keberhasilan pendidikan memerlukan sebuah alat ukur. Menurut Wantah (2010:132) menyusun sebuah alat ukur diperlukan sebuah kecermatan dan juga

harus berdasarkan aturan-aturan yang telah disetujui oleh ahli pengukuran di bidang Pendidikan, karena hasil perhitungan yang telah dilakukan akan menjadi sebuah patokan dalam menetapkan klasifikasi peserta didik. Semakin tinggi kemampuan seorang peserta didik dalam memahami sebuah pelajaran maka semakin tinggi juga peserta didik mencapai kompetensi yang telah ditetapkan, semakin rendah kemampuan peserta didik memahami pelajaran, maka semakin kecil juga peluang untuk mencapai kompetensi yang telah ditetapkan

Sudijono (2011:201). juga menjelaskan bahwa dalam evaluasi hasil belajar terdapat dua teknik untuk mengukur keberhasilan peserta didik mencapai kompetensi yaitu tes dan non tes. Teknik tes adalah teknik yang paling sering digunakan oleh seorang pengajar untuk memperoleh informasi sejauh mana peserta didiknya menguasai sebuah materi. Dalam menyusun instrumen tes pengajar harus memiliki kemampuan untuk menyusun instrumen, sehingga tes yang dilakukan benar-benar dapat mengukur kemampuan peserta didiknya atau tes yang dilakukan adalah tes yang berkualitas. Tes dikatakan berkualitas apabila tes tersebut valid, bersifat reliabel, bersifat objektif, dan bersifat praktis.

Dalam penyusunan sebuah tes yang berkualitas analisis kualitas butir soal diperlukan oleh seorang pengajar. Analisis butir soal memiliki peranan penting dalam sebuah tes yang berkualitas. Arifin (2014:246) mengungkapkan belajar, analisis butir soal dirancang untuk mengetahui kekurangan dalam sebuah butir soal tes yang dibuat sehingga soal tes dapat digunakan untuk tes berikutnya, selain itu analisis butir soal perlu dilakukan untuk mengetahui tingkat kesulitan butir soal tes. Arikunto (2016: 222) berpendapat bahwa soal yang baik adalah soal yang tidak terlalu mudah maupun soal yang tidak terlalu sulit. Analisis butir soal juga digunakan untuk mengetahui kelayakan sebuah butir soal yang digunakan untuk tes dan

juga sebagai salah satu control hasil prestasi belajar peserta didik.

Dari uraian diatas analisis butir soal berkaitan erat dengan peningkatan kualitas pembelajaran serta tingkat kompetensi pengajar. Dengan dilakukannya analisis butir soal komponen pembelajaran dapat diketahui berfungsi sebagaimana fungsinya atau tidak. Pengajar juga dapat mengetahui perkembangan hasil belajar siswa. Hal-hal tersebut dapat diraih apabila dalam penyusunan sebuah tes dilakukan dengan baik dengan melaksanakan analisis butir soal. Karena soal yang baik akan memberikan kualitas yang baik pula dalam hasil pembelajaran.

Hasil wawancara terhadap salah satu pengajar bahasa Jepang di SMA Negeri 1 Driyorejo, kegiatan pembelajaran sudah berjalan sesuai dengan harapan pengajar. Peserta didik tidak pasif saat pembelajaran berlangsung, apabila ada hal yang tidak di mengerti oleh peserta didik, peserta didik juga sangat aktif dalam bertanya. Dalam kegiatan pembelajaran pengajar juga menerapkan beberapa metode metode dan media pembelajaran. Untuk metode yang sering digunakan oleh pengajar yaitu metode diskusi dan juga metode pemecahan masalah atau *problem solving*, dan juga pengajar sering menggunakan media *visual* berupa *power point* dan tidak jarang juga menggunakan media *audio visual* berupa video percakapan, namun sayangnya pengajar belum pernah melakukan kegiatan menganalisis kualitas butir soal, yang artinya butir soal dari tes tersebut belum diketahui kualitasnya.

Berdasarkan latar belakang tersebut maka dilakukan penelitian terkait analisis kualitas butir soal dengan judul “Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Bahasa Jepang Kelas X Bahasa SMA Negeri 1 Driyorejo Tahun Ajaran 2018/2019”. Berdasarkan permasalahan tersebut dapat dirumuskan “Bagaimana kualitas butir soal Ulangan Akhir Semester Gasal mata pelajaran bahasa Jepang kelas X Bahasa SMA Negeri 1 Driyorejo Tahun Ajaran 2018/2019 berdasarkan validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektifitas pengecoh”.

Sehingga berdasarkan latar belakang dan rumusan masalah diatas tujuan dilakukannya penelitian ini agar dapat diketahui kualitas butir soal Ujian Akhir Semester Gasal mata pelajaran bahasa Jepang kelas X Bahasa SMA Negeri 1 Driyorejo tahun ajaran 2018/2019 berdasarkan validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektifitas pengecoh.

Dengan dilakukannya analisis kualitas butir soal manfaat yang diharapkan dari kegiatan ini dapat menjadi sebuah dorongan untuk pengajar khususnya bahasa Jepang untuk terus meningkatkan kualitas butir soal yang dibuat dan dengan meningkatnya kualitas butir soal diharapkan

siswa mendapat dorongan lebih untuk belajar Bahasa Jepang.

## **KAJIAN TEORI**

Penelitian ini dilakukan dengan menggunakan penelitian yang relevan, yaitu penelitian yang dilakukan oleh R. Antonius Mulyono Raharjo dengan judul “Analisis Butir Soal Kenaikan Kelas XI Tahun Pelajaran 2012/2013 Buatan Dinas Pendidikan Kota Semarang”. Selain itu penelitian yang relevan dengan penelitian ini yaitu dengan penelitian Masitho Arini berjudul “Analisis Kualitas Butir Soal Ujian Sekolah Mata Pelajaran Ekonomi kelas XII SMA Negeri 1 Taman Tahun Ajaran 2015/2016”. Dalam hal ini penelitian yang dibahas oleh peneliti sama-sama mengenai analisis butir soal. Disini perbedaan yang ada untuk penelitian sebelumnya terbatas pada tingkat kesukaran, dan daya pembeda, dan juga terbatas pada bentuk soal pilihan ganda, sedangkan dalam penelitian ini peneliti menganalisis kualitas soal berdasarkan lima indeks yaitu validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektifitas pengecoh, dan juga peneliti menganalisis semua bentuk soal pilihan ganda dan soal uraian.

## **EVALUASI HASIL BELAJAR**

Dalam sebuah pembelajaran, evaluasi hasil belajar termasuk sebuah kegiatan yang harus dilakukan oleh seorang pengajar. Yoshiaki (1998:196) memaparkan bahwa evaluasi merupakan sebuah proses dalam menentukan nilai suatu objek dengan cara membandingkan objek seperti siswa, guru, metode belajar, kurikulum, dan lain lain, dengan sasaran atau target pembelajaran. Kegiatan tersebut dilakukan dengan tujuan untuk mengumpulkan informasi agar seorang pengajar bisa melakukan perbaikan bagi proses pembelajaran selanjutnya.

Dengan adanya evaluasi hasil belajar pengajar dapat mengetahui informasi mengenai seberapa jauh pembelajaran yang dilakukan berjalan sesuai dengan tujuan yang sudah ditetapkan sebelumnya. Arikunto (2016: 3) juga mengungkapkan, melakukan sebuah evaluasi pada dasarnya melakukannya dua kegiatan yaitu penilaian dan pengukuran. Dalam sebuah ilmu pendidikan penilaian adalah suatu usaha untuk mengetahui atau mengukur seberapa besar keberhasilan kegiatan pendidikan dengan ukuran baik atau buruk dan penilaian bersifat kuantitatif, sedangkan mengukur pada dasarnya merupakan kegiatan membandingkan sesuatu hal dengan dengan suatu ukuran dan pengukuran bersifat kuantitatif.

Dari pendapat diatas dapat dilihat bahwa evaluasi merupakan kegiatan tersusun yang mana di dalamnya

terdapat kegiatan mengukur dan menilai suatu objek dengan tujuan mengumpulkan informasi dan hasilnya dapat dipakai sebagai acuan untuk proses pembelajaran selanjutnya.

### **ANALISIS BUTIR SOAL**

Untuk mengetahui seberapa baik kualitas suatu soal maka diperlukan pengkajian soal tersebut atau biasa disebut analisis butir soal. Arikunto (2016: 220) menjelaskan analisis butir soal merupakan sebuah kegiatan yang tersusun yang dilakukan untuk mencari informasi-informasi khusus terkait butir soal yang dibuat. Kegiatan analisis butir soal diharapkan dapat digunakan untuk perbaikan dan penyempurnaan soal agar kualitas soal yang dibuat semakin baik.

Sudjana (2012: 135) juga memaparkan bahwa analisis butir soal merupakan kegiatan mengidentifikasi dan mengkaji soal yang sangat diperlukan untuk memperoleh butir soal yang baik atau memadai.

Dari pendapat ahli diatas dapat dilihat bahwa analisis butir soal merupakan kegiatan mengidentifikasi dan mengkaji soal yang penting dilakukan seorang pengajar agar pengajar dapat memperoleh soal dengan kualitas yang baik.

### **VALIDITAS**

Arikunto (2016: 72) menjelaskan sebuah data atau informasi bisa dikatakan valid apabila sesuai dengan keadaan senyatanya. Apabila data yang dihasilkan dari sebuah instrumen dinyatakan valid maka bisa dinyatakan bahwa instrumen tersebut valid. Dalam sebuah tes, tes tersebut akan dinyatakan baik/valid apabila tes tersebut dapat mengukur apa yang akan diukur.

Selain itu Arikunto juga menggolongkan validitas menjadi 4 jenis validitas: (1) Validitas Isi (*content validity*) atau disebut juga validitas kurikuler dimana sebuah tes akan dinyatakan mempunyai validitas isi apabila tes tersebut mempunyai tujuan untuk mengukur tujuan khusus tertentu, dimana tujuan khusus tersebut sejajar dengan materi isi pelajaran yang diajarkan; (2) Validitas Konstruksi (*construct validity*) merupakan validitas dimana tes mempunyai validitas konstruksi apabila butir-butir soal yang dibuat pada tes tersebut mengukur setiap aspek berfikir seperti yang sudah tertera pada Indikator sebagai contoh jika rumusan pada indikator tertulis “siswa dapat menjelaskan perbedaan penggunaan partikel は dan が dalam sebuah kalimat” maka pertanyaan yang akan ditulis dalam tes adalah sebuah perintah untuk menuliskan perbedaan antara 2 partikel tersebut; (3) Validitas “ada sekarang” (*concurrent validity*) Validitas “ada sekarang”

atau disebut juga dengan validitas empiris merupakan validitas dimana tes yang dibuat memiliki hasil yang sesuai dengan pengalaman yang sudah terjadi dimasa lampau. Dengan kata lain hasil tes akan dibandingkan dengan hasil tes yang telah dilakukan sebagai contoh ketika seorang pengajar akan menilai apakah soal yang dibuatnya sudah valid atau belum maka diperlukan sebuah pembandingan atau kriterium yang sekarang datanya sudah dimiliki; (4) Validitas prediksi (*predictive validity*) merupakan validitas dimana tes yang disusun memiliki kemampuan untuk memprediksi apa yang akan terjadi dimasa depan sebagai contoh soal *Try Out* digunakan untuk mengukur kemampuan siswa sebelum menghadapi Ujian Nasional. Disini dari hasil tes dapat digunakan untuk memprediksi hasil pada Ujian Nasional yang akan datang. Siswa yang hasil *Try Out*nya baik diharapkan pada Ujian Nasional nilai yang diperoleh juga baik. Setelah hasil Ujian Nasional keluar maka hasil tersebut dibandingkan dengan hasil *Try Out* sebelumnya Apabila siswa yang nilai *Try Out* yang didapat baik dan Ujian Nasional juga mendapat hasil yang baik maka soal *Try Out* tersebut memiliki validitas prediksi, namun apabila hasil yang didapatkan bertolak belakang maka soal *Try Out* tersebut tidak memiliki validitas prediksi.

### **RELIABILITAS**

Arifin (2014: 258) menyebutkan bahwa reliabilitas merupakan sebuah tingkatan derajat konsistensi dari sebuah instrumen. Arikunto (2016: 100) juga menjelaskan tingkat reliabilitas sebuah instrumen ditentukan dari tingkat ke validtan dari sebuah instrumen, Sebuah instrumen dinyatakan mempunyai tingkat reliabilitas yang tinggi apabila instrumen tersebut dapat menghasilkan data yang sesuai dengan apa yang ada.

Selain itu Arikunto (2016: 105) juga menyebutkan beberapa hal yang dapat mempengaruhi hasil dari sebuah tes. Hasil dari sebuah tes dapat dipengaruhi dari tes itu sendiri seperti kualitas dan banyaknya jumlah butir soal tes. Banyaknya jumlah butir soal tes dapat mempengaruhi validitas dan reliabilitas dari sebuah tes. Tes dengan soal yang lebih banyak maka tingkat validitas dan reliabilitas akan semakin tinggi. Jumlah dari peserta tes juga dapat mempengaruhi hasil tes, semakin banyak peserta tes semakin beragam pula jawaban yang ada. Dari keberagaman jawaban tersebut dapat menggambarkan besar kecilnya reliabilitas. Selain itu penyelenggaraan tes juga dapat mempengaruhi hasil tes.

### **TINGKAT KESUKARAN SOAL**

Arikunto (2016: 222) mengungkapkan bahwa tingkat kesukaran soal juga berpengaruh terhadap kualitas suatu soal. Soal yang tidak terlalu mudah dan soal yang tidak terlalu sulit merupakan soal yang baik. Apabila soal yang dibuat oleh seorang pengajar terlalu sulit hal tersebut akan

berpengaruh pada siswa, siswa akan menjadi putus asa dan berfikir bahwa soal tersebut memang berada di luar kemampuannya, dan juga apabila soal yang dibuat terlalu mudah soal tersebut tidak dapat mengembangkan pemikiran dan kemampuan siswa tersebut.

Arikunto (2016: 225) juga mengklasifikasikan tingkat mudah sulitnya suatu soal atau disebut indeks kesukaran menjadi beberapa bagian soal dengan indeks kesukaran (P) 0,00 sampai dengan 0,30 merupakan soal yang sulit, soal dengan indeks kesukaran (P) 0,31 sampai dengan 0,70 termasuk soal golongan sedang, dan soal dengan indeks kesukaran (P) 0,70 sampai dengan 1,00 merupakan soal golongan mudah.

### **DAYA PEMBEDA**

Arifin (2014:273) menjabarkan daya pembeda merupakan sebuah perhitungan yang bertujuan untuk mengkategorikan yang mana siswa yang menguasai sebuah kompetensi dan siswa yang belum atau kurang menguasai kompetensi tersebut. Arikunto (2016:225) juga menjelaskan bahwa daya pembeda adalah kemampuan sebuah pertanyaan untuk membedakan peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah.

Arikunto (2016:226) juga memaparkan cara melakukan analisis daya pembeda yaitu dengan menyusun skor jawaban siswa dari yang paling tinggi ke yang paling rendah, kemudian membagi siswa dalam 2 kelompok sama besar, siswa yang memiliki skor diatas sampai menengah dikelompokkan dalam kelompok atas (*upper group*) sedangkan setengahnya atau siswa yang memiliki skor rendah dikelompokkan dalam kelompok bawah (*lower group*).

### **EFEKTIFITAS PENGECOH**

Arifin (2014:277) menjelaskan bahwa pengecoh merupakan jawaban yang salah dari sebuah pertanyaan. Pengecoh dikatakan efektif ketika pengecoh yang dibuat berhasil menjalankan fungsinya sebagaimana fungsi dari pengecoh tersebut. Selain itu Arifin (2014:280) juga mengungkapkan kriteria persentase Indeks Pengecoh (IP) berdasarkan beberapa kriteria yaitu apabila indeks pengecoh (IP) 76% sampai dengan 125% maka indeks pengecoh pilihan jawaban tersebut dikategorikan sangat baik, apabila indeks pengecoh (IP) 51% sampai dengan 75% atau 126% sampai dengan 150% maka indeks pengecoh pilihan jawaban tersebut dikategorikan baik, jika indeks pengecoh (IP) 26% sampai dengan 50% atau 151% sampai dengan 175% maka indeks pengecoh pilihan jawaban tersebut dikategorikan kurang baik, jika indeks pengecoh (IP) 0% sampai dengan 25% atau 176% sampai dengan 200% maka indeks pengecoh pilihan jawaban

tersebut dikategorikan jelek, dan apabila lebih dari 200% maka dikategorikan sangat jelek.

Arifin (2014:279) juga memaparkan sebuah pengecoh disebut berhasil melaksanakan fungsinya dengan baik ketika sebuah pengecoh yang ada dipilih oleh siswa, paling sedikit 5% dari seluruh jumlah siswa yang mengikuti tes tersebut. Artinya disini efektifitas pengecoh mengkaji dan menghitung seberapa besar fungsi dari sebuah pengecoh itu berjalan, apabila pengecoh yang telah dibuat berhasil melaksanakan fungsinya dengan baik maka pengecoh tersebut bisa dipakai kembali.

### **METODE**

Dalam penelitian ini, digunakan pendekatan kuantitatif deskriptif, yang memiliki tujuan menganalisis dan mendeskripsikan mengenai kualitas butir soal tes yang diamati. Kocimaheni (2017) Penelitian kuantitatif deskriptif juga termasuk penelitian kuantitatif *non-experimental*.

Penelitian yang dilakukan dibagi menjadi beberapa tahap. Pertama pada tahap mengumpulkan data, teknik yang digunakan adalah teknik dokumentasi karena data diamati dan dituliskan ke dalam bentuk tabel. Data yang dimaksud diambil dari lembar jawaban seluruh siswa berupa jawaban dari siswa pada setiap butir soal yang ada sebagai persiapan awal sebelum melakukan perhitungan. Pada tahap selanjutnya data dihitung berdasarkan item-item yang akan dianalisis, dan dilakukan 5 buah analisis yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektifitas pengecoh. Untuk perhitungan digunakan bantuan dari aplikasi *IBM SPSS Statistic 26* dan *Microsoft Excel 2020*. Data yang telah dihitung kemudian dianalisis menurut kaidah setiap item dan dideskripsikan sesuai hasil yang ada.

Dalam penelitian ini subjek yang digunakan yaitu seluruh siswa kelas X Bahasa SMA Negeri 1 Driyorejo tahun ajaran 2018/2019 dan objek yang diteliti dalam penelitian ini adalah soal Ulangan Akhir Semester Bahasa Jepang.

### **UJI VALIDITAS**

Dalam menghitung validitas digunakan dua perhitungan yang berbeda pada setiap jenis soal, disini untuk soal pilihan ganda digunakan perhitungan koefisien korelasi biserial ( $Y_{pbi}$ ) dan untuk soal uraian digunakan perhitungan korelasi produk moment karena pada soal uraian nilai yang didapat pada setiap butir soal berupa angka kasar. Oleh karena itu untuk menghitung angka kasar perhitungan yang paling cocok digunakan adalah perhitungan korelasi produk moment.

Dari hasil perhitungan yang telah didapatkan kemudian mengkonsultasikan hasil perhitungan tersebut ke dalam  $r$ -tabel pada taraf signifikan 5%.

### **UJI RELIABILITAS**

Dalam menghitung tingkat reliabilitas digunakan perhitungan rumus *Alpha Cronbach*. Rumus ini digunakan karena rumus tersebut bisa diterapkan dalam dua jenis soal yang berbeda. Rumus *Alpha Cronbach* dapat digunakan dalam menghitung reliabilitas nilai angka kasar.

Setelah hasil perhitungan diperoleh kemudian hasil tersebut dianalisis dengan menggunakan patokan interpretasi koefisien reliabilitas tes yang telah disampaikan oleh Sudijono (2011:209). Ketika koefisien reliabilitas ( $r_{11}$ ) pada sebuah tes lebih besar atau sama dengan 0,70 maka tes yang dianalisis dinyatakan reliabel (*reliable*), apabila hasil dari perhitungan tersebut lebih kecil dari 0,70 maka soal tes tersebut dapat dinyatakan belum reliabel (*un-reliable*).

### **UJI TINGKAT KESUKARAN**

Tingkat kesukaran dapat dihitung dengan perhitungan yang telah dijelaskan oleh Arikunto (2016:223) yaitu dengan cara menghitung jumlah siswa yang menjawab benar soal tersebut dibagi dengan jumlah seluruh peserta tes.

Setelah hasil perhitungan atau disebut dengan indeks kesukaran sudah diketahui, selanjutnya harus dikelompokkan berdasarkan tingkat indeks kesukaran yang telah dijelaskan oleh Arikunto (2016: 225).

### **UJI DAYA PEMBEDA**

Seperti menghitung tingkat kesukaran soal, sebelum melakukan perhitungan Daya Pembeda (DP) ditentukan terlebih dahulu kelompok atas dan kelompok bawah. Perhitungan daya pembeda dapat dilakukan dengan penjelasan sebagai berikut: jumlah siswa kelompok atas dikurangi dengan jumlah siswa kelompok bawah, kemudian dibagi dengan jumlah siswa kelompok atas yang dapat menjawab soal dengan benar (nilai ideal kelompok atas) dikurangi dengan siswa kelompok bawah yang dapat menjawab soal dengan benar (nilai ideal kelompok bawah).

Kemudian apabila hasil perhitungan telah diperoleh, hasil tersebut harus dikelompokkan berdasarkan kriteria indeks pembeda yang dijabarkan oleh Arikunto (2016:232). Apabila nilai daya pembeda (D) bernilai 0,00 sampai dengan 0,20 menandakan daya beda soal tersebut memiliki daya beda jelek, 0,21 sampai dengan 0,40 menandakan daya beda soal tersebut cukup baik, 0,41 sampai dengan 0,70 menandakan dayabeda soal tersebut

baik, apabila nilai daya pembeda dari soal tersebut 0,71 sampai dengan 1,00 maka daya daya pembeda soal tersebut sangat baik, namun ketika nilai daya pembeda negatif hal tersebut menandakan tidak ada pembeda pada soal tersebut.

### **UJI EFEKTIFITAS PENGECOH**

Berbeda dengan analisis yang lain efektifitas pengecoh dihitung dengan *Microsoft Excel 2020*. Hal tersebut dikarenakan aplikasi *IBM SPSS Statistic 26* belum dapat digunakan untuk membantu proses perhitungan. Data yang digunakan untuk menghitung efektifitas pengecoh juga berbeda dengan analisis-analisis sebelumnya. Yang membedakan adalah apabila empat perhitungan yang lain membutuhkan data nilai dari setiap butir soal, untuk efektifitas membutuhkan data dari setiap pilihan jawaban yang dipilih oleh peserta tes. Sebagai contoh untuk menghitung validitas dibutuhkan data jumlah siswa yang menjawab benar soal nomor 1, sedangkan untuk menghitung efektifitas pengecoh data yang dibutuhkan dihitung dari berapa jumlah siswa yang menjawab A pada soal nomor 1, kemudian berapa siswa yang menjawab B pada soal nomor 1, dan seterusnya.

Mengitung persentase Indeks Pengecoh (IP) pada setiap butir jawaban dilakukan dengan cara membagi jumlah siswa yang menjawab pilihan jawaban tersebut dibagi dengan jumlah seluruh peserta tes dikali seratus persen (100%).

Setelah hasil persentase indeks pengecoh ditemukan kemudian dikelompokkan berdasarkan tingkat Indeks Pengecoh (IP) yang telah di jelaskan oleh Arifin (2014:280). Kemudian setelah dilakukan pengelompokkan kemudian disimpulkan berdasarkan skala *Likert*. Apabila tiga buah pengecoh berfungsi maka efektifitas pengecoh soal tersebut digolongkan baik, apabila dua buah pengecoh berfungsi maka efektifitas pengecoh soal tersebut digolongkan cukup, dan ketika satu pengecoh berfungsi maka efektifitas pengecoh soal tersebut digolongkan kurang baik. Apabila semua siswa menjawab benar maka pengecoh pada soal tersebut tidak berfungsi.

### **HASIL**

Hasil yang diperoleh dari analisis butir soal Ujian Akhir Semester Gasal mata pelajaran bahasa Jepang kelas X Bahasa SMA Negeri 1 Driyorejo tahun ajaran 2018/2019 sebagai berikut:

### **UJI VALIDITAS**

Dari hasil perhitungan yang telah dilakukan pada setiap butir soal, dari 10 soal pilihan ganda yang ada 8 soal (80%) dinyatakan valid dan 2 soal (20%) dinyatakan tidak valid. Untuk hasil perhitungan soal uraian dari 15 soal

**Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Bahasa Jepang Kelas X Bahasa SMA Negeri 1 Driyorejo Tahun Ajaran 2018/2019**

yang ada, 13 soal (87%) dinyatakan valid dan 2 soal (13%) dinyatakan tidak valid.

Tabel 1. Persebaran Soal Pilihan Ganda Kelas X berdasarkan analisis Validitas

No	Indeks Validitas	Nomor Soal	Jumlah Item	Persentase
1	$\geq 0,361$	1,2,3,4,5, 7,9,10	8	80%
2	$< 0,361$	6,8	2	20%

Tabel 2. Persebaran Soal Uraian Kelas X berdasarkan analisis Validitas

No	Indeks Validitas	Nomor Soal	Jumlah Item	Persentase
1	$\geq 0,361$	1,2,3,4,5, 6,7,8,10,11,12,13,14,15	13	87%
2	$< 0,361$	6,9	2	13%

**UJI RELIABILITAS**

Setelah melakukan perhitungan dengan patokan indeks reliabilitas yang telah diungkapkan oleh Sudijono (2011:209) untuk soal pilihan ganda setelah dihitung dengan *IBM SPSS Statistic 26* koefisien reliabilitas ( $r_{11}$ ) yang diperoleh sebesar 0,70, karena koefisien reliable yang diperoleh 0,70 maka soal pilihan ganda termasuk soal yang reliabel (*reliable*). Untuk soal uraian koefisien yang diperoleh setelah dihitung dengan *IBM SPSS Statistic 26* yaitu sebesar 0,92, karena 0,92 lebih besar dari 0,70 hal ini menandakan soal uraian termasuk soal yang reliabel (*reliable*).

**UJI TINGKAT KESUKARAN**

Berdasarkan perhitungan indeks tingkat kesukaran yang telah dilakukan dan diklasifikasikan menggunakan klasifikasi indeks kesukaran yang dijelaskan oleh Arikunto (2016: 225), hasil yang diperoleh sebagai berikut: untuk soal pilihan ganda dari 10 soal, 7 soal (70%) memiliki indeks kesukaran mudah, dan 3 soal (30%) memiliki indeks kesukaran sedang. Untuk soal uraian dari 15 soal, 9 soal (60%) memiliki indeks kesukaran mudah, dan 6 soal (40%) memiliki indeks kesukaran sedang.

Tabel 3. Persebaran Soal Pilihan Ganda Kelas X berdasarkan analisis Tingkat Kesukaran Soal.

No	Indeks Kesukaran	Nomor Soal	Jumlah Item	Persentase
1	0,00 – 0,30 (Sukar)	-	0	0%
2	0,31 – 0,70 (Sedang)	7,8,9	3	30%
3	0,71 – 1,00 (Mudah)	1,2,3,4, 5,6,10	7	70%

Tabel 4. Persebaran Soal Uraian Kelas X berdasarkan Tingkat Kesukaran Soal.

No	Indeks Kesukaran	Nomor Soal	Jumlah Item	Persentase
1	0,00 – 0,30 (Sukar)	-	0	0%
2	0,31 – 0,70 (Sedang)	3,5,6,8, 9,10	6	40%
3	0,71 – 1,00 (Mudah)	1,2,4,7, 11,12, 13,14, 15	9	60%

**UJI DAYA PEMBEDA**

Dari hasil perhitungan yang telah dilakukan dan dianalisis sesuai dengan kriteria yang diungkapkan oleh Arikunto (2016:232) hasil yang diperoleh sebagai berikut. Untuk soal pilihan ganda dari 10 soal, 8 soal (80%) tergolong soal dengan indeks pembeda yang baik dan 2 soal (20%) tergolong soal dengan indeks pembeda cukup. Untuk soal uraian dari 15 soal yang ada 9 soal (60%) tergolong soal dengan indeks pembeda sangat baik, 4 soal (27%) tergolong soal dengan indeks pembeda baik, dan satu soal (13%) tergolong soal dengan indeks pembeda cukup.

**Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Bahasa Jepang Kelas X Bahasa SMA Negeri 1 Driyorejo Tahun Ajaran 2018/2019**

Tabel 5. Persebaran Soal Pilihan Ganda Kelas X berdasarkan analisis Daya Pembeda.

No	Indeks pembeda	Nomor Soal	Jumlah Item	Persentase
1	Tanda Negatif	-	0	0
2	< 0,20 Lemah	-	0	0
3	0,20 – 0,39 Cukup	6,8	2	20%
4	0,40 – 0,69 Baik	1,2,3,4,5, 7,9,10	8	80%
5	0,70 – 1,00 Baik Sekali	-	0	0

Tabel 6. Persebaran Soal Uraian Kelas X berdasarkan analisis Daya Pembeda

No	Indeks Pembeda	Nomor Soal	Jumlah Item	Persentase
1	Tanda Negatif	-	0	0
2	< 0,20 Lemah	-	0	0
3	0,20 – 0,39 Cukup	6,10	2	13%
4	0,40 – 0,69 Baik	3,5,8,9,	4	27%
5	0,70 – 1,00 Baik Sekali	1,2,4,7,11, 12,13,14, 15	9	60%

**UJI EFEKTIFITAS PENGECOH**

Berdasarkan perhitungan efektifitas pengecoh yang telah dilakukan dan telah disesuaikan dengan kriteria yang diungkapkan oleh Arifin (2014:280) hasil yang diperoleh sebagai berikut: untuk soal pilihan ganda dari 10 soal, satu

soal (10%) tergolong soal dengan efektifitas pengecoh sangat baik, 4 soal (40%) tergolong soal dengan efektifitas pengecoh yang baik, 4 soal (40%) tergolong soal dengan efektifitas pengecoh yang cukup, dan satu soal (10%) tergolong soal dengan efektifitas pengecoh yang kurang baik.

Tabel 7. Persebaran Soal pilihan ganda Kelas X berdasarkan analisis Efektifitas Pengecoh.

No	Efektifitas Pengecoh	Nomor Soal	Jumlah Item	Persentase
1	Sangat Baik	7	1	10%
2	Baik	5,6,8,9	4	40%
3	Cukup Baik	1,2,4,10	4	40%
4	Kurang Baik	3	1	10%
5	Tidak Baik	-	0	0%

**PEMBAHASAN**

Berdasarkan penelitian yang dilakukan di SMA Negeri 1 Driyorejo dengan tujuan menganalisis kualitas butir soal berdasarkan indeks validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektifitas pengecoh.

Dari hasil perhitungan dengan menggunakan bantuan aplikasi *IBM SPSS Statistic 26*, dari 10 soal pilihan ganda 8 soal mendapatkan hasil valid (80%), sedangkan untuk soal uraian, dari 15 soal uraian 13 soal dinyatakan valid (87).

Validitas dihitung dengan menggunakan rumus koefisien korelasi point biserial ( $Y_{pbi}$ ), Hasil dari perhitungan tersebut akan dikonsultasikan ke dalam  $r$ -tabel pada taraf signifikansi 5% dari jumlah seluruh siswa yang diteliti. Apabila hasil yang di dapat dari perhitungan tersebut lebih besar daripada  $r$ -tabel maka soal tersebut dapat dinyatakan sebagai soal yang valid.

Arifin (2014: 247) mengemukakan bahwa instrumen yang digunakan oleh pengajar, dan jawaban dari peserta didik merupakan faktor-faktor yang dapat mempengaruhi tingkat kevaliditan sebuah soal.

Untuk perhitungan reliabilitas, dalam penelitian ini digunakan rumus alpha untuk menghitung koefisien reliabilitas ( $r_{11}$ ) dibantu dengan aplikasi *IBM SPSS Statistic 26*, setelah hasil perhitungan didapatkan, hasil tersebut akan di intrepretasikan ke patokan yang diungkapkan oleh



Sudijono (2011:209). Apabila hasil koefisien reliabilitas ( $r_{11}$ ) lebih besar dari 0,70 maka tes tersebut dinyatakan reliabel (*reliable*), apabila hasil dari perhitungan tersebut lebih kecil dari 0,70 maka soal tes tersebut dapat dinyatakan belum reliabel (*un-reliable*).

Dari hasil perhitungan dengan bantuan aplikasi *IBM SPSS Statistic 26*, hasil yang didapatkan untuk tes pilihan ganda yaitu sebesar 0,70 dan untuk tes uraian hasil soal uraian sebesar 0,92. Apabila diinterpretasikan ke dalam patokan yang disampaikan oleh Sudijono (2011:209) maka untuk tes pilihan ganda, karena hasil yang didapatkan sama dengan 0,70 maka dapat dinyatakan bahwa tes tersebut reliabel (*reliable*), dan untuk tes uraian karena hasil yang didapatkan lebih dari 0,70 maka dapat dinyatakan sudah reliabel (*reliable*).

Seperti yang dijelaskan Arikunto (2013: 101) tes yang tidak reliabel dapat diperbaiki dengan cara menambahkan jumlah butir soal yang valid, karena apabila jumlah soal yang valid semakin banyak maka koefisien reliabilitas juga akan semakin tinggi.

Tingkat kesukaran dapat dihitung dengan rumus yang telah dijelaskan oleh Arikunto (2016:223) yaitu dengan cara menghitung jumlah siswa yang menjawab benar soal tersebut dibagi dengan jumlah seluruh peserta tes. Setelah hasil perhitungan atau disebut dengan indeks kesukaran sudah diketahui, selanjutnya harus dikelompokkan berdasarkan tingkat indeks kesukaran yang telah dijelaskan oleh Arikunto, apabila tingkat kesukaran (P) 0,00 sampai dengan 0,30 soal tersebut merupakan soal golongan sukar, tingkat kesukaran (P) 0,31 sampai dengan 0,70 maka soal tersebut termasuk soal sedang, tingkat kesukaran 0,71 sampai dengan 1,00 maka termasuk soal golongan mudah.

Hasil perhitungan dengan bantuan aplikasi *IBM SPSS Statistic 26* untuk soal pilihan ganda tiga soal yaitu nomor 7, 8, dan 9 mendapat skor (0,53), (0,57), dan (0,50). Apabila dikelompokkan berdasarkan tingkat indeks kesukaran, maka tiga soal tersebut merupakan soal golongan sedang. Untuk tujuh soal yaitu nomor 1, 2, 3, 4, 5, 6, 10 mendapatkan skor (0,90), (0,87), (0,97), (0,73), (0,77), (0,83), (0,87). Apabila dikelompokkan berdasarkan tingkat indeks kesukaran, maka tujuh soal tersebut merupakan soal golongan mudah. Untuk keseluruhan soal pilihan ganda tiga soal 30% termasuk soal dengan golongan sedang, tujuh soal 70% termasuk soal dengan golongan mudah.

Hasil perhitungan dengan bantuan aplikasi *IBM SPSS Statistic 26* untuk soal uraian enam nomor yaitu nomor 3, 5, 6, 8, 9, 10 mendapatkan skor (0,57), (0,57), (0,55),

(0,53), (0,52), (0,50). Apabila dikelompokkan berdasarkan tingkat indeks kesukaran, maka enam soal tersebut merupakan soal golongan sedang. Untuk sembilan soal yaitu nomor 1, 2, 4, 7, 11, 12, 13, 14, dan 15 mendapatkan skor (0,75), (0,72), (0,73), (0,72), (0,78), (0,80), (0,78), (0,80), (0,85). Apabila dikelompokkan berdasarkan tingkat indeks kesukaran, maka sembilan soal tersebut merupakan soal golongan mudah. Untuk keseluruhan soal uraian enam soal 40% termasuk soal dengan golongan sedang, sembilan soal 60% termasuk soal dengan golongan mudah.

Daya pembeda soal dihitung dengan menggunakan *IBM SPSS Statistic 26*, setelah hasil perhitungan daya pembeda diperoleh, hasil daya pembeda tersebut digolongkan menurut kriteria yang telah diungkapkan oleh Arikunto (2016:232). Apabila nilai daya pembeda (D) bernilai 0,00 sampai 0,20 menandakan daya pembeda soal tersebut memiliki daya beda jelek, 0,21 sampai 0,40 menandakan daya beda soal tersebut cukup baik, 0,41 sampai 0,70 menandakan daya beda soal tersebut baik, apabila nilai daya pembeda dari soal tersebut 0,71 sampai dengan 1,00 maka daya pembeda soal tersebut sangat baik, namun ketika nilai daya pembeda negatif hal tersebut menandakan tidak ada pembeda pada soal tersebut.

Hasil perhitungan dengan menggunakan bantuan aplikasi *IBM SPSS Statistic 26* untuk soal pilihan ganda, dua soal yaitu nomor 6 dan nomor 8 mendapat skor 0,29 dan 0,32. Apabila dianalisis menurut kriteria, hal ini menunjukkan dua soal tersebut termasuk soal dengan daya pembeda cukup. Selanjutnya delapan soal yaitu nomor 1,2,3,4,5,7,9, dan 10 mendapat skor (0,47), (0,59), (0,60), (0,57), (0,57), (0,68), (0,53), (0,54). Apabila dianalisis menurut kriteria, hal ini menunjukkan delapan soal tersebut termasuk soal yang tergolong soal dengan indeks pembeda baik. Untuk keseluruhan soal pilihan ganda, dua soal 20% tergolong soal dengan indeks pembeda daya pembeda cukup, dan delapan soal 80% tergolong soal dengan indeks pembeda yang baik.

Hasil perhitungan dengan bantuan aplikasi *IBM SPSS Statistic 26* untuk soal uraian, dua soal yaitu nomor 6 dan nomor 10 mendapat skor 0,32 dan 0,32. Apabila dianalisis menurut kriteria, hal ini menunjukkan dua soal tersebut termasuk soal yang memiliki daya pembeda cukup, empat soal yaitu nomor 3, 5, 8, 9 mendapat skor (0,69), (0,69), (0,63), (0,69). Apabila dianalisis menurut kriteria, hal ini menunjukkan empat soal tersebut termasuk soal yang tergolong soal dengan indeks pembeda baik, dan Sembilan soal yaitu soal nomor 1, 2, 4, 7, 11, 12, 13, 14, 15 mendapat skor (0,75), (0,72), (0,75), (0,72), (0,87), (0,78), (0,78), (0,80), (0,85). Apabila dianalisis menurut kriteria, hal ini menunjukkan sembilan soal tersebut termasuk soal dengan daya pembeda sangat baik. Untuk keseluruhan soal uraian dua soal 13% tergolong soal dengan daya pembeda cukup,

dan empat soal 27% tergolong soal dengan daya pembeda yang baik, dan sembilan soal 60% tergolong soal dengan daya pembeda yang sangat baik.

Dari hasil perhitungan dengan menggunakan *Microsoft Excel 2020*, hasil yang diperoleh yaitu 2 soal (20%) tergolong soal dengan efektifitas pengecoh yang sangat baik, 3 soal (30%) tergolong soal dengan efektifitas pengecoh baik, 4 soal (40%) tergolong soal dengan efektifitas pengecoh cukup dan satu soal (10%) tergolong soal dengan efektifitas pengecoh kurang baik.

Sudjana (2011: 110) mengungkapkan ketika efektifitas pengecoh suatu soal tergolong kurang baik maupun negatif artinya pengecoh yang dibuat oleh pengajar belum dapat menarik minat siswa yang belum maupun yang tidak menguasai materi untuk memilih pengecoh tersebut. Soal yang termasuk golongan cukup maupun tidak baik, soal tersebut tidak harus dibuang. Pengajar hanya perlu mengganti dengan opsi jawaban yang lainnya,

## **SIMPULAN DAN SARAN**

### **SIMPULAN**

Dari hasil analisis butir soal Ujian Akhir Semester Gasal bahasa Jepang SMA Negeri 1 Driyorejo simpulan yang dapat diambil sebagai berikut:

- (1) Hasil dari analisis validitas ditunjukkan bahwa soal dengan pilihan ganda, mendapatkan hasil valid berjumlah 8 soal (80%), sedangkan untuk soal uraian yang mendapatkan hasil valid berjumlah 13 soal (87%). Dengan hasil tersebut validitas soal ujian akhir semester memiliki validitas yang baik.
- (2) Hasil analisis reliabilitas mendapatkan hasil 0,70 hal ini menunjukkan bahwa soal tersebut reliabel karena apabila hasil koefisien reliabilitas sama dengan 0,70 maka soal tersebut dikatakan reliabel. Untuk soal uraian mendapatkan hasil 0,92 hal ini menunjukkan bahwa soal tersebut sudah reliabel.
- (3) Hasil dari analisis kesukaran soal menunjukkan, dari 10 soal pilihan ganda 7 soal digolongkan mudah (70%), dan 3 soal dengan digolongkan sedang (30%). Untuk soal uraian dari 15 soal uraian 9 soal dapat digolongkan mudah (60%) dan 6 soal digolongkan sedang (40%).
- (4) Hasil dari analisis daya beda, untuk soal pilihan ganda mendapatkan hasil 8 soal dengan daya beda baik (80%) dan dua soal berdaya beda cukup (20%), sedangkan soal uraian mendapatkan hasil sembilan soal dengan

daya beda baik sekali (60%), 4 soal berdaya beda baik (27%), dan satu soal berdaya beda cukup (13%).

- (5) Dari segi efektifitas pengecoh 2 soal mempunyai efektifitas pengecoh sangat baik (10%), 3 soal mempunyai efektifitas pengecoh yang baik (20%), 4 soal mempunyai efektifitas pengecoh cukup (40%) dan satu soal mempunyai efektifitas pengecoh yang kurang baik (10%)

### **SARAN**

Dari hasil analisis diharapkan kepada pengajar untuk meningkatkan kualitas soal dengan melakukan analisis butir soal agar standard kualitas soal menjadi lebih baik. Selain itu pihak pengajar harus didukung juga oleh pihak sekolah agar guru lebih termotivasi untuk melakukan evaluasi kualitas butir soal karena apabila standar soal yang ada di sekolah lebih baik, maka standar sekolah juga akan lebih tinggi. Untuk peneliti selanjutnya diharapkan pada saat menganalisis kualitas butir soal tidak hanya soal UAS, melainkan masuk ke ranah yang lebih dalam seperti soal Ulangan Harian, dan dalam menghitung menggunakan dua rumus yang berbeda, agar dapat dilihat perbandingan hasil ke dua rumus tersebut, Dengan demikian hasil yang didapatkan dapat menjadi hasil yang benar-benar valid, selain itu setelah membaca penelitian ini diharapkan agar menarik minat pengajar/mahasiswa untuk melaksanakan penelitian terkait analisis butir soal.

### **DAFTAR PUSTAKA**

- Arifin, Zainal. 2014. *Evaluasi Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Arikunto, Suharsimi. 2006. *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: PT. Rineka Cipta.
- Arikunto, Suharsimi. 2016. *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.
- Arini, Masitho. 2015. Analisis Kualitas Butir Soal Ujian Sekolah Mata Pelajaran Ekonomi kelas XII SMA Negeri 1 Taman Tahun Ajaran 2015/2016. Diakses pada Januari 2021
- Kocimaheni, Amira Agustin. 2017. *Kecenderungan Penelitian Mahasiswa Program Study Bahasa Jepang Fakultas Bahasa dan Seni Universitas Negeri Surabaya*. Jurnal Asa Online.

(<https://journal.unesa.ac.id/index.php/asa/article/view/2474>) diakses pada Desember 2020

- Nurgiyantoro, Burhan. 2012. *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta: B PFE
- Prasetyo, Joko. 2018. *Problematika Guru dalam Mengimplementasikan Penilaian Kurikulum 2013 pada Siswa Kelas IV di SDN IV Tempursari*. Jurnal.  
(<https://scholar.google.co.id/citations?user=JdYo-wwAAAAJ&hl=en>) diakses pada Januari 2021
- Raharjo, R. Antonius Mulyono S. 2013. *Analisis Butir Soal Kenaikan Kelas XI Tahun Pelajaran 2012/2013 Buatan Dinas Pendidikan Kota Semarang*. Skripsi. Semarang. Diakses pada Desember 2020.
- Sudijono, Anas. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Rajagrafindo Persada.
- Sudjana, Nana. 2012. *Penelitian Hasil Proses Belajar Mengajar*. Bandung: PT. Remaja Rosdakarya.
- UNESA. 2000. *Pedoman Penelitian Artikel Jurnal*, Surabaya: Lembaga Penelitian Universitas Negeri Surabaya.
- Wantah, Maria J. 2010. "Validitas dan Reliabilitas Instrumen Asesmen Anak TK Berkebutuhan Khusus". *Jurnal Penelitian dan Evaluasi Pendidikan*,
- Yanagisawa, Yoshiaki dan Eriko Ishii. 1998. *Japanese Language Resource 1000 Book (Nihongo Kyouiku Juuyouyougo)*. Baber Press.