

ANALISIS KUALITAS SOAL PREDIKSI UJIAN NASIONAL BAHASA INGGRIS MADRASAH ALIYAH TAHUN 2019

(Hasil Kerja Peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019)

Wachidun

Widyaiswara Ahli Muda
Balai Diklat Keagamaan Denpasar
Email: idunwachidun@gmail.com

Diterima: 16 Maret 2021 | Direvisi: 8 April 2021 | Disetujui: 14 Juni 2021

Abstrak

Penelitian ini bertujuan untuk mengetahui Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh soal prediksi Ujian Nasional Mata Pelajaran Bahasa Inggris Hasil Kerja Peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun 2019. Metode pengumpulan data dilakukan dengan menggunakan metode dokumentasi untuk mendapatkan data yang meliputi soal, kunci jawaban, kisi-kisi soal dan lembar jawab siswa. Teknik analisis data yang digunakan adalah teknik deskriptif kuantitatif. Hasil penelitian ini menunjukkan bahwa ditinjau dari: (1) Validitas, soal yang valid berjumlah 22 butir (44%), dan tidak valid 28 butir (56%). (2) Reliabilitas, soal memiliki Reliabilitas tinggi dengan koefisien Reliabilitas sebesar 0,77 (3) Tingkat Kesukaran, soal yang tergolong sangat sukar berjumlah 3 butir (6%), sukar berjumlah 11 butir (22%), sedang 28 butir (56%), mudah 4 butir (8%), dan sangat mudah 4 butir (8%). (4) Daya Pembeda, butir soal dengan Daya Pembeda sangat jelek 18 butir (36%), Daya Pembeda jelek 4 butir (8%), Daya Pembeda sedang 12 butir (24%), Daya Pembeda baik 12 butir (24%) dan Daya Pembeda sangat baik 4 butir (8%). (5) Efektivitas Pengecoh, Butir soal dengan Efektivitas Pengecoh sangat baik 6 butir (12%), Efektivitas Pengecoh baik 5 butir (10%), Efektivitas Pengecoh buruk 18 butir (26%), dan Efektivitas Pengecoh sangat buruk 21 butir (42%).

Kata kunci: Analisis Butir Soal, Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, Efektivitas Pengecoh

Abstract

This study aims to determine the validity, reliability, level of difficulty, item discrimination, and distracter effectiveness of the prediction of the National Examination for English subjects made by the participants of Substantive Technical Training in the National Madrasah Aliyah English Language Examination in 2019. Data collection methods were carried out using the documentation method to collect the data that includes the test questions, key answers, and student answer sheets. The data analysis technique used is quantitative descriptive technique. The results of this study indicate that in terms of: (1) Validity, valid questions are 22 items (44%), and 28 items are invalid (56%). (2) Reliability, questions have a high reliability with a reliability coefficient of 0.77 (3) Difficulty level, questions that are classified as very difficult amounted to 3 items (6%), difficult amounted to 11 items (22%), moderate were 28 items (56%), easy 4 items (8%), and very easy 4 items (8%). (4) Item discrimination, items with very poor discrimination index are 18 items (36%), poor discrimination index 4 items (8%), moderate discrimination index 12 (24%), good discrimination index 12 (24%) and very good discrimination 4 items (8%). (5) Distracter Effectiveness, items with very good distracter effectiveness are 6 items (12%), the good ones are 5 items (10%), the poor ones are 18 items (26%), and the poorest ones are 21 items (42%).

Keywords: *Item analysis, Validity, Reliability, Difficulty Level, Item Discrimination, Distracter Effectiveness*

PENDAHULUAN

Kunci keberhasilan belajar peserta didik dan mutu lulusan yang berkualitas sangat dipengaruhi oleh guru yang profesional dalam pembelajaran. Dengan demikian seorang guru yang profesional akan melaksanakan tugas utamanya yakni merencanakan pembelajaran, melaksanakan pembelajaran dan menilai proses serta hasil pembelajaran dengan baik pula. Penilaian proses maupun hasil pembelajaran yang dilakukan untuk mengetahui kemampuan siswa dan untuk mengetahui tingkat keberhasilan pembelajaran sering dikenal dengan istilah evaluasi. Menurut Anas Sudijono (2015: 5), evaluasi adalah kegiatan atau proses untuk menilai sesuatu dan untuk dapat menentukan nilai dari sesuatu yang sedang dinilai tersebut, dilakukanlah pengukuran. Wujud dari pengukuran tersebut adalah pengujian yang dalam dunia pendidikan dikenal dengan istilah tes.

Masih menurut Sudijono (2015: 16), evaluasi dilakukan dengan tujuan untuk memperoleh data pembuktian, yang akan menjadi petunjuk sampai di mana tingkat kemampuan dan tingkat keberhasilan peserta didik dalam pencapaian tujuan-tujuan kurikuler setelah mereka menempuh proses pembelajaran dalam jangka waktu yang telah ditentukan serta untuk mengukur dan menilai sampai di manakah efektivitas mengajar dan metode-metode mengajar yang telah diterapkan atau dilaksanakan oleh pendidik. Sedangkan Sukardi (2011) menyatakan bahwa tujuan evaluasi dalam kaitannya dengan belajar mengajar antara lain: (1) Menilai ketercapaian (*attainment*) tujuan, (2) Mengukur macam-macam aspek belajar yang bervariasi, (3) Sebagai sarana (*means*) untuk mengetahui apa yang siswa telah ketahui, (4) Memotivasi belajar siswa, (5) Menyediakan informasi untuk tujuan bimbingan dan konseling, (6) Menjadikan hasil evaluasi sebagai dasar perubahan kurikulum.

Salah satu teknik yang digunakan untuk melakukan evaluasi terhadap hasil belajar siswa yaitu dengan tes. Anas Sudijono (2015: 67) menyatakan bahwa tes adalah cara atau prosedur yang perlu ditempuh dalam rangka pengukuran dan penilaian di bidang pendidikan, yang berbentuk pemberian tugas atau serangkaian tugas berupa pertanyaan-pertanyaan yang harus dijawab atau perintah-perintah yang harus dikerjakan oleh testee, sehingga dapat dihasilkan nilai yang melambangkan tingkah laku atau prestasi testee.

Masih menurut Anas Sudijono (2015: 67) suatu tes yang diujikan kepada siswa harus memiliki kualitas yang baik. Tes yang berkualitas dapat memberikan informasi yang sesungguhnya mengenai

hasil belajar siswa, sehingga informasi tersebut dapat digunakan untuk mengukur tingkat perkembangan atau kemajuan yang telah dicapai oleh peserta didik setelah mereka menempuh proses belajar mengajar dalam jangka waktu tertentu. Di samping itu, informasi yang diperoleh dari tes yang berkualitas dapat digunakan untuk mengetahui sudah seberapa jauh program pengajaran yang telah ditentukan dapat tercapai.

Menurut Nana Sudjana (2014: 12), suatu alat penilaian (tes) dikatakan mempunyai kualitas yang baik apabila alat tersebut memenuhi dua hal, yakni ketepatan atau validitas dan ketetapan atau keajegan atau reliabilitas. Anas Sudijono (2015: 370) menjelaskan bahwa penganalisisan terhadap butir-butir soal tes hasil belajar agar memiliki kualitas yang tinggi dapat dilakukan dari tiga segi yaitu dari segi tingkat kesukaran itemnya, dari segi daya pembeda itemnya dan dari segi fungsi distraktornya (fungsi pengecoh).

Soal tes yang berkualitas tersusun atas butir-butir soal yang berkualitas pula. Agar dapat mengetahui kualitas suatu soal tes, maka perlu dilakukan analisis terhadap kualitas soal. Analisis kualitas soal merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu soal tes, baik secara keseluruhan maupun butir soal yang menjadi bagian dari soal tes tersebut. Analisis kualitas soal yang dimaksudkan adalah analisis yang ditinjau dari masing-masing aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Analisis validitas dan reliabilitas dapat digunakan untuk mengetahui kualitas soal secara keseluruhan, sedangkan analisis tingkat kesukaran, daya pembeda dan efektivitas pengecoh digunakan untuk mengetahui kualitas butir soal.

Sementara itu Sumarna Surapranata (2009: 10-11) menyatakan bahwa salah satu tujuan dilakukannya analisis terhadap kualitas soal adalah untuk meningkatkan kualitas soal, yaitu apakah suatu soal telah dapat diterima, perlu diperbaiki karena memiliki beberapa kelemahan atau tidak digunakan sama sekali karena tidak berfungsi. Akan tetapi, masih banyak guru yang belum melakukan analisis terhadap soal tes yang dibuat karena menganggap bahwa menganalisis kualitas soal tes memerlukan waktu yang lama dan juga membutuhkan banyak tenaga. Seperti yang dinyatakan oleh Widoyoko (2014: 130) tentang pentingnya guru melakukan analisis terhadap soal tes yang dibuat. Dampak dari para guru tidak melakukan analisis terhadap soal yang dibuat adalah banyak butir soal yang digunakan dalam tes tidak dapat menghasilkan data yang benar

atau akurat tentang hasil belajar siswa. Apabila keputusan yang diambil didasarkan pada data yang tidak benar atau tidak akurat, yang disebabkan oleh instrumen yang digunakan untuk mengumpulkan data tidak disusun secara baik, maka keputusan tersebut merupakan keputusan yang tidak dapat dipertanggungjawabkan.

Balai Diklat Keagamaan Denpasar sebagai unit pelaksana teknis (UPT) di bidang kediklatan melaksanakan Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah pada tanggal 12-17 Agustus 2019. Peserta diklat berjumlah 30 orang peserta yang merupakan guru Bahasa Inggris dari tiga provinsi, yaitu Bali, Nusa Tenggara Barat, dan Nusa Tenggara Timur. Sesuai dengan nama diklatnya maka mata diklat-mata diklat yang terdapat dalam diklat tersebut adalah materi substansi yang berkaitan dengan ujian nasional Bahasa Inggris. Salah satunya adalah praktik merancang soal ujian nasional Bahasa Inggris Madrasah Aliyah. Jadi peserta diklat yang berjumlah 30 orang peserta dibagi menjadi beberapa kelompok dan diberi tugas untuk menyusun beberapa item soal prediksi UN Bahasa Inggris Madrasah Aliyah. Kegiatan menyusun soal diawali dengan menyusun kisi-kisi soal terlebih dahulu sebelum menuliskan menjadi soal. Dalam penulisan butir soal para peserta diklat telah dibekali dengan pengetahuan kaidah penulisan butir soal pilihan ganda. Analisis secara kualitatif dilakukan dengan mempresentasikan dan ditanggapi oleh kelompok lain, serta mendapat konfirmasi dari Widyaiswara yang mengajar. Sedangkan analisis secara kuantitatif belum bisa dilakukan karena mengingat waktu yang sangat terbatas dan tahun ajaran baru baru saja akan dimulai sehingga materi atau soal-soal yang akan diujikan kepada siswa kelas XII belum diajarkan.

Agar dapat mengetahui kualitas soal prediksi Ujian Nasional Mata Pelajaran Bahasa Inggris yang dibuat oleh peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun 2019 yang ditinjau dari aspek Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh, perlu diadakan penelitian yang berjudul "Analisis Kualitas Soal prediksi Ujian Nasional Mata Pelajaran Bahasa Inggris yang dibuat oleh peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun 2019".

Rumusan masalah dalam penelitian ini

1. Bagaimanakah kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek validitas soal?

2. Bagaimanakah kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek reliabilitas soal?
3. Bagaimanakah kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek tingkat kesukaran?
4. Bagaimanakah kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek daya pembeda soal?
5. Bagaimanakah kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek kualitas pengecoh/opsi jawaban?

Dari rumusan masalah di atas, dapat kita rumuskan beberapa hipotesis pada penelitian ini, yaitu:

1. Kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek validitas soal adalah valid.
2. Kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek reliabilitas soal berkategori reliabel.
3. Kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek tingkat kesukaran berkategori baik.
4. Kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek daya pembeda soal adalah berkategori baik.
5. Kualitas soal prediksi Ujian Nasional Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah Tahun 2019 ditinjau dari aspek kualitas pengecoh/opsi jawaban adalah berkategori baik/efektif.

Tujuan dari penelitian ini yaitu untuk mengetahui Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh soal prediksi Ujian Nasional Mata Pelajaran Bahasa Inggris yang dibuat oleh peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun

2019. Sedangkan manfaat Hasil penelitian ini diharapkan dapat memberi manfaat bagi berbagai pihak. Bagi guru penelitian ini dapat untuk mengetahui kualitas soal yang dibuat, sehingga guru memiliki acuan untuk melakukan tindak lanjut terhadap soal yang dibuatnya. Bagi peneliti yaitu untuk menambah wawasan dan pengalaman mengenai penerapan ilmu evaluasi pendidikan khususnya pada bidang analisis kualitas soal.

KAJIAN TEORI

1. Analisis Butir Soal

Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Tugas melakukan evaluasi terhadap alat pengukuran yang telah digunakan untuk mengukur keberhasilan belajar peserta didik pada umumnya dilupakan oleh evaluator. Menurut Nana Sudjana (2006: 135), "Analisis butir soal atau analisis item adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai". Menurut Daryanto (2007: 177), "Analisis soal adalah suatu prosedur sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun".

Tujuan penelaahan butir soal adalah untuk mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu untuk digunakan. Di samping itu, tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif serta untuk mengetahui informasi diagnostik pada peserta didik apakah mereka sudah atau belum memahami materi yang telah diajarkan.

Soal yang bermutu adalah soal yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuannya, di antaranya dapat menentukan peserta didik mana yang sudah atau belum menguasai materi yang diajarkan guru. Salah satu cara memperbaiki proses belajar-mengajar yang paling efektif adalah dengan cara mengevaluasi tes hasil belajar yang diperoleh dari proses belajar-mengajar itu sendiri. Dengan kata lain, hasil tes tersebut kita olah sedemikian rupa sehingga hasil dari pengolahan itu dapat diketahui komponen manakah dari proses belajar-mengajar itu yang masih lemah. Pengolahan tes hasil belajar dalam rangka memperbaiki proses belajar-mengajar salah satunya adalah dengan melakukan analisis butir soal.

2. Teknik Analisis Butir Soal

Hasil suatu analisis soal dapat digunakan untuk mengetahui apakah soal akan berfungsi dengan baik atau tidak. Pada dasarnya cara yang ditempuh dalam analisis soal adalah melalui telaah (analisis kualitatif) dan analisis berdasarkan data hasil uji coba (analisis kuantitatif) yang meliputi validitas, reliabilitas, daya beda, tingkat kesukaran dan efektifitas distraktor. Analisis secara kualitatif terhadap butir soal pada dasarnya adalah penelaahan butir soal ditinjau dari segi kaidah penulisan soal yaitu (1) isi atau materi; (2) konstruksi; dan (3) bahasa.

Analisis soal secara kualitatif dilakukan sebelum suatu soal diujikan. Analisis ini belum memberikan gambaran tentang karakteristik psikometri soal. Oleh karena itu, untuk membuktikan bahwa soal-soal itu sudah baik perlu diujicobakan terhadap sejumlah siswa. Jawaban siswa terhadap soal-soal itu dijadikan dasar untuk analisis kuantitatif. Ada 2 macam karakteristik soal yang ditinjau dalam analisis kuantitatif yaitu tingkat kesukaran dan daya beda. Tingkat kesukaran adalah proporsi jumlah peserta tes yang menjawab benar dengan jumlah peserta seluruhnya. Daya pembeda adalah kemampuan suatu soal untuk membedakan antara siswa yang mampu dengan siswa yang kurang mampu.

Suatu tes dapat dikatakan sebagai alat pengukur yang baik jika memenuhi kriteria berikut ini, antara lain:

a. Validitas

Valid sering diartikan sebagai kesahihan. Suatu alat ukur disebut valid bilamana alat ukur tersebut memiliki isi yang layak mengukur objek yang harus diukur dan sesuai dengan kriteria tertentu. Dalam hal ini ada kesesuaian antara alat ukur dengan fungsi pengukuran dan sarana pengukuran. Menurut Ridlo (2005), valid merupakan kualitas yang menunjukkan hubungan antara suatu pengukuran dengan tujuan belajar. Sedangkan menurut Supriyadi (2005) validitas tes mempunyai arti bahwa tes harus mengukur apa yang semestinya harus diukur. Suatu instrument dikatakan valid apabila mampu mengukur apa yang seharusnya diukur. Lebih lanjut Arikunto (2002) mengatakan bahwa validitas dapat dibagi menjadi 4 macam yaitu validitas isi, validitas konstruksi, validitas ramalan dan validitas "ada sekarang."

b. Reliabilitas

Reliabel berarti dapat dipercaya. Reliabilitas (keandalan) adalah ketetapan atau ketelitian suatu

alat evaluasi. Suatu tes atau alat evaluasi dikatakan andal jika ia dapat dipercaya, konsisten atau stabil dan produktif (Arifin, 2009). Tes reliabel dapat tidak valid, tetapi tes yang valid biasanya reliabel. Di antara hal yang mempengaruhi hasil tes adalah banyaknya soal dalam tes dan kualitas butir soal. Semakin panjang tes biasanya makin reliabel. Cara mencari reliabilitas antara lain, 1) metode parallel (*equivalent*), 2) metode tes ulang (*test-retest method*), dan 3) metode belah dua (*split-half method*).

3. Tingkat Kesukaran

Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya, soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya. Ada beberapa alasan untuk menyatakan tingkat kesukaran soal. Bisa saja tingkat kesukaran soal ditentukan oleh kedalaman soal dan kompleksitas.

Dalam hal ini, bermutu tidaknya butir soal-soal tes hasil belajar dapat diketahui dari derajat kesukaran soal yang dimiliki oleh masing-masing butir soal tersebut. Butir-butir soal tes belajar dapat dikatakan sebagai butir soal yang baik, jika butir soal tersebut tidak terlalu mudah juga tidak terlalu sulit. Dengan kata lain, taraf kesukaran soal tersebut adalah sedang/cukup. Angka yang dapat memberi petunjuk mengenai tingkat kesukaran soal itu dikenal dengan istilah "*Difficulty indek*", dalam evaluasi belajar umumnya dilambangkan huruf P (*propotion*). Besarnya tingkat kesukaran antara 0,00 sampai dengan 1,0. Menurut Arikunto (2002) klasifikasi indeks kesukaran adalah sebagai berikut:

Soal dengan P antara 0,00 sampai 0,30 adalah soal sukar

Soal dengan P antara 0,31 sampai 0,70 adalah soal sedang

Soal dengan P antara 0,71 sampai 1,00 adalah soal mudah

Sedangkan dalam penelitian ini klasifikasi indeks kesukaran yang dipakai adalah sebagai berikut.

Besarnya Nilai P	Kategori Tingkat Kesukaran
$0 < P \leq 0,15$	Sangat Sukar
$0,15 < P \leq 0,3$	Sukar
$0,3 < P \leq 0,7$	Sedang
$0,7 < P \leq 0,85$	Mudah
$0,85 < P \leq 1$	Sangat Mudah

4. Daya Pembeda

Menurut Arikunto (2002) daya pembeda soal adalah kemampuan soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah). Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi (D). Seperti halnya indeks kesukaran, indeks diskriminasi ini berkisar antara 0,00 sampai 1,00. Hanya bedanya, indeks kesukaran tidak mengenal tanda negatif (-), tetapi pada indeks diskriminasi ada tanda negatif. Tanda negatif pada indeks diskriminasi digunakan jika sesuatu soal "terbalik" menunjukkan kualitas testee. Yaitu anak pandai disebut bodoh dan anak bodoh disebut pandai. Dengan menggunakan program Iteman, Anates atau yang lain dapat menggambarkan tingkat kemampuan soal dalam membedakan antara peserta didik yang sudah memahami materi yang diujikan dengan peserta didik yang belum atau tidak memahami materi yang diujikan. Menurut Arikunto (2002) klasifikasi daya pembeda soal adalah sebagai berikut:

D = 0,00 - 0,20 daya pembeda soal adalah jelek

D = 0,21 - 0,40 daya pembeda soal adalah cukup

D = 0,41 - 0,70 daya pembeda soal adalah baik

D = 0,71 - 1,00 daya pembeda soal adalah baik sekali

D = negatif daya pembeda soal adalah sangat jelek

Sedangkan dalam penelitian ini, klasifikasi daya pembeda yang dipakai adalah:

Besarnya Nilai D	Kategori Daya Pembeda
$D \leq 0$	Sangat Jelek
$0,00 < D \leq 0,2$	Jelek
$0,2 < D \leq 0,4$	Sedang
$0,4 < D \leq 0,8$	Mudah
$0,8 < D \leq 1$	Sangat Mudah

5. Kualitas Pengecoh/Opsi

Option atau alternatif jawaban dalam soal jumlahnya berkisar antara 3 sampai 5 opsi. Selain itu juga terdapat kemungkinan satu jawaban benar dari ke lima alternatif jawaban tersebut. Sedangkan sisanya merupakan jawaban yang salah. Jawaban yang salah itulah yang biasa dikenal dengan istilah distraktor atau pengecoh. Menganalisis fungsi distraktor sering dikenal dengan istilah lain yaitu menganalisis dengan pola penyebaran jawaban soal. Adapun yang dianalisis dengan pola penyebaran soal adalah suatu pola yang menggambarkan bagaimana testee menentukan pilihan jawabannya terhadap kemungkinan jawaban yang telah dipasangkan pada setiap butir soal.

Menurut Daryanto (2001) yang dimaksud pola jawaban soal adalah distribusi peserta tes dalam hal menentukan jawaban pada soal bentuk pilihan ganda. Dari pola jawaban soal dapat ditentukan apakah pengecoh (distraktor) dapat berfungsi sebagai pengecoh yang baik atau tidak. Pengecoh yang tidak dipilih sama sekali oleh peserta tes, berarti pengecoh tersebut buruk. Sebaliknya, pengecoh dapat dikatakan baik jika pengecoh tersebut mempunyai daya tarik yang besar bagi peserta tes yang kurang memahami konsep atau kurang menguasai bahan untuk dipilih sebagai jawaban soal. Suatu pengecoh (distraktor) dapat dikatakan berfungsi baik jika paling sedikit dipilih 5% peserta tes.

Metode Penelitian

1. Jenis Penelitian

Jenis penelitian ini adalah penelitian deskriptif kuantitatif. Artinya penelitian ini dilakukan secara kuantitatif tetapi tidak untuk menerima atau menolak hipotesis, melainkan untuk menjelaskan keadaan yang apa adanya sesuai dengan keadaan objek yang diteliti. Dengan penelitian deskriptif digunakan pengumpulan data untuk mengetahui keadaan objek yang diteliti. Penelitian ini berusaha melaporkan atau mendeskripsikan keadaan objek yang diteliti secara apa adanya, dalam hal ini yaitu kriteria soal prediksi Ujian Nasional mata pelajaran Bahasa Inggris yang dibuat oleh peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun 2019 termasuk baik atau kurang baik yang diteliti secara kuantitatif (tingkat kesukaran, daya pembeda, dan keefektifan pengecoh).

2. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah seluruh siswa kelas XII IPA MAN Karangasem Tahun Ajaran 2019/2020 yang terdiri dari 25 siswa kelas XII IPA-1 dan 23 siswa kelas XII IPA-2. Karena jumlah keseluruhan populasi yang kurang dari 100 orang maka peneliti menggunakan seluruh populasi sebagai sampel dalam penelitian ini.

3. Metode pengumpulan data

Teknik yang digunakan dalam pengumpulan data adalah teknik dokumentasi yaitu untuk mendapatkan data berupa lembar soal prediksi Ujian Nasional mata pelajaran Bahasa Inggris yang dibuat oleh peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun 2019, lembar kunci jawaban, dan lembar jawaban peserta didik.

4. Analisis Data

Data tentang validitas, realibilitas, kualitas tingkat kesukaran, daya pembeda dan efektifitas

pengecoh dianalisis menggunakan aplikasi Anates versi 4.0.9

HASIL DAN PEMBAHASAN

A. Hasil

Data yang dikumpulkan dalam penelitian adalah hasil analisis jawaban 48 siswa yang mengerjakan soal prediksi UN 2019 menggunakan Anates Versi 4.0.9. Hasil analisis dapat disajikan sebagai berikut:

1. Validitas Soal

Berdasarkan hasil analisis Validitas butir soal Prediksi UN Bahasa Inggris MA buatan peserta diklat UN Bahasa Inggris MA tahun 2019 diketahui terdapat 22 butir soal valid atau sebesar 44% dan butir soal tidak valid berjumlah 28 butir atau 56%. Analisis Validitas butir soal dapat ditelusuri dengan menggunakan rumus korelasi point biserial (r_{pb}). Indeks korelasi point biserial yang diperoleh dari hasil perhitungan dikonsultasikan dengan r tabel pada taraf signifikansi 5% sesuai jumlah subjek atau jumlah peserta tes. Subjek dalam penelitian ini berjumlah 48 siswa, sehingga dengan $n=48$ didapatkan hasil r tabel sebesar 0,2403. Jika koefisien korelasi point biserial lebih besar atau sama dengan r tabel, maka butir soal tersebut dinyatakan valid, tetapi jika koefisien korelasi point biserial lebih kecil dari r tabel, butir soal dinyatakan tidak valid.

Sebagai tindak lanjut dari hasil analisis Validitas butir soal, butir soal yang valid dapat digunakan kembali pada tes berikutnya atau disimpan di bank soal. Butir soal yang tidak valid diperbaiki dengan disesuaikan pada indikator pencapaian maupun disesuaikan dengan teknik penyusunan butir soal.

2. Reliabilitas

Hasil perhitungan Reliabilitas soal Prediksi UN Bahasa Inggris MA buatan peserta diklat UN Bahasa Inggris MA tahun 2019 menunjukkan bahwa soal memiliki koefisien Reliabilitas (r_{11}) sebesar 0,77. Reliabilitas soal dapat dihitung dengan menggunakan rumus KR-20, dengan interpretasi apabila koefisien Reliabilitas lebih dari atau sama dengan 0,70 maka soal tersebut dinyatakan reliabel atau memiliki Reliabilitas tinggi dan apabila koefisien Reliabilitas kurang dari 0,70 maka soal termasuk dalam kategori tidak reliabel atau memiliki Reliabilitas rendah. Hasil tersebut berarti bahwa soal Prediksi UN Bahasa Inggris MA buatan peserta diklat UN Bahasa Inggris MA tahun 2019 termasuk dalam kategori soal dengan Reliabilitas tinggi atau reliabel karena r_{11} lebih dari 0,70.

3. Tingkat kesukaran

Dari hasil analisis data tentang tingkat kesukaran dapat ditampilkan data sebagai berikut :

Tabel 4.2 Distribusi Frekuensi dan Persentase Tingkat Kesukaran Soal prediksi UN Bahasa Inggris Madrasah Aliyah Hasil Kerja Peserta Diklat DTS UN Bahasa Inggris tahun 2019

No	Kriteria	No soal	Jumlah	Persentase
1	Sangat Mudah	8, 11, 12, 37	4	8%
2	Mudah	9, 16, 17, 26,	4	8%
3	Sedang	2,3,4,5,6,7,10,13,14,15,18,22,23,27,31, 32,36, 38,39,40,41,42,43,45,47,48	28	56%
4	Sukar	1, 19, 20, 24, 25, 30, 34, 35, 46, 49, 50	11	22%
5	Sangat Sukar	21, 28, 44	3	6%

4. Daya Pembeda

Distribusi ke 50 soal tersebut berdasarkan daya pembeda sebagai berikut:

Tabel 4.4. Distribusi Frekuensi dan Persentase Daya Pembeda

No	Kriteria	No soal	Jumlah	Persentase
1	Sangat Jelek	6, 8, 11, 12, 13, 19, 20, 21, 23, 24, 25, 28, 30, 44, 45, 46, 49, 50	18	36%
2	Jelek	9, 34, 35, 47	4	8%
3	Cukup	1,4,7,10,14,15,26,36,37,39,40,43	12	24%
4	Baik	2, 3, 5, 16, 17, 18, 22, 32, 33, 38, 41, 48	12	24%
5	Sangat Baik	27, 29, 31, 42	4	8%

5. Kualitas Pengecoh

Hasil analisis data tentang efektivitas pengecoh dapat disajikan data sebagai berikut ;

No	Kategori	Jumlah	Dalam %
----	----------	--------	---------

1	Sangat Baik	6	12
2	Baik	5	10
3	Buruk	18	36
4	Sangat Buruk	21	42

B. Pembahasan

Berdasarkan hasil analisis kuantitatif butir soal prediksi UN Bahasa Inggris Madrasah Aliyah Hasil Kerja Peserta Diklat DTS UN Bahasa Inggris tahun 2019 dalam penelitian ini dengan jumlah peserta tes sebanyak 48 peserta didik dan jumlah jumlah butir soal sebanyak 50 butir soal memperoleh karakteristik butir soal yang meliputi: tingkat kesukaran, daya pembeda, dan efektivitas opsi.

Berdasarkan table 4.2 di atas, dapat kita simpulkan bahwa persentase terbesar adalah soal dengan kategori sedang, yaitu 28 butir (56%), sedangkan soal dengan kategori mudah dan sangat mudah masing-masing 4 butir (8%), sedangkan sukar 11 butir (22%) dan sangat sukar 3 butir (6%). Khusus butir soal dengan kategori sedang bisa langsung disimpan dalam buku bank soal, sedangkan butir soal dengan kategori mudah dan sangat mudah serta sukar dan sangat sukar perlu dikaji dan diidentifikasi penyebabnya. Sebagai contoh, butir soal no 8,11, 12 dan 37 dengan kategori sangat mudah sehingga hampir seluruh peserta tes menjawab betul pada soal-soal tersebut. Kemendiknas (2010) dalam buku Panduan Penulisan Butir Soal menjelaskan bahwa beberapa kemungkinan terkait soal yang sangat mudah, yaitu (1) Sebagian besar siswa/peserta tes telah paham dengan materi yang diujikan, dan (2) pengecoh butir soal tidak berfungsi. Sedangkan untuk butir soal nomor 21, 28 dan 44 hanya 2 dan 7 siswa yang menjawab benar pada soal nomor tersebut. Beberapa kemungkinan penyebab rendahnya jumlah siswa yang menjawab benar pada soal tersebut adalah (1) Butir soal tersebut "mungkin" salah kunci jawaban, (2) Butir soal tersebut memiliki 2 atau lebih jawaban yang benar, (3) Materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya, (4) Materi yang diukur tidak cocok ditanyakan menggunakan bentuk soal yang diberikan, dan (5) Pernyataan atau kalimat soal terlalu panjang atau kompleks. Setelah soal-soal tersebut diperbaiki perlu kiranya diujicobakan lagi sebelum nantinya dipergunakan lagi.

Terkait dengan proporsi jumlah soal mudah, sedang, dan sukar belum ada aturan baku yang menetapkan jumlah pastinya. Namun begitu, secara umum proporsi jumlahnya mengikuti kurva normal, yaitu 2-6-2 (20% soal mudah, 60% soal sedang, dan 20% soal sukar), atau 2,5-5-2,5 (25% soal mudah, 50% soal sedang, 25% soal sukar), atau 3-4-3 (30%

soal mudah, 40% soal sedang, dan 30% soal sukar). Intinya, jumlah soal dengan kategori sedang jumlahnya harus lebih banyak dari soal dengan kategori mudah dan sukar.

Melihat data hasil analisis tingkat kesukaran di atas, maka kualitas soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif Ujian Nasional Bahasa Inggris Madrasah Aliyah tahun 2019 ditinjau dari tingkat kesukaran soal dapat dikatakan dalam kategori baik, di mana jumlah item soal dengan kategori sedang 56%, mudah dan sangat mudah 16%, sementara soal dengan kategori sukar dan sangat sukar 28%.

Yang kedua terkait dengan daya pembeda. Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara siswa yang telah menguasai materi yang ditanyakan dan siswa yang tidak/kurang/belum menguasai materi yang ditanyakan. Berikut ini hasil analisis daya pembeda pada soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah 2019.

Mengacu pada data Tabel 4.3 di atas, terdapat 36% soal dengan daya pembeda sangat jelek dan 8% soal dengan daya pembeda jelek. Khusus untuk soal-soal yang berada pada kategori ini harus dibuang, diganti, atau direvisi terlebih dahulu sebelum digunakan kembali pada ujian yang akan datang. Sedangkan 56% soal yang lain bisa dimasukkan dalam buku bank soal untuk dapat dipakai lagi pada ujian yang akan datang.

Seperti telah dijelaskan di atas bahwa daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara siswa yang telah menguasai materi yang ditanyakan dan siswa yang tidak/kurang/belum menguasai materi yang ditanyakan. Berarti untuk butir soal dengan kategori jelek dan sangat jelek yang berjumlah 44% harus dikaji dan identifikasi penyebabnya. Kemendiknas (2010) memberikan beberapa kemungkinan penyebab soal dengan daya pembeda yang jelek, di antaranya adalah:

1. Kunci jawaban butir soal itu tidak tepat.
2. Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar
3. Kompetensi yang diukur tidak jelas
4. Pengecoh tidak berfungsi
5. Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak
6. Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya

Terakhir adalah terkait kualitas pengecoh. Seperti yang tersaji dalam Tabel 4.5 tentang hasil analisis kualitas pengecoh dapat dirangkum bahwa

18 butir soal (36%) memiliki pengecoh/opsi jawaban dalam kategori buruk dan 21 butir soal (42%) sangat buruk, 5 butir soal (10%) memiliki kualitas pengecoh dengan kategori baik dan 6 butir soal (12%) dengan kategori sangat baik, Untuk butir soal dengan kategori sangat baik, butir soal tersebut bisa langsung disimpan dalam bank soal dan dijaga kerahasiannya agar bisa dipergunakan lagi untuk waktu yang akan datang. Sedangkan untuk butir soal dengan kategori baik, soal-soal tersebut dapat disimpan dalam bank soal dengan syarat opsi yang tidak berfungsi direvisi. Untuk butir soal dengan kategori buruk maka soal-soal tersebut tidak dapat disimpan dalam bank soal. Soal tersebut harus direvisi sampai memenuhi kriteria soal yang baik. Sedangkan untuk soal dengan kategori sangat buruk maka soal-soal tersebut tidak dapat disimpan dalam bank soal. Soal tersebut harus direvisi sampai memenuhi kriteria soal yang baik atau soal tersebut dibuang dan diganti dengan soal yang baru.

Mengacu pada data hasil analisis kualitas pengecoh di atas, dapat disimpulkan bahwa kualitas soal-soal prediksi UN Bahasa Inggris Madrasah Aliyah buatan peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 dinyatakan belum baik karena jumlah butir soal dengan kategori buruk dan sangat buruk jauh lebih banyak daripada jumlah butir soal dengan kategori baik dan sangat baik, yaitu 22% diterima berbanding 78% yang harus dibuang atau diganti.

Hasil Analisis Validitas butir, Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh

Dari data hasil penelitian dan pembahasan dapat dirangkum hasil analisis validitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh dalam Tabel 4.4 berikut:

Tabel 4.3 Rekap Hasil Analisis Validitas butir, Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh

No Butir	Validitas	Tingkat Kesukaran	Daya Pembeda	Kualitas Pengecoh	Keterangan
1	-	Sukar	Cukup	Sangat Baik	Dibuang
2	Sangat Signifikan	Sedang	Baik	Sangat Buruk	Diperbaiki
3	Sangat Signifikan	Sedang	Baik	Buruk	Diperbaiki
4	-	Sedang	Cukup	Sangat Buruk	Dibuang

5	Sangat Signifikan	Sedang	Baik	Buruk	Diperbaiki
6	-	Sedang	Sangat Jelek	Buruk	Dibuang
7	-	Sedang	Cukup	Buruk	Dibuang
8	-	Sangat Mudah	Sangat Jelek	Sangat Buruk	Dibuang
9	-	Mudah	Jelek	Sangat Buruk	Dibuang
10	Signifikan	Sedang	Cukup	Buruk	Diperbaiki
11	-	Sangat Mudah	Sangat Jelek	Sangat Buruk	Dibuang
12	-	Sangat Mudah	Sangat Jelek	Sangat Buruk	Dibuang
13	-	Sedang	Sangat Jelek	Buruk	Dibuang
14	-	Sedang	Cukup	Sangat Buruk	Dibuang
15	-	Sedang	Cukup	Sangat Buruk	Dibuang
16	Sangat Signifikan	Mudah	Baik	Buruk	Dibuang
17	Sangat Signifikan	Mudah	Baik	Sangat Baik	Diperbaiki
18	Signifikan	Sedang	Baik	Buruk	Diperbaiki
19	-	Sukar	Sangat Jelek	Sangat Buruk	Dibuang
20	-	Sukar	Sangat Jelek	Sangat Buruk	Dibuang
21	-	Sangat Sukar	Sangat Jelek	Buruk	Dibuang
22	Sangat Signifikan	Sedang	Baik	Buruk	Diperbaiki
23	-	Sedang	Sangat Jelek	Sangat Buruk	Dibuang
24	-	Sukar	Sangat Jelek	Sangat Buruk	Dibuang
25	-	Sukar	Sangat Jelek	Buruk	Dibuang
26	-	Mudah	Cukup	Buruk	Dibuang
27	Sangat Signifikan	Sedang	Baik	Sangat Baik	Disimpan
28	-	Sangat Sukar	Sangat Jelek	Sangat Buruk	Dibuang
29	Sangat Signifikan	Sedang	Sangat Baik	Buruk	Diperbaiki
30	-	Sukar	Sangat Jelek	Sangat Buruk	Dibuang

31	Sangat Signifikan	Sedang	Sangat Baik	Sangat Buruk	Diperbaiki
32	Signifikan	Sedang	Baik	Buruk	Diperbaiki
33	Sangat Signifikan	Sedang	Baik	Sangat Buruk	Diperbaiki
34	-	Sukar	Jelek	Sangat Buruk	Dibuang
35	-	Sukar	Jelek	Buruk	Dibuang
36	Signifikan	Sedang	Cukup	Baik	Disimpan
37	Sangat Signifikan	Sangat Mudah	Cukup	Sangat Buruk	Dibuang
38	Sangat Signifikan	Sedang	Baik	Sangat Buruk	Diperbaiki
39	Signifikan	Sedang	Cukup	Sangat Buruk	Diperbaiki
40	-	Sedang	Cukup	Baik	Diperbaiki
41	Sangat Signifikan	Sedang	Baik	Sangat Buruk	Diperbaiki
42	Sangat Signifikan	Sedang	Sangat Baik	Baik	Disimpan
43	Sangat Signifikan	Sedang	Cukup	Sangat Baik	Disimpan
44	-	Sangat Sukar	Sangat Jelek	Buruk	Dibuang
45	-	Sedang	Sangat Jelek	Baik	Dibuang
46	-	Sukar	Sangat Jelek	Buruk	Dibuang
47	-	Sedang	Jelek	Sangat Baik	Dibuang
48	Sangat Signifikan	Sedang	Baik	Sangat Baik	Disimpan
49	-	Sukar	Sangat Jelek	Buruk	Dibuang
50	-	Sukar	Sangat Jelek	Baik	Dibuang
Total	Disimpan: 4 butir (8%) Diperbaiki: 15 butir (30%) Dibuang/diganti: 31 butir (62%)				

Butir soal dapat disimpan di dalam bank soal atau digunakan kembali apabila memenuhi empat kriteria yaitu ditinjau dari Validitas butir soal termasuk butir soal valid, ditinjau dari Tingkat Kesukaran termasuk dalam kategori Tingkat Kesukaran sedang, ditinjau dari Daya Pembeda

termasuk kategori Daya Pembeda sedang, baik dan baik sekali dan ditinjau dari Efektivitas Pengecoh termasuk butir soal dengan pengecoh baik dan sangat baik. Butir soal prediksi UN Bahasa Inggris Madrasah Aliyah buatan peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 yang dapat disimpan atau digunakan kembali berjumlah 4 butir atau 8%. Butir soal tersebut perlu dijaga kerahasiaannya untuk dapat dikeluarkan kembali pada ujian berikutnya.

Butir soal yang hanya memenuhi tiga dari keempat kriteria, sebaiknya diperbaiki. Butir soal prediksi UN Bahasa Inggris Madrasah Aliyah buatan peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 yang memenuhi tiga kriteria berjumlah 15 butir atau 30%. Butir soal yang sudah diperbaiki dapat diujikan kembali untuk mengetahui apakah butir soal tersebut sudah menjalankan fungsinya atau belum. Butir soal yang hanya memenuhi dua kriteria sebaiknya tidak digunakan kembali. Butir soal-soal prediksi UN Bahasa Inggris Madrasah Aliyah buatan peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 yang hanya memenuhi dua kriteria berjumlah 31 butir atau 62%. Butir soal tersebut sebaiknya tidak digunakan kembali, karena apabila hendak digunakan kembali memerlukan perbaikan yang signifikan.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian dan pembahasan pada bagian sebelumnya dapat disimpulkan bahwa:

1. Kualitas soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 ditinjau dari aspek tingkat kesukaran soal berada pada kategori baik, yaitu 56% soal dengan kategori sedang.
2. Kualitas soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 ditinjau dari aspek daya pembeda soal berada pada kategori baik. Hal ini dikarenakan sebanyak 56% butir item mampu membedakan peserta tes yang berkemampuan tinggi dengan peserta tes/siswa yang berkemampuan rendah.
3. Kualitas soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 ditinjau dari aspek kualitas pengecoh/opsi berada pada kategori tidak baik, yaitu hanya 5 butir soal (10%) dalam kategori baik, 6 butir soal dengan kategori sangat baik, 18 butir soal (36%) berkategori buruk dan 21 butir soal (42%)

dengan kategori sangat buruk. Ini berarti hanya 11 butir soal (22%) yang dapat disimpan dalam bank soal, sedangkan 39 butir (78%) tidak dapat disimpan dan harus dibuang atau diganti.

4. Kualitas soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 ditinjau dari aspek validitas belum bisa dinyatakan sebagai soal yang baik karena hanya 21 butir soal (42%) yang valid, sedangkan 29 butir soal (58%) tidak valid.
5. Kualitas soal prediksi UN Bahasa Inggris hasil kerja peserta Diklat Teknis Substantif UN Bahasa Inggris Madrasah Aliyah tahun 2019 ditinjau dari aspek realibilitas termasuk soal yang baik karena koefisien realibilitasnya di atas 0,7, yaitu sebesar 0,77.

Saran

Saran yang dapat dikemukakan berdasarkan kesimpulan yang diperoleh dari hasil penelitian ini, yaitu:

1. Kepala Madrasah perlu kiranya mengadakan pelatihan, workshop atau bimbingan teknis penulisan soal agar guru lebih memahami pedoman dalam membuat soal yang baik dan sebagai sarana untuk meningkatkan kemampuan dan keterampilan dalam menulis soal, khususnya pada aspek materi dan bahasa.
2. Guru hendaknya melakukan administrasi terhadap butir-butir soal yang telah teruji kualitasnya sebagai langkah awal pembuatan bank soal, sehingga nantinya guru dapat merakit soal Ujian Sekolah yang berasal dari butir-butir soal yang baik.
3. Kerja sama antarguru bidang studi antarsekolah dalam forum-forum ilmiah perlu ditingkatkan, misalnya dalam MGMP, agar guru dapat berbagi pengalaman dan mengetahui perubahan-perubahan terkini terkait dengan situasi pendidikan yang dinamis sehingga guru dapat semakin profesional dalam menjalankan tugas profesinya, salah satunya mengenai pengukuran hasil belajar peserta didik.

DAFTAR PUSTAKA

- Arikunto, S. 2002. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arifin, Z. 2009. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.

- Daryanto. 2001. *Evaluasi Pendidikan*. Jakarta: Rineka Cipta.
- Jahja, U. dkk. 1996. *Bahan Penataran Pengujian Pendidikan*. Jakarta: Balitbang Dikbud.
- Kemendiknas. 2010. Panduan Penulisan Butir Soal. Jakarta: Direktorat Pembinaan SMP Ditjen Manajemen Pendidikan Dasar Dan Menengah
- Ridlo, S. 2005. Diklat Bahan Ajar; *Evaluasi Pembelajaran Biologi*. Semarang: Jurusan Biologi FMIPA UNNES.
- Sudijono, A. 2015. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Sudjana, N. (2014). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Sukardi. 2011. *Evaluasi Pendidikan: Prinsip & Operasionalnya*. Jakarta: Bumi Aksara
- Surapranata, S. (2009). *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*. Bandung: PT Remaja Rosdakarya.
- Supriyadi. 2005. *Kajian Penilaian Pencapaian Belajar Fisika*. Malang: Universitas Negeri Malang.
- Widoyoko, E.P. (2014). *Penilaian Hasil Pembelajaran di Sekolah*. Yogyakarta: Pustaka Pelajar.