

Validitas dan Reliabilitas dalam Analisis Butir Soal

Farida¹, Anna Musyarofah²

¹ IAIN Syaikh Abdurrahman Siddik Bangka Belitung

² IAIN Syaikh Abdurrahman Siddik Bangka Belitung

Info Artikel :

Diterima 13 April, 2021

Direvisi 27 Mei, 2021

Dipublikasikan 03 Juni 2021

Kata Kunci

Validitas

Reliabilitas

Analisis Butir Soal

Keywords:

Validity

Reliability

Item analysis

Abstrak

Kegiatan menganalisis butir soal adalah kegiatan yang harus dilakukan guru dalam rangka meningkatkan mutu soal yang telah disusun dan bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Dengan analisis soal dapat diperoleh informasi tentang kejelekan sebuah soal dan petunjuk untuk mengadakan perbaikan. Soal ulangan akhir semester (UAS) bahasa arab buatan guru di MA Al-Islam Kemuja belum pernah dianalisis baik secara kualitas dan kuantitas. Penelitian ini bertujuan untuk mengetahui kualitas butir soal dari tingkat validitas, reliabilitas, tingkat kesukaran serta daya beda soal pada UAS semester genap mata pelajaran bahasa arab kelas X di MA Al-Islam Kemuja tahun ajaran 2019/2020.

Penelitian ini merupakan penelitian kuantitatif, sedangkan pendekatan dalam penelitian ini adalah pendekatan deskriptif. Metode pengumpulan data yang digunakan dalam penelitian ini yaitu dokumentasi.

Hasil penelitian menunjukkan bahwa: hasil analisis validitas menunjukkan 11 soal kategori soal yang valid dan 24 soal dalam kategori tidak valid. pada hasil analisis reliabilitas menunjukkan besarnya korelasi koefisien $r_{11}=0,669$ ini menunjukkan reliabilitas butir soal memiliki kategori soal yang rendah reliabilitasnya. Pada tingkat kesukaran menunjukkan soal yang termasuk kategori sukar berjumlah 10 butir soal (28%), sedang berjumlah 23 butir soal (66%) dan mudah berjumlah 2 butir soal (6%). Sedangkan hasil analisis daya beda butir soal menunjukkan daya pembedanya tidak baik berjumlah 19 butir (54%), butir soal yang daya pembedanya cukup baik berjumlah 3 butir (9%) dan butir soal yang daya pembedanya baik berjumlah 13 butir (37%).

مستخلص البحث

عمل تحليل بنود الأسئلة هو عمل الذي يجب أن يقوم المعلم في تحسين جودة الأسئلة التي إعدادها ويهدف إلى التعرف على الأسئلة الجيدة، غير جيد و السيئة. بتحليل الأسئلة يمكن الحصول على معلومات عن القبح السؤال و تعليمات لإجراء التحسينات. الأسئلة في الإمتحان النهائي لمادة اللغة العربية في المدرسة العالية الإسلامية "الإسلام" كموجا لم يتم تحليلها من الجودة والكمية. وهذا البحث لمعرفة عن الجودة الأسئلة من درجة الصدق، درجة الثبات، درجة الصعوبة و درجة التمييز في الإمتحان النهائي لمادة اللغة العربية في المدرسة الثانوية الإسلامية "الإسلام" كموجا سنة دراسية 2019\2020. هذا البحث هو بحث كمي، وأما النهج في هذا البحث هو نهج وصفي. و طريقة جمع البيانات المستخدمة في هذا البحث هو

الوثائق. ظهرت من نتائج البحث: أن من درجة الصدق هناك ١١ سؤالاً صحيحة و أسئلة غير صحيحة ٢٤ سؤالاً. من درجة الثبات هي الأسئلة التي لها ثبات منخفض لأن معامل الثبات أقل من ٧٠,٠٠. من درجة الصعوبة ينظر في الفئة الصعبة، هناك ١٠ أسئلة (٢٨%) ، في الفئة المتوسطة هناك ٢٣ سؤالاً (٦٦%) و في الفئة السهلة هناك ٢ سؤالين (٦%). من درجة التمييز، أظهر أن له درجة التمييز الضعيف كانت ١٩ (٥٤%) ، الأسئلة لها درجة التمييز الكافي كانت ٣ أسئلة (٩%) سؤالاً، و الأسئلة لها درجة التمييز الجيد كانت ١٣ (٣٧%) سؤالاً.



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2019 by author.

Koresponden:

Farida, Anna Musyarofah
Faridarida3012@gmail.com, musyarofah@iainsasbabel.ac.id

Pendahuluan

Pendidikan merupakan usaha yang dilakukan dengan sengaja dan terencana untuk mengembangkan potensi dalam diri individu. pendidikan secara sederhana dapat merujuk pada Kamus Besar Bahasa Indonesia (KBBI), merupakan proses pengubahan sikap dan tingkah laku seseorang atau sekelompok orang dalam upaya mendewasakan manusia melalui upaya pengajaran dan pelatihan.¹ Dalam pasal 1 ayat (1) Undang-Undang RI Nomor 2 tahun 1989 tentang sistem pendidikan nasional, pendidikan diartikan sebagai usaha sadar untuk menyiapkan peserta didik melalui kegiatan bimbingan, pengajaran dan atau latihan bagi peranannya di masa yang akan datang.²

Dua komponen penting dalam dunia pendidikan diantaranya adalah guru dan siswa. Guru yang mengarahkan siswa dalam belajar dan siswa yang menjadi objek dalam pembelajaran tersebut. Dalam jurnal Abdimas Unwahas, Guza mengatakan bahwa standar kompetensi guru meliputi 4 kompetensi utama, yaitu kompetensi pedagogik, kompetensi kepribadian, kompetensi sosial, dan kompetensi profesional.³ Kompetensi pedagogik merupakan kemampuan guru dalam pengelolaan pembelajaran peserta didik yang sekurang-kurangnya meliputi beberapa hal berikut: a) pemahaman wawasan atau landasan pendidikan; b) pemahaman terhadap peserta didik; c) pengembangan kurikulum/silabus; d) perencanaan pembelajaran; e) pelaksanaan pembelajaran yang mendidik dan dialogis; f) pemanfaatan teknologi pembelajaran; g) evaluasi hasil belajar (EHB); h) pengembangan peserta didik untuk mengaktualisasikan berbagai potensi yang dimilikinya.⁴

¹ Febry Fahreza dan Rabiatul Rahmi, "Peningkatan Keterampilan Sosial Melalui Metode Role Playing Pada Pembelajaran IPS Di Kelas IV SD Negeri Pasi Pinang Kabupaten Aceh Barat", *Bina Gogik Volume 5 No 1*, Maret 2018.

² Ahmad Syar'i, *Filsafat pendidikan Islam*, (Jakarta: Pustaka Firdaus: 2005), hlm. 5

³ Happy Fitria, Muhammad Kristiawan, dkk, "Upaya Peningkatan Kompetensi Guru Melalui Pelatihan Penelitian Tindakan Kelas", *Jurnal Abdimas Unwahas*, Vol. 4, No. 1, April 2019, hlm. 15.

⁴ Winda Kusuma Wardani, "Analisis Kompetensi Pedagogik Guru Dalam Proses Pembelajaran di SD Muhammadiyah 15 Surakarta", *Skripsi Program Studi Pendidikan Guru Sekolah Dasar Universitas Muhammadiyah Surakarta*, 2017, hlm. 3

Undang-undang nomor 20 Tahun 2003 pasal 39 ayat 2 tentang sistem pendidikan nasional menyatakan bahwa pendidik adalah tenaga profesional yang bertugas merencanakan dan melaksanakan proses pembelajaran, menilai hasil pembelajaran, melakukan bimbingan dan pelatihan, serta melakukan penelitian dan pengabdian kepada masyarakat, terutama bagi pendidikan pada perguruan tinggi. Dengan demikian, salah satu kompetensi yang harus dimiliki seorang pendidik adalah kemampuan mengadakan evaluasi, baik dalam proses pembelajaran maupun penilaian hasil belajar.⁵

Dalam al-quran terdapat ayat yang menyebutkan tentang perlunya mengadakan evaluasi, diantaranya dalam surat Al-Ankabut ayat 2 dan 3:

أَحْسِبَ النَّاسُ أَنْ يُتْرَكُوا أَنْ يَقُولُوا ءَامَنَّا وَهُمْ لَا يُفْتَنُونَ (٢)

وَلَقَدْ فَتَنَّا الَّذِينَ مِنْ قَبْلِهِمْ فَلَيَعْلَمَنَّ اللَّهُ الَّذِينَ صَدَقُوا وَلَيَعْلَمَنَّ الْكَاذِبِينَ (٣)

Artinya: "apakah manusia mengira bahwa mereka akan dibiarkan hanya dengan mengatakan "kami telah beriman", dan mereka tidak diuji? dan sungguh, kami telah menguji orang-orang sebelum mereka, maka Allah pasti mengetahui orang-orang yang benar dan pasti mengetahui orang-orang yang dusta". (QS. Al-Ankabut: 2-3).⁶

Berdasarkan ayat di atas Allah SWT mengadakan ujian atau evaluasi kepada setiap makhluk-Nya untuk mengetahui sejauh mana kadar keimanan mereka. Dalam dunia pendidikan, diadakannya evaluasi adalah untuk mengetahui sejauh mana mutu dari komponen-komponen pendidikan tersebut. Misalnya seorang guru mengadakan evaluasi terhadap peserta didik, tujuannya adalah untuk mengetahui sejauh mana kemampuan peserta didik tersebut.⁷

Evaluasi atau penilaian merupakan konsep payung yang memayungi beberapa konsep dibawahnya, misalnya pengukuran dan tes. kata evaluasi berasal dari bahasa inggris yaitu *evaluation*, dan dalam bahasa arab *التقدير* dalam bahasa indonesia yang berarti penilaian. sedangkan secara istilah, sebagaimana yang dikemukakan oleh Edwin Wandt dan Gerald W.Brown yang dikutip dalam buku Pengantar Evaluasi Pendidikan oleh Anas Sudijono: *evaluation refer to the act or process to determining the value of something*, Yang berarti evaluasi itu menunjuk kepada atau mengandung pengertian: suatu tindakan atau suatu proses untuk menentukan nilai dari sesuatu.⁸

⁵ Rusydi Ananda dkk, *Evaluasi Pembelajaran*, (Bandung: Citapustaka Media, 2014), hlm.1.

⁶ Lihat, di Mushaf Al-Qur'an Terjemah, Al-Qur'an Terjemah Edisi tahun 2002, (Jakarta: Al-Huda (kelompok gema insani), 1426 H, hlm. 397.

⁷ Rahmatia Lessa Ulfah, "Analisis Butir Soal Tes Pilihan Ganda Penilaian Tengah Semester Genap Akidah Akhlak Di MTs Al-Hidayah Rawadenok 2017/2018", *Skripsi Jurusan Pendidikan Agama Islam Universitas Muhammadiyah Jakarta 1439 H/ 2018 M*.

⁸ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta: Rajawali Pers, 2016), hlm. 1

Penilaian adalah proses memberikan atau menentukan nilai kepada objek tertentu berdasarkan suatu kriteria tertentu,⁹ penilain dilakukan dengan tujuan untuk menilai proses dan hasil belajar siswa di sekolah, mendiagnosis kesulitan belajar siswa dan menentukan kenaikan kelas atau kelulusan.¹⁰ Fungsi penilaian adalah sebagai alat untuk mengetahui tercapai tidaknya tujuan instruksional, memberikan umpan balik bagi perbaikan proses belajar mengajar, meningkatkan motivasi belajar siswa, dan sebagai dasar dalam menyusun laporan kemajuan belajar siswa kepada para orang tuanya.¹¹

Secara garis besar, maka alat evaluasi yang digunakan dapat digolongkan menjadi dua macam, yaitu tes dan nontes.¹² Teknik tes dapat dilakukan baik secara tertulis maupun tidak tertulis. Teknik non-tes digunakan untuk menilai sikap, keterampilan, tingkah laku, dan sebagainya.¹³

Tes dalam pendidikan adalah alat penilaian atau metode penilaian yang sistematis, sah, dapat dipercaya dan objektif untuk menentukan kecakapan, keterampilan, dan tingkat pengetahuan siswa terhadap bahan ajar, berupa suatu tugas atau persoalan yang harus diselesaikan oleh seorang siswa atau sekelompok siswa.¹⁴ Berdasarkan objektivitas penskoran, tes dibagi kedalam tes objektif dan tes subjektif.

Tes objektif dapat berupa tes benar salah, tes menjodohkan, tes melengkapi, tes isian serta tes pilihan ganda.¹⁵ Sedangkan Tes subjektif seperti tes uraian dan tes essay. Tes pilihan ganda (*multiple choice*) adalah satu bentuk teks yang mempunyai jawaban yang benar atau paling tepat.¹⁶ Pada tes ini, kebenaran jawaban bersifat mutlak. Jawaban hanya mempunyai dua kemungkinan yaitu benar dan salah.¹⁷

Menurut Nurgiyantoro yang dikutip oleh Rohmah Sholehah, Tes bahasa dalam kaitannya atau sebagai bagian pembelajaran bahasa, baik bahasa pertama, bahasa kedua, maupun bahasa asing, merupakan alat yang dipakai untuk mencoba mengukur seberapa banyak siswa telah menguasai

⁹ Nana Sudjana, *penilaian hasil proses belajar mengajar*, (Bandung: PT. Remaja Rosdakarya, 1989), hlm. 3

¹⁰ Rohmah Sholehah, "Analisis Butir Soal Tes UTS (Ulangan Tengah Semester) Mata Pelajaran Bahasa Arab Siswa Kelas VIII MTs Negeri I Semarang Tahun Pelajaran 2013/2014", *Skripsi* Jurusan Pendidikan Bahasa Arab Universitas Negeri Semarang, 2015.

¹¹ Nana Sudjana, *penilaian hasil proses...*, hlm. 4

¹² Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 2008), hlm. 39

¹³ Adea Wulan H. Z., Risa Aristia, "Jenis-Jenis Instrumen dalam Evaluasi Pembelajaran", *Artikel* Universitas Muhammadiyah Sidoarjo, 2018.

¹⁴ Ismet Basuki Dan Hariyanto, *Asesmen Pembelajaran*, (Bandung: PT Remaja Rosdakarya, 2014), hlm. 22

¹⁵ Anas Sudijono, *Pengantar Evaluasi...*, hlm. 107

¹⁶ Nana Sudjana, *penilaian hasil proses ...*, hlm. 267.

¹⁷ Purwanto, *Evaluasi Hasil...*, hlm. 188.

bahasa yang dipelajari. Tes bahasa, khususnya tes bahasa arab merupakan alat untuk mengukur kemampuan dan performasi berbahasa arab siswa.¹⁸

Kegiatan menganalisis butir soal (*item analysis*) adalah kegiatan yang harus dilakukan guru dalam rangka meningkatkan mutu soal yang telah disusun. Dan bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Dengan analisis soal dapat diperoleh informasi tentang kejelekan sebuah soal dan “petunjuk” untuk mengadakan perbaikan.¹⁹

Dalam hal ini penulis akan membahas tentang kajian isi tes, yaitu menganalisis tingkat validitas, reliabilitas, tingkat kesukaran, daya beda serta fungsi pengecohnya. Menurut nurgiyantoro yang dikutip oleh Rohmah Sholehah, Sesuai dengan namanya, tes buatan guru adalah tes yang dibuat oleh guru itu sendiri. Tes tersebut dimaksudkan untuk mengukur tingkat keberhasilan siswa mencapai tujuan setelah berlangsungnya proses pengajaran yang dikelola oleh guru yang bersangkutan.²⁰

Pembahasan

1. Analisis Butir Soal

a. Pengertian analisis butir soal

Menurut Arikunto, “analisis soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun”. Nana Sudjana menambahkan bahwa, Analisis butir soal atau analisis item adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai.²¹ Analisis soal juga dilakukan untuk mengetahui berfungsi tidaknya sebuah soal. Analisis pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif (*qualitatif control*) dan analisis kuantitatif (*quantitatif control*).²² Analisis kualitatif sering pula dinamakan sebagai validitas logis (*logical validity*) yang dilakukan sebelum soal digunakan untuk melihat berfungsi tidaknya sebuah soal. Analisis soal secara kuantitatif sering pula dinamakan sebagai validitas empiris (*empirical validity*) yang dilakukan untuk melihat lebih berfungsi tidaknya sebuah soal, setelah itu diujicobakan kepada sampel yang representatif.²³

Dari beberapa pendapat para ahli di atas, penulis menyimpulkan bahwa analisis butir soal adalah sebuah proses yang harus ditempuh dan dilaksanakan oleh seorang guru untuk mengetahui sejauh mana kualitas butir soal yang diberikan dan dengan adanya kegiatan analisis ini bisa membantu para guru terkait butir soal yang bagus dan yang layak untuk dipertahankan dan butir soal yang harus dibuang.

b. Fungsi analisis butir soal

Arikunto (2006:205) dalam jurnal Analisis Butir Soal Pilihan Ganda Dari Aspek Kebahasaan mengatakan bahwa fungsi analisis soal yaitu: membantu kita dalam mengidentifikasi butir-butir soal yang jelek, memperoleh informasi yang akan dapat

¹⁸ Rohmah Sholehah, “Analisis Butir Soal Tes UTS (Ulangan Tengah Semester) Mata Pelajaran Bahasa Arab Siswa Kelas VIII MTs Negeri I Semarang Tahun Pelajaran 2013/2014”, *Skripsi* Jurusan Pendidikan Bahasa Arab Universitas Negeri Semarang, 2015.

¹⁹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi...*, hlm. 232

²⁰ Rohmah Sholehah, “Analisis Butir Soal Tes UTS (Ulangan Tengah Semester) Mata Pelajaran Bahasa Arab Siswa Kelas VIII MTs Negeri I Semarang Tahun Pelajaran 2013/2014”, *Skripsi* Jurusan Pendidikan Bahasa Arab Universitas Negeri Semarang, 2015.

²¹ Nana sudjana, hlm. 135

²² Surapranata, Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes Implementasi Kurikulum 2004, (Bandung: PT. Remaja Rosdakarya, 2005), hlm. 1.

²³ Ibid., hlm. 1

digunakan untuk menyempurnakan soal-soal untuk kepentingan lanjut, memperoleh gambaran secara selintas tentang keadaan yang kita susun.²⁴

c. Tujuan analisis butir soal

Tujuan analisis butir soal menurut Thorndik dan Hagn (1977) sebagaimana dikutip purwanto (2001:118) dalam jurnal Analisis Butir Soal Pilihan Ganda dari Aspek Kebahasaan adalah sebagai berikut:

- 1) Jawaban-jawaban soal itu merupakan informasi diagnostik untuk meneliti pelajaran dari kelas itu dan kegagalan-kegagalan belajarnya, serta selanjutnya untuk membimbing belajar yang lebih baik.
- 2) Jawaban-jawaban terhadap soal yang terpisah dan perbaikan soal-soal didasarkan atas jawaban-jawaban itu merupakan basis penyiapan tes-tes yang lebih baik untuk tahun berikutnya.²⁵

d. Karakteristik Butir Soal

Dalam menilai suatu butir soal atau melihat kualitas butir soal apakah sudah baik atau belum, maka ada hal-hal yang harus dipenuhi oleh butir soal agar butir soal tersebut berkualitas. Dan setiap butir akan diperiksa mutunya dalam empat kategori sebagai berikut:

1) Validitas

Validitas adalah sejauh mana tes itu mengukur apa yang dimaksudkan untuk diukur. Validitas alat ukur tidak terganggu.²⁶

a) Validitas tampilan

Validitas tampilan lebih berhubungan dengan bagaimana tanggapan dari pihak awam mengenai alat ukur tersebut. Bila kita mengacu ke konsep validitas yang telah dibahas sebelumnya, validitas tipe ini kurang tepat dikategorikan ke dalam tipe validitas karena makna dari validitas tampilan tidak terkait dengan kemampuan mengukur dari suatu alat ukur. Ada kemungkinan validitas tipe ini tidak terlalu ilmiah dan hanya berdasarkan kebiasaan yang ada, misalnya format penyusunan pilihan-pilihan dalam soal pilihan ganda.

Seiring dengan perkembangan teori belajar bahasa asing, validitas tampilan mengalami perkembangan dari waktu ke waktu. Misalnya, pada awalnya orang awam beranggapan bahwa penggunaan bahasa itu secara lisan sehingga alat ukurnya pun selalu dilaksanakan secara lisan. Waktu itu sementara orang mungkin beranggapan bahwa kemampuan berkomunikasi dalam bahasa Arab yang diukur secara tertulis dianggap tidak valid. Namun, akhir-akhir ini tes untuk mengukur kemampuan bahasa Arab dalam bentuk tulisan sudah mulai populer.²⁷

b) Validitas Isi

Validitas isi digunakan untuk mengukur sejauh mana tes mencerminkan apa yang akan diukur dari kemampuan siswa sehubungan dengan tujuan pembelajaran yang ingin dicapai. Validitas isi mengukur lebih pada ranah kognitif siswa seperti yang tercantum dalam kurikulum. Penilaian harus dicocokkan dengan Tujuan Instruksional Khusus (ITK). Karena itu isi tes hendaknya sesuai dengan pokok-pokok bahasan yang

²⁴ Nurjannah dan Noni Marliansih, Analisis Butir Soal Pilihan Ganda Dari Aspek Kebahasaan, faktor jurnal ilmu kependidikan, vol II No. 1 maret 2015, hlm. 72

²⁵ Ibid., hlm. 72

²⁶ Helli Ihsan Validitas Isi Alat Ukur Penelitian Konsep dan Panduan Penilaiannya, Pedagogia: Jurnal Ilmu Pendidikan, hlm. 266

²⁷ Izuddin Musthafa dan Acep Hermawan, Metodologi Penelitian Bahasa Arab (Konsep dasar Strategi, Metode, Teknik), PT. Remaja Rosdakarya, Bandung: hlm 162

diberikan kepada siswa. Cara penilaian validitas isi dapat dilakukan oleh beberapa judges kalau bias 3 atau 5 orang, usahakan jangan meminta judges dalam jumlah genap agar tidak sulit menarik kesimpulan jika ada dua pendapat yang berlainan bias ada penengah.²⁸

c) Validitas prediktif

Validitas ini menggambarkan sejauh mana hasil tes dari suatu alat ukur mempunyai korelasi dengan suatu keberhasilan di masa mendatang. Dengan kata lain, suatu alat ukur yang mempunyai validitas prediktif dapat digunakan untuk memprediksi apakah seseorang akan lebih berhasil atau kurang berhasil dalam belajar sesuatu.

Dalam pembelajaran bahasa arab suatu keberhasilan seseorang dapat diprediksi oleh suatu alat ukur walaupun belum terjadi proses belajar. Sebagai contoh, keberhasilan belajar bahasa arab mahasiswa dapat diprediksi dengan penggunaan strategi belajar.

Kemampuan alat ukur untuk memprediksi dapat dibuktikan setelah terjadi suatu proses belajar. Waktu yang digunakan untuk membuktikan kemampuan memprediksi bisa dalam waktu yang singkat dan bisa terbukti dalam waktu yang lama hingga 5 tahun. Misalnya, masuk perguruan tinggi islam di luar negeri atau beberapa perguruan tinggi islam di indonesia mesyaratkan kemampuan berbahasa arab dengan nilai tertentu yang diperoleh dari TOAFL (*Test Of Arabic as a Foreign Language/ ikhtibar al-lughah al-'arabiyyah ka al-lughah al-ajnabiyyah*). Asumsi yang mendasari kebijakan ini adalah bahwa TOAFL bukan hanya mengukur kemampuan berbahasa arab saja namun alat ini juga dikembangkan untuk mengukur kemampuan berpikir, yang sangat diperlukan untuk belajar di perguruan tinggi.

Bila ternyata terdapat korelasi yang baik antara nilai dari tes bakat kebahasaan yang diukur sebelum terjadi poses belajar dengan nilai pengetahuan dan keterampilan dari bahasa arab yang dipelajari setelah dilakukan tes kebahasaan, alat ukur bakat kebahasaan tersebut mempunyai validitas prediktif yang lazim digunakan untuk alat analisis korelasi dan regresi.

Jenis validitas ini ditetapkan dengan cara mengkorelasikan skor tes yang sedang disusun dengan kriteria yang menyangkut hasil karya atau prestasinya di masa yang akan datang.²⁹

d) Validitas konstruk

Validitas konstruk atau *construct validity* diperlukan alat ukur yang mempunyai beberapa indikator dalam mengukur konstruk. Jika ada alat ukur yang mempunyai beberapa aspek dan setiap aspek diukur dengan beberapa indikator, indikator yang sejenis harus berasosiasi positif satu dengan lainnya. Sebaliknya, indikator-indikator tersebut harus berasosiasi negatif dengan indikator lainnya jika indikator tersebut mengukur aspek yang berbeda atau berlawanan.

Penggunaan analisis faktor sering juga berfungsi sebagai cara untuk mendapatkan aspek validitas konstruk. Walaupun validitas konstruk sering dikaitkan dengan alat ukur yang terkait dengan psikolinguistik, misalnya persepsi, sikap kebahasaan, motivasi dalam belajar bahasa asing, validitas konstruk sering pula digunakan untuk alat tes pengetahuan dan keterampilan kebahasaan. Jika alat ukur tersebut hanya mengukur satu

²⁸ Komarudin dan Sarkadi, *Evaluasi Pembelajaran*, Rizqita publishing & printing, Yogyakarta: hlm. 129

²⁹ Amat Jaedun, *Validitas dan Penetapan Validitas Instrumen*, pascasarjana UNY

aspek saja, misalnya kosakata, validitas konstruksinya dapat diukur dengan mengevaluasi dengan semua butir soalnya. Jika semua butir soalnya mengukur penguasaan kosakata, alat ukur tersebut sudah memenuhi aspek validitas konstruk

Contoh, seorang peneliti mengembangkan suatu alat untuk mengukur gaya belajar bahasa arab berdasarkan teori bahwa belajar bahasa dapat dikelompokkan ke dalam gaya mandiri dan kerja kelompok. Bila alat ukur tersebut mempunyai 10 indikator untuk gaya belajar mandiri, dan 10 indikator untuk kerja kelompok, sepuluh indikator yang pertama harus saling berasosiasi atau berkorelasi positif, dan sepuluh indikator kedua juga harus saling berkorelasi positif.

Karena kedua kelompok tersebut mengukur aspek yang berlawanan, kedua kelompok indikator tersebut harus berkorelasi negatif. Jika ada indikator yang mengukur aspek yang sama namun berkorelasi negatif, indikator tersebut harus dibuang dari kuesioner yang dipakai. Jika dianalisis dengan analisis faktor, indikator-indikator tersebut akan dikelompokkan dengan sendirinya secara statistik sehingga masing-masing kelompok akan berkorelasi tinggi.³⁰

2) Reliabilitas

Reliabilitas berasal dari kata *reliability* berarti sejauh mana hasil suatu pengukuran dapat dipercaya. Suatu hasil pengukuran dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama, diperoleh hasil pengukuran yang relatif sama, selama aspek yang diukur dalam diri subyek memang belum berubah. Nur (1987: 47) menyatakan dalam jurnal Validitas dan Reliabilitas Suatu Instrumen Penelitian bahwa reliabilitas ukuran menyangkut seberapa jauh skor deviasi individu, atau skor-z, relatif konsisten apabila dilakukan pengulangan pengadministrasian dengan tes yang sama atau tes yang ekuivalen. Azwar (2003 : 176) menyatakan dalam jurnal Validitas dan Reliabilitas Suatu Instrumen Penelitian bahwa reliabilitas merupakan salah-satu ciri atau karakter utama instrumen pengukuran yang baik. Arifin (1991: 122) menyatakan dalam jurnal Validitas dan Reliabilitas Suatu Instrumen Penelitian bahwa suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda.³¹

Memperhatikan pendapat di atas, maka reliabilitas merupakan koefisien yang menunjukkan sejauh mana suatu instrumen/ alat pengukur dapat dipercaya, artinya apabila suatu instrumen digunakan berulang-ulang untuk mengukur sesuatu yang sama, maka hasilnya relatif stabil atau konsisten. Secara empiris tinggi rendahnya reliabilitas ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas, besarnya koefisien reliabilitas berkisar antara 0 sampai dengan 1, dimana semakin tinggi angka reliabilitas berarti semakin konsisten hasil pengukuran, akan tetapi secara empiris koefisien reliabilitas yang mencapai angka 1 jarang dijumpai.

Menurut Allen dan Yen dalam jurnal Reliabilitas Instrumen Pendidikan bahwa koefisien reliabilitas besarnya ditentukan oleh satu dikurangi dengan perbandingan varians kesalahan pengukuran dan varians skor tampak. Atas dasar demikian semakin kecil varians kesalahan pengukuran, maka akan semakin tinggi koefisien reliabilitas, sehingga varians tampak atas dasar demikian semakin kecil varians kesalahan pengukuran, maka akan semakin tinggi koefisien reliabilitas, sehingga varians tampak Atas dasar demikian semakin

³⁰ Izuddin Musthafa dan Acep Hermawan, Metodologi Penelitian Bahasa Arab (Konsep dasar Strategi, Metode, Teknik), PT. Remaja Rosdakarya Bandung: hlm. 165

³¹ Zulkifli Matondang, Validitas Dan Reliabilitas Suatu Instrumen Penelitian, Jurnal Tabularasa PPS UNIMED Vol. 6 No. 1, Juni 2009

kecil varians kesalahan pengukuran, maka akan semakin tinggi koefisien reliabilitas, sehingga varians tampak dapat digunakan sebagai atribut koefisien reliabilitas yang sesungguhnya.³²

Rumus yang dapat digunakan untuk mencari reliabilitas menurut Arikunto (2011:87) adalah:

Spearman-Brown:

Rumus ini digunakan dalam menghitung besarnya reliabilitas berhubung dengan penambahan banyaknya butir soal.

Keterangan:

$r_{1/2/2}$ = korelasi antara skor-skor setiap belahan tes.

r_{11} = koefisien reliabilitas yang sudah disesuaikan

Rumus Flanagan

Rumus yang penghitungannya menggunakan belah dua ganjil – genap.

$$r_{11} = 2 \left(1 - \frac{s_{12}}{s_{12} + s_{22} + s_{t2}} \right)$$

Keterangan:

r_{11} = reliabilitas tes

s_{12} = varians belahan pertama (1) yang dalam hal ini varians skor item ganjil

s_{22} = varians belahan kedua (2) yaitu varians skor item genap

s_{t2} = varians total yaitu varians skor total

Rumus Alpha:

Rumus ini digunakan untuk mencari soal uraian.

$$r_{11} = \frac{n}{n-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

Keterangan:

r_{11} = reliabilitas yang dicari

$\sum \sigma_i^2$ = jumlah varians skor tiap-tiap item

σ_t^2 = varians total

Penelitian reliabilitas ini akan menggunakan rumus K – R. 20 untuk soal bentuk pilihan ganda, dan rumus Alpha untuk soal bentuk uraian.

Untuk interpretasi reliabilitas instrumen adalah sebagai berikut:

0,80 – 1,00 = tinggi

0,60 – 0,80 = cukup

0,40 – 0,60 = agak rendah

0,20 – 0,40 = rendah

0,00 – 0,20 = sangat rendah (Iskandarwassid, Sunendar 2009:188)³³

³² Muhammad Khumaedi, Reliabilitas Instrumen Pendidikan, Jurnal pendidikan tehnik mesin vol. 12, No. 1 Juni 2012, hlm. 26.

Kesimpulan

Analisis butir soal adalah sebuah proses yang harus ditempuh dan dilaksanakan oleh seorang guru untuk mengetahui sejauh mana kualitas butir soal yang diberikan dan dengan adanya kegiatan analisis ini bisa membantu para guru terkait butir soal yang bagus dan yang layak untuk dipertahankan dan butir soal yang harus dibuang.

Reliabilitas mempermasalahkan sejauh mana hasil suatu pengukuran dapat dipercaya. Suatu hasil pengukuran hanya dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama, diperoleh hasil pengukuran yang relatif sama. Penentuan koefisien reliabilitas instrumen untuk skor butir dikotomi digunakan rumus KR-20, sedang untuk skor politomi digunakan rumus Alpha. Interpretasi terhadap koefisien reliabilitas merupakan intrepretasi relatif, artinya tidak ada batasan mutlak yang menunjukkan berapa angka koefisien minimal yang harus dicapai agar suatu pengukuran dapat disebut reliabel. Namun, memberikan informasi tentang hubungan varians skor teramati dengan varians skor sejati kelompok individu.

Apabila ingin melakukan suatu penelitian misalnya analisis terhadap soal baik ulangan semester ataupun ulangan tengah semester maka sebaiknya harus melakukan analisis pada kriteria-kriteria yang telah ditetapkan seperti harus memenuhi derajat validitas dan reliabilitas.

Referensi

- Ananda, Rusydi dkk. *Evaluasi Pembelajaran*. (Bandung: Citapustaka Media, 2014).
- Arikunto, Suharsimi. *Dasar-Dasar Evaluasi Pendidikan*. (Jakarta: Bumi Aksara, 2008).
- Basuki, Ismet Dan Hariyanto. *Asesmen Pembelajaran*. (Bandung: PT Remaja Rosdakarya, 2014).
- Komarudin dan Sarkadi, *Evaluasi Pembelajaran*. (Yogyakarta: Rizqita publishing & printing, 2017)
- Musthafa, Izuddin dan Acep Hermawan. *Metodologi Penelitian Bahasa Arab (Konsep dasar Strategi, Metode, Teknik)*. Bandung: PT. Remaja Rosdaarya
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*. (Jakarta: Rajawali Pers, 2016).
- Sudjana, Nana. *penilaian hasil proses belajar mengajar*. (Bandung" PT. Remaja Rosdakarya, 1989).
- Surapranata. *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*. (Bandung: PT. Remaja Rosdakarya, 2005).
- Syar'i, Ahmad. *Filsafat pendidikan Islam*. (Jakarta: Pustaka Firdaus: 2005).
- Aristia, Risa Dan Adea Wulan H. Z. "Jenis-Jenis Instrumen dalam Evaluasi Pembelajaran". *Artikel Universitas Muhammadiyah Sidoarjo*. 2018.

³³ Wawan Gunawan, Analisis Butir Soal Latihan Buku "ta'lim Al Lughah Al Arabiyah" karangan dr. d. hidayat kelas x madrasah aliyah negeri 01 kota magelang , *skripsi* diajukan dalam rangka penyelesaian studi strata i untuk memperoleh gelar sarjana pendidikan, hlm. 26-27

- Fahreza, Febry dan Rabiatul Rahmi. "Peningkatan Keterampilan Sosial Melalui Metode Role Playing Pada Pembelajaran IPS Di Kelas IV SD Negeri Pasi Pinang Kabupaten Aceh Barat". *Bina Gogik Volume 5 No 1*, Maret 2018.
- Fitria, Happy, Muhammad Kristiawan, dkk, "Upaya Peningkatan Kompetensi Guru Melalui Pelatihan Penelitian Tindakan Kelas". *Jurnal Abdimas Unwahas*, Vol. 4, No. 1, April 2019.
- Gunawan , Wawan. Analisis Butir Soal Latihan Buku "ta'lim Al Lughah Al Arabiyah" karangan dr. d. hidayat kelas x madrasah aliyah negeri 01 kota magelang. *skripsi* diajukan dalam rangka penyelesaian studi strata i untuk memperoleh gelar sarjana Pendidikan.
- Ihsan, Helli. Validitas Isi Alat Ukur Penelitian Konsep dan Panduan Penilaiannya, *Pedagogia:Jurnal Ilmu Pendidikan*, hlm. 266
- Khumaedi Muhammad. Reliabilitas Instrumen Pendidikan, *Jurnal pendidikan tekhnik mesin* vol. 12, No. 1 Juni 2012.
- Lihat, di Mushaf Al-Qur'an Terjemah, Al-Qur'an Terjemah Edisi tahun 2002, (Jakarta: Al-Huda (kelompok gema insani), 1426 H.
- Nurjannah dan Noni Marlianingsih, Analisis Butir Soal Pilihan Ganda Dari Aspek Kebahasaan, faktor. *jurnal ilmu kependidikan*. vol II No. 1 maret 2015.
- Matondang , Zulkifli. Validitas Dan Reliabilitas Suatu Instrumen Penelitian. *Jurnal Tabularasa PPS UNIMED* Vol. 6 No. 1, Juni 2009.
- Ulfah, Rahmatia Lessa. "Analisis Butir Soal Tes Pilihan Ganda Penilaian Tengah Semester Genap Akidah Akhlak Di MTs Al-Hidayah Rawadenok 2017/2018". *Skripsi* Jurusan Pendidikan Agama Islam Universitas Muhammadiyah Jakarta 1439 H/ 2018 M.
- Sholeha, Rohmah. "Analisis Butir Soal Tes UTS (Ulangan Tengah Semester) Mata Pelajaran Bahasa Arab Siswa Kelas VIII MTs Negeri I Semarang Tahun Pelajaran 2013/2014". *Skripsi* Jurusan Pendidikan Bahasa Arab Universitas Negeri Semarang, 2015.
- Wardani, Winda Kusuma. "Analisis Kompetensi Pedagogik Guru Dalam Proses Pembelajaran di SD Muhammadiyah 15 Surakarta". *Skripsi* Program Studi Pendidikan Guru Sekolah Dasar Universitas Muhammadiyah Surakarta, 2017