
Validitas Isi dan Reliabilitas Instrumen Penilaian Harian Mata Pelajaran Bahasa Arab MI Manarul Islam Malang

Alfi Maghfirotnunisa¹, Nur Qomari², Asbarin³

^{1,2,3}Pascasarjana Pendidikan Bahasa Arab Universitas Islam Negeri Maulana Malik Ibrahim Malang

Corresponding E-mail: alfaalfi546@gmail.com

Abstrak: Tujuan dari penelitian ini adalah : 1) untuk mengetahui validitas isi pada mata pelajaran Bahasa Arab pada Penilaian Harian (PH) di MI Manarul Islam Malang. 2) untuk mengetahui reabilitas pada mata pelajaran Bahasa Arab pada Penilaian Harian (PH) pada bentuk soal multiple choice dan isian di MI Manarul Islam Malang. Penelitian ini merupakan jenis penelitian kuantitatif dengan teknik pengumpulan data melalui tes dari sampel penelitian sejumlah 36 siswa-siswi kelas 4A dan 4B tahun ajaran 2022/2023 serta analisis pengujian butir-butir instrumen tes yang terdiri dari uji validitas dan uji reliabilitas dievaluasi menggunakan SPSS versi 22. Pada penilaian harian mata pelajaran bahasa asing yaitu Bahasa Arab, lembar jawaban siswa digunakan untuk menilai validitas butir soal dan reliabilitas tes instrumen tes (Arab). Hasil dari penelitian ini adalah dimana butir soal pada penilaian harian mata pelajaran bahasa Arab MI Manarul Islam Malang pada uji validitas soal berbentuk multiple choice yang terdiri dari 20 soal, 12 diantaranya tergolong valid sementara 8 soal lainnya tergolong tidak valid. Sedangkan soal berbentuk isian yang terdiri dari 10 soal tergolong valid secara keseluruhan. Pada uji realibilitas soal berbentuk multiple choice dinyatakan memiliki reabilitas yang tinggi. Sementara hasil uji reabilitas soal berbentuk isian dinyatakan memiliki reabilitas yang sangat tinggi.

Kata Kunci: Validitas; reliabilitas; penilaian harian

***Abstract:** The aims of this research are: 1) to find out the content validity of the Arabic subject in the Daily Assessment (PH) at MI Manarul Islam Malang. 2) to determine the reliability of the Arabic language subject in the Daily Assessment (PH) in the form of multiple choice questions and entries at MI Manarul Islam Malang. This research is a type of quantitative research with data collection techniques through tests from a research sample of 36 students in grades 4A and 4B for the 2022/2023 academic year as well as analysis of testing the items of the test instrument which consists of a validity test and reliability test evaluated using SPSS version 22. In the daily assessment of a foreign language subject, namely Arabic, student answer sheets are used to assess the validity of the item items and the reliability of the test instrument test (Arabic). The results of this study are where the items in the daily assessment of the MI Manarul Islam Malang Arabic subject in the multiple choice question validity test consist of 20 questions, 12 of which are classified as valid while the other 8 items are classified as invalid. While the questions in the form of fields consisting of 10 questions are classified as valid as a whole. In the reliability test, the multiplechoice questions were declared to have high reliability. While the results of the reliability test in the form of filled in items were stated to have very high reliability.*

Keywords: validity; reliability; daily assessment

PENDAHULUAN

Pendidikan adalah upaya sadar dan dipikirkan dengan matang untuk membantu orang mencapai potensi penuh mereka. Dalam Kamus Besar Bahasa Indonesia (KBBI), pendidikan merupakan tindakan merubah sikap dan perilaku individu atau sekelompok orang untuk membantu individu menjadi dewasa melalui pengajaran dan pelatihan. Pendidikan dijelaskan dalam Pasal 1(1) Undang-Undang Sistem Pendidikan Nasional Republik Indonesia Nomor 2 Tahun 1989 sebagai upaya sadar untuk mempersiapkan

peserta didik untuk peran mereka di masa depan melalui kegiatan pengawasan, pengajaran dan/atau pelatihan. Dalam rangka mencerdaskan kehidupan bangsa, pendidikan nasional bertendensi sebagai perkembangan keterampilan serta membentuk peradaban berwatak adil dan bermartabat. Tujuan pendidikan nasional adalah membantu peserta didik mencapai potensi dirinya sebagai manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat jasmani dan rohani, memperoleh ilmu pengetahuan dan keterampilan, berakhlak mulia, mandiri, dan menjadi warga negara yang demokratis¹.

Guru dan siswa merupakan dua elemen penting dalam dunia pendidikan. Siswa yang menjadi objek dalam proses pembelajaran dan guru yang mengarahkan pembelajaran siswanya Guza mengklaim dalam majalah *Abdimas Unwahas* bahwa empat keterampilan utama yang membentuk standar kompetensi guru adalah kompetensi pedagogik, kompetensi kepribadian, kompetensi sosial dan kompetensi profesional². Kemampuan pelatih untuk mengarahkan pembelajaran siswa setidaknya harus mencakup: pemahaman tentang prinsip atau wawasan pedagogis; tahu siswa; pengembangan kurikulum atau kurikulum; rencana pembelajaran; pelaksanaan pembelajaran pedagogis dan dialogis; penggunaan teknologi pembelajaran; Evaluasi Hasil Belajar (EHB); dan membantu siswa untuk mencapai berbagai potensi mereka hanyalah beberapa contoh. Salah satu langkah krusial dalam menyelenggarakan pendidikan madrasah adalah penilaian hasil belajar siswa³. Proses pengumpulan dan analisis data atau informasi untuk tujuan membandingkan hasil belajar siswa dengan persyaratan kompetensi lulusan yang telah ditetapkan disebut penilaian. Berikut adalah contoh kegiatan yang digunakan untuk menilai hasil belajar madrasah: 1) Daily Assessment (PH), tes yang digunakan untuk mengukur tingkat kompetensi siswa setelah menyelesaikan satu atau lebih kompetensi inti; 2) Penilaian Akhir Semester (PAS) Tes yang digunakan untuk mengetahui seberapa baik kompetensi setiap mahasiswa telah dikembangkan pada akhir semester. 3) Penilaian Akhir Tahun (PAT), tes untuk menilai tingkat kompetensi yang dicapai mahasiswa pada akhir semester; dan 4) Ujian Akhir Sekolah Agama Islam (UAMBN), yaitu tes yang digunakan untuk mengukur kemampuan siswa di akhir suatu kelas.

Penilaian sangat penting dalam pendidikan sebagai suatu bukti perkembangan strategi yang mendukung pengembangan kualitas individu yang kompetitif di era global saat ini. Untuk alasan ini, evaluasi dan penilaian adalah bagian penting dari satu proses dalam pengaturan pendidikan apa pun. Evaluasi dilakukan setidaknya untuk dua alasan: pertama, untuk menentukan/menunjukkan tingkat pemahaman evaluator setelah proses, dan kedua, untuk menggunakan keputusan dari evaluasi yang dijadikan tolok ukur untuk meningkatkan sebuah sistem atau sebuah proses. Ini juga sesuai perspektif House⁴ bahwa adanya tujuan evaluasi bukan untuk membuktikan hasil saja, tetapi juga untuk mengonversi kekurangan yang ada. Dalam bukunya, House menambahkan jika pengambil keputusan berupa penerimaan informasi dari penilaian untuk membantu mereka membuat

¹ UU RI no 20 tahun 2003 tentang sisdiknas & Peraturan Pemerintah RI no 47 2008 tentang Wajib Belajar

² 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

³ 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

⁴ House, E.R. (2012). Introduction. In Kellaghan, T., & Stufflebeam, D. L. (Eds.). *International Handbook of Educational Evaluation: Part One: Perspectives/Part Two: Practice* (Vol. 9). Springer Science & Business Media.

keputusan. Kemampuan siswa dan efektivitas metode pengajaran yang digunakan oleh guru yang bersangkutan dapat dipelajari oleh guru dengan menawarkan penilaian. Teknik lain untuk menilai kemampuan seseorang setelah belajar adalah asesmen. Guru juga dapat menilai keberhasilan kurikulum dan teknik yang digunakan berdasarkan hasil evaluasi.

Sejauh mana suatu tes mengukur apa yang hendak diukur disebut validitas, sifat pertama dari tes yang baik⁵. Misalnya, tes yang hanya menilai kemampuan matematika dianggap sebagai tes matematika yang valid. Hal senada diungkapkan oleh Azwar⁶ yang menyatakan bahwa validitas berasal dari kata validitas, yang mengacu pada kebenaran suatu tes atau skala dalam melakukan suatu fungsi pengukuran. Validitas isi, validitas konstruk, dan validitas terkait kriteria adalah tiga kategori di mana validitas secara tradisional dibagi⁷. Validitas mengacu pada masalah apakah tes yang dimaksudkan dapat secara akurat mengukur apa yang dimaksudkan untuk diukur. validitas menurut Taherdoost⁸ dapat didefinisikan sebagai seberapa jauh instrumen tersebut mengukur apa yang akan dijadikan ukuran. Menetapkan validitas instrumen dapat dilihat seperti pembangunan sebuah argumen yang dikuatkan dengan bukti. Sebaik apakah instrumen mengukur apa yang ingin diukur⁹. Adapun reliabilitas item terkait dengan masalah kepercayaan. Keandalan menggambarkan bahwa tes tersebut secara konsisten mengukur sesuatu yang reliabel atau dapat dipercaya¹⁰. Hubungan antara validitas dan reliabilitas dapat dijelaskan dengan mengatakan bahwa validitas itu penting sedangkan reliabilitas diperlukan karena reliabilitas mendukung validitas¹¹.

Dalam penelitian ini, peneliti ingin menguji validitas dan reliabilitas butir soal yang dipakai untuk mengukur kemampuan siswa pada penilaian harian di MI Manarul Islam Malang.

METODE

Penelitian ini termasuk pada jenis penelitian kuantitatif karena subjeknya merupakan jenis subjek kuantitatif. Menurut Margono¹², penelitian kuantitatif adalah metode pengumpulan pengetahuan yang menggunakan data numerik untuk mempelajari lebih lanjut tentang topik yang menarik bagi peneliti. Adapun populasi pada penelitian ini merupakan siswa-siswi kelas 4 Madrasah Ibtidaiyah Manarul Islam Malang tahun ajaran 2022/2023 sedangkan sampel yang digunakan pada penelitian ini merupakan siswa-siswi kelas 4A dan 4B tahun ajaran 2022/2023 dengan jumlah 36 anak di MI Manarul Islam Malang. Pengumpulan data pada penelitian ini menggunakan tes yang mana merupakan hasil dari Penilaian Harian (PH). Setelah semua data terkumpul dari responden atau sumber data lainnya, analisis data kuantitatif merupakan langkah yang perlu dilakukan. Untuk mengatasi perumusan masalah, tindakan-tindakan ini meliputi pengelompokan

⁵ Brown, H. Douglas. 2004. *Language Assessment. Principles and Classroom Practices*. San Fransisco : Pearson Education, Inc.

⁶ Azwar, Saifuddin. 2012. *Reliabilitas dan Validitas*. Yogyakarta : Pustaka Pelajar

⁷ Azwar, Saifuddin. 2012. *Reliabilitas dan Validitas*. Yogyakarta : Pustaka Pelajar

⁸ Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research

⁹ Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*. 3:119-120

¹⁰ Bajpai, S., & Bajpai, R. (2014). Goodness of measurement: reliability and validity. *International Journal of Medical Science and Public Health*, 3(2), 112-116

¹¹ Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Insan Madani: Yogyakarta.

¹² Margono, S. 1997. *Metodologi Penelitian Pendidikan*. Jakarta: PT Rineka Cipta

data menurut variabel dan jenis responden, tabulasi data, melihat data, dan melakukan perhitungan¹³. Sedangkan pengujian butir-butir instrumen tes yang terdiri dari uji validitas dan uji reliabilitas dievaluasi secara kuantitatif pada penilaian harian mata pelajaran bahasa asing yaitu Bahasa Arab untuk menilai validitas butir soal dan reliabilitas tes instrumen tes (Arab) dengan menggunakan SPSS versi 22.

PEMBAHASAN

A. Validitas

Ketepatan interpretasi yang diambil dari nilai tes atau alat penilaian lainnya disebut sebagai validitas¹⁴. Jika suatu alat evaluasi berhasil mengukur objek yang akan dievaluasi, maka dianggap valid. Dengan kata lain, ketika menggunakan alat ukur, validitas dan “ketepatan” saling berkaitan¹⁵. Cukup sulit mencari kata pengganti valid dalam bahasa Indonesia, sehingga diubah menjadi sah, sedangkan validitas diubah menjadi valid, namun ada terjemahan lain, seperti cocok dan akurat. ' atau 'hati-hati' dan 'benar'. (Kamus Bahasa Indonesia).

B. Macam-macam Validitas

1. Validitas Internal

Validitas logis adalah nama lain dari validitas internal (validitas internal). Kata "logis" dalam frasa "validitas logis" berasal dari kata "logika" yang menunjukkan akal atau rasionalitas. Pengertian validitas logis suatu alat penilaian ini mengacu pada kondisi suatu instrumen yang memenuhi standar berdasarkan logika dan hasil yang beralasan¹⁶. Suatu kriteria disebut sebagai validitas logis dari suatu instrumen evaluasi, berdasarkan mana instrumen tersebut sesuai dengan standar yang valid. Instrumen yang dimaksud telah dirancang dengan baik dengan menggunakan teori dan peraturan yang ada, sehingga dianggap memenuhi persyaratan.

2. Validitas Eksternal

Validitas empiris adalah istilah lain dari validitas eksternal (validitas eksternal). Validitas empiris diartikan sebagai validitas yang diperoleh berdasarkan observasi lapangan. Apabila pemeriksaan observasi lapangan menunjukkan bahwa suatu tes hasil belajar telah berhasil mengukur hasil belajar yang perlu diverifikasi dengan tes hasil belajar, maka dianggap memiliki validitas empiris¹⁷.

C. Realibitas

¹³ Noor, Juliansyah. 2011. Metodologi Penelitian: Skripsi, Tesis, Disertasi, dan Karya Ilmiah. Jakarta: Kencana

¹⁴ Pramono, Sigit. 2014. Panduan Evaluasi Kegiatan Belajar Mengajar. Yogyakarta: Diva Press

¹⁵ Widoyoko, Eko Putra. 2012. Evaluasi Program Pembelajaran (Panduan Praktis bagi Pendidik dan Calon Pendidik). Yogyakarta: Pustaka Pelajar

¹⁶ Arikunto, Suharsimi. 2012. Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara

¹⁷ Sudijono, Anas. 2011. Pengantar Statistik Pendidikan. Jakarta: PT Rajagrafindo Persada

Istilah "reliability" dalam bahasa Inggris berasal dari kata "reliable", yang dalam bahasa Indonesia berarti "dapat dipercaya". Menurut Widoyoko¹⁸, suatu instrumen tes dianggap reliabel jika secara konsisten memberikan hasil yang tetap atau konsisten setelah berbagai pengujian. Secara umum, alat uji yang andal adalah sah, tetapi tidak selalu demikian.

D. Macam-macam Realibitas

1. Uji-Uji Ulang Reliabilitas

Tingkat reliabilitas ini menentukan seberapa konsisten hasil tes dari waktu ke waktu. Test-retest menunjukkan variasi hasil dari melakukan tes penilaian tunggal, apakah itu dilakukan sekali, dua kali, atau lebih. Jelas bahwa hasil dari setiap pengujian yang dilakukan seringkali menghasilkan hasil yang sama.

2. Reliabilitas ekuivalen

artinya dibuat dua instrumen yang hampir identik sebelum diujikan pada responden yang sama. Instrumen ekuivalen adalah dua tes yang mengukur variabel yang sama, memiliki jumlah item, format, dan kesukaran yang sama, tetapi menggunakan pertanyaan atau pernyataan yang berbeda.

3. Keandalan Internal

adalah ukuran konsistensi dalam suatu organisasi. Jenis reliabilitas yang didasarkan pada konsistensi di setiap item tes benchmark disebut konsistensi internal dan memenuhi tujuan ini. Data dari kumpulan data tunggal atau hanya sekali selama

E. Faktor yang Mempengaruhi Realibilitas

1. Panjang tes mengacu pada jumlah total pertanyaan dalam tes. Karena lebih banyak sampel diukur dan ada proporsi jawaban benar yang lebih besar, semakin banyak pertanyaan yang diajukan tes, semakin tinggi keandalannya.
2. Derajat penyebaran skor meningkatkan reliabilitas karena koefisien reliabilitas yang lebih baik dicapai ketika siswa tetap pada posisi relatif yang sama dari satu kelompok tes ke tes berikutnya.
3. Kesulitan Tes (Indeks Kesulitan), Kesulitan Tes Siswa, dan Kemudahan Tes Siswa semuanya menghasilkan reliabilitas yang rendah.

F. Tes

Tes adalah alat ukur, khususnya alat untuk mengumpulkan data tentang sifat-sifat suatu benda¹⁹. Tes merupakan strategi atau pendekatan untuk melakukan kegiatan pengukuran yang menuntut siswa untuk menyelesaikan serangkaian tugas atau menanggapi serangkaian pertanyaan. Tes sering dapat dibagi menjadi dua kategori: tes subjektif atau deskriptif dan tes objektif. Di masa lalu, tes deskriptif datang sebelum tes objektif.

¹⁸ Widoyoko, Eko Putra. 2012. *Evaluasi Program Pembelajaran (Panduan Praktis bagi Pendidik dan Calon Pendidik)*. Yogyakarta: Pustaka Pelajar

¹⁹ Pramono, Sigit. 2014. *Panduan Evaluasi Kegiatan Belajar Mengajar*. Yogyakarta: Diva Press

1. Tes Objektif

Salah satu jenis tes hasil belajar yang dapat dilakukan siswa adalah tes objektif, yang terdiri dari soal-soal dengan beberapa alternatif solusi yang dapat dipilih siswa²⁰.

2. Tes Subjektif

Evaluasi subjektif biasanya disajikan sebagai deskripsi (esai). Tes deskriptif adalah pertanyaan dimana dalam menjawabnya, siswa harus mendeskripsikan, menjelaskan, mendiskusikan, membandingkan, berdebat, dll²¹. Tes deskriptif berbeda dari tes lain karena mengharuskan siswa untuk mempersiapkan jawaban atas pertanyaan daripada meminta pembuat pertanyaan memberikannya. Hanya pertanyaan atau tugas yang termasuk dalam kategori ini, dan siswa harus mempertimbangkan jawaban mereka dengan cermat.

G. Bahasa Arab

Jika dianalogikan dengan bahasa asing yang lain, substansi yang dimiliki bahasa Arab lebih lengkap dan unggul menurut Salim²². Keutuhan dan kesempurnaannya membuatnya unik. Salah satu ciri bahasa Arab adalah kemampuannya untuk mengubah konsep yang kompleks menjadi sederhana, konsep yang tidak jelas menjadi jelas, dan konsep yang jelek menjadi indah, di antara transformasi lainnya. Adapun tujuan mempelajari Bahasa Arab sebagai berikut:

1. Menguasai empat keterampilan berbahasa yaitu Mendengar (Istima), Berbicara (Kalam), Membaca (Qira'ah) dan Menulis untuk berkomunikasi baik secara lisan maupun tulisan (Kitabah) dalam bahasa Arab.
2. Menumbuhkan kesadaran akan nilai bahasa Arab sebagai bahasa asing yang seharusnya menjadi bahasa pengantar utama, khususnya dalam menggali asal-usul ajaran Islam.
3. Memperluas cakrawala budaya dan memahami bagaimana bahasa dan budaya berinteraksi. Oleh karena itu, mahasiswa diharapkan memiliki komitmen terhadap keragaman budaya dan memiliki pemahaman antarbudaya.

HASIL

Dalam proses memperoleh hasil yang diinginkan, dalam hal ini validitas dan reliabilitas, peeneliti menggunakan bantuan SPSS Versi 22. Hasil pengujian pertama ini mengacu pada validitas butir soal yang mana pemberian intrepretasi pada uji validitas item tes bentuk *multiple choice* ditentukan dengan patokan sebagai berikut :

1. Apabila $t\text{-hitung} > t\text{-tabel}$ berarti item yang diuji adalah valid.

²⁰ Sudijono, Anas. 2011. Pengantar Statistik Pendidikan. Jakarta: PT Rajagrafindo Persada

²¹ Ratnawulan, Elis dan Rusdiana, A. 2015. Evaluasi Pembelajaran. Bandung: CV Pustaka Setia

²² Salim, Latifah. 2015. JurnalAdabiyah: Peranan Bahasa Arab Terhadap Ilmu Pengetahuan. Jurnal Vol. 15 Nomor 2

2. Apabila $t_{hitung} < t_{tabel}$ berarti item yang diuji adalah tidak valid.

BUTIR SOAL MULTIPLE CHOICE	HASIL UJI		INTERPRETASI
	t_{hitung}	$t_{tabel} (N=36)$	
1	0,268	0,329	Tidak Valid
2	0,474	0,329	Valid
3	0,056	0,329	Tidak Valid
4	0,0	0,329	Tidak Valid
5	0,0	0,329	Tidak Valid
6	0,395	0,329	Valid
7	0,0	0,329	Tidak Valid
8	0,0	0,329	Tidak Valid
9	0,0	0,329	Tidak Valid
10	-0,049	0,329	Tidak Valid
11	0,576	0,329	Valid
12	0,715	0,329	Valid
13	0,651	0,329	Valid
14	0,659	0,329	Valid
15	0,818	0,329	Valid
16	0,616	0,329	Valid
17	0,475	0,329	Valid
18	0,557	0,329	Valid
19	0,580	0,329	Valid
20	0,484	0,329	Valid

Setelah mendapatkan interpretasi terhadap butir-butir item tes dengan taraf signifikan 0,05 berbentuk *multiple choice*, maka dapat diklasifikasikan sebagai berikut ini:

1. Butir item tes yang tergolong valid sebanyak 12 item atau butir soal secara keseluruhan.
2. Sedangkan butir item tes yang tergolong invalid sebesar 8 item dari 20 soal secara keseluruhan.

BUTIR SOAL ISIAN	HASIL UJI		INTERPRETASI
	t_{hitung}	$t_{tabel} (N=36)$	
1	0,700	0,329	Valid
2	0,583	0,329	Valid
3	0,683	0,329	Valid
4	0,749	0,329	Valid
5	0,678	0,329	Valid
6	0,743	0,329	Valid
7	0,659	0,329	Valid
8	0,798	0,329	Valid
9	0,658	0,329	Valid
10	0,799	0,329	Valid

Setelah mendapatkan interpretasi terhadap butir-butir item tes berbentuk Isian dengan taraf signifikan 0,05, maka dapat diklasifikasikan sebagai berikut ini:

1. Butir item tes yang tergolong valid sebanyak 10 item soal secara keseluruhan.
2. Sedangkan butir item tes yang tergolong invalid tidak terdeteksi pada soal tes dalam bentuk Isian.

Kedua, setelah melakukan ujia validitas kini peneliti menggunakan uji reliabilitas untuk mengetahui apakah butir soal pada PH (Penilaian Harian) terbukti reliabel atau tidak. Pemberian intrepretasi pada Uji Reabilitas item tes ditentukan dengan patokan sebagai berikut:

Koefisien Korelasi	Kriteria Reliabilitas
0,81 < r ≤ 1,00	Sangat Tinggi
0,61 < r ≤ 0,80	Tinggi
0,41 < r ≤ 0,60	Cukup
0,21 < r ≤ 0,40	Rendah
0,00 < r ≤ 0,21	Sangat Rendah

Realibilitas Tes *Multiple Choice*

Reliability Statistics

Cronbach's Alpha	N of Items
.791	20

Maka interpretasi yang dapat disimpulkan dari tabel di atas bahwa item tes *multiple choicw* dapat dinyatakan bahwa tes hasil belajar mata pelajaran Bahasa Asing (Arab) dengan menyajikan 20 butir item tes *multiple choice* adalah reliabel atau memiliki tingkat reliabilitas yang tinggi dibuktikan dengan nilai Cronbach's Alpha sebesar 0,791 sehingga bisa dinyatakan mempunyai reliabilitas yang tinggi.

Realibilitas Tes Isian

Reliability Statistics

Cronbach's Alpha	N of Items
.883	10

Maka interpretasi yang dapat disimpulkan dari tabel di atas bahwa item tes isian dapat dinyatakan bahwa tes hasil belajar mata pelajaran Bahasa Asing (Arab) dengan menyajikan 10 butir item tes isian adalah reliabel atau memiliki tingkat reliabilitas yang sangat tinggi dibuktikan dengan nilai Cronbach's Alpha sebesar 0,883 sehingga bisa dinyatakan mempunyai reliabilitas yang sangat tinggi.

KESIMPULAN

Setelah dilakukan analisis terhadap butir-butir item tes mata pelajaran Bahasa Arab pada Penilaian Harian (PH) sebagaimana yang telah dijelaskan pada di atas, maka secara keseluruhan dapat disimpulkan sebagai berikut:

1. Ditinjau dari segi uji validitas dapat disimpulkan bahwa tes mata pelajaran Bahasa Asing (Arab) pada penilaian harian setelah dilakukan analisis secara

kuanitatif dengan menggunakan format penelaah soal bentuk *multiple choice* dan isian hasilnya adalah sebagai berikut: pada soal *multiple choice* yang berjumlah 20 soal pilihan ganda, terdapat 8 yang tidak valid dan sebanyak 12 soal yang memiliki validitas. Sedangkan pada soal berbentuk isian yang soalnya terdiri dari 10 soal tergolong valid secara keseluruhan.

2. Ditinjau dari segi uji reabilitas dapat disimpulkan bahwa tes mata pelajaran Bahasa Asing (Arab) pada penilaian harian setelah dilakukan analisis secara kuantitatif dengan menggunakan format penelaah soal bentuk *multiple choice* dan isian hasilnya adalah sebagai berikut: pada soal *multiple choice* yang berjumlah 20 soal pilihan ganda memiliki nilai Cronbach's Alpha sebesar 0,791 sehingga bisa dinyatakan mempunyai reliabilitas yang tinggi. Sedangkan pada soal berbentuk isian yang soalnya terdiri dari 10 soal memiliki nilai Cronbach's Alpha sebesar 0,883 sehingga bisa dinyatakan mempunyai reliabilitas yang sangat tinggi.

Adapun saran yang ditujukan kepada guru bahasa Arab MI Manarul Islam Malang selaku penyusun butir soal pada Penilaian Harian (PH), untuk memperbaiki kembali butir soal pada PH dengan lebih memperhatikan pembuatan butir soal dari segi tingkat kesukaran, maupun mengganti soal yang telah ada dengan soal baru, sehingga diperoleh butir soal yang valid dan reliabel secara keseluruhan.

DAFTAR REFERENSI

- Arikunto, Suharsimi. 2012. Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara
- Azwar, Saifuddin. 2012. Reliabilitas dan Validitas. Yogyakarta : Pustaka Pelajar
- Bajpai, S., & Bajpai, R. (2014). Goodness of measurement: reliability and validity. *International Journal of Medical Science and Public Health*, 3(2), 112-116
- Brown, H. Douglas. 2004. *Language Assessment. Principles and Classroom Practices*. San Fransisco: Pearson Education, Inc.
- 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
- House, E.R. (2012). Introduction. In Kellaghan, T., & Stufflebeam, D. L. (Eds.). *International Handbook of Educational Evaluation: Part One: Perspectives/Part Two: Practice (Vol. 9)*. Springer Science & Business Media.
- Margono, S. 1997. *Metodologi Penelitian Pendidikan*. Jakarta: PT Rineka Cipta
- Noor, Juliansyah. 2011. *Metodologi Penelitian: Skripsi, Tesis, Disertasi, dan Karya Ilmiah*. Jakarta: Kencana
- Ratnawulan, Elis dan Rusdiana, A. 2015. *Evaluasi Pembelajaran*. Bandung: CV Pustaka Setia
- Pramono, Sigit. 2014. *Panduan Evaluasi Kegiatan Belajar Mengajar*. Yogyakarta: Diva Press
- Salim, Latifah. 2015. *Jurnal Adabiyah: Peranan Bahasa Arab Terhadap Ilmu Pengetahuan*. Jurnal Vol. 15 Nomor 2
- Sudijono, Anas. 2011. *Pengantar Statistik Pendidikan*. Jakarta: PT Rajagrafindo Persada
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Insan Madani: Yogyakarta.
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*. 3:119-120

- Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research
- UU RI no 20 tahun 2003 tentang sisdiknas & Peraturan Pemerintah RI no 47 2008 tentang Wajib Belajar
- Widoyoko, Eko Putra. 2012. Evaluasi Program Pembelajaran (Panduan Praktis bagi Pendidik dan Calon Pendidik). Yogyakarta: Pustaka Pelajar
- 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.