EDUKASIA ISLAMIKA
Jurnal Pendidikan Islam
Vol. 4 No. 2, Desember 2019, hlm. 143-170
P-ISSN: 2548-723X; E-ISSN: 2548-5822

Estimasi G-Study dalam Pengembangan Instrumen Penilaian Tahsinul Qur'an

Irfa Ma'alina Li'illiyyina, Badrun Kartowagiran
Pascasarjana Universitas Negeri Yogyakarta
allinar55@gmail.com

DOI: https://doi.org/10.28918/jei.v4i2.2296		
Received: July 20, 2019	*Revised: August 19, 2019*	*Approved: November 27, 2019*

Abstrak
Tulisan ini memfokuskan kepada tes unjuk kerja (*performance* test) dalam penilaian ujian sertifikasi *tahsin al-qur'an* menggunakan satu *rater* pada tiap kelompok asistensi. Proses penilaian yang melibatkan satu *rater* pada tesunjuk kerja (*performance* test) berpeluang besar melahirkan ketidak konsistenan *rater* dalam memberikan penilaian pada individu yang berbeda dengan kemampuan yang sama. Tulisan ini mencoba mengurai masalah tersebut dengan pengembangan istrumen untuk dua *rater* atau lebih. Proses penyusunan ini menggunakan analisis kuantitatif dengan pendekatan *generalizability theory* serta menggunakan bantuan program *Edu-G version 6.1*. Penelitian ini menunjukkan bahwa uji instrumen penilaian *tahsinul qur'an* dikatakan valid dengan nilai index aiken > 0.67 pada tiap indikatornya. Uji reliabilitas pada instrumen ini menunjukan nilai koefisiean reliabilitas > 0.80 baik pada ujicoba sampel terbatas dan sampel luas.Berdasarkan hal-hal tersebut, instrumen penilaian *tahsinul qur'an* dapat dikatakan valid dan reliabel untuk mengukur *tahsin al-qur'an* dan dapat meminimalisir subjektivitas *rater* dalam melakukan penilaian.

Kata Kunci: *Instrumen, Tahsin al-Quran, Generalizability Theory, Edu-G*

DOI: https://doi.org/10.28918/jei.v4i2.2296

Abstract

This paper focuses on performance tests (performance tests) in the assessment of the Tahsin al-Qur'an certification exam using one rater in each assistance group. The assessment process that involves a rater on a performance test (performance test) has a great chance of giving birth to inconsistencies in giving rater assessments to different individuals with the same ability. This paper tries to solve this problem by developing instruments for two or more rater. This compilation process uses a quantitative analysis with generalizability theory approach and uses the help of the Edu G version 6.1 program. This study shows that the examination of the tahsinulquran assessment instrument is said to be valid with aiken index value> 0.67 on each indicator. The reliability test on this instrument shows the reliability coefficient value> 0.80 for both limited and large sample trials. Based on these things, the assessment instrument of tahsinulquran can be said to be valid and reliable to measure the tahsin al-quran, and can minimize the subjectivity of the rater in making an assessment.

Keywords: *Instrument, Tahsin al-Quran, Generalizability Theory, EduG*

PENDAHULUAN

Menjawab perubahan jaman adalah sebuah keniscayaan. Semua aspek kehidupan akan mengalaminya tidak terkecuali Lembaga Pendidikan. Lembaga Pendidikan selalu dituntut sigap dalam menjawab perubahan-perubahan yang berhubungan erat dengan Pendidikan (Arifa, 2017: 27–28). Pada kurun awal 2000-an, Lembaga Pendidikan Tinggi Islam di Indonesia berlomba-lomba menjawab perubahan jaman dengan beralih status, dari STAIN ke IAIN, dari IAIN ke UIN (Adinugraha, Hidayanti, & Riyadi, 2018: 3). Perubahan tersebut, dalam paradigm keilmuan, memiliki tujuan besar untuk menghilangkan dikotomi keilmuan (Mustopa, 2019). Sebelumnya, dikotomi tersebut sangat terasa. Ada sebuah wacana yang berkembang bahwa

Lembaga Pendidikan Tinggi Islam hanya mempelajari ilmu-ilmu Agama. Akibatnya, input mahasiswanya lebih besar dari sekolah-sekolah yang berafiliasi keagamaan, seperti Madrasah maupun pesantren.

Selain itu, dalam tradisi keilmuan dikotomi tersebut telah membawa stagnasi keilmuan sehingga kejumu dan wacana tidak berkembang (Arifa, 2017: 30). Hal tersebut dikarenakan pemahaman doktrin agama yang dilakukan masih dalam *religious oriented*, sedangkan untuk mendapatkan pemahaman yang lebih komprehensif, diperlukan juga pemahaman dengan pendekatan dan paradigm lainnya (Setiawan, 2016: 72). Oleh karena itu, perubahan tersebut selain untuk menghapus dikotomi keilmuan, juga bertujuan untuk menjadikan Lembaga Pendidikan tinggi Islam lebih kompetitif, modern dan berkualitas (Idris, 2009: 22).

Perubahan adalah sebuah keniscayaan. Meski demikian, perubahan tentu memiliki implikasi, baik progress maupun regres. Kendati memiliki dua implikasi yang tidak dapat dihindarkan, perubahan akan selalu dialami dan implikasi tersebut selalu ada mengiringinya (Muslim, 2012: 138). Begitu juga yang terjadi dalam perubahan di Lembaga Pendidikan Tinggi Islam. Salah satu yang terlihat adalah animo masyarakat sebagai input Mahasiswa mulai bervarian dan tidak hanya berafiliasi dari Madrasah atau Pesantren. Selain itu, di Lembaga Pendidikan Tinggi Islam juga mulai dibuka program studi non-agama, seperti Matematika, Biologi, Komputer, Kedokteran dan sebagainya.

Salah satu dampak nyata dari perubahan yang baru dirasakan adalah mengenai Baca Tulis al-Quran. Kemampuan Mahasiswa Lembaga Pendidikan Tinggi Islam sudah tidak terukur dan tidak terkontrol. (Mustopa,

DOI: https://doi.org/10.28918/jei.v4i2.2296

2019) Ketidak-ukuran dan ketidak-kontrolan dari kemampuan Mahasiswa dalam Baca Tulis al-Quran sudah disadari dan antisipasi oleh masing-masing Lembaga. Salah satunya adalah lahirnya program asistensi PKTQ, yaitu program bimbingan bagi mahasiswa yang belum mampu membaca al-Quran dengan baik. Program asistensi tersebut dapat dijumpai di Fakultas Tarbiyah dan Ilmu Keguruan (FTIK), UIN Sunan Kalijaga. Oleh karena itu, Program PKTQ menyasar seluruh mahasiswa Fakultas Ilmu Tarbiyah dan Keguruan tanpa terkecuali (TIM 10 PKTQ, 2015: 11).

Standarisasi program PKTQ ditunjukkan dengan adanya ujian sertifikasi sebagai syarat KKN dan *munaqosyah*. Hasil ujian yang digunakan untuk sertifikasi menjadi penting agar kualitas sebuah kemampuan yang diujikan mencapai standar atau kriteria yang ditentukan (Mardapi, 2012: 120). Oleh karena itu, ujian sertifikasi merupakan kegiatan menguji kemampuan mahasiswa dalam membaca Al-Qur'an yang tidak terlepas dari proses pengukuran dan penilaian terhadap penguasaan kompetensi yang ditentukan dalam program Pengembangan Kepribadian dan Tahsin al-Qur'an. Selain itu, unsure pengukuran dan penilaian dalam ujian sertifikasi PKTQ dapat dilihat dari pelaksanaan ujiannya, yaitu dua *season*, yang pertama adalah ujian tulis dengan rentang waktu 120 menit. Soal ujian terdiri dari 30 pilihan ganda, 15 isian, dan 5 uraian. Sesi kedua, peserta akan mengikuti ujian lisan yang mana komponen ujiannya terdiri dari tes hafalan juz 30, tahsin, dan tajwid (Haramain, 2017).

Berdasarkan silabus PKTQ, tiga komponen yang menjadi ranah penilaian ujian sertifikasi PKTQ menjadi standar kompetensi yang harus dimiliki oleh mahasiswa. Uji kemampuan dalam penguasaan komponen

DOI: https://doi.org/10.28918/jei.v4i2.2296

tersebut dapat dilihat dari instrument penilaian dalam ujian lisan. Instrumen penilaian dalam ujian lisan yang berupa penilaian unjuk kerja (*performance* test) tidak lepas dari penilaian *rater/* penguji. Begitu pula pada penilaian ujian lisan sertifikasi PKTQ yang menggunakan satu *rater* pada tiap kelompok asistensi (Rohmah, 2017). Namun, penilaian dengan satu *rater* secara teoritik meninggalkan persoalan serius. Proses penilaian yang melibatkan satu *rater* pada tes unjuk kerja (*performance* test) berpeluang besar melahirkan ketidak konsistenan *rater* dalam memberikan penilaian pada individu yang berbeda dengan kemampuan yang sama. Untuk menghindari subjektifitas *rater* tersebut, menurut Azwar dibutuhkan minimal dua orang pemberi *rating* atau *rater* (Azwar, 2016: 113).

Dengan demikian, pengukuran kualitas dengan dua *rater* dapat dipertimbangkan hasil akhirnya dikarenakan nilai yang didapatkan bersumber dari kombinasi dua *rater* yang berbeda dengan objek yang sama. Penggunaan dua *rater* adalah sebuah upaya untuk menjawab problematika subjektifitas penilaian dalam proses uji sertifikasi Baca Tulis al-Quran yang ada di PKTQ. Meski demikian, subjektifitas masing-masing *rater* tidak dapat dihilangkan begitu saja, namun dapat diupayakan untuk diminimalisirkan. Minimalisir tersebut dapat dicapai jika masing-masing instrument yang dipakai oleh masing-masing *rater* sama dan berstandard. Oleh karena itu, tantangannya adalah membuat instrument penilaian yang berstandard dan dapat dipakai oleh kedua *rater*. Jika tidak ada instrument standard, proses penilaian dengan dua *rater* tetap akan sia-sia. Oleh karena itu, tulisan ini memfokuskan kepada pengembangan instrumen penilaian. Instrumen yang

DOI: https://doi.org/10.28918/jei.v4i2.2296

berstandard dan dapat dipakai oleh masing-masing *rater*. Instrumen berstandard adalah instrumen yang valid dan reliabel.

Prosedur yang digunakan dalam mengembangkan instrument penelitian ini merujuk pada pengembangan instrument tes yang ditawarkan oleh Djemari Mardapi dan disesuaikan dengan kebutuhan penilaian *tahsin al-qur'an.* Prosedur ini terdiri dari 7 tahap yaitu: menyusun spesifiksi tes, menulis soal tes, menelaah tes, melakukan uji coba tes, menganalisis instrument tes, memperbaiki tes dan merakit tes (Lusiana& Lestari, 2013: 3; Mardapi, 2012: 110). Uji coba dalam penelitian ini dilakukan sebanyak dua kali. Uji coba pertama merupakan uji coba sampel terbatas yang bertujuan mengetahui karakteristik instrumen dan uji coba sampel luas dilakukan untuk mengetahui kelayakan instrumen. Pengambilan sampel penelitian sebagai subjek coba dilakukan dengan teknik *Convenience sampling* dimana peneliti memilih partisipan karena kesediaanya untuk diteliti (Creswell, 2015: 294). Subjek uji coba dalam penelitian ini adalah mahasiswa asistensi *Tahsin al-Qur'an* sebagai peserta tes dan penguji sebagai *rater*, masing-masing berjumlah 13 orang mahasiswa dengan 3 orang *rater* pada uji coba sampel terbatas, dan 90 orang mahasiswa dengan 9 orang *rater* pada uji coba sampel luas.

Analisis data dilakukan secara kuantitatif. Analisis kuantitatif dilakukan untuk mengetahui hasil dari uji coba instrument penilaian *Tahsin al-Qur'an* secara empiric menggunakan pendekatan *generalizability theory*, data hasil uji coba dianalisis dengan menggunakan bantuan program *Edu G version 6.1*. Program *Edu G* merupakan aplikasi program berdasarkan teori

DOI: https://doi.org/10.28918/jei.v4i2.2296

generalisasi yang dikembangkan oleh Cronbach sebagai pengembangan dari teori tes klasik (Graham, Hebert, Sandbank, & Harris, 2016: 2). Teori ini memberikan kerangka kerja untuk memisahkan faktor-faktor yang mendasari berbagai desain pengukuran. Teori ini menjelaskan bahwa reliabilitas pengukuran yang secara konseptual dilakukan dengan mengidentifikasi sebanyak mungkin sumber potensial yang berkontribusi terhadap variasi skor yang memberikan perkiraan statistic besarnya sumber variasi skor tersebut.

Menurut Zhehan Jiang (2018), teori generalisasi atau teori G pada dasarnya adalah pendekatan terhadap perkiraan presisi pengukuran dalam situasi dimana pengukuran adalah subjek pada banyaknya sumber kesalahan. Pendekatan ini memungkinkan mengetahui informasi tentang kontribusi kesalahan yang digunakan untuk memperbaiki prosedur pengukuran di masa yang akan datang. Hal ini dikarenakan teori ini memungkinkan peneliti untuk menerapkan desain pengumpulan data yang berbeda dan memanipulasi ukuran sampel faset untuk mendapatkan berbagai rancangan pengukuran alternatif dan perkiraan nilai reliabilitsnya. Teori G memberikan perkiraan besarnya variasi-variasi skor, *true score*, atau *universe* yang diinginkan, dan variasi-variasi kesalahan yang tidak diinginkan. Hal tersebut berlaku untuk setiap sumber variasi misalnya seperti individu, item/ tugas, dan *rater* dan kemungkinan kombinasi dari semua variasi tersebut.

Estimasi instrument dengan *generalizability theory* sudah pernah dilakukan oleh Lumaurridlo (2019), dimana nilai Koefisien yang didapat adalah 0,81. Nilai tersebut diperoleh dari 4 aspek yang dijadikan dalam penilaian munaqasah. Sama halnya dengan penelitian Abdullah Faiz

DOI: https://doi.org/10.28918/jei.v4i2.2296

mengenai pengembangan instrument penilaian *tahfidz Al-Qur'an* di FITK UNSIQ Wonosobo. Tujuan penelitian tersebut untuk mengukur kualitas hafalan Al-Qur'an dalam mata kuliah *Tahfidz Al-Qur'an* di Fakultas IlmuTarbiyah dan Keguruan (FITK) Universitas Sains Al-Qur'an (UNSIQ) Wonosobo. Persamaan penelitian ini adalah terdapat pada objek penelitiannya yaitu mengembangkan instrument penilaian dalam bentuk tes unjuk kerja (*performance* test) membaca Al-Qur'an dan pendekatan analisis data yang digunakan, yaitu dengan *generalizability theory*. Namun, perbedaan mendasar penelitian Faiz adalah penggunaan prosedur Delphi dengan program SPSS (Faiz, 2011: 68–69).

Penelitian oleh Bahrudin dan Kumaidi mengenai pengembangan model asesmen *musabaqah tilawah Al-qur'an* (MTQ) cabang tilawah juga perlu dipertimbangkan. Penelitian tersebut menjelaskan model asesmen yang dikembangkan bertujuan untuk menilai MTQ. Salah satu alasan yang dikemukakan oleh Bahrudin dan Kumaidi bahwa penilaian MTQ selama ini tidak sesuai dengan teori asesmen (Bahrudin & Kumaidi, 2014: 157). Penelitian yang dilakukan oleh Bahrudin dan Kumaidi tersebut dalam mengembangkan instrument penilaian MTQ menggunakan model pengembangan Borg and Gall (Bahrudin & Kumaidi, 2014: 110).

Beberapa penelitian diatas dapat mengilustrasikan posisi tulisan ini dalam penelitian yang menggunakan pendekatan *generalizability theory*. Meskipun pendekatan sama dengan Lumaurridlo dan Faiz, untuk Lumaurridlo tentunya objek kajian sudah berbeda. Sedangkan untuk Faiz tentunya penelitian penulis berbeda pada prosedur dan program yang dipakai.

DOI: https://doi.org/10.28918/jei.v4i2.2296

Sedangkan penelitian Baharudin dan Kumaidi pada model pengembangan dimana penelitian ini mengikuti model yang ditawarkan oleh Djemari Mardapi. Sedangkan penyusunan instrumen juga pernah dilakukan oleh Zuhaida (2018). Tujuan dari penelitian ini ialah melakukan penyusunan instrument analisis PCK bagi guru IPA madrasah tsanawiyah yang terintegrasi konten Islami. Namun, analisis yang dilakukan adalah deskriptif-kualitatif dengan penggunaan SPSS.

PENGEMBANGAN MODEL PENILAIAN TAHSINUL QUR'AN

Pengembangan model yang dimaksud adalah pengembangan instrumen penilaian *Tahsin al-Qur'an* yang digunakan sebagai alat penilaian dalam ujian lisan sertifikasi program Pengembangan Kepribadian dan *Tahsin al-Qur'an* (PKTQ) di Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta. Penelitian ini dilakukan pada kurun waktu 09 Januari 10 April 2017. Beberapa aspek dan hasil kajian penelitiannya seputar Estimasi Validitas, kriteria penilaian, ujicoba sampel terbatas dan ujicoba sampel luas.

Estimasi Validitas

Validasi bertujuan untuk mengetahui sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya (Matondang, 2009: 89). Adapun dalam penelitian ini validitas yang digunakan adalah validitas isi (*conten validity*). Setelah penyusunan instrument unjuk kerja penilaian *Tahsin al-Qur'an*, maka instrument tersebut diajukan kepada tim ahli (*expert judgement*). Telaah secara kuantitatif dilakukan oleh 3 orang ahli yaitu ahli pengukuran, ahli ilmu qur'an dan ahli *qiro'at*. Tugas ahli dalam hal

DOI: https://doi.org/10.28918/jei.v4i2.2296

ini melihat kesesuaian indicator dengan tujuan pengembangan instrumen, kesesuaian indicator dengan cakupan materi atau kesesuaian teori, melihat kesesuaian instrument dengan indicator butir, melihat kebenaran konsep butir, melihat kebenaran isi, kebenaran kunci (Retnawati, 2016: 5). Ahli memberikan validasi berupa penilaian secara kuantitatif dengan memberikan skor pada lembar validasi mengenai kesesuaian Kisi-kisi tes, lembar penilaian dan rubric penilaian *Tahsin al-Qur'an* yang telah tersusun dengan tujuan dan spesifikasi tes *Tahsin al-Qur'an*. Berdasarkan skor yang diberikan oleh 3 ahli terhadap 18 butir instrumen, peneliti menghitung hasil kesepakatan ahli dengan *formula index Aiken's* untuk mengetahui validitas isi instrument penilaian *Tahsin al-Qur'an*.

Nomor Butir	*Indeks Aiken*	Keterangan
1	0.77	Valid
2	0.77	Valid
3	0.77	Valid
4	0.88	Valid
5	0.88	Valid
6	1.00	Valid
7	0.77	Valid
8	0.88	Valid
9	0.88	Valid
10	0.88	Valid
11	0.88	Valid
12	0.88	Valid
13	0.88	Valid
14	0.88	Valid
15	0.88	Valid
16	0.77	Valid
17	0.88	Valid

DOI: https://doi.org/10.28918/jei.v4i2.2296

Nomor Butir	*Indeks Aiken*	Keterangan
18	0.88	Valid

Tabel 1. Hasil Perhitungan Formula Aiken's

Hasil perhitungan indeks Aiken diatas kemudian disesuaikan dengan batas minimal nilai V pada tabel Aiken. Berdasarkan tabel Aiken, nilai minimal pada 18 butir instrument dengan 4 *rating* kategori dan menggunakan tingkat kesalahan sebesar 5% adalah V = 0.67 (Aiken, 1985, hal. 134). Hasil perhitungan indeks Aiken yang terdapat pada tabel 3 menunjukkan nilai indeksnya lebih dari 0.67 pada tiap butir, oleh karena itu dapat disimpulkan 18 butir instrument penilaian *Tahsin al-Qur'an* valid dalam mengukur kemampuan membaca Al-Qur'an secara *tahsin*.

Setelah diestimasi validitas isi, kemudian dilakukan uji coba. Uji coba penelitian dilaksanakan dua tahap, yaitu uji coba lapangan pada sampel terbatas dan uji coba penelitian pada sampel luas yang diselenggarakan di Fakultas Ilmu Tarbiyyah dan Keguruan UIN Sunan Kalijaga Yogyakarta. Uji coba sampel terbatas dilaksanakan pada tanggal 8 juni 2017 yang diikuti oleh 13 orang mahasiswa asistensi dan 3 orang penguji. Hasil uji coba dianalisis untuk mengetahui desain pengukuran yang sesuai dan reliabilitas instrumen. Selanjutnya, instrument digunakan pada pelaksanaan ujian kelayakan tahsin al-qur'an yang dilaksanakan pada tanggal 17 juni 2017. Ujian kelayakan ini diikuti oleh 90 orang mahasiswa asistensi dan 9 orang penguji.

Menentukan Kriteria Penilaian

Penilaian *Tahsin al-Qur'an* merupakan prosedur dalam menentukan kelulusan pada program sertifikasi Pengembangan Kepribadian dan *Tahsin*

DOI: https://doi.org/10.28918/jei.v4i2.2296

al-Qur'an (PKTQ). Oleh karena itu, proses penilaian terdapat batas nilai untuk menentukan kelulusan. Penetapan batas nilai dan hasil akhir dari penilaian ini, selain untuk mengetahui mahasiswa yang lulus dan tidak lulus, juga bertujuan untuk mengetahui sejauh mana kualitas bacaan Al-Qur'an mahasiswa. Pengetahuan atas kualitas bacaan al-Quran, tentunya, sangat membantu dalam proses pengkategorian kualitas bacaan mahasiswa. Lebih lanjut, kategorisasi diperlukan untuk mengetahui criteria kualitas membaca al-qur'an mahasiswa agar dapat digunakan sebagai bahan pertimbangan bagi kemajuan pembelajaran *Tahsin al-Qur'an* dan tercapainya tujuan program PKTQ.

Berdasarkan hal tersebut, maka penentuan criteria penilaian dilakukan dengan prosedur penilaian kombinasi. Penilaian kombinasi merupakan penentuan criteria penilaian dengan mengkombinasikan antara penilaian absolut (*criterion referenced* evaluation) dan penilaian relatif (*norm-referenced evaluation*) (Idris, 2007: 26). Penilaian absolut (*criterion referenced evaluation*) adalah penilaian yang berpatokan pada domain penilaian yang sudah ditentukan tanpa membandingkan kemampuan antara individu. Oleh karena itu, penilaian ini sering digunakan untuk menentukan kelulusan pada kriteria domain tertentu. Sedangkan penilaian relatif (*norm-referenced evaluation*) adalah penilaian dengan membandingkan kemampuan individu dengan kelompoknya. Berbeda dengan penilaian absolut, penilaian relative sering digunakan untuk melihat performa atau kemampuan individu pada kelompoknya. Pada dasarnya, penilaian relative tidak membandingkan pada tingkat kesulitan tes, tetapi membandingkan performa individu pada

DOI: https://doi.org/10.28918/jei.v4i2.2296

kelompok test tertentu. Oleh karena itu, pemakaian penilaian kombinasi dari kedua prosedur tersebut merupakan jalan keluar untuk menghadapi kemungkinan tidak dapatnya menerapkan prosedur penilaian absolute ataupun relative secara murni (Azwar, 2016: 171).

Adapun penilaian kombinasi yang diterapkan untuk menentukan criteria penilaian *Tahsin al-Qur'an* dalam penelitian ini adalah *pertama*, menetapkan terlebih dulu skor sebagai criteria atau batas kelulusan yang harus dicapai oleh mahasiswa. *Kedua*, menetapkan skor relative menggunakan skor standar dengan menggunakan mean dan simpangan baku pada keseluruhan skor yang diperoleh. Skor yang ditetapkan sebagai batas kelulusan adalah>75 (TIM 10 PKTQ, 2017). Skor tersebut ditentukan berdasarkan kriteria yang digunakan oleh Tim penilai Program Pengembangan dan Kepribadian *Tahsin al-Qur'an* (PKTQ) dalam menentukan standar kelulusan bagi mahasiswa yang mengikuti ujian sertifikasi *Tahsin al-Qur'an*.

Berdasarkan ketentuan di atas, mahasiswa yang mendapatkan skor 75 dan dibawah skor tersebut dinyatakan tidak lulus. Selanjutnya, mahasiswa yang memperoleh skor diatas nilai tersebut akan diberi nilai relative dengan menggunakan skor standar berdasarkan pada mean dan simpangan baku. Selanjutnya, dalam menetapkan norma penilaian, nilai tersebut diklasifikasikan menjadi lima kategori dengan mengkonversikan skor menjadi lima skala (Azwar, 2015: 163).

Kelas Interval	**Kategori**
(M + 1,50s) < X	Sangatbaik

DOI: https://doi.org/10.28918/jei.v4i2.2296

(M + 0,50s) < X ≤ (M + 1,50s)	Baik
(M – 0,50s) < X ≤ (M + 1,50s)	Cukup
(M – 1,50s) < X ≤ (M - 0,50s)	Kurang
X ≤ (M - 1,50s)	Sangatkurang

Tabel 2. Konversiskoraktual pada lima skala

Keterangan:
M = Mean (ΣfX/N)
s = Simpanganbaku ($s^2 = \Sigma f(X - M)^2$/(N-1))
X = Skoraktual

Berdasarkan pada prosedur penilaian tersebut, mahasiswa peserta tes yang mendapat nilai 75 dan kurang dari 75 secara otomatis berada pada kategori sangat kurang, sedangkan mahasiswa yang memperoleh nilai diatasnya akan berada pada kategori sangat baik, baik, cukup, dan kurang bergantung pada posisi rentang nilai dari hasil yang diperoleh oleh mahasiswa peserta tes.

Analisis Hasil Uji Coba Terbatas

Setelah mendapatkan criteria penilaian sebagaimana diatas, maka tahap selanjutnya adalah melakukan uji coba kelompok terbatas. Pada dasarnya, uji coba terbatas bertujuan untuk memperoleh informasi mengenai keandalan instrumen dan varians yang berkonstribusi pada nilai koefisien yang diperoleh dari hasil analisis *G-study*. Informasi tersebut digunakan sebagai landasan untuk uji coba instrumen pada sampel yang lebih luas.

Uji coba terbatas dilakukan dengan simulasi ujian di Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta. Peserta ujian yang

DOI: https://doi.org/10.28918/jei.v4i2.2296

mengikuti simulasi ini adalah mahasiswa fakultasilmutarbiyah dan keguruan yang mengikuti program tahsin dari 1 kelompok asistensi yang berjumlah 13 mahasiswa. Adapun penentuan subjek uji coba terbatas dilakukan secara acak. Detailnya, penilaian dalam uji coba terbatas diberikan oleh 3 orang *rater* dari tim penguji, masing-masing *rater* menilai peserta ujian secara independen. Selain itu, uji coba kelompok terbatas instrument penilaian *Tahsin al-Qur'an* melibatkan 3 varaiasi facet yaitu: mahasiswa, *rater*, dan butir. Ketiga facet tersebut disilangkan, yaitu setiap *rater* menilai mahasiswa pada tiap butir instrumen yang sama. Oleh karena itu desain observasi *G-study* yang digunakan adalah (MxRxB). Variasi dari tiga faset ini dapat menghasilkan estimasi tujuh komponen varians yang berbeda: M, R, B, MR, MB, RB dan MRB.

Source	SS	Df	MS	Varian	%
M	115.910	12	9.659	0.141	28 %
R	7.960	2	3.979	0.009	1.8 %
B	26.770	17	1.575	0.029	5.3 %
MR	42.598	24	1.775	0.089	18.9 %
MB	81.786	204	0.401	0.081	15.7 %
RB	6.197	34	0.182	0.002	0.4 %
MRB	63.248	408	0.155	0.155	29.8 %
Total	344.462	701			100%

Tabel 3. Estimasi Komponenvariansi Mahasiswa, *Rater*, dan Butir Penilaian *Tahsin al-Qur'an* dengan desain G studi M X R X B

Berdasarkan hasil analisis varian dapat diketahui bahwa terdapat 3 variabel yang memiliki nilai persentase tinggi dalam mempengaruhi varian kesalahan pengukuran. Varian paling tinggi adalah interaksi mahasiswa

DOI: https://doi.org/10.28918/jei.v4i2.2296

sebesar 28.0 %. Kemudian lebih lanjut pada varian mahasiswa yang disilangkan dengan *rater* (MR). Varian (MR) memiliki kontribusi sebesar 18.9% pada kesalahan pengukuran. Angka tersebut memiliki selisih yang tidak terlalu besar dengan varian mahasiswa yang disilangkan dengan butir instrument (MB). Varian tersebut memiliki nilai kontribusi sebesar 15.7%. Berdasarkan hasil ini dapat dikatakan bahwa terdapat kesenjangan pemahaman pada *rater* dalam memberikan penilaian. Artinya, *rater* belum memahami rubric penilaian yang digunakan untuk memberikan skor pada mahasiswa. Hal tersebut berbeda terjadi pada interaksi *rater* dan butir yang memiliki kontribusi rendah yaitu 0.4% atau tidak terjadi kesenjangan. Dengan kata lain, penilaian dapat dilakukan secara objektif.

Setelah mengetahui kesenjangan antar varian yang member kontribusi kesalahan dalam penilaian, selanjutnya dilakukan uji reliabilitas untuk mengetahui sejauh mana penilaian menggunakan instrumen yang dikembangkan dengan bentuk facet MxRxB tersebut reliabel. Oleh karena itu, Analisa tahap ini dilakukan menggunakan *G-study* dengan tujuan mengetahui nilai koefisien *G-study* yang menunjukkan nilai reliabilitas dan kebermaknaan penggunaan instrument penilaian *Tahsin al-Qur'an* dari hasil uji coba pada sampel terbatas. Pada akhirnya, Hasil *G-study* memberikan informasi tentang nilai koefisien G *relative* dan G *absolute* yang dijadikan rujukan untuk mengetahui tingkat reliabilitas instrumen.

Source of variance	Relative		Absolute	
M	Coef_G	SE	Coef_G	SE
	0.82	0.18	0.80	0.19

Tabel 4. Koefisien G pada Uji coba terbatas instrument penilian *Tahsin al-Qur'an*

DOI: https://doi.org/10.28918/jei.v4i2.2296

Tabel 4 menunjukkan nilai koefisien relatif dan nilai koefisien absolute penggunaan instrument penilaian tahsin Al-Qur'an yang masing-masing sebesar 0,82 dengan standar eror sebesar 0,18 (18%) dan 0,80 dengan standar eror sebesar 0.19 (19%). Berdasarkan hasil ini, secara keseluruhan pengembangan instrument penilaian *Tahsin al-Qur'an* dapat dikatakan reliabel. Lebih jelasnya, instrument penilaian *Tahsin al-Qur'an* dapat digunakan pada pengukuran yang bersifat relatif, yaitu membandingkan kemampuan membaca al-qur'an mahasiswa baik yang memiliki kemampuan tinggi ataupun rendah. Oleh karena itu, nilai koefisien yang digunakan dari hasil analisis *G-study* adalah coeficien G relative sebesar 0.82. Berdasarkan nilai tersebut, instrument penilaian *Tahsin al-Qur'an* dapat dikatakan reliabel. Berdasarkan pendapat Jean Cardinet (2010: 5) setidaknya koefisien reliabilitas > 0,8 sehingga instrument tersebut dapat diterima untuk digunakan pada penilaian dengan facet yang lebih luas pada uji selanjutnya.

Tahap berikutnya dalam analisis dengan menggunakan pendekatan *generalizability theory* adalah *D-study*. *D-study* dilakukan untuk menghitung reliabilitas dan *standar error* pada instrument penilaian untuk mengidentifikasi skema pengambilan sampel faset yang meminimalkan kesalahan pengukuran. Pada penelitian ini, dilakukan penambahan dan pengurangan *rater* untuk mengetahui nilai koefisien yang lebih tinggi dengan jumlah sampel pada faset tertentu. Hasil perhitungan *D-study* diperoleh lima nilai coefisien yang berbeda. Hal ini dikarenakan jumlah *rater* pada masing-masing penilaian dinaikan sebanyak satu kali dengan jumlah item yang sama. Sehingga hasil nilai akan berbeda pada setiap kenaikan jumlah *rater*.

DOI: https://doi.org/10.28918/jei.v4i2.2296

No	SAMPLE SIZE			GENERALIZABILITY		SelisihKoefisienEduG
	M INF	**R INF**	**B INF**	**COEF**	**SEM**	
1	13	1	18	0.60	0.31	0.09
2	13	2	18	0.75	0.22	0.04
3	13	3	18	0.82	0.18	0.02
4	13	4	18	0.86	0.16	0.02
5	13	5	18	0.88	0.14	0.01
6	13	6	18	0.90	0.13	

Tabel 5. Estimasi *D-study* Instrumen Penilaian *Tahsin al-Qur'an* Pada Uji coba Sampel Terbatas

Tabel 5 diatas menginformasikan bahwa penambahan jumlah rater yang berbeda pada penggunaan instrument penilaian *Tahsin al-Qur'an* menghasilkan nilai koefisien dan SEM yang berbeda pula. Nilai koefisien yang diperoleh pada instrument penilaian tahsin al-qur'an dengan rater 1 sampai dengan 6 dimana jumlah item 18 secara berturut-turut adalah 0.60, 0.75, 0.82, 0.86, 0.88, dan 0.90. Sedangkan untuk nilai SEM secara berturut-turut adalah0.31, 0.22, 0.18, 0.16, 0.14, dan 0.13.

Berdasarkan hasil tersebut, dapat dikatakan semakin banyak jumlah rater yang digunakan dengan jumlah item yang sama maka nilai coefisiennya akan semakin tinggi dan nilai SEM menjadi semakin rendah. Oleh karena itu, untuk mengurangi tingkat kesalahan dalam pengukuran dan memaksimalkan penilaian dalam menggunakan instrument penilaian *Tahsin al-Qur'an* diperlukan lebih dari 1 orang *rater* untuk menilai secara objektif. Setidaknya, *rater* dalam penilaian *Tahsin al-Qur'an* berjumlah 3 orang. Dengan demikian, melihat nilai coefisien G yang diperoleh diatas, dapat dikatakan bahwa nilai reliabilitas cukup stabil.

DOI: https://doi.org/10.28918/jei.v4i2.2296

Analisis Hasil Uji Coba Sampel Luas

Setelah reliabilitas instrument terbukti, uji coba dengan sampel luas dilakukan pada facet yang berbeda. Uji coba sampel luas dilaksanakan pada saat ujian kelayakan *Tahsin al-Qur'an*. Adapun jumlah peserta pada ujian kelayakan *Tahsin al-Qur'an* skala luas diiukuti oleh 90 peserta yang terbagi dalam 3 kelas. Setiap kelas berisi 30 mahasiswa yang diuji oleh 3 penguji pada masing-masing kelas. Penambahan facet berupa kelas menjadikan desain pengukuran yang berbeda pada uji coba tahap sebelumnya.

Uji penelitian dengan sampel luas instrument penilaian *Tahsin al-Qur'an* melibatkan 4 varaiasi facet yaitu: Kelas (K), Rater (R), Mahasiswa (M), dan Butir (B). Adapun dari keempat facet, tiga diantaranya merupakan rangkaian bersarang tunggal (*nested*) yaitu: Mahasiswa bersarang di *rater*, *rater* bersarang di kelas. Selanjutnya, facet butir disilangkan dengan setiap aspek lainnya. Langkah detailnya adalah semua mahasiswa disetiap kelas dinilai oleh *rater* siapapun menggunakan instrumen yang sama dalam proses penilaian *Tahsin al-Qur'an*. Sehingga, desain observasinya adalah (M:R:K) B.

Source	SS	Df	MS	Varian	%
K	158.044	2	79.022	0.043	6.4 %
R:K	38.044	6	6.341	0.006	1.4 %
M:R:K	757.298	261	2.902	0.150	34.3 %
B	219.883	17	12.934	0.035	9.3 %
KB	114.526	34	3.368	0.033	4.4 %
RB:K	43.533	102	0.427	0.008	1.6 %
MB:R:K	893.002	4437	0.201	0.201	42.8 %
Total	2224.331	4859			100%

Tabel 6. Estimasi Komponen variansi Kelas, Mahasiswa, Penilai, dan Butir Penilaian *Tahsin al-Qur'an* dengan Desain G Studi (M:R:K) B

DOI: https://doi.org/10.28918/jei.v4i2.2296

Berdasarkan tabel 6 dari hasil analisis varian uji coba lapangan dengan sampel luas dapat diketahui kontribusi yang sangat tinggi pada efek interaksi antara mahasiswa dan butir (MB:R:K) yaitu sebesar 42.8%. Kemudian pada mahasiswa yang bersarang di *rater* dan kelas sebesar 34.3 %. Hasil tersebut menunjukkan bahwa mahasiswa menjadi penyumbang kesalahan dalam penilaian. Selain dari hasil analisis varian, hasil estimasi dari *G-study* menunjukkan penyebab masalah lebih lanjut yang dapat diketahui berdasarkan pada sumber kesalahan.

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
K	0.02991	R:K	0.00212	54.2	0.00212	54.2
		M:R:K	0.00179	45.8	0.00179	45.8

Tabel 7. Kesalahan Varian pada Hasil *G-Study*

Berdasarkan tabel 7 diatas, sumber kesalahan relatif pada mahasiswa yang bersarang pada *rater* dan kelas (M:R:K) sebesar 45.8% dan penyumbang kesalahan tertinggi adalah pada *rater* yang bersarang pada kelas (R:K) yaitu sebesar 54.2%. Hasil tersebut menunjukkan masalah dalam pengukuran terjadi pada *rater* yang menilai mahasiswa dalam tiap kelas. Untuk mengetahui *rater* di kelas mana yang bermasalah dapat diketahui melalui hasil perhitungan *mean*.

K KELAS	Mean	Variance	Std. Dev.
1	3.741	0.429	0.653
2	3.578	0.350	0.592
3	4.015	0.497	0.705
Grand mean		3.778	

Variance	0.458
Standard Dev	0.677

Tabel 8. Hasil M*ean* Varian K

Berdasarkan tabel 8 diatas dapat diketahui bahwa nilai *mean* pada kelas ke-3 memiliki nilai yang lebih besar dari *grand mean* yaitu 4.015. Hal ini menunjukkan masalah terjadi pada *rater* dalam ruangan ujian ke-3. Selanjutnya untuk mengetahui *rater* mana yang bermasalah dalam ruang kelas ke-3 dapat diketahui melalui hasil perhitungan *mean* varian R:K.

K KELAS	R RATER	Mean	Variance	Std. Dev.
1	1	3.865	0.458	0.677
1	2	3.707	0.400	0.632
1	3	3.652	0.405	0.636
2	1	3.533	0.334	0.578
2	2	3.520	0.305	0.552
2	3	3.681	0.395	0.628
3	1	3.972	0.538	0.734
3	2	4.154	0.489	0.700
3	3	3.920	0.433	0.658

Tabel 9. Hasil *mean* varian R:K

Tabel 9 diatas memberikan informasi bahwa nilai *mean* untuk ketiga *rater* yang berada dalam ruang kelas ke-3 lebih besar dari nilai *grand mean*, dan nilai yang lebih tinggi adalah pada *rater* ke-2. Hal ini membuktikan bahwa penyumbang kesalahan dalam pengukuran terdapat pada *rater* di ruang ujian ke-3 khususnya pada *rater* kedua. Hal ini dikarenakan *rater* dalam ruang ke-tiga pada uji coba sampel luas berbeda dengan *rater* pada uji coba sebelumnya. Sedangkan pada interaksi *rater* dan butir kontribusi

DOI: https://doi.org/10.28918/jei.v4i2.2296

efeknya rendah yaitu sebesar 1.6% dan interaksi *rater* dalam kelas sebesar 1.4%. Hal ini berarti secara keseluruhan *rater* sudah memahami penggunaan instrument penilaian *Tahsin al-Qur'an* yang dikembangkan dan menggunakannya secara sepaham sehingga penggunaan instrument dapat digunakan secara maksimal. Selanjutnya untuk mengetahui reliabilitas penilaian pada facet ini, dapat diketahui dari nilai koefisien G yang dihasilkan dari analisis *G-study*.

Source of variance	Relative		Absolute	
K	Coef_G	SE	Coef_G	SE
	0.88	0.06	0.88	0.06

Tabel 10. Koefisien G pada Uji Coba Sampel Luas Instrument Penilian *Tahsin al-Qur'an*

Table 10 diatas menunjukkan bahwa nilai koefisien G relative digunakan sebagai penentu nilai reliabilitas instrument penilaian al-qur'an yang digunakan sebagai alat penilaian yang bersifat relatif. Berdasarkan tabel 9 dapat diketahui juga bahwa nilai koefisien relatif pada butir instrument sebesar 0.88 dan nilai standar eror relative sebesar 0.06 (6%) sehingga dapat dikatakan bahwa instrument yang digunakan pada penilaian dengan facet ini reliabel.

Penambahan dan pengurangan facet dalam bentuk penilaian diatas dimungkinkan. Salah satu alasannnya adalah untuk mengetahui bentuk penilaian dengan facet tertentu yang lebih efisien dan reliabel, baik dengan cara mengurangi peserta tes maupun *rater*. Lebih lanjut, efisien dan reliabelnya dapat diketahui dengan melakukan analasisi *D-study.* Analisis ini

mengestimasi pengurangan atau penambahan jumlah facet berdasarkan hasil *G-study* yang telah dianalisis sebelumnya. Sehingga dapat diketahui nilai koefisien reliabilitas pada rancangan tertentu.

Analisis *D-study* pada facet ini dilakukan dengan penambahan faset kelas dimana antara varian mahasiswa bersarang pada rater, dan rater bersarang pada kelas. Pada analisis ini dilakukan penambahan *rater* sebanyak 1 kali dan pengurangan jumlah mahasiswa sebanyak 2 kali pada jumlah butir dan kelas yang sama. Hal ini dilakukan untuk mengetahui desain pengukuran yang baik dalam melakukan penilaian *Tahsin al-Qur'an.*

No	SAMPLE SIZE				GENERALIZABILITY		Selisih Koefisien Edu G
	K INF	**M INF**	**R INF**	**B**	**COEF**	**SEM**	
1	3	30	1	18	0.79	0.11	0.03
2	3	30	2	18	0.88	0.08	0.02
3	3	30	3	18	0.88	0.06	
4	3	15	1	18	0.72	0.13	0.04
5	3	15	2	18	0.84	0.09	0.01
6	3	15	3	18	0.89	0.08	

Tabel 11. Estimasi *D-study* Instrumen Penilaian *Tahsin al-Qur'an* Pada Uji coba Sampel Luas

Dari tabel 11 dapat diketahui bahwa penambahan jumlah *rater* dan mengurangi jumlah varian mahasiswa yang berbeda pada penggunaan instrument penilaian *Tahsin al-Qur'an* menunjukkan nilai koefisien dan SEM yang berbeda pula. Nilai koefisien yang diperoleh pada instrument penilaian *Tahsin al-Qur'an* pada 3 kelas dengan rater sebanyak 1, 2, dan 3 orang yang menilai mahasiswa sejumlah 30 orang secara berturut-turut adalah 0.79, 0.88, 0.88. Sedangkan untuk nilai SEM secara berturut-turut adalah 0.11, 0.08, dan

DOI: https://doi.org/10.28918/jei.v4i2.2296

0.06. Sedangkan besar nilai koefisien pada *rater* sejumlah 1, 2, dan 3 orang yang menilai 15 orang mahasiswa secara berturut-turut adalah 0.72, 0.84, dan 0.89. dan untuk nilai SEM secara berturut-turut adalah 0.13, 0.09, dan 0.08.

Berdasarkan hasil tersebut, secara keseluruhan nilai koefisien reliabilitas mengalami peningkatan pada setiap penambahan jumlah *rater*. Hal ini berarti penambahan *rater* pada beberapa keadaan dianjurkan untuk meningkatkan nilai reliabilitas dalam penggunaan instrumen penilaian ini. Adanya pengurangan terhadap jumlah mahasiswa pada masing-masing ruangan dari 30 menjadi 15 orang memberikan pengaruh terhadap besar nilai koefisien reliabilitas, tetapi tidak terlalu signifikan jika dilihat dari selisih besar nilai koefisien pada masing-masing *facet* yang relatif kecil. Hal ini dapat dikatakan semakin banyak jumlah *rater* yang digunakan dalam penilaian maka semakin tinggi nilai koefisien G nya sehingga semakin reliable penilaiannya.

Adapun nilai koefisien tertinggi dan SEM terendah pada penilaian terhadap mahasiswa sejumlah 30 orang adalah pada penggunaan *rater* sejumlah 2 dan 3 orang dengan nilai koefisien sebesar 0.88 dan SEM 0.08. Sedangkan pada penilaian terhadap 15 orang mahasiswa nilai koefisien tertinggi pada penggunaan *rater* sebanyak 3 orang dengan nilai koefisien sebesar 0.88 dan SEM 0.08. Selain itu, penggunaan *rater* sebanyak 2 orang pada penilaian terhadap jumlah mahasiswa yang berbeda sudah memenuhi nilai reliabilitas yaitu 0.88 pada penilaian 30 orang mahasiswa dan 0.84 pada penilaian terhadap 15 orang mahasiswa. Berdasarkan hasil ini maka banyaknya *rater* yang dianjurkan dalampenilaian *Tahsin al-Qur'an* dengan

menggunakan instrumen yang dikembangkan adalah sebanyak 2 orang untuk nilai kestabilan diatas 0.80. Apabila memungkinkan, maka penilaian dapat dilakukan lebih dari 2 *rater* untuk memaksimalkan penilaian *Tahsin al-Qur'an* dengan instrumen yang dikembangkan.

SIMPULAN

Berdasarkan pembahasan diatas, tulisan ini menyimpulkan bahwa hasil dari uji instrumen penilaian *tahsinul qur'an* dikatakan valid dengan nilai index aiken > 0.67 pada tiap indikatornya. Uji reliabilitas pada instrumen ini menunjukan nilai koefisiean reliabilitas > 0.80 baik pada ujicoba sampel terbatas dan sampel luas yang berarti instrumen tersebut dapat dikatakan reliabel. Penyumbang kesalahan pengukuran dalam estimasi ini terdapat pada *rater* di ruangujian ke-3 khususnya pada *rater* kedua, hal ini dapat diminimalisir dengan mengurangi setengah dari jumlah sampel mahasiswa yang diuji Berdasarkan hal-hal tersebut, instrumen penilaian *tahsinul qur'an* dapat dikatakan valid dan reliabel untuk mengukur *tahsin al-qur'an* dan dapat meminimalisir subjektivitas *rater* dalam melakukan penilaian.

DAFTAR PUSTAKA

Adinugraha, Ema Hidayanti, Agus Riyadi, H. H., Hidayanti, E., & Riyadi, A. (2018). Fenomena Integrasi Ilmu di Perguruan Tinggi Keagamaan Islam Negeri: Analisis Terhadap Konsep Unity of Sciences di UIN Walisongo Semarang. HIKMATUNA: Journal for Integrative Islamic Studies, 4(1), 1–16. https://doi.org/10.28918/hikmatuna.v4i1.1267

DOI: https://doi.org/10.28918/jei.v4i2.2296

Aiken, L. (1985). Three Coefficients for Analyzing the Reliability and Validity of Rating. Educational and Psychological Measurement, (45), 131–142.

Arifa, L. N. (2017). Perubahan STAIN/ IAIN Menjadi UIN Sebagai Bentuk Pengembangan Pendidikan Tinggi Islam (Contoh Kasus Perubahan STAIN menjadi UIN Malang Perspektif Manajemen Perubahan Kurt Lewin). Vicratina, Vol 01(02), 27–42.

Azwar, S. (2015). Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar (4 ed.). Yogyakarta: Pustaka Pelajar.

Azwar, S. (2016). Reliabilitas dan Validitas (4 ed.). Yogyakarta: Pustaka Pelajar.

Bahrudin, & Kumaidi. (2014). Model Asesmen Musabaqah Tilawah al-Quran (MTQ) Cabang Tilawah. Jurnal Penelitian dan Evaluasi Pendidikan, Vol. 18 (2), 153–167.

Creswell, J. W. (2015). Riset Pendidikan: Perencanaan, Pelaksanaan, dan Evaluasi Riset Kualitatif & Kuantitatif. Yogyakarta: Pustaka Pelajar.

Faiz, A. (2011). Pengembangan instrumen penilaian tahfidz Al-Qur'an di FITK UNSIQ Wonosobo (Tesis). UNY Yogyakarta, Yogyakarta.

Graham, S., Hebert, M., Paige Sandbank, M., & Harris, K. R. (2016). Assessing the Writing Achievement of Young Struggling Writers: Application of Generalizability Theory. Learning Disability Quarterly, 39(2), 72–82. https://doi.org/10.1177/0731948714555019

Haramain, M. I. N. (2017, Maret 26). Pra-Observasi Ujian Sertikasi PKTQ [Panduan Wawancara].

Idris, D. (2007). Teknik Penilaian Pembelajaran Dengan Menggunakan Passing Grade. JMSK: Jurnal Matematika, Statistika & Komputasi, Vol. 4 (No. 2), 26–29.

DOI: https://doi.org/10.28918/jei.v4i2.2296

Jiang, Z., & Skorupski, W. (2018). A Bayesian approach to estimating variance components within a multivariate generalizability theory framework. Behavior Research Methods, 50(6), 2193–2214. https://doi.org/10.3758/s13428-017-0986-3

Lumaurridlo. (2019). Estimasi Keandalan Penilaian Munaqosah. Jurnal Tawadhu, Vol. 3 (1), 665–673.

Lusiana, D., & Lestari, W. (2013). Instrumen Penilaian Afektif Pendidikan Karakter Bangsa Mata Pelajaran PKN SMK. Journal of Education Research and Evaluation, Vol. 2(1), 1–6.

Mardapi, D. (2012). Pengukuran penilaian & Evaluasi pendidikan. Yogyakarta: Nuha Medika.

Matondang, Z. (2009). Validitas dan Reliabilitas Suatu Instrumen Penelitian. Jurnal Tabularasa, Vol.6 (No. 1), 87–97.

Muh. Idris. (2009). STAIN/ IAIN Menuju UIN (Perspektif Pemikiran Pendidikan A. Malik Fadjar). Jurnal Iqra', Vol. 3 (1), 21–36.

Muslim, A. (2012). Ashobiyah Ibn Khaldun: Konsep Perubahan Sosial Di Indonesia. Sulesana, Vol. 7 (2), 138–148.

Mustopa. (2019, November 15). Menakar Kemampuan Baca Tulis Al-Qur'an Mahasiswa UIN. Diambil dari Lajnah Pentashihan Mushaf al-Quran, Badan Litbang dan Diklat Kementerian Agama Republik Indonesia website: https://lajnah.kemenag.go.id/berita/513-menakar-kemampuan-baca-tulis-al-qur-an-mahasiswa-uin

Retnawati, H. (2016). Validitas reliabilitas & karakteristik butir panduan untuk peneliti, mahasiswa, dan psikometrian. Yogyakarta: Nuha Medika.

Rohmah, M. (2017, April 15). Hasil Observasi [Panduan Observasi].

DOI: https://doi.org/10.28918/jei.v4i2.2296

Setiawan, A. (2016). Hermeneutika al-Quran "Mahzab Yogya": Telaah atas Teori Ma'na-Cum-Maghza dalam Penafsiran al-Quran. Jurnal Studi Ilmu-Ilmu al-Qur'an dan Hadis, Vol. 17 (No. 1), 69–96.

TIM 10 PKTQ. (2015). Pengembangan Kepribadian dan Tahsinul Quran. Yogyakarta: PKTQ FTIK UIN Sunan Kalijaga.

TIM 10 PKTQ. (2017). Petunjuk Teknis Ujian Sertifikasi PKTQ. PKTQ FTIK UIN Sunan Kalijaga.

Zuhaida, A. (2018). Penyusunan Instrumen Analisis Pedagogical Content Knowledge Guru IPA Madrasah Tsanawiyah Terintegrasi Konten Islami. Edukasia Islamika, 234–248. https://doi.org/10.28918/jei.v3i2.1690.