

P2M

Laporan Penelitian Individual

KUALITAS TES MASUK SEBAGAI PREDIKTOR KEBERHASILAN BELAJAR MAHASISWA

(STUDI TENTANG PENDEKATAN ANALISIS
TES MASUK BAHASA ARAB DAN BAHASA
INGGRIS SERTA AKURASINYA DALAM
MEMPREDIKSI KEBERHASILAN BELAJAR
MAHASISWA IAIN WALISONGO)

Prof.Dr.H. Ibnu Hadjar, M.Ed.

NIP. 19580507 198402 1002

Penelitian ini dibiayai dengan Anggaran DIPA IAIN Walisongo



**LEMBAGA PENELITIAN DAN PENGABDIAN
KEPADA MASYARAKAT**

**IAIN WALISONGO SEMARANG
TAHUN 2013**

**LAPORAN HASIL
PENELITIAN INDIVIDUAL**

**KUALITAS TES MASUK SEBAGAI PREDIKTOR
KEBERHASILAN BELAJAR MAHASISWA (STUDI TENTANG
PENDEKATAN ANALISIS TES MASUK BAHASA ARAB DAN
BAHASA INGGRIS SERTA AKURASINYA DALAM
MEMPREDIKSI KEBERHASILAN BELAJAR MAHASISWA
IAIN WALISONGO)**



Oleh:

Prof. Dr. H. Ibnu Hadjar, M. Ed
NIP. 19580507 198402 1002

**DIBIYAI DENGAN ANGGARAN DIPA
IAIN WALISONGO
TAHUN 2013**



**KEMENTERIAN AGAMA
INSTITUT AGAMA ISLAM NEGERI WALISONGO
LEMBAGA PENELITIAN DAN PENGABDIAN
KEPADA MASYARAKAT**

Jl. Walisongo No. 3-5 Telp./Fax.7615923 Semarang 50185

SURAT KETERANGAN

No. In.06.0/P.1/TL.01/ 243 /2013

Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat IAIN Walisongo Semarang, dengan ini menerangkan bahwa penelitian Individual yang berjudul:

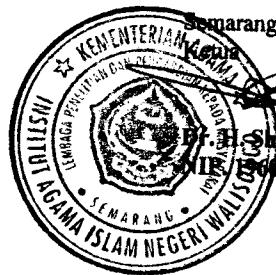
**KUALITAS TES MASUK SEBAGAI PREDIKTOR
KEBERHASILAN BELAJAR MAHASISWA (STUDI TENTANG
PENDEKATAN ANALISIS TES MASUK BAHASA ARAB DAN
BAHASA INGGRIS SERTA AKURASINYA DALAM
MEMPREDIKSI KEBERHASILAN BELAJAR MAHASISWA
IAIN WALISONGO)**

adalah benar-benar merupakan hasil penelitian yang dilaksanakan oleh:

Nama : Prof.Dr.H. Ibnu Hadjar, M.Ed.
NIP : 19580507 198402 1002
Pangkat/Jabatan: Pembina Utama Madya (IV/d) / Guru Besar
Pekerjaan : Dosen Fakultas Ilmu Tarbiyah dan Keguruan

Demikian surat keterangan ini kami buat untuk dipergunakan sebagaimana mestinya.

Semarang, 10 September 2013



[Signature]
Prof. H. Sholihan, M. Ag.
NIP. 19600604 199403 1004

KATA PENGANTAR

Bismi Allāhi al-Rahmāni al-Rahīmi. Dengan mengucapkan *Alhamdu li Allāh*, peneliti memanjatkan puji syukur ke hadirat Allah SWT. Atas berkah dan karunia-Nya, peneliti dapat menyelesaikan penulisan laporan penelitian tentang tes masuk dan hasil belajar mahasiswa di IAIN Walisongo ini sesuai dengan rencana.

Penelitian ini tidak akan pernah terselesaikan tanpa adanya bantuan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, peneliti merasa berkewajiban menyampaikan terima kasih kepada mereka. *Pertama*, peneliti mengucapkan terima kasih kepada Rektor IAIN Walisongo Semarang, Prof. Dr. H. Muhibbin, M.Ag., dan Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LP2M) beserta seluruh jajarannya atas kesempatan dan dukungan dana yang telah diberikan kepada peneliti.

Kepada Wakil Rektor I Bidang Akademik dan Pengembangan, Kepala Pusat Komputer dan Informasi, Kepala Bagian Akademik dan Kemahasiswaan beserta jajarannya juga peneliti ucapkan terima kasih atas bantuannya dalam mengumpulkan data yang diperlukan untuk penelitian ini. Tanpa bantuan mereka, penelitian ini tidak akan pernah terselesaikan. Kepada rekan-rekan sejawat yang ikut memberikan saran dan kritik terhadap draft awal dari laporan penelitian ini juga penulis sampaikan terima kasih.

Terakhir, kepada Istri tercinta, Dr. Umul Baroroh, M.Ag., serta ketiga ananda tersayang, Nabil Hajar, Nurman Hajar, dan Nadien Hajar peneliti sampaikan terima kasih pula atas pengorbanan waktu, tenaga, dan dorongannya yang tak pernah terputus dalam rangka penyelesaian studi dan penulisan disertasi ini. Rasanya tanpa mereka semua tersebut, peneliti tidak akan pernah mampu menyelesaikannya.

Akhirnya, bagaimanapun maksimal usaha yang telah peneliti lakukan, peneliti menyadari bahwa penelitian ini masih memerlukan kritik dan saran dari pembaca. Namun demikian, betapapun kecilnya, peneliti

berharap disertasi ini memberikan sumbangan dalam pengayaan literatur tentang keberagaman, sikap terhadap korupsi, dan perilaku menyontek di kalangan mahasiswa muslim. Peneliti juga berharap semoga disertasi ini bermanfaat bagi pembaca. Amîn.

Semarang, 10 September 2013

Peneliti,

Ibnu Hadjar

DAFTAR ISI

HALAMAN JUDUL	i
PENGESAHAN/KETERANGAN KETUA LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Masalah Penelitian	6
C. Signifikansi Penelitian	7
D. Sistematika Penulisan Laporan	8
BAB II TES MASUK DAN HASIL BELAJAR DI PERGURUAN TINGGI	11
A. Deskripsi Teori	11
1. Tes Masuk Perguruan Tinggi	12
2. Kualitas Tes: Validitas Tes sebagai Indikator	15
3. Teori Tes	17
4. Kemampuan Dasar dan Tes Seleksi	28
5. Hasil Belajar	31
B. Kajian Pustaka	36
BAB III METODE	39
A. Subjek Penelitian	39
B. Objek Penelitian	42
C. Variabel	44
D. Teknik Pengumpulan Data	45
	v

E. Teknik Analisis Data	47
BAB IV HASIL PENELITIAN	55
A. Kualitas Tes Masuk IAIN Walisongo 2012/2013	55
1. Deskripsi Umum	52
2. Kalibrasi Tes Masuk Bahasa Arab A	56
3. Kalibrasi Tes Masuk Bahasa Arab B	68
4. Kalibrasi Tes Masuk Bahasa Inggris A	76
5. Kalibrasi Tes Masuk Bahasa Inggris B	84
B. Uji Hipotesis: Tes Masuk sebagai Prediktor Hasil Belajar	90
C. Pembahasan	97
D. Keterbatasan Penelitian	100
BAB V PENUTUP	103
A. Kesimpulan	103
B. Saran	105
DAFTAR PUSTAKA	107
LAMPIRAN-LAMPIRAN	110

BABI PENDAHULUAN

A. Latar Belakang

Setiap menjelang tahun akademik baru, ribuan lulusan pendidikan menengah berebut kursi belajar di perguruan tinggi dengan cara mendaftarkan diri untuk bisa melanjutkan ke perguruan tinggi sesuai dengan minatnya. Karena keterbatasan daya tampung yang dimilikinya, sebagian perguruan tinggi tidak menerima semua pendaftar yang memenuhi persyaratan yang telah mereka tetapkan. Alih-alih, mereka melakukan seleksi untuk memilih sebagian pendaftar untuk diterima menjadi calon mahasiswa baru. Seleksi tersebut dilakukan dengan berbagai cara yang dianggap mampu memilih calon yang “baik,” sesuai dengan kriteria yang ditetapkan, baik secara akademis maupun non akademis. Seleksi dimaksudkan untuk menjaring calon mahasiswa yang diduga akan mampu menyelesaikan studi di perguruan tinggi dengan baik.

Berdasarkan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi, sistem penerimaan mahasiswa baru program sarjana pada perguruan tinggi dilakukan melalui seleksi secara nasional dan bentuk lain. Khusus untuk perguruan tinggi negeri, seleksi calon mahasiswa program sarjana pada seluruh perguruan tinggi negeri dilakukan oleh pemerintah secara nasional dan diikuti oleh peserta dari seluruh Indonesia dalam satu sistem yang terpadu.¹ Seleksi ini didasarkan pada hasil belajar calon saat belajar di sekolah menengah dan

¹Panitia SNMPTN 2013, *Informasi Umum SNMPTN 2013* (diakses dari: <http://www.snmptn.ac.id/informasi.html#pengantar>, pada 12 Oktober 2013).

prestasi lain, baik akademik maupun non akademik. Namun demikian, tes yang digunakan untuk seleksi “dirancang untuk mengukur kemampuan umum yang diduga menentukan keberhasilan calon mahasiswa di semua program studi, yakni kemampuan penalaran tingkat tinggi (higher order thinking), yang meliputi potensi akademik, penguasaan bidang studi dasar, bidang saintek dan/atau bidang social dan humaniora.”² Di samping itu, masing-masing perguruan tinggi negeri diberi kewenangan untuk melakukan seleksi sendiri dalam bentuk lain. Di antara bentuk seleksi tersebut ada yang dilakukan secara bersama secara nasional dan ada yang dilakukan secara sendiri oleh perguruan tinggi yang bersangkutan.

IAIN Walisongo sebagai salah satu perguruan tinggi negeri menggunakan berbagai sistem dalam melakukan seleksi calon mahasiswa baru yang akan diterimanya. Untuk program sarjana, sistem seleksi yang digunakan pada tahun akademik 2013/2014 adalah SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri), SBMPTAIN (Seleksi Bersama Masuk Perguruan Tinggi Agama Islam Negeri), SMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri), SMPTAIN (Seleksi Nasional Masuk Perguruan Tinggi Agama Islam Negeri), dan Ujian Masuk Mandiri. Hasil belajar dari mahasiswa yang diseleksi melalui sistem-sistem tersebut saat ini belum bisa diketahui karena baru memasuki semester I. Sedangkan untuk tahun

² Panitia SBMPTN 2013, *INFORMASI Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) Tahun 2013* (diakses dari: <http://www.sbmptn.or.id/utama.php?hmenu=3&hitem=1&dikirim=1>, pada 12 Oktober 2013).

akademik 2012/2013 sistem seleksi yang digunakan oleh IAIN Walisongo adalah SMPTAIN, UM Mandiri, dan PSSB. Kecuali PSSB, sistem seleksi tersebut menggunakan ujian/tes tertulis sebagai dasar pijakan untuk penentuan/pemilihan calon mahasiswa baru. Hasil belajar dari mahasiswa yang diseleksi dengan sistem tersebut saat ini (semester Genap 2012/2013) sudah dapat diketahui karena mereka telah belajar selama satu semester. Khusus hasil belajar bahasa asing (Arab dan Inggris), hasil belajar tersebut telah mencerminkan seluruh perkuliahan bahasa asing tersebut karena seluruhnya (I, II, dan III) telah diberikan tuntas pada semester gasal (meskipun mahasiswa baru mengikuti salah satu dari keduanya. Karena itu, hasil tes masuk yang mereka tempuh saat mengikuti seleksi calon mahasiswa baru dapat dihubungkan dengan hasil belajar mereka setelah menjadi mahasiswa selama satu semester.

Ujian/tes masuk merupakan kegiatan akademis yang setiap tahun dilakukan oleh perguruan tinggi, termasuk IAIN Walisongo, untuk menyeleksi calon mahasiswa baru yang akan diterimanya. Tes tersebut diharapkan dapat digunakan untuk menjaring calon mahasiswa yang potensial memiliki peluang keberhasilan belajar yang tinggi. Dengan demikian, tes tersebut berfungsi sebagai alat untuk memprediksi keberhasilan belajar calon mahasiswa.³ Dengan fungsi ini, diduga mereka yang mendapatkan skor tinggi dalam tes masuk juga akan mendapatkan hasil belajar yang baik, dan sebaliknya. Oleh karena itu,

³Sumadi Suryabrata, *Seleksi Calon Mahasiswa Baru Perguruan Tinggi: Yang Sekarang dan Kemungkinan untuk Masa Mendatang* (Yogyakarta: Andi Offset, 1989).

hanya mereka yang memperoleh skor tinggi pada tes masuk yang diterima sebagai calon mahasiswa. Setidaknya, bila peserta tes melebihi kuota jumlah calon yang akan diterima sebagai mahasiswa baru, maka seleksi didasarkan pada rangking skor yang diperoleh para peserta, dengan prioritas dari yang tertinggi.

Fungsi prediktif tes masuk tersebut akan terlaksana dengan baik manakala tes tersebut memiliki “kualitas” yang baik pula. Secara umum, kualitas sebuah tes dapat diukur melalui tingkat validitas dan reliabilitasnya. Validitas berkaitan dengan kemampuan tes untuk mengukur kemampuan atau kompetensi subjek berkenaan dengan penguasaan materi yang hendak diukur. Sedangkan reliabilitas berkenaan dengan keandalan tes tersebut untuk dapat menghasilkan ukuran yang stabil.⁴ Kedua kriteria kualitas tes tersebut harus merupakan satu kesatuann yang tidak dapat dipisahkan.

Lebih lanjut, karena sebuah tes terdiri dari butir-butir sebagai komponennya, kualitas tes tersebut sangat tergantung dan tidak dapat dilepaskan dari kualitas butir-butirnya. Oleh karena itu, untuk mengetahui kualitas tes tersebut juga perlu menguji kualitas butir-butir. Kualitas butir ditunjukkan oleh seberapa jauh butir mampu membedakan tingkat kemampuan testee (*daya beda*), tingkat kesulitan yang proporsional (*tingkat kesulitan*), dan berfungsinya opsi setiap butir untuk mengecoh peserta tes yang tidak memiliki kemampuan untuk menjawab dengan benar butir tersebut.⁵

⁴N.E. Grounlund, & R.L. Linn, *Measurement and Evaluation in Teaching*, (New York: Macmillan Pub. Co, 1985).

⁵Purwanto, *Evaluasi Hasil Belajar* (Yogyakarta : Pustaka Pelajar, 2009).

Tes ujian masuk mandiri IAIN Walisongo dibuat secara khusus untuk digunakan sebagai alat seleksi calon mahasiswa pada setiap kali tes. Adapun mata ujian yang dijadikan alat untuk seleksi pada ujian mandiri adalah Bahasa Indonesia, Bahasa Arab, Bahasa Inggris, Pengetahuan Agama Islam, dan Pengetahuan Umum. Seleksi didasarkan pada skor kumulatif yang dicapai oleh peserta dari semua mata tes. Meskipun tes yang digunakan sama, seleksi dilakukan untuk calon mahasiswa pada masing-masing program studi sesuai dengan kuota atau daya tampung yang tersedia.

Dalam kenyataan, setiap tahun dibuat tes yang berbeda, meskipun memiliki tujuan yang sama. Tidak ada upaya untuk membuat bank soal yang dapat digunakan secara ulang sewaktu-waktu dengan tingkat kepercayaan yang tinggi. Tes tersebut tidak pernah dianalisis secara sistematis sehingga tidak pernah diketahui kualitasnya. Karena itu, tes tersebut juga tidak dapat diketahui kemampuan prediktifnya untuk meramalkan keberhasilan mahasiswa.

Berdasarkan hal tersebut di atas kiranya tes yang digunakan untuk seleksi calon mahasiswa perlu dievaluasi untuk mengetahui seberapa jauh dapat memenuhi kriteria kualitas yang "standar" untuk mengukur "potensi" belajar calon mahasiswa. Buku ini merupakan laporan dari hasil penelitian yang dimaksudkan untuk menguji kualitas tes masuk tersebut, khususnya untuk mata tes dan matakuliah Bahasa Arab dan Bahasa Inggris. Pembatasan pada kedua mata tes tersebut karena keduanya berfungsi untuk mengukur kompetensi instrumental yang diperlukan dalam memperlancar penyelesaian studinya di IAIN Walisongo. Tidak seperti hasil mata tes yang lain, hasil tes keduanya

dijadikan standar kemampuan mahasiswa sehingga calon mahasiswa yang mencapai skor minimal 60 dibebaskan dari kewajiban mengikuti kuliah Bahasa Arab I dan Bahasa Inggris I apabila diterima dan kuliah di IAIN Walisongo.

Lebih lanjut, karena fungsi prediktifnya, evaluasi terhadap kualitas tes tersebut juga dilakukan terhadap kemampuan prediktifnya. Dalam hal ini, hasil tes masuk tersebut diuji kemampuannya untuk memprediksi hasil belajar kedua bahasa asing yang telah dicapai oleh mahasiswa setelah belajar di IAIN Walisongo. Dengan pengujian ini diharapkan kualitas tes tersebut dapat diketahui secara internal maupun eksternal.

B. Rumusan dan Pembatasan Masalah

Berdasarkan latar belakang masalah tersebut di atas, penelitian ini dimaksudkan untuk menguji kualitas tes masuk IAIN Walisongo, khususnya mata tes Bahasa Arab dan Bahasa Inggris. Secara rinci, masalah yang akan dicari jawabnya dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kualitas Tes Masuk Bahasa Arab dan Bahasa Inggris IAIN Walisongo ditinjau dari teori tes klasik dan modern?
2. Bagaimana kualitas Tes Masuk Bahasa Arab dan Bahasa Inggris IAIN Walisongo ditinjau dari kemampuan prediksinya untuk menentukan hasil belajar mahasiswa?

Dalam penelitian ini fokus masalah dibatasi pada tes masuk IAIN Walisongo tahun akademik 2012/2013 yang diselenggarakan melalui jalur tes tulis mandiri yang dilaksanakan sepenuhnya oleh IAIN Walisongo. Sementara itu, tes melalui jalur lain (SMPTAIN) tidak

diteliti karena datanya tidak tersedia di IAIN Walisongo, karena diselenggarakan secara nasional dan berbasis di kampus PTAIN lain. Sementara jalur PSSB tidak menggunakan tes masuk, tetapi menggunakan nilai raport yang diperoleh di Madrasah atau sekolah menengah.

Lebih lanjut, hasil belajar mahasiswa dibatasi hanya pada mata kuliah Bahasa Arab (I, II, dan III) dan Bahasa Inggris (I, II, dan III) yang ditempuh oleh mahasiswa angkatan 2012/2013 yang belajar pada semester Gasal 2012/2013. Pembatasan ini dilakukan karena materi kedua matakuliah tersebut terkait dengan materi tes masuk, baik secara sekuensial maupun substansial. Bahkan bagi calon mahasiswa yang telah memperoleh skor yang "baik" dalam seleksi tes masuk "diakui" telah lulus matakuliah tersebut sehingga tidak diwajibkan untuk mengambil matakuliah tersebut ketika lulus seleksi dan menjadi mahasiswa di IAIN Walisongo.

C. Signifikansi Penelitian

Hasil penelitian ini diharapkan memberikan masukan bagi IAIN Walisongo berkenaan dengan kualitas tes masuk yang digunakan untuk seleksi calon mahasiswa baru program sarjana. Informasi tentang kualitas tersebut dapat dijadikan pijakan dalam rangka memperbaiki sistem seleksi, sehingga dapat dirumuskan kebijakan yang betul-betul mampu menjaring calon mahasiswa yang berpotensi bagus untuk keberhasilan studinya di IAIN Walisongo.

Di samping itu, penelitian ini juga memberikan informasi teoritis tentang pengembangan tes, terutama yang terkait dengan validitas

prediktif berdasarkan teori tes yang digunakan untuk menganalisis butir-butirnya. Hasil penelitian ini berfungsi untuk membandingkan keandalan teori tes klasik dan modern dalam kaitan dengan seleksi butir dan keandalannya.

D. Sistematika Pelaporan Penelitian

Untuk memudahkan pembaca memahami penelitian ini, laporan secara diorganisasikan sistematis ke dalam lima bab. Bab 1, Pendahuluan, menyajikan latar belakang mengapa peneliti memilih permasalahan yang menjadi fokus penelitian ini, rumusan permasalahan, signifikansi penelitian secara teoretis maupun praktis, serta sistematika laporan.

Bab 2 menyajikan kerangka teoretis yang menjadi landasan konseptual penelitian. Bab ini diawali dengan kajian teoritis tentang kualitas tes, baik berdasarkan teori klasik maupun modern dan hasil belajar berdasarkan sumber-sumber pustaka yang relevan. Selanjutnya, bab ini menyajikan hasil kajian terhadap hasil-hasil penelitian terkait yang pernah dilakukan para peneliti di masa lalu.

Informasi tentang bagaimana penelitian ini dilakukan disajikan dalam Bab 3. Bab ini menyajikan metode yang digunakan untuk mengumpulkan data maupun analisisnya. Bab ini diawali dengan penyajian tujuan penelitian, dilanjutkan dengan pendekatan yang digunakan sebagai pijakan dan acuan dalam pelaksanaan penelitian. Selanjutnya, bab ini menyajikan sumber dari mana data diperoleh, termasuk penentuan sampel dan instrumen yang digunakan. Bab ini diakhiri dengan penyajian teknik analisis data yang digunakan setelah data terkumpul, sesuai dengan permasalahan penelitian. Dengan

demikian, diharapkan bab ini dapat dijadikan acuan untuk menguji validitas eksternal maupun internal dari hasil penelitian ini.

Bab 4 menyajikan hasil-hasil penelitian yang merupakan hasil dari analisis data serta pembahasannya. Bagian ini diharapkan dapat memberikan pemahaman yang komprehensif dari hasil penelitian.

Laporan ini diakhiri dengan Bab 5, Penutup, yang menyajikan kesimpulan dari temuan yang merupakan jawaban singkat dari permasalahan yang disajikan. adalah sebagai penutup. Berdasarkan kesimpulan tersebut bab ini menyajikan saran yang sebaiknya atau seharusnya dilakukan oleh pihak terkait, baik yang bersifat praktis (untuk keperluan kehidupan) maupun teoretis (untuk pengembangan ilmu lebih lanjut melalui penelitian lanjutan).

BAB II TES MASUK DAN HASIL BELAJAR DI PERGURUAN TINGGI

A. Deskripsi Teori

1. Tes Masuk Perguruan Tinggi

Tes merupakan “alat atau prosedur yang sistematis untuk mengukur sampel perilaku,”¹ yang dapat berupa apa yang dikatakan, dilakukan, atau diciptakan oleh testee.² Tes berisi pertanyaan yang jawabannya digunakan untuk menentukan seberapa jauh seseorang menunjukkan kinerjanya, baik dalam kaitannya dengan orang lain maupun dengan ranah tugas tertentu. Pada umumnya di perguruan tinggi tes digunakan untuk mengukur keberhasilan belajar maupun seleksi calon mahasiswa baru.

Berdasarkan karakteristik pengukurannya, tes masuk dapat dikategorikan sebagai pengukuran kinerja maksimum. Hal ini karena tes tersebut menggunakan prosedur untuk menentukan kemampuan individu, yakni seberapa jauh kinerja individu tersebut jika ia termotivasi untuk mencapai skor setinggi mungkin.³ Hasil pengukuran dengan tes ini menunjukkan seberapa jauh individu dapat mengerjakan jika ia mencurahkan seluruh upayanya.

Berdasarkan penggunaan hasil pengukurannya, tes masuk dapat dikategorikan sebagai *aptitude test*. Meskipun soal-soal yang

¹ N.E. Grounlund & R.L. Linn, *Measurement and Evaluation in Teaching* (New York: Macmillan Pub. Co., 1985), h. 5.

² R.L. Johnson, J.A. Penny, & B. Gordon *Assessing Performance: Designing, Scoring, and Validating Performance Tasks* (New York: The Guilford Press, 2009), h. 3.

³ Grounlund & Linn, *op. cit.*, h. 11.

dibuat berdasarkan materi yang dipelajari sebelumnya (tingkat pendidikan menengah dan dasar yang telah ditempuh peserta tes masuk), tes masuk dimaksudkan untuk memprediksi mengukur seberapa jauh hasil belajar

Dalam pendidikan, tes berfungsi untuk mengukur perilaku atau karakteristik psikologis tersembunyi yang dimiliki testee (subyek yang masih calon atau yang sudah menjadi peserta didik).⁴ Karena keterbatasan, tes hanya mengukur sampel perilaku tersembunyi. Karena hanya sampel, kesalahan hasil ukur tes dalam menggambarkan perilaku yang sebenarnya tidak dapat dihindari. Tingkat kesalahan tersebut sangat tergantung pada kualitas tes yang digunakan untuk mengukurnya.

Kualitas tes dapat dilihat dari validitasnya, yakni seberapa jauh tes mampu mengukur perilaku yang hendak diukur.⁵ Pengukuran validitas dapat dilakukan secara teoritis/konseptual dan secara empiris. Secara teoritis, validitas diuji dengan melihat seberapa jauh butir-butir tes telah sesuai dengan dimensi dan indikator konseptual perilaku yang hendak diukur. Uji ini dilakukan dengan cara penilaian ahli (*expert judgement*), yang ditunjukkan oleh langkah-langkah penyusunannya. Sedangkan secara empiris, validitas tes diukur dari seberapa jauh butir-butir tersebut didukung data empiris dalam mengukur objeknya, perilaku yang hendak diukur. Uji ini dilakukan

⁴ *Ibid.*

⁵ Anne Anastasi & Susana Urbina, *Psychological Testing*, (Upper Saddle River: Prentice Hall, 1997).

berdasarkan data empiris, yakni hasil tes yang dilakukan kepada subjek yang hendak diukur.

Validitas tes secara empiris dibedakan menjadi dua: internal dan eksternal. Validitas internal menguji seberapa jauh data empiris butir-butir tes sesuai dengan kriteria internal butir itu sendiri. Sedangkan validitas eksternal menguji seberapa jauh hasil tes sesuai dengan kriteria yang ada di luar tes, baik melalui uji perbedaan antar kelompok maupun hubungan dengan kriteria.

Uji validitas internal menekankan pada pengembangan tes sehingga dapat diperoleh butir-butir yang valid dengan kriteria butir-butir itu sendiri. Dengan uji ini diharapkan dapat diketahui apakah tes “berkualitas” secara empiris. Uji ini dapat dilakukan dengan menggunakan dua pendekatan atau teori tes. Teori tes memberikan kerangka untuk melihat proses pengembangan tes⁶ karena memfokuskan pada spesifikasi aturan untuk mengubah respon testee terhadap butir tes ke dalam perkiraan sifat yang diasumsikan mendasari respon yang teramati.

Ada dua pendekatan yang dikenal, yaitu: klasik (disebut *Classical Test Theory/CTT*) dan modern (disebut *Item Response Theory/IRT*). Keduanya memiliki tujuan yang sama, yaitu mengetes orang, dengan fokus menempatkan individu pada rentang dimensi tersembunyi tertentu.⁷ Dalam pendidikan, karakteristik tersembunyi/

⁶Linda Crocker & James Algina, *Introduction to Classical and Modern Test Theory* (Forth Worth: Holt, Rinehart and Winston, 1986).

⁷Partchev, Ivaelo, 2004, *A Visual Guide to Item Response Theory* (Jena: Friedrich Scheiler Universitat).

latent trait yang diukur adalah kemampuan testee atau *ability* dalam menjawab pertanyaan/merespon butir-butir tes.

Dalam teori klasik, uji tes dilakukan untuk memperoleh informasi tentang tingkat kesulitan dan validitas butir, sebagai kriteria internal. Tingkat kesulitan (p) merupakan proporsi peserta tes yang menjawab benar suatu butir, sehingga nilai akan merentang dari 0,0 – 1,0. Tingkat kesulitan merupakan fungsi dari tiga hal: kerumitan tugas, kemampuan testee, dan kualitas pembelajaran sebelumnya, yang secara bersama mempengaruhi persentase jawaban benar menjadi tinggi atau rendah. Berdasarkan hal ini, butir dan testee saling bergantung, di mana tingkat kemampuan testee selalu dikaitkan dengan kesulitan butir. Dengan kata lain, tingkat kesulitan butir ditentukan melalui jawaban testee, sementara tingkat kemampuan testee ditentukan melalui taraf kesulitan butir. Tingkat kesulitan tes akan rendah bila peserta tes memiliki kemampuan rendah, dan akan tinggi bila kemampuan testee tinggi. Demikian juga, kemampuan testee akan rendah bila tingkat kesulitan tes rendah, dan akan tinggi bila tingkat kesulitan tes juga tinggi.

Sedangkan validitas butir mengukur unidimensionalitas kemampuan testee, dengan melihat hubungan antara skor masing-masing butir dengan skor keseluruhan butir.⁸ Validitas butir juga menunjukkan kemampuan butir untuk membedakan kemampuan testee, sehingga juga disebut daya beda. Karena secara teknis menggunakan teknik statistik korelasi, indeks validitas butir ini merentang dari -1,0 – 1,0. Hanya butir yang memiliki nilai positif

⁸ Naga, *loc. Cit.*

sedang atau tinggi yang dianggap valid karena menunjukkan adanya kebersamaan dengan butir yang lain dalam memberikan informasi tentang karakteristik testee.

Dalam teori tes modern, pengembangan tes didasarkan pada dua asumsi, unidimensional (hanya ada satu kemampuan yang diukur oleh butir yang membentuk tes) dan fungsi karakteristik butir mencerminkan hubungan antara variabel tak teramati (kemampuan) dan teramati (respon butir). Fungsi karakteristik butir merupakan ekspresi matematis yang menghubungkan peluang menjawab benar suatu butir dan karakteristik butir⁹. Karakteristik kemampuan testee dilambangkan dengan θ (θ), yang nilainya terstandar, dengan rerata = 0 dan simpang baku = 1. Sedangkan karakteristik kesulitan butir dilambangkan dengan b , yang sebagaimana kemampuan testee, nilainya terstandar, dengan rerata = 0 dan simpang baku = 1. Dengan karakteristik yang demikian, tingkat kemampuan dapat dihadapkan dengan tingkat kesulitan butir, tanpa terpengaruh/mempengaruhi satu sama lain.

2. Kualitas Tes: Validitas Tes sebagai Indikator

Dalam bidang pendidikan, tes merupakan bagian yang sangat penting karena tidak bisa dipisahkan dari seluruh proses. Tes adalah alat atau prosedur yang sistematis untuk mengukur suatu sampel perilaku.¹⁰ Pengukuran tersebut merupakan cara untuk memberikan angka, sesuai dengan aturan, sebagai representasi kuantitatif bagi karakteristik sesuatu. Pada dasarnya tujuan utama pembuatan tes

⁹ Crocker & Algina, *Op.cit.*

¹⁰ *Ibid.*, h. 5.

adalah untuk mengukur sifat atau karakteristik psikologis yang tersembunyi pada obyek, yang berkaitan dengan aspek kognitif, afektif, dan psikomotor. Sebagai obyek pengukuran dalam pendidikan adalah siswa, mahasiswa, guru, dosen, dan lain-lain. Tes sebagai alat ukur di bidang pendidikan merupakan sejumlah pertanyaan yang memiliki jawaban benar atau salah sehingga ia termasuk kategori instrumen performa/kinerja maksimum (*maximum performance test*).¹¹

Untuk mendapatkan informasi yang tepat mengenai karakteristik obyek pengukuran tersebut diperlukan alat ukur/tes yang baik, yakni yang mampu menghasilkan informasi yang mengandung kesalahan sekecil mungkin. Karena itu, untuk mendapatkan tes yang baik, alat ukur tersebut perlu dilihat mutu/kualitasnya. Semakin tinggi kualitas suatu tes, semakin kecil kesalahan yang dikandung oleh tes tersebut, dan sebaliknya.

Kriteria yang paling penting dalam kaitannya dengan kualitas tes sebagai alat ukur adalah *validitas*, yakni seberapa jauh tes tersebut mampu mengukur apa yang hendak diukur.¹² Validitas memeriksa secara langsung seberapa baik tes tersebut memenuhi fungsinya. Validitas dapat diuji secara *teoritis*, dengan melihat seberapa jauh butir-butir tes sesuai dengan dimensi atau indikator konseptual, atau secara *empiris*, dengan melihat seberapa jauh kesesuaian butir-butir

¹¹ Howard B. Lyman, *Test Score and What They Mean* (Englewood Cliffs: Prentice-Hall, 1978), h.18-20.

¹² Anne Anastasi & Susana Urbina, *Psychological Testing* (Upper Saddle River: Prentice Hall, ed. Ke-7, 1997), h. 9.

tes untuk mengukur objeknya didukung oleh data empiris. Validitas empiris dibedakan menjadi dua: *eksternal* dan *internal*. Dalam validitas eksternal, butir-butir tes di uji kesesuaiannya dengan kriteria yang ada di luar instrumen itu sendiri, melalui uji perbedaan kelompok atau hubungan dengan kriteria. Sedangkan didasarkan pada data empiris dari butir-butir tes itu sendiri, tanpa dikaitkan dengan kriteria lain.

Dalam penelitian yang diusulkan ini, uji validitas tes akan difokuskan pada validitas internal. Hal ini karena uji tersebut menekan pada pengembangan tes sehingga dapat diperoleh butir-butir tes yang valid dengan kriteria butir-butir tes itu sendiri. Dengan uji tersebut diharapkan dapat diperoleh tes yang “berkwalitas,” terutama sebagai alat prediksi keberhasilan belajar mahasiswa di IAIN Walisongo Semarang.

3. Teori Tes

Sebagaimana disebutkan diatas, tes merupakan alat untuk mengukur karakteristik tersembunyi yang dimiliki yang dimiliki testee. Pengukuran merupakan proses pemberian angka sebagai deskripsi dari tingkat karakteristik yang dimiliki oleh individu atau testee.¹³ Aturan-aturan tentang penetapan korespondensi antara angka dan jumlah atau besarnya karakteristik tersebut telah menjadi konsen dari teori pengukuran atau tes. Teori tes memfokuskan perhatiannya pada spesifikasi aturan-aturan untuk mengubah respon seseorang terhadap suatu butir tes ke dalam perkiraan sifat yang diasumsikan mendasari respon yang teramati tersebut. Karena itu, teori tes

¹³ Grounlund & Linn, *Op. Cit.*, h. 5.

memberikan kerangka umum untuk melihat proses pengembangan tes.¹⁴ Dalam teori tes dikenal ada dua macam pendekatan yaitu atau teori: *klasik* dan *modern* (teori kedua ini dikenal dengan sebutan *Item Response Theory*/teori respon butir).

Dalam pengembangan butir tes, kedua pendekatan (teori tes) menggunakan langkah-langkah pembuatan tes yang sama. Keduanya memerlukan suatu kelompok butir yang lebih besar jumlahnya dari pada yang diperlukan pada versi terakhir. Hal ini diperlukan untuk penghitungan statistik butir yang digunakan untuk memilih dan mengevaluasi bentuk alternatif dari tes versi terakhir. Meskipun demikian, kedua teori memiliki asumsi yang berbeda tentang tes. Dalam teori klasik, peneliti/pengembang tes berangkat dari satu atau lebih konstruk yang harus ditentukan dan dapat di ukur. Karena itu, butir-butir soal diasumsikan mengukur konstruk tunggal yang didukung oleh teori substantif. Sedangkan dalam teori modern diasumsikan bahwa butir-butir tes ditulis untuk mengukur sifat atau karakteristik tunggal yang tersembunyi.¹⁵ Lebih lanjut, perbedaan antara keduanya, sebagai konsekuensi, juga terletak pada pengolahan hasil tes dan interpretasinya.

Bagaimanakah aplikasi kedua teori tes, klasik dan modern, untuk mengembangkan atau menguji suatu tes secara empiris? Penelitian yang diusulkan ini dimaksudkan untuk membandingkan dan menyajikan hasil analisis tes ujian masuk IAIN Walisongo

¹⁴ Croker & Algina, *Op. Cit.*, h. 7.

¹⁵ Charles L. Hulin, Fritz Drasgow, & Charles K. Parsons, *Item Response Theory: Application to Psychological Measurement* (Homewood: Dow Jones-Irwin, 1983), h. 75-79.

Semarang dengan menggunakan kedua teori tes tersebut. Analisis tersebut didasarkan pada tes yang sama dan data yang sama pula. Sebagai dasar berfikir dalam analisis tersebut, terlebih dahulu akan dibahas tentang kedua teori tes secara terpisah. Pembahasan akan difokuskan pada konsep tingkat kesulitan, kemampuan testee, dan interdependensi testee-butir. Dari pembahasan ketiga aspek tersebut diharapkan dapat diketahui perbedaan konseptual antara keduanya.

a. Pendekatan Klasik

Dalam pendekatan klasik, hasil tes dianalisis dengan menggunakan dua teknik statistik untuk memperoleh informasi tentang *tingkat kesulitas butir* dan *validitas butir*, yang keduanya mencerminkan perbandingan dan kriteria internal. Tingkat kesulitan butir adalah proporsi testee yang berhasil menjawab benar suatu butir.¹⁶ Indeks tingkat kesulitan tersebut dilambangkan dengan p , yang dapat merentang antara 0,0, yang menunjukkan tingkat kesulitan yang sangat rendah (dimana tak seorang testee pun yang mampu menjawab dengan benar), sampai 1,0, yang menunjukkan tingkat kesulitan yang sangat tinggi (dimana semua testee dapat menjawab dengan benar). Nilai antara kedua ujung tersebut menunjukkan bahwa sebagian testee ada yang menjawab butir yang bersangkutan dengan benar dan sebagian yang lain menjawab dengan salah. Pada dasarnya, semakin tinggi nilai p (semakin besar proporsi testee yang menjawab dengan benar), semakin mudah butir tersebut bagi kelompok testee, dan sebaliknya. Meskipun secara teknis

¹⁶Naga, *Op. Cit.*, h. 48.

merupakan presentase jawaban benar dari seluruh testee, tingkat kesulitan butir merupakan persentase jawaban benar dari seluruh testee, tingkat kesulitan butir merupakan fungsi dari tiga hal: kerumitan tugas, kemampuan testee, dan kualitas pengajaran yang mendahuluinya.¹⁷ Kombinasi dari ketiga faktor tersebut mempengaruhi persentase jawaban benar menjadi tinggi atau rendah.

Sedang validitas butir merupakan uji tes untuk mengukur unidimensionalitas kemampuan testee. Karena tes mengukur hal yang sama, maka seharusnya kedudukan relatif testee pada setiap butir tes adalah sama. Testee yang memiliki kemampuan relatif tinggi pada suatu butir seharusnya juga memiliki kemampuan relatif tinggi pula pada semua butir. Dengan kata lain, validitas ini melihat hubungan antara skor masing-masing butir dengan skor keseluruhan butir.¹⁸ Karena itu, secara teknis *indeks validitas butir* merupakan *koefisien korelasi* antara skor suatu butir dan skor total, kumulatif dari seluruh butir. Pada dasarnya indeks tersebut menunjukkan apakah variasi skor butir yang diperoleh konsisten dengan variasi skor total. Dengan kata lain, validitas ini menunjukkan sumbangan butir untuk menerangkan variasi skor total. Indeks atau nilai koefisien tersebut juga sebagai petunjuk tentang kemampuan butir yang bersangkutan untuk menunjukkan daya beda kemampuan testee. Semakin tinggi nilai koefisien, semakin tinggi kemampuan daya bedanya, dan sebaliknya.

¹⁷ Anthony J. Nitko, *Educational Assesment of Students* (Englewood Cliffs: Prentice Hall, 1996), h. 120.

¹⁸ Naga, *op. cit.* h. 75-76.

Dengan demikian, dapat disimpulkan bahwa dalam pendekatan klasik butir dan testee saling tergantung, di mana kemampuan testee selalu dikaitkan dengan alat ukur/butir. Hal ini karena taraf kesulitan butir ditentukan melalui jawaban testee, sehingga bila testee memiliki kemampuan tinggi, butir menjadi mudah, dan sebaliknya. Demikian juga, kemampuan testee ditentukan melalui taraf kesulitan butir, sehingga bila butir sukar, kemampuan testee menjadi rendah.

Teori tes klasik memiliki beberapa keunggulan.¹⁹ Pertama, teori ini memiliki konsep yang mudah dipahami, terutama bila dibandingkan dengan teori modern, karena kesederhanaannya. Kedua, teori ini tidak memerlukan sampel testee yang besar karena secara statistik sudah dianggap cukup dengan melibatkan minimum 30 orang testee sehingga tidak banyak memerlukan biaya administrasi tes dan pengolahan datanya. Ketiga, karena sederhananya konsep yang digunakan, maka analisis dengan teori ini dapat dilakukan dengan hanya menggunakan bantuan kalkulator sederhana, bahkan dengan perhitungan manual. Karena sifat-sifat keunggulan tersebut, maka analisis dengan teori tes klasik ini telah banyak digunakan oleh para peneliti dan praktisi pendidikan, baik dalam skala kecil untuk kelas, maupun skala besar untuk sampel yang besar.

¹⁹ Djemari Murdapi, *Analisis Butir Dengan Teori Klasik dan Teori Respon Butir* (Yogyakarta: IKIP Yogyakarta, 1994), h. 4.

Namun demikian, teori tes klasik ini juga memiliki beberapa kelemahan yang mendasar.²⁰ Pertama, dalam teori ini karakteristik peserta tes (kemampuan) dan karakteristik butir tidak dapat dipisahkan karena masing-masing hanya dapat ditafsirkan dalam konteks yang lainya. Dalam hal ini tingkat kemampuan testee tergantung pada kesukaran butir. Pada butir yang sulit kemampuan testee menjadi rendah, sedang pada butir yang mudah kemampuan testee menjadi tinggi. Begitu juga, tingkat kesukaran butir tergantung pada kemampuan testee. Bila testee berkemampuan tinggi, maka butir menjadi mudah, sedang bila testee berkemampuan rendah, maka butir menjadi sulit. Dengan kata lain, tes tersebut tergantung pada kelompok (*group dependent*). Hal ini membawa konsekuensi pada kesulitan untuk membandingkan kemampuan testee yang mengambil dua tes yang berbeda, karena keduanya memiliki skala yang berbeda dan tidak ada hubungan fungsional. Ketiga, konsep reliabilitas yang mendasarkan pada galat baku pengukuran (*standard error of measurement*) merupakan kelemahan lain karena berlaku secara umum untuk semua testee dan tidak membedakan variasi kemampuan testee. Keempat, teori tersebut terlalu berorientasi pada tes secara keseluruhan, bukan pada butir sehingga tidak ada dasar untuk memperkirakan kemampuan testee dalam mengerjakan suatu butir tertentu. Kemampuan testee diukur dengan keseluruhan butir secara kumulatif.

²⁰ Ronald K. Hambleton, H. Swaminathan, dan H. Jane Rogers, *Fundamentals of Item Response Theory* (Newbury Park: Sage Publication, 1991), h. 2-5.

Karena kelemahan-kelemahan yang terdapat dalam teori klasik tersebut, para ahli pengukuran telah mencari alternatif pemecahan, dengan mengemukakan teori yang memasukkan karakteristik butir yang tidak tergantung pada kelompok, skor yang mendeskripsikan kemampuan testee pada tingkat butir dari pada tingkat tes secara keseluruhan, model yang tidak mensyaratkan tes paralel untuk mengukur reliabilitasnya, dan model yang memberikan pengukuran yang akurat untuk masing-masing skor kemampuan.²¹ Semua karakteristik ini dapat terwujud dalam teori alternatif, yaitu teori tes modern atau *item response theory*.

b. Pendekatan Item Response Theory

Dalam rangka mengatasi kelemahan teori klasik, para ahli pengukuran mengembangkan teori baru, yang dikenal dengan sebutan Teori Respon Butir atau *Item Response Theory*.²² Salah satu karakteristik penting dari teori ini adalah orientasinya pada tingkat butir, yakni membuat pernyataan yang jelas mengenai hubungan antara kemungkinan menjawab suatu butir dengan benar dan kemampuan siswa atau tingkat pencapaian.²³

Sebagai mana dalam pendekatan klasik, dalam pendekatan modern pengembangan skala tes didasarkan pada kriteria internal.

²¹ *Ibid.*, h. 5.

²² Isaac I. Bejar, "Introducing to Item Response Models and Their Assumption," dalam Ronald K. Hambleton (Ed.) *Application of Item Response Theory* (Burnaby: Educational Institut of Britnish Columbia, 1983: 1-23), h.1.

²³ *Ibid.*, h. 3.

Pendekatan ini didasarkan pada tiga asumsi.²⁴ *Pertama*, butir-butir tes hanya mengukur satu sifat/karakteristik bersama. Asumsi ini juga disebut *unidimensionalitas*.²⁵ *Kedua*, regresi skor butir pada skor sifat tersembunyi mengikuti pola umum fungsi logistik. Dengan kata lain, fungsi karakteristik butir mencerminkan hubungan yang sebenarnya antara variabel yang tak-teramati (kemampuan) dan variabel yang teramati (respon butir).²⁶ *Ketiga*, fungsi logistik tertentu menentukan *Kurva Karakteristik Butir* atau *Item Characteristics Curve* bagi suatu butir, yang ditentukan oleh satu, dua, atau tiga parameter. Analisis ini melibatkan penggunaan Kurva Karakteristik Butir, yang menggambarkan peluang menjawab benar suatu butir sebagai fungsi dari karakteristik tersembunyi (disimbolkan dengan θ) yang mendasari kinerja terhadap butir tes. Kurva tersebut (di asumsikan berbentuk S) memperlihatkan bahwa semakin meningkat karakteristik tersembunyi, semakin meningkat peluang untuk menjawab dengan benar suatu butir sehingga ia menunjukkan bagaimana peluang menjawab dengan benar tergantung pada karakteristik tersembunyi.²⁷

²⁴ Rose E. Traub, "A Priori Consideration in Choosing an Item Response Model" dalam Roland K. Hambleton (Ed.), *Application of Item Response Theory* (Burnaby: Educational Research Institut of British Columbia, 1983: 57-70), h. 58.

²⁵ Roderick P. McDonald, *Test Theory: A Unified Treatment* (Mahwah: Lawrence Erlbaum Association, 1999), h. 167.

²⁶ Hulin, dkk., *Op. Cit.*, h. 14

²⁷ Linda Crocker & James Algina, *Introduction to Classical and Modern Test*, h. 340-341.

Fungsi karakteristik butir merupakan ekspresi matematis yang menghubungkan peluang menjawab dengan benar suatu butir dengan kemampuan yang diukur dengan tes tersebut dan karakteristik butir. Kemampuan yang merupakan karakteristik tersebut tersembunyi dari seseorang, yang diperkirakan dari respon orang tersebut terhadap butir tes, dilambangkan dengan theta (θ), yang mengacu pada nilai yang digeneralisasikan pada kontinum sifat laten.²⁸ Theta (θ) merupakan nilai standard dengan mean = 0 dan Standard Deviasi = 1.

Sesuai dengan jumlah karakteristik butir yang menjadi konsen, ada tiga model dalam Teori Respon Butir yang paling dikenal, yaitu: *model logistik satu, dua, dan tiga parameter*. Model logistik satu parameter²⁹ mengasumsikan bahwa hanya ada satu karakteristik tersembunyi dari butir yang mempengaruhi kinerja testee. Dengan kata lain, kemungkinan suatu jawaban yang benar terhadap suatu butir merupakan fungsi dari parameter yang mewakili tingkat kesulitan butir dan yang mewakili kemampuan testee.³⁰ Yakni tingkat kesulitan butir (dilambangkan dengan parameter b). parameter b untuk suatu butir merupakan suatu titik pada skala kemampuan di mana peluang menjawab benar adalah 0,5 (50 persen). Ia merupakan lokasi parameter, yang menunjukkan posisi Kurva Karakteristik Butir dalam hubungannya dengan skala

²⁸ Hulin, dkk., *Op. Cit.*, h. 15.

²⁹ Sebagai penghormatan terhadap pengembangannya model ini juga dikenal dengan model Rasch. Lihat Ja-Eric Gustafsson, *An Introduction to Rasch's Measurement Model* (Princeton: Educational Testing Service, 1981), h.6.

³⁰ *Ibid.*, h. 7.

kemampuan. Dalam kemampuan suatu kelompok testee yang dtransformasikan sehingga nilai mean 0 dan standrad deviasi 1, nilai b akan terletak bervariasi di sekitar 0 (biasanya antara -2.0 sampai +2.0). semakin besar nilai b , semakin sulit butir yang bersangkutan.³¹

Untuk model logistik dua parameter, di samping parameter tingkat kesulitan juga memasukan daya beda butir (dilambangkan dengan a). parameter ini merupakan proporsional dari letuk (*slope*) kurva karakteristik butir pada titik b pada skala kemampuan. Nilai parameter ini biasanya berkisar antara 0 sampai 2, meskipun dimungkinkan lebih besar dari 2. Semakin tinggi nilai a , semakin curam fungsi karakteristik, sehingga semakin besar daya bedanya sejalan dengan fungsi kemampuan. Dengan kata lain, parameter ini menunjukkan peluang menjawab dengan benar sejalan dengan perubahan kemampuan testee.

Terakhir, model logistik tiga parameter melibatkan parameter *tingkat peluang palsu* (*psedeo-chance-level*) atau *peluang menebak*, dilambangkan dengan c , di samping kedua parameter yang dibahas sebelumnya. Hal ini terjadi karena soal pilihan ganda atau benar-salah memiliki format yang memungkinkan respon melalui tebakan. Parameter ini memberikan (kemungkinan) nilai lebih besar dari nol pada asimptot bawah dari suatu Kurva Karakteristik Butir. Nilai tersebut merepresentasikan peluang peserta ujian dengan kemampuan rendah menjawab butir tersebut dengan benar karena

³¹ Hambleton, dkk., *Op. Cit.*, h. 14
26

adanya faktor tebakan dalam kinerja tes pada respon terpilih dari butir tes tersebut. Nilai c merupakan proporsi dari peluang menjawab dengan benar suatu butir sejalan dengan menurunnya kemampuan testee. Semakin tinggi nilai c , semakin besar peluang menjawab butir tersebut dengan benar, bahkan bagi setiap nilai θ yang sangat rendah sekalipun.

Dari uraian diatas dapat disimpulkan bahwa pendekatan tes modern atau *Item Response Theory* berusaha mengatasi kelemahan teori klasik. Dengan pendekatan ini, tingkat kemampuan testee dan tingkat kesulitan butir tidak saling tergantung. Tingkat kesulitan butir tidak dipengaruhi oleh tingkat kemampuan testee, dan sebaliknya, tingkat kemampuan testee tidak dipengaruhi oleh tingkat kesulitan butir. Dengan kata lain, meskipun tingkat kesulitan butir berbeda, kemampuan testee tetap. Begitu juga, meskipun tingkat kemampuan testee bervariasi, tingkat kesulitan butir tetap. Perbedaan bagi testee yang tingkat kemampuannya berbeda adalah peluang untuk menjawab dengan benar butir tertentu. Bagi testee yang memiliki kemampuan tinggi memiliki peluang yang lebih besar dari pada testee yang memiliki kemampuan yang lebih rendah untuk menjawab suatu butir dengan tingkat kesulitan tertentu.

Dalam penelitian ini, model yang dipilih adalah model satu parameter atau Model Rasch. Dengan menyadari beberapa kelemahan yang dimilikinya,³² pemilihan model ini karena

³² Traub, *Op. Cit.*, h. 63-64.

merupakan model yang paling sederhana, karena hanya melibatkan suatu karakteristik butir saja dan paling banyak digunakan.³³ Model ini mengasumsikan bahwa karakteristik tersembunyi untuk sebuah butir tes yang mempengaruhi peluang testee menjawab dengan benar adalah tingkat kesulitannya.

4. Kemampuan Dasar dan Tes Seleksi

Sebagaimana disebutkan sebelumnya, hasil belajar sangat dipengaruhi oleh faktor internal, yang di antaranya berupa pengetahuan faktual, kemampuan intelektual, dan strategi belajar, yang telah dimiliki pebelajar melalui pengalaman sebelumnya. Karena itu, hasil belajar tidak bisa dipisahkan dari pengalaman sebelumnya. Di samping itu, adanya organisasi isi kurikulum yang vertikal juga menuntut penguasaan awal (*entry behavior*) sebelum mengalami proses pembelajaran tertentu.

Kemampuan awal yang diperlukan sebagai dasar untuk mempelajari suatu tingkat materi prasyarat atau *prerequisite*. Prasyarat mengidentifikasi kualitas yang diperlukan oleh pebelajar untuk mencapai tujuan kurikulum atau untuk diaplikasikan dalam proses pembelajaran. Prasyarat tersebut diperlukan dengan asumsi bahwa ia merupakan bagian yang mendahului suatu materi dari suatu rangkaian yang sifatnya sekuensial. Karena itu, prasyarat tersebut besar sekali pengaruhnya pada keberhasilan belajar.

³³ Mardapi, *Op. Cit.*, h. 25.

Menurut Pratt,³⁴ ada tiga macam prasyarat yang diperlukan dalam rangka mencapai tujuan kurikulum tertentu. Pertama, prasyarat kognitif, yaitu pengetahuan, ketrampilan, atau kemampuan lain yang harus dikuasai atau dimiliki oleh pebelajar agar dapat mencapai tujuan pembelajaran tertentu. Prasyarat ini harus dikuasai pebelajar sebelum ia memulai belajar materi yang memerlukan prasyarat tersebut. Kedua, prasyarat biologis, yakni yang berkaitan dengan keadaan fisik, seperti umur, kesehatan, dan sebagainya. Ketiga, prasyarat akademik, yang berkaitan dengan program akademik yang harus dimiliki sebelum mengikuti suatu program akademik tertentu. Ia biasanya terkait dengan tingkatan yang diasumsikan isinya bersifat sekuensial.

Dalam suatu pembelajaran, informasi awal yang mencerminkan penguasaan prasyarat sangat diperlukan dalam rangka merencanakan dan menyesuaikan input dengan kurikulum.³⁵ Informasi tersebut biasanya diketahui melalui tes awal yang di desain untuk menentukan apakah pebelajar telah memenuhi kemampuan, pengetahuan, dan ketrampilan yang diperlukan untuk mencapai tujuan yang di inginkan. Sebagaimana tes lain, tes awal merupakan alat atau prosedur yang sistematis untuk mengukur suatu sampel perilaku³⁶ yang dimiliki pebelajar sebelum mengikuti proses. Tes sebagai alat ukur di bidang pendidikan merupakan sejumlah pertanyaan yang memiliki jawaban benar atau salah sehingga ia

³⁴ David Pratt, *Curriculum : Design and development* (San Diego : Harcourt Brace Jovanovich, 1980).

³⁵ Wile & Bondi, *Ibid.*

³⁶ Grounlund & Linn, *Measurement and Evaluation in Teaching*, h. 5.

termasuk kategori instrumen performa maksimum (*maximum performance*).³⁷

Dalam konteks proses pembelajaran di IAIN Walisongo secara umum, informasi awal tersebut diperoleh melalui tes ujian masuk calon mahasiswa baru. Atas dasar hasil tes awal dapat ditentukan/diperkirakan apakah pebelajar dapat mengikuti suatu program pembelajaran yang ditetapkan atau tidak. Mereka yang mencapai skor yang tinggi dianggap lebih siap mengikuti pembelajaran dari pada mereka yang mendapatkan skor rendah. Karena itu, bila harus memilih karena pendaftar melebihi kapasitas daya tampung, maka Institut akan memilih calon yang memiliki nilai tinggi pada tes masuk.

Untuk mendapatkan informasi yang tepat mengenai karakteristik obyek pengukuran tersebut diperlukan alat ukur/tes yang baik, yakni yang mampu menghasilkan informasi yang mengandung kesalahan sekecil mungkin. Karena itu, mendapatkan tes yang baik, alat ukur tersebut perlu dilihat mutunya. Kriteria yang paling penting dalam kaitanya dengan kualitas tes sebagai alat ukur adalah *validitas*, yakni seberapa jauh tes tersebut mampu mengukur apa yang hendak di ukur.³⁸ Validitas memeriksa secara langsung seberapa tes tersebut memenuhi fungsinya. Validitas tes dapat diuji secara *teoritis*, melihat seberapa jauh butir-butir tes sesuai dengan dimensi atau indikator konseptual, atau secara *empiris*, melihat

³⁷Howard B. Lyman, *Test Score and What They Mean* (Englewood Cliffs: Prentice-Hall, 1979), h. 18-20.

³⁸Anne Anatasi & Susana Urbina, *Psychological Testing* (Upper Saddle River: Prentice-Hall, ed ke-7, 1997), h. 9.

seberapa jauh kesesuaian butir-butir tes untuk mengukur objeknya didukung oleh data empiris. Validitas empiris dibedakan menjadi dua: *eksternal* dan *internal*. dalam validitas eksternal, butir-butir tes diuji kesesuaiannya dengan kriteria yang ada di luar instrumen itu sendiri, melalui perbedaan kelompok atau hubungan dengan kriteria. Sedangkan dalam validitas internal, pemeriksaan validitas didasarkan pada data empiris dari butir-butir tes itu sendiri, tanpa dikaitkan dengan kriteria lain.

Dalam penelitian ini, hasil tes masuk diperlukan untuk memprediksi keberhasilan belajar mahasiswa. Karena itu, kriteria validitas eksternalnya adalah hasil belajar mereka. Namun demikian, karena hasil tes masuk tersebut dapat diperlakukan secara beragam, maka keragaman tersebut diperkirakan juga mempengaruhi validitas eksternalnya. Penelitian ini didisain untuk mengungkap pengaruh dari beragam perlakuan terhadap (hasil) tes tersebut dalam memprediksi keberhasilan belajar mahasiswa, sebagai bukti validitas eksternalnya. Dengan cara ini diharapkan dapat diperoleh ragam yang paling valid (memiliki efektifitas yang tinggi dalam memprediksi kriteria [hasil belajar]).

5. Hasil Belajar

Tujuan utama pembelajaran adalah menciptakan siswa dapat belajar. Belajar merupakan proses perubahan perilaku yang terjadi pada diri siswa. Proses perubahan tersebut terjadi karena adanya interaksi antara kekuatan internal (kesadaran atau kognisi) dan kekuatan eksternal yang berupa lingkungan, tantangan, kesempatan).

Proses perubahan tersebut meliputi struktur perseptual kognitif, struktural penilaian moral dan kemauan, dan pola motorik untuk menghadapi kondisi obyektif.

Perubahan tersebut tidak dapat di amati secara langsung, tetapi harus melalui penampilan dalam wujud pengetahuan, sikap, dan ketrampilan yang dimiliki siswa.³⁹ Ia diketahui melaluiobservasi terhadap perilaku siswa, yang bersifat; permanen.⁴⁰ Perilaku sebagai akibat dari proses tersebut seringkali juga di sebut hasil belajar atau *learning outcomes*, yang dapat di kenali melalui kinerja atau *performace* siswa. Kinerja tersebut secara operasional mencakup tindakan, perasaan, dan pikiran, yang diharapkan berkembang pada diri siwa sebagai hasil dari proses belajar.⁴¹

Kinerja siswa, sebagai petunjuk tentang hasil belajar, sangat beragam, yang dapat dikategorikan berdasarkan ciri-ciri yang dimilikinya secara umum. Romiszowski membedakan hasil belajar menjadi dua: pengetahuan dan ketrampilan.⁴² Pengetahuan berkenaan dengan informasi yang tersimpan dalam otak pembelajar setelah mengalami proses belajar. Pengetahuan tersebut dibedakan menjadi empat, yaitu: fakta, prosedur, konsep, dan prinsip.

³⁹Robert M. Gagné, *The Condition of Learning* (New York: Holt Rinehlt, & Wins ton, 1997)

⁴⁰ Robert M. Gagné, dan Marcy P. Driscoll, *Essential of Learning for Intruction* (Englewood Criffs, NJ: Prentice hall, 1988), h. 2-3.

⁴¹ David R. Krathwohl, Benjamin S. Bloom, dan Bertram B. Masia, *Taxonomy of Educational Obyectives,Book II: Affective Domain* (London: Longman Group, 1973), h. 5.

⁴² A. J. Romiszowski, *Designing intructional system* (new York: Kognan Page, 1981).

Sedangkan ketrampilan berkaitan dengan tindakan fisik maupun intelektual dalam mencapai tujuan sebagai akibat dari proses belajar. Hasil belajar yang berupa ketrampilan tersebut dapat dibagi menjadi empat, yaitu berupa ketrampilan kognitif, reaktif, motorik, dan interaktif.

Gagné dan Driscoll mengelompokan hasil belajar ke dalam lima kategori,⁴³ yaitu (1) informasi verbal atau *declarative knowledge*: pengetahuan tentang yang mungkin dapat dikategorikan sebagai: fakta, prinsip, atau generalisasi; (2) ketrampilan intelektual: kemampuan melakukan sesuatu secara intelektual yang menceritakan atau mendemostrasikan, tidak hanya sekedar mengetahui. Ini mencakup: diskriminasi, konsep kongkrit, konsep tertentu, aturan dan aturan tingkat tinggi. (3) strategi kognitif: cara yang membimbing kehadiran, belajar, mengingat dan berfikir. Ia mengatur cara seseorang dalam menghadapi lingkungan dengan cara mempengaruhi proses internal; (4) sikap: kondisi internal yang dicapai yang mempengaruhi pilihan tindakan pribadi terhadap sesuatu, orang atau kejadian tertentu, dan (5) ketrampilan motorik: kemampuan yang melibatkan kegiatan fisik.

Sementara itu, Bloom dkk. mengkategorikan hasil belajar ke dalam tiga ranah atau domain, yaitu: (1) kognitif, yang menekankan pada penguasaan atau menghasilkan kembali sesuatu yang telah dipelajari dan yang melibatkan pemecahan tugas-tugas intelektual dimana individu harus menentukan masalah esensial, kemudian menyusun mengkombinasikan dengan ide, metode, atau prosedur

⁴³ Gagné & Driscoll, *Op .Cit.* h. 44-59.

yang telah dipelajari sebelumnya; (2) afektik, yang menekankan perasaan, emosi, atau derajat penerimaan atau penolakan, dan (3) psikomotor: tujuan yang menekankan ketrampilan otot atau gerak, manipulasi material dan obyek, atau yang memerlukan koordinasi saraf otot. Ketiga ranah tersebut terbagi lagi dalam sub ranah yang lebih rinci.⁴⁴

Teori tentang dimensi hasil belajar tersebut masih banyak lagi yang telah dikemukakan oleh para ahli, dengan berbagai variasinya. Untuk tujuan praktis, Grounland mengelompokkan berbagai macam dimensi hasil belajar yang dikemukakan para ahli ke dalam sembilan dimensi, yaitu: pengetahuan, pemahaman, aplikasi, ketrampilan, berfikir, ketrampilan umum, sikap, minat, apresiasi, dan penyesuaian. Dari kesembilan dimensi tersebut nampaknya hanya yang tiga yang pertama yang banyak dilakukan pengukuran.⁴⁵

Lebih lanjut, hasil belajar sebagai salah satu aspek dari hasil pembelajaran dapat dilihat dari tiga segi: keefektifan, efisiensi, dan daya tarik. Keefektifan merupakan taraf hasil belajar yang dicapai pembelajaran yang disimbolkan dalam bentuk skor yang diperoleh melalui tes yang dilakukan setelah proses pembelajaran. Skor tersebut memberikan informasi tentang kuantitas pengetahuan dan ketrampilan yang telah diperoleh pembelajar sebagai bukti dari adanya perubahan perilaku. Dalam penelitian ini, hasil belajar tersebut berupa indeks prestasi yang diperoleh mahasiswa setelah mengikuti

⁴⁴Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Jump, *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain*, (New York: Longmans, 1956).

⁴⁵Grounland & Linn, *Op. Cit.*

pembelajaran selama tiga semester. Karena itu, hasil belajar tersebut merupakan gabungan dari berbagai mata kuliah.

Hasil belajar yang berupa berbagai dimensi tersebut dipengaruhi oleh berbagai faktor, baik internal (dari dalam diri pebelajar) maupun eksternal (dari luar diri pebelajar).⁴⁶ Dalam penelitian ini, perhatian akan difokuskan pada faktor internal karena besar pengaruhnya pada proses perubahan perilaku. Faktor internal tersebut terdiri atas tiga macam, yaitu: pengetahuan faktual (sesuatu yang telah diketahui), kemampuan intelektual (sesuatu yang telah dipelajari agar dapat dimunculkan kembali), dan strategi belajar. Di samping itu, keberhasilan belajar juga bersumber pada keingintahuan, keinginan berhasil, dan percaya diri. Proses belajar tidak keingintahuan, keinginan berhasil, dan percaya diri. Proses belajar tidak akan berhasil tanpa faktor-faktor tersebut.

Atas dasar tersebut diatas, hasil belajar dalam penelitian ini adalah kemampuan yang diperoleh mahasiswa setelah mengikuti suatu proses pembelajaran yang memungkinkan terjadinya proses belajar. Hasil belajar tersebut mencakup dimensi pengetahuan, pemahaman, dan aplikasi dalam kaitan dengan pengetahuan dan ketrampilan dalam berbagai mata kuliah yang telah ditempuh oleh mahasiswa selama tiga semester. Hasil belajar tersebut berupa nilai indeks prestasi kumulatif, yang diperoleh melalui pengukuran masing-masing mata kuliah. Sebagai simbol, indeks prestasi tersebut mencerminkan tingkat kemampuan, perubahan perilaku mahasiswa. Hasil tersebut dipengaruhi oleh, diantaranya, faktor internal yang yang

⁴⁶Gagné dan Driscoll, *Op. Cit.*

berkaitan dengan pengalaman belajar sebelumnya maupun suasana batin.

B. Kajian Penelitian Terkait

Penelitian tentang tes masuk IAIN Walisongo telah dilakukan oleh beberapa peneliti di masa lalu. Hadjar⁴⁷ (2001), misalnya, telah meneliti kualitas tes ujian masuk IAIN Walisongo dengan menggunakan pendekatan teori tes klasik dan *item response theory* model *Rash*. Penelitian tersebut menemukan bahwa lebih dari separoh butir-butir tes dari semua mata ujian masuk tidak memenuhi semua kriteria butir tes yang baik (daya beda/validitas butir, tingkat kesulitan, dan daya kecoh opsi).

Dalam penelitian lain, Hadjar⁴⁸ (2004) telah menguji beberapa model seleksi calon mahasiswa baru IAIN Walisongo Semarang. Pengujian dilakukan dengan mengukur keandalan prediktif pembobotan hasil tes masuk sebagai prediktor keberhasilan belajar mahasiswa secara umum, dengan IP kumulatif sebagai indikatornya. Dari hasil uji tersebut disimpulkan bahwa model terbaik untuk memprediksi hasil belajar mahasiswa adalah dengan memberikan bobot yang sama pada semua (lima) mata tes ujian dikombinasi dengan skor standar. Temuan lain dari penelitian tersebut adalah bahwa skor tes pengetahuan umum merupakan mata tes yang paling besar kemampuan prediksinya untuk meramalkan keberhasilan studi mahasiswa, dengan sumbangan 25 persen dari varian

⁴⁷Ibnu Hadjar, "Uji kualitas ujian masuk IAIN Walisongo Semarang dengan pendekatan klasik dan item response theory," *Jurnal Penelitian Walisongo*, 9(19), 2001, h. 84-92.

⁴⁸Ibnu Hadjar, "Pengembangan model seleksi calon mahasiswa baru IAIN Walisongo Semarang", *Jurnal Penelitian Walisongo*, 12(1), 2004, h. 47-59.

hasil belajar. Sementara empat mata tes yang lain hanya memberikan tambahan sumbangan sebesar 5 persen.

Sebelum penelitian tersebut, Soediyono dkk.⁴⁹ telah lebih dahulu menyelidiki kemampuan prediksi tes masuk untuk meramalkan hasil belajar mahasiswa Fakultas Tarbiyah. Hasil penelitian memperlihatkan bahwa hanya mata ujian Pancasila yang secara signifikan mampu memprediksi keberhasilan mahasiswa Fakultas Tarbiyah. Sedangkan mata tes yang lain, Bahasa Arab, Bahasa Indonesia, Bahasa Inggris dan Ilmu Pengetahuan Agama tidak signifikan untuk memprediksi keberhasilan belajar mahasiswa.

Meskipun terbatas, hasil-hasil penelitian tersebut menunjukkan bahwa studi tentang tes masuk calon mahasiswa dan hasil belajar mahasiswa belum dilakukan secara komprehensif, karena dilakukan secara terpisah, tidak ada kaitan satu sama lain. Karena tes masuk diadakan setiap tahun, perlu ada masukan yang didasarkan data empiris sebagai pijakan untuk menyiapkan tes masuk yang “berkualitas” agar dapat diperoleh calon mahasiswa yang “berkualitas” sehingga dapat pula dihasilkan lulusan yang “berkualitas”. Penelitian yang diusulkan ini didisain untuk memenuhi harapan tersebut karena didisain secara komprehensif, menguji kualitas dan sekaligus kemampuan prediktif tes masuk IAIN Walisongo.

⁴⁹ Soediyono, dkk., 1995, *Tes Masuk sebagai Prediktor Keberhasilan Belajar*, (Semarang: Fakultas Tarbiyah, 1995).

BAB III

METODE

Sesuai dengan permasalahan sebagaimana telah disajikan dalam bab sebelumnya, penelitian ini dirancang menggunakan pendekatan kuantitatif, dengan fokus hasil tes masuk IAIN Walisongo dan hasil belajar. Pendekatan ini memiliki kelebihan karena menekankan praktik berbasis-bukti empiris.¹ Untuk menjawab permasalahan yang pertama, uji kualitas tes masuk secara internal, penelitian ini menggunakan disain deskriptif,² dengan memerlukan variabel hasil tes masuk terlepas dari variabel yang lain. Sedangkan untuk menjawab permasalahan kedua, uji kualitas tes masuk secara eksternal, penelitian ini menggunakan disain korelasional sederhana,³ dengan memperlakukan hasil tes masuk sebagai prediktor dan hasil belajar matakuliah sebagai kriteria. Adapun langkah-langkah metodologis dalam rangka pengumpulan dan analisis data untuk menjawab permasalahan tersebut secara rinci dijelaskan dalam bagian-bagian berikut ini.

A. Subjek Penelitian.

Sejalan dengan tujuan pertama, yakni untuk meguji kualitas tes masuk secara internal, penelitian ini melibatkan subjek dari seluruh pendaftar calon mahasiswa baru tahun akademik 2012/2013 melalui jalur seleksi mandiri, yang diselenggarakan oleh IAIN Walisongo.

¹ W.E. Martin & K.D. Bridgmon, *Quantitative and statistical research methods: From hypothesis to results*, (San Francisco, 2012), h. 3.

² I Hadjar, *Dasar-dasar metodologi penelitian kuantitatif dalam pendidikan*, (Jakarta: RajGrafindo Persada, 1999), h. 274.274.

³ *Ibid.*, h. 278-289.

Dengan demikian, dalam penelitian ini tidak dilakukan pemilihan sampel karena melibatkan seluruh individu dalam populasi target. Jumlah calon mahasiswa yang mengikuti tes masuk melalui jalur tersebut adalah 1.242 orang. Dari jumlah tersebut, hanya 1.235 orang testee yang mengerjakan tes secara lengkap (baik Bahasa Arab maupun Bahasa Inggris).

Karena tes yang diujikan untuk masing-masing mata tes ada dua versi, subyek pun dibagi menjadi dua kelompok. Kelompok peserta dengan nomor tes ganjil, berjumlah 621 orang, diberi dan mengerjakan tes versi A dan kelompok peserta dengan nomor tes genap, berjumlah 621 orang, mengerjakan tes versi B. Dari jumlah tersebut, peserta yang mengerjakan tes masuk Bahasa Arab versi A dan Bahasa Inggris versi A hanya 614, sedangkan 7 orang yang lain tidak mengerjakannya sehingga dalam analisis hanya melibatkan yang menjawab pertanyaan dalam tes. Sedangkan untuk jumlah peserta yang mengerjakan tes masuk Bahasa Arab versi A dan Bahasa Inggris versi B adalah 621 orang testee.

Selanjutnya, sesuai dengan tujuan kedua, untuk menguji kemampuan prediksi skor tes masuk untuk memperkirakan hasil belajar mahasiswa, penelitian ini melibatkan mahasiswa yang diterima melalui tes masuk jalur mandiri pada tahun akademik 2012/2013 dan sekaligus telah mengikuti matakuliah Bahasa Arab dan/atau Bahasa Inggris pada semester Gasal 2012/2013. Berdasarkan data yang ada dalam Sistem Informasi Akademik, mahasiswa angkatan 2012/2013 yang masuk melalui tes jalur mandiri dan mengikuti kuliah bahasa asing adalah sebagai berikut:

Tabel 1. Jumlah subjek penelitian berdasarkan matakuliah dan kelompok tes masuk

No	Tes Masuk/Versi	Matakuliah	Jumlah Subjek
1	Bahasa Arab A	Bahasa Arab	108
2	Bahasa Arab B	Bahasa Arab	87
3	Bahasa Inggris A	Bahasa Inggris	173
4	Bahasa Inggris B	Bahasa Inggris	186
Jumlah			554

Jumlah seluruh subjek mahasiswa yang dilibatkan dalam penelitian ini adalah 554 mahasiswa, 195 orang di antaranya mengikuti matakuliah Bahasa Arab dan 359 yang lain mengikuti matakuliah Bahasa Inggris pada semester Gasal 2012/2013. Karena paket matakuliah bahasa asing yang diwajibkan untuk diambil masing-masing mahasiswa hanya satu, maka tidak ada mahasiswa yang sekaligus mengambil kedua matakuliah bahasa asing tersebut. Sesuai paket matakuliah semester Gasal, Bahasa Arab hanya ditawarkan untuk dan diikuti oleh mahasiswa dari Fakultas Dakwah dan Komunikasi (FDK) serta Fakultas Syariah dan Ekonomi Islam (FSIE). Sedangkan matakuliah Bahasa Inggris hanya ditawarkan untuk dan diikuti oleh mahasiswa Fakultas Ilmu Tarbiyah dan Keguruan (FITK) dan Fakultas Ushuludin (FU). Meskipun masing-masing matakuliah Bahasa diikuti oleh mahasiswa yang berasal dari fakultas yang berbeda, siabus/materi yang dipelajari dan tes yang digunakan untuk mengukur keberhasilan belajar mereka tidak berbeda. Hal ini karena penyelenggaraan perkuliahan bahasa dilakukan oleh Pusat Bahasa dan Budaya (sekarang Pusat Pengembangan Bahasa), meskipun pelaksanaannya di fakultas masing-masing.

B. Obyek

Sesuai dengan tujuan, ada dua macam objek yang menjadi fokus penelitian ini. Sesuai dengan tujuan penelitian pertama, uji kualitas tes masuk secara internal, yang menjadi obyek penelitian ini adalah hasil tes masuk IAIN Walisongo Semarang yang digunakan untuk menyeleksi calon mahasiswa baru pada tahun akademik 2012/2013. Sedangkan untuk tujuan penelitian yang kedua, objek penelitian adalah skor tes masuk dan skor hasil belajar mahasiswa.

Untuk uji kualitas tes masuk, penelitian ini menggunakan tes yang tersedia dan telah diujikan pada seleksi calon mahasiswa baru tahun akademik 2012/2013, yang dibuat oleh tim yang dibentuk oleh Rektor. Peneliti tidak mengembangkan tes sendiri, karena sudah di luar wewenangnya. Dengan demikian, penelitian ini bersifat evaluatif, yakni untuk mengevaluasi tes yang tersedia dan telah digunakan, berdasarkan kriteria internal maupun eksternalnya. Penggunaan data tes yang tersedia tersebut dilakukan atas ijin Rektor IAIN Walisongo Semarang.

Meskipun sesuai dengan pedoman akademik⁴ tes tersebut meliputi lima mata ujian masuk, dalam penelitian ini hanya dua mata ujian yang diteliti, yaitu Bahasa Arab dan Bahasa Inggris. Kedua tes, sebagaimana tes yang lain, didesain secara obyektif, di mana dalam merespon tes tersebut peserta tes tinggal memilih salah satu alternatif jawaban yang tersedia yang dianggap paling benar. Jenis tes ini memiliki kelebihan terutama dalam penskoran karena hasilnya tidak dipengaruhi oleh pendapat penilai, tetapi lebih ditentukan oleh tingkat kemampuan

⁴ IAIN Walisongo Semarang, *Pedoman Akademik IAIN Walisongo Semarang* (Semarang: IAIN Walisongo, 2012-2013), h. 19.

testee.⁵ Untuk lebih jelasnya, deskripsi masing- masing tes ujian adalah sebagai berikut:

1. Tes Bahasa Arab. Tes untuk mata ujian ini terdiri dari 40 butir soal obyektif. Setiap butir diikuti oleh 5 alternatif jawaban sehingga untuk meresponnya testee diminta memilih salah satu jawaban yang benar. Ada dua macam/versi tes, A dan B, yang masing-masing diujikan bagi peserta dengan nomor tes masuk ganjil dan nomor genap. Namun demikian, kedua tes memiliki butir yang sama, hanya dengan urutan nomor yang berbeda. Kedua tes berisi butir yang mengukur penguasaan kosa kata (9 butir [22,5 persen]), kemampuan membaca (18 butir [45 persen]), penguasaan tatabahasa/*nahwu* dan *shorf* (11 butir [27,5 persen]), tarjamah (1 butir [2,5 persen]), dan pengetahuan umum (1 butir [2,5 persen]). Hal ini menunjukkan bahwa tes tersebut lebih didominasi oleh butir-butir tes membaca dari pada butir-butir yang lain.
2. Tes Bahasa Inggris. Sebagaimana tes Bahasa Arab, tes untuk mata ujian ini terdiri dari 40 butir soal obyektif. Setiap butir diikuti oleh 5 alternatif jawaban sehingga untuk meresponnya testee diminta memilih salah satu jawaban yang benar. Tes Bahasa Inggris juga terdiri dari dua macam model tes, A dan B. Berbeda dari tes Bahasa Arab, kedua macam tes Bahasa Inggris tersebut berisi butir yang berbeda satu sama lain dari segi isi. Namun demikian, kedua mengukur kemampan yang sama, penguasaan kosa kata, kemampuan membaca, dan tatabahasa. Kedua macam tes memiliki

⁵ Robert L. Linn & Nourman E. Grounlund, *Measurement and Assessment in Teaching* (Englewood Clift: Prentice Hall, 1995), h. 20.

jumlah butir yang berbeda untuk komponen kemampuan membaca (masing-masing untuk Versi A dan B: 7 butir [17,5 persen] dan 2 butir [5,0 persen]) dan komponen membaca (masing-masing untuk Versi A dan B: 13 butir [32,5 persen] dan 18 butir [45,0 persen]). Sedangkan untuk komponen tatabahawa/grammar keduanya memiliki jumlah butir yang sama, yaitu 20 butir (50,0 persen). Kedua tes tidak berisi butir untuk mengukur kemampuan terjemah. Distribusi butir tes tersebut enunjukkan bahwa kedua macam tes Bahasa Inggris didominasi oleh butir tes untuk mengukur penguasaan tatabahasa atau gramatika.

Sedangkan untuk tujuan penelitian kedua, uji kualitas tes secara eksternal, objek penelitian ini adalah skor hasil tes masuk Bahasa Arab dan Bahasa Inggris (masing-masing sebagai satu kesatuan, tidak dilihat perbutir) dan skor hasil belajar Bahasa arab dan Inggris yang dicapai oleh mahasiswa.

C. Variabel

Sesuai dengan permasalahan pertama, penelitian ini akan fokus pada variabel hasil tes masuk bagi calon mahasiswa baru tahun akademik 2012/2013, khususnya mata tes Bahasa Arab dan Bahasa Inggris. Di samping itu, sesuai dengan permasalahan kedua, penelitian ini melibatkan variabel skor hasil tes masuk dan skor hasil belajar bahasa Arab dan Bahasa Inggris. Skor tes masuk merupakan skor yang diperoleh peserta tes atas jawaban yang benar terhadap pertanyaan soal ujian tersebut. Berdasarkan cara menghitungnya, skor tes masuk ini dibedakan dua: klasik dan modern. Skor tes klasik merupakan jumlah butir soal yang dijawab benar oleh peserta tes, dengan kemungkinan

rentang antara 0 (tidak ada butir yang dijawab dengan benar) sampai 40 (semua butir dijawab dengan benar). Sedangkan skor tes klasik dihitung dengan menggunakan teori tes klasik satu parameter yang menghasilkan skor standar teta (θ), dengan kemungkinan rentangan yang tak terbatas, baik + maupun -.

Sementara itu, hasil belajar Bahasa Arab dan Inggris merupakan nilai matakuliah yang diperoleh oleh mahasiswa dari hasil tes setelah mengikuti perkuliahan bahasa tersebut selama satu semester. Nilai tersebut merupakan nilai rata-rata dari ketiga matakuliah Bahasa Arab dan Inggris I (*istima'* dan *muhadatsah/listening* dan *speaking*), II (*qawa'id* dan *kitabah/gramatika* dan *writing*), dan III (*mufrodah* dan *muthola'ah/vocabulary* dan *reading comprehension*). Matakuliah tersebut diikuti oleh mahasiswa secara paket dalam satu semester dengan kurikulum dan tes yang sama. Perkuliahan diselenggarakan oleh Pusat Pengembangan Bahasa. Nilai yang diperoleh memiliki kemungkinan rentang dari 0,0 sampai 4,0; sesuai dengan sistem evaluasi yang berlaku di IAIN Walisongo.

D. Teknik Pengumpulan Data

Karena penelitian ini bertujuan untuk mengevaluasi tes yang telah digunakan untuk seleksi calon mahasiswa baru, peneliti tidak mengembangkan instrumen atau tes sendiri. Alih-alih, peneliti menggunakan instrumen/tes yang telah dikembangkan/dibuat oleh panitia penerimaan calon mahasiswa baru tahun akademik 2012/2013. Tes yang diteliti dibatasi pada mata tes Bahasa Arab dan Bahasa Inggris. Untuk tujuan evaluasi kualitas tes masuk (validitas dan reliabilitas

instrumen), hasil tes seleksi calon mahasiswa baru diperlakukan sebagai variabel tunggal, tanpa dikaitkan dengan variabel lain. Sebagaimana dikemukakan sebelumnya, masing-masing mata tes tersebut memiliki dua macam/versi tes, Versi A dan versi B. Setiap versi tes terdiri dari 40 butir tes, yang dirancang untuk mengukur penguasaan kosa kata, kemampuan membaca teks, kemampuan menterjemahkan, dan penguasaan gramatika atau Qowaid. Seluruh butir soal dibuat dalam bentuk obyektif dengan 5 pilihan jawaban dan dilengkapi dengan kunci jawaban untuk alternatif yang paling benar.

Lebih lanjut, peneliti juga tidak melakukan pengumpulan data sendiri secara langsung dari para subyek penelitian. Alih-alih, peneliti menggunakan data yang telah tersedia yang sudah terekam dalam bentuk dokumen/file soft copy. Data tentang hasil tes masuk calon mahasiswa baru tahun 2012/2013 diperoleh dari Bagian Akademik IAIN Walisongo.

Sesuai dengan obyek penelitian ini, pengumpulan data tidak dilakukan oleh peneliti, tetapi oleh panitia ujian masuk yang dibentuk oleh Rektor. Karena sifatnya yang evaluatif, maka peneliti tidak mengembangkan instrumen sendiri, tetapi menggunakan instrumen yang telah ada. Demikian juga, peneliti tidak mengumpulkan data sendiri, tetapi hanya menggunakan data yang telah tersedia, yakni hasil ujian yang telah dilaksanakan oleh panitia. Di samping itu, pelaksanaan pengumpulan data juga tidak dilakukan sendiri oleh peneliti sehingga administrasinya tidak dapat dikendalikan olehnya. Namun demikian, hasil tes tersebut diasumsikan akan mencerminkan karakteristik testee yang sebenarnya karena tes tersebut termasuk kategori performa

maksimum (*maximum performance test*).⁶ Hal ini karena penyelenggaraan tersebut bersifat alami, dimana testee diasumsikan menggunakan seluruh kemampuannya untuk mengerjakan tes ujian. Prosedur pelaksanaan tes juga dilakukan secara ketat berdasarkan ketentuan/tata tertib yang telah ditetapkan. Tidak ada catatan yang mengurangi kepercayaan terhadap hasil tes. Dengan demikian, hasilnya akan terhindar dari bias peneliti.

E. Teknik Analisis Data.

Sesuai dengan tujuan penelitian, data terkumpul selanjutnya dianalisis dalam dua tahap analisis: uji kualitas tes dan uji prediktif tes. Untuk lebih jelasnya masing-masing uji tersebut adalah sebagai berikut.

1. Analisis kualitas tes masuk

Analisis uji kualitas tes masuk Bahasa Arab dan Inggris dilakukan terhadap data respon peserta tes pada masing-masing butir tes. Dalam analisis ini, pertama-tama data dianalisis dengan menggunakan pendekatan teori tes klasik. Selanjutnya data dianalisis dengan menggunakan pendekatan teori tes modern atau *Item Response Theory*. Meskipun demikian, analisis kedua tersebut bukan merupakan lanjutan dari analisis pertama, karena keduanya merupakan alternatif pendekatan yang berbeda, tetapi menggunakan data yang sama. Lebih lanjut, karena alat bantu analisis (program komputer) yang dipakai tidak memerlukan data matang (berupa skor), maka sebelum analisis tidak akan dilakukan penskoran terhadap hasil tes. Alih-alih, hanya data mentah (respon asli dari

⁶ Howard B. Lyman, *Test Scores and What They Mean* (Englewood Cliffs: Prentice-hall, 1978), h. 18.

tastee yang berupa pilihan dari opsi jawaban yang tersedia) yang dijadikan masukan analisis dan skor untuk masing- masing butir maupun masing- masing testee akan diperoleh setelah proses analisis. Untuk lebih jelasnya, teknik analisis yang digunakan untuk masing- masing pendekatan adalah sebagai berikut:

a. Pendekatan Klasik

Sesuai dengan fokus dari pendekatan ini, data yang terkumpul dianalisis dengan tiga teknik analisis: proporsi, korelasi, dan persentase. Teknik proporsi digunakan untuk mengungkap tingkat kesukaran butir (p).⁷ Secara operasional, penghitungannya dilakukan dengan cara mencari proporsi testee yang menjawab dengan benar suatu butir. Karena itu, nilai indeks tingkat kesukaran (p) yang diperoleh akan berkisar antara 0,0 sampai 1,0. Semakin tinggi nilai indeks, semakin mudah suatu butir, dan sebaliknya.⁸ Hasil analisis ini selanjutnya digunakan untuk menentukan apakah butir yang bersangkutan baik (dapat diterima) atau tidak baik (tidak diterima). Sebagai kriteria, suatu butir dikatakan baik bila indeks tingkat kesukarannya (p) berada dalam kisaran 0,2 sampai 0,8 dan sebaliknya bila berada di luar kisaran tersebut.

Sedangkan teknik analisis korelasi digunakan untuk mengungkap tingkat validitas atau daya beda, yang dihitung

⁷G. Fulcher & F. Davidson, *Language testing and assessment: An advanced resource book*, (London: Routledge, 2007), h. 102.

⁸ Karena konsep ini, dalam Bahasa Arab disebut *mu'āmil al-suhūlah* atau "indeks kemudahan." Lihat M.A.K. Muhammad, *Ikhtibārāt al-Lughoh*, (Riyādh: Jāmi'ah al-Mālik Su'ūd, 1996), h. 318.

melalui hubungan antara skor masing-masing butir dengan skor total (seluruh butir). Karena skor jawaban testee untuk masing-masing butir bersifat dikotomis dan skor total kontinum, maka teknik analisis yang digunakan adalah korelasi *point-biserial* (r_{pbis}),⁹ yang merupakan kriteria internal. Nilai koefisien korelasi point-biserial (r_{pbis}) dapat berkisar antara 0,0 (tidak mampu membedakan kemampuan testee), sampai 1,0 (secara mutlak mampu membedakan kemampuan testee). Nilai tersebut dapat dalam arah positif atau negatif. Arah positif menunjukkan bahwa testee yang memiliki skor tinggi cenderung dapat menjawab dengan benar butir yang bersangkutan, sedang testee yang memiliki skor total rendah cenderung tidak dapat menjawab dengan benar. Sedang nilai koefisien dengan arah negatif menunjukkan kecenderungan yang sebaliknya. Butir tes yang “baik” atau “valid” adalah butir yang memiliki nilai koefisien positif tinggi (signifikan). Nilai koefisien korelasi yang rendah (tidak signifikan) atau negatif menunjukkan bahwa butir tes tersebut tidak valid karena tidak memberikan sumbangan pada nilai total. Karena itu, dalam pengembangan tes, hanya butir yang memiliki nilai koefisien yang tinggi positif yang dipilih untuk digunakan, sedang yang memiliki nilai rendah atau negatif harus dibuang atau direvisi. Dalam penelitian ini, kriteria butir yang baik atau dapat diterima atau dianggap “valid” adalah bila memiliki nilai $r_{pbis} \geq 0,3$ dan sebaliknya tidak diterima atau “tidak valid” bila nilai $r_{pbis} < 0,3$.

⁹ *Ibid.*, h. 103.

Lebih lanjut, secara keseluruhan tes tersebut juga dianalisis untuk mengetahui reliabilitasnya, yakni seberapa jauh tes tersebut mampu menghasilkan skor yang “mendekati” sama bagi individu bila dilakukan pengulangan.¹⁰ Dalam penelitian ini, teknik analisis yang digunakan untuk mengetahui tingkat reliabilitas tersebut adalah Kuder-Richadson/K-R 20.¹¹ Teknik ini banyak digunakan karena menghasilkan perkiraan rerata dari semua kemungkinan teknik belah-tengah yang diperoleh dari skor tes secara independen.

Untuk penghitungan tingkat kesulitan butir dan indeks validitas atau daya beda butir tersebut digunakan program komputer (software) S-Stats (Program Aplikasi Statistik Walisongo).¹²

b. Pendekatan Item Response Theory

Setelah dianalisis dengan pendekatan klasik, data tes selanjutnya dianalisis dengan menggunakan pendekatan modern atau *item response theory* atau *Latent Trait Theory*.¹³ Dalam penelitian ini, analisis terhadap respon butir-butir tes masuk menggunakan model satu parameter atau *Rasch Model* (mengukur parameter kesulitan butir tes dan kemampuan testee). Analisis ini

¹⁰ M.D. Lodico, D.T. Spoulding & K.H. Voegtle, *Methods in educational research: From theory to practice*, (San Francisco: Josey-Bass, 2006), h. 87.

¹¹ Fulcher & Davidson, *Op. cit.*, h. 107.

¹² Ibnu Hadjar, *W-Stats, Program Aplikasi Statistik Walisongo* (Semarang: IAIN Walisongo, 2013).

¹³ C. Magno, “Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data,” *The International Journal of Educational and Psychological Assessment*, 1 (1), pp. 1-11, 2009), h. 1.

diharapkan dapat mengungkapkan tingkat kesulitan butir (b) dan kemampuan testee (θ).¹⁴ Indeks kesulitan butir (b) akan dinyatakan dalam nilai standart deviasi. Pada dasarnya, semakin besar nilai indeks, semakin sulit suatu butir, dan sebaliknya. Demikian juga, indeks kemampuan testee dinyatakan dalam nilai standart dalam rentangan 0 sebagai titik tengah dan 1 sebagai simpang bakunya (standart deviasi). Hanya saja posisi masing-masing testee tidak distandarkan dengan kelompoknya (testee yang lain), tetapi distandarkan dalam kaitan dengan capaiannya dan tingkat kesukaran butir.

Lebih lanjut, untuk menentukan apakah suatu butir dikatakan baik atau tidak, hasil analisis tersebut diuji lebih lanjut untuk mengetahui apakah butir yang bersangkutan sesuai dengan model parameter atau tidak. Untuk itu digunakan uji kesesuaian/pencocokan, dengan menggunakan uji χ^2 (*kai kuadrat*) (yang menunjukkan apakah ada kesesuaian antara data empiris dengan model data yang diidealkan). Sebagai kriteria dalam uji ini, suatu butir dikatakan baik atau "*fit*" bila nilai χ^2 (*kai kuadrat*) tidak signifikan pada taraf 5 persen (yang berarti tidak ada perbedaan antara model dengan data empiris) dan tidak baik atau "*unfit*" bila nilai χ^2 (*kai kuadrat*) signifikan pada taraf 5 persen (yang berarti ada perbedaan antara model dengan data empiris) .

¹⁴R.K. Hambleton, H. Swaminathan, & H.J. Rogers, *Fundamentals of item response theory*. (Newbury Park: Sage, 1991), h. 32.

Untuk penghitungan nilai parameter (orang dan butir) serta prosedur pencocokan tersebut digunakan program komputer RASCAL, versi 3.¹⁵

2. Analisis uji prediktif

Sesuai dengan permasalahan kedua, data hasil tes masuk Bahasa Arab dan bahasa Inggris selanjutnya dianalisis dalam kaitannya dengan hasil belajar Bahasa Arab dan Bahasa Inggris yang diperoleh mahasiswa setelah belajar kedua bahasa asing di IAIN Walisongo. Analisis ini dimaksudkan untuk menguji validitas prediktif¹⁶ tes masuk tersebut, terutama berkaitan dengan hasil belajar. Teknik analisis yang digunakan adalah regresi sederhana,¹⁷ dengan melibatkan satu variabel prediktor (hasil tes masuk) dan satu variabel kriteria (hasil belajar). Dengan teknik analisis ini akan dihasilkan persamaan regresi/ Y' (sebagai dasar prediksi hasil belajara berdasarkan skor tes masuk), indeks nilai F /koefisien regresi dan taraf signifikansi/ p (sebagai dasar untuk melakukan generalisasi), serta koefisien determinasi/ R^2 (sebagai dasar menghitung besaran kontribusi skor tes masuk dalam menjelaskan varian nilai hasil belajar). Hasil analisis dinyatakan signifikan bila nilai $p \leq 0,05$. Analisis dilakukan secara terpisah untuk masing-masing bahasa asing. Lebih lanjut, karena analisis tes masuk menghasilkan dua macam skor dengan pendekatan teori tes yang berbeda, maka analisis

¹⁵Assessment Systems Corporation, *Rasch Model Item Calibration Program-RASCAL*, Version 3.00 (1988).

¹⁶Lodico, Spoulding & Voegtler, *Op. cit.*, h. 94.

¹⁷E.J. Pedhazur, *Multiple regression in behavioral research: Explanation and prediction*, (New York: Holt, Rinehart, & Winston, 1982), h. 14-18.

juga dilakukan secara terpisah untuk masing-masing skor tersebut. Hal ini memberikan kemungkinan untuk membandingkan tingkat efektivitas kedua macam skor dalam memprediksi hasil belajar.

Untuk perhitungan dalam analisis regresi ini digunakan program statistik SPSS 16.0.

BAB IV

HASIL PENELITIAN

Tujuan utama penelitian ini adalah untuk menguji kualitas tes masuk Bahasa Arab dan Bahasa Inggris yang digunakan untuk seleksi calon mahasiswa baru tahun akademik 2012/2013 secara internal. Di samping itu, ia juga digunakan untuk menguji kualitas tes tersebut secara eksternal, yakni kemampuan prediksi hasil tes masuk tersebut untuk memperkirakan hasil belajar Bahasa Arab dan Bahasa Inggris mahasiswa IAIN Walisongo. Untuk tujuan tersebut, data yang terkumpul dianalisis dengan menggunakan teknik statistik yang sesuai dengan aspek kualitas yang diuji, sebagaimana secara rinci telah disebutkan dalam bab sebelumnya. Hasil analisis data tersebut secara berturut-turut disajikan sebagai berikut ini.

A. Kualitas Tes Masuk 2012/2013: Teori Klasik dan Modern

1. Deskripsi Umum

Sebelum disajikan hasil analisis data penelitian, terlebih dahulu akan ini disajikan gambaran secara umum tentang tes yang menjadi objek dan fokus dari penelitian ini. Penyajian ini didasarkan hasil analisis terhadap masing-masing tes yang dijadikan alat seleksi calon mahasiswa baru tahun akademik 2012/2013. Analisis terutama diarahkan pada isi dari masing-masing butir tes, dilihat materi butir tes yang digunakan untuk mengukur kompetensi testee dalam berbagai aspek bahasa Arab dan Inggris yang terkandung dalam butir-butir tes. Secara umum, isi atau materi tersebut dikelompokkan ke dalam aspek-aspek bahasa Arab dan Bahasa Inggris, seperti kosa kata, membaca, tata bahasa atau komposisi, tarjamah, dan

pengetahuan umum. Secara deskriptif, ringkasan hasil analisis 40 butir tes dari masing-masing tes tersebut disajikan dalam tabel berikut ini.

Tabel 2. Distribusi jumlah (persen) butir soal untuk masing-masing versi berdasarkan aspek bahasa

Aspek Bahasa	Jumlah Butir untuk Tes Bahasa/Versi			
	Arab A	Arab B	Inggris A	Inggris B
Kosa kata	9 (22,5%)	9 (22,5%)	7 (17,5%)	2 (5,0%)
Membaca	18 (45%)	18 (45%)	13 (32,5%)	18 (45,0%)
Tata bahasa /komposisi	11 (27,5%)	11 (27,5%)	20 (50,0%)	20 (50,0%)
Tarjamah	1 (2,5%)	1 (2,5%)	-	-
Pengetahuan umum	1(2,5%)	1(2,5%)	-	-
Jumlah butir (persen)	40 (100,0%)	40 (100,0%)	40 (100,0%)	40 (100,0%)

Tabel tersebut memperlihatkan bahwa tes Bahasa Arab menguji kemampuan testee dalam lima aspek bahasa, yaitu kosa kata, membaca, tata bahasa atau komposisi, tarjamah, dan pengetahuan umum. Sedangkan tes Bahasa Inggris menguji hanya tiga aspek bahasa, yaitu kosa kata, membaca, dan tata bahasa atau komposisi. Kedua versi tes Bahasa Arab (versi A dan B) berisi butir tes dengan komposisi yang sama proporsinya untuk masing-masing aspek bahasa. Hal ini dapat terjadi karena versi tes bahasa Arab ini hanya berbeda dalam pengurutan nomor tes. Kedua versi tes didominasi butir yang digunakan untuk mengukur kemampuan membaca testee (45 persen atau 18 dari 40 butir), diikuti penguasaan tata bahasa/kompoisisi (27,5 persen atau 11 dari 40 butir). Sedangkan butir yang mengukur aspek terjemah dan pengetahuan umum, masing-masing hanya 1 butir atau 2,5 persen).

Sementara itu, kedua versi tes bahasa Inggris memiliki komposisi yang berbeda, kecuali untuk aspek tata bahasa atau kosposisi, yang memiliki jumlah butir yang sama. Pengukuran aspek ini mendominasi butir-butir tes bahasa Inggris, yakni 20 atau 50 persen dari 40 butir tes secara keseluruhan. Pengukuran aspek membaca untuk kedua versi tes memiliki jumlah butir yang berbeda, yaitu terdiri dari 13 butir atau 32,5 persen untuk versi A dan 18 butir atau 45 persen untuk versi tes B. Sedangkan pengukuran aspek kosa kata masing-masing terdiri dari 7 butir atau 17,5 persen untuk versi A dan 2 butir atau 5 persen untuk versi tes B. Tidak sebagaimana kedua tes bahasa Arab yang persis sama isinya, kedua tes bahasa Inggris berisi pertanyaan yang berbeda.

Selanjutnya, untuk memberikan gambaran bagaimana kemampuan subjek penelitian/peserta tes dalam menjawab pertanyaan yang terdapat dalam tes, data hasil tes dari seluruh peserta dianalisis secara deskriptif untuk mengetahui kecenderungan secara umum dari kemampuan/penguasaan bahasa Arab dan Inggris yang mereka miliki. Ringkasan hasil analisis ini disajikan dalam tabel berikut ini.

Tabel 3. Deskripsi skor tes masuk berdasarkan analisis teori tes klasik dan modern untuk masing-masing versi tes

No.	Nama Tes/Versi	N	Skor klasik		Skor Modern	
			Rerata	SB	Rerata	SB
1	Tes Bahasa Arab A	614	11,46	7,44	-1,52	1,96
2	Tes Bahasa Arab B	621	12,29	7,84	-1,34	1,89
3	Tes Bahasa Inggris A	614	9,29	4,25	-1,60	1,25
4	Tes Bahasa Inggris B	621	12,72	5,68	-1,21	1,51

Tabel tersebut di atas memperlihatkan bahwa berdasarkan pendekatan teori tes klasik kecenderungan kemampuan peserta tes masuk tergolong sangat rendah (kurang dari sepertiga dari skor maksimal teoritis [dari rentangan skor teoritis 0 – 40]), dengan penyebaran skor yang heterogin. Bahkan kecenderungan tersebut juga jauh lebih rendah dari batas bawah tingkat kelulusan hasil tes yang dapat diakui sama dengan pencapaian matakuliah Bahasa Arab (yakni 60% dari 40 = 25). Tabel tersebut juga memperlihatkan perbedaan kemampuan Bahasa Arab peserta tes nomor ganjil (mengerjakan tes versi A) dan nomer genap (mengerjakan tes versi B). Meskipun butir-butir tes sama, dengan urutan nomor yang berbeda, hasil tes Bahasa Arab menunjukkan bahwa kelompok peserta tes versi A memiliki kemampuan yang lebih rendah dari pada kelompok peserta tes versi B.

Sementara itu, untuk tes bahasa Inggris, secara relatif kemampuan peserta tes nomor ganjil (mengerjakan tes versi A) lebih rendah dari pada kemampuan peserta tes nomor genap (mengerjakan tes versi B). Hanya saja, karena kedua versi tes berisi butir-butir soal yang berbeda, perbedaan hasil tes tersebut tidak bisa serta merta secara mutlak sebagai petunjuk tentang adanya perbedaan kemampuan antara kedua kelompok peserta tes. Hal ini bisa jadi karena adanya perbedaan tingkat kesulitan tes yang berbeda sehingga menghasilkan tingkat kemampuan yang berbeda.

Sementara itu, berdasarkan pendekatan tes modern (versi Rash) tabel tersebut di atas memperlihatkan bahwa kecenderungan kemampuan peserta tes (θ) untuk semua mata tes dan versi tes jauh di

bawah rerata tingkat kesulitan butir-butir tes (yang secara standar memiliki tingkat kesulitan rata-rata, $b = 0,00$). Hal ini menunjukkan bahwa soal-soal tes tersebut sangat sulit bagi peserta tes, sehingga probabilitas peserta untuk dapat menjawab soal dengan benar sangat rendah.

Berbeda dari kecenderungan kemampuan peserta tes yang sangat rendah, kemampuan hasil belajar mahasiswa dalam matakuliah Bahasa Arab dan Bahasa Inggris tergolong bagus (sekitar 3,0 dari rentang teoritis 0,00 – 4,00). Hal ini ditunjukkan oleh rerata nilai matakuliah yang dicapai oleh mahasiswa, sebagaimana secara ringkas disajikan dalam tabel di bawah ini. Hasil tersebut juga menunjukkan bahwa kemampuan (hasil belajar) mahasiswa dalam Bahasa Arab cenderung lebih baik dari pada kemampuan (hasil belajar) Bahasa Inggris. Sementara itu, penyebaran kemampuan mahasiswa (yang ditunjukkan oleh nilai simpang baku) juga relatif sama antar kelompok matakuliah. Secara ringkas kecenderungan tersebut disajikan dalam tabel berikut.

Tabel 4. Deskripsi data hasil belajar matakuliah bahasa berdasarkan kelompok versi tes masuk

No.	Mata kuliah	Kelompok Tes Masuk	N	Rerata	SB
1	Bahasa Arab	Tes Bahasa Arab A	108	3,06	0,80
		Tes Bahasa Arab B	87	3,00	0,73
2	Bahasa Inggris	Tes Bahasa Inggris A	176	2,81	0,76
		Tes Bahasa Inggris B	189	2,84	0,71

2. Kalibrasi Tes Masuk Bahasa Arab A

Untuk mengetahui kualitas masing-masing tes masuk, data hasil tes masing-masing butir dari masing-masing tes (dan versinya) dianalisis dengan menggunakan pendekatan tes klasik dan modern. Hasil analisis disajikan dalam bagian berikut ini. Analisis dengan menggunakan pendekatan klasik dimaksudkan untuk menghasilkan indeks tingkat kesulitan butir, validitas butir, fungsi pengecoh masing-masing pilihan jawaban butir. Selanjutnya, berdasarkan kriteria yang telah ditentukan diambil kesimpulan apakah butir tersebut dinyatakan baik (bila semua kriteria terpenuhi) atau tidak baik (bila setidaknya salah satu kriteria tidak terpenuhi). Secara lengkap, hasil analisis tersebut dapat dilihat dalam lampiran. Sedangkan secara ringkas hasil tersebut disajikan berikut ini.

a. Hasil Analisis Butir: Pendekatan Klasik dan Modern

Dalam rangka untuk menguji apakah butir-butir tes masuk sebagai calon mahasiswa baru memiliki kualitas yang baik atau tidak, data respon peserta tes masuk Bahasa Arab A pertama-tama dianalisis dengan menggunakan pendekatan teori tes klasik untuk menghasilkan indeks tingkat kesulitan butir, koefisien korelasi antara skor butir dan skor total (seluruh butir), fungsi pengecoh (kesimpulan dari hasil analisis). Tabel 5 berikut ini menyajikan ringkasan hasil analisis tersebut.

Tabel 5. Hasil Analisis Butir Tes Bahasa Arab versi A dengan Teori Klasik

NO. BUTIR	Tingkat Kesulitan (p)	Korelasi butir-total (r_{pb})	Kesimpulan berdasarkan			Kesimpulan Akhir ¹
			Tingkat Kesulitan ²	Validitas ³	Fungsi Pengecoh ⁴	
1	0,31	0,22	Baik	Tdk. Valid	Baik	Tdk. Baik
2	0,18	0,36	Tdk. Baik	Valid	Baik	Tdk. Baik
3	0,19	0,34	Tdk. Baik	Valid	Baik	Tdk. Baik
4	0,50	0,53	Baik	Valid	Baik	Baik
5	0,31	0,40	Baik	Valid	Baik	Baik
6	0,51	0,48	Baik	Valid	Baik	Baik
7	0,49	0,53	Baik	Valid	Baik	Baik
8	0,26	0,46	Baik	Valid	Baik	Baik
9	0,26	0,51	Baik	Valid	Baik	Baik
10	0,12	0,35	Tdk. Baik	Valid	Baik	Tdk. Baik
11	0,08	0,15	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
12	0,16	0,49	Tdk. Baik	Valid	Baik	Tdk. Baik
13	0,27	0,50	Baik	Valid	Baik	Baik
14	0,54	0,47	Baik	Valid	Baik	Baik
15	0,44	0,47	Baik	Valid	Baik	Baik
16	0,20	0,36	Tdk. Baik	Valid	Baik	Tdk. Baik
17	0,26	0,43	Baik	Valid	Baik	Baik
18	0,15	0,14	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
19	0,19	0,27	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
20	0,68	0,53	Baik	Valid	Baik	Baik
21	0,21	0,33	Baik	Valid	Baik	Baik
22	0,21	0,45	Baik	Valid	Baik	Baik
23	0,17	0,43	Tdk. Baik	Valid	Baik	Tdk. Baik
24	0,24	0,50	Baik	Valid	Baik	Baik
25	0,06	0,00	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
26	0,17	0,24	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
27	0,45	0,62	Baik	Valid	Baik	Baik
28	0,25	0,60	Baik	Valid	Baik	Baik
29	0,25	0,57	Baik	Valid	Baik	Baik
30	0,39	0,56	Baik	Valid	Baik	Baik

¹ Kesimpulan: "Baik" apabila semua kriteria hasil analisis butir Baik atau Valid.

² Tingkat Kesulitan: "Baik" apabila $p > 0,2$ dan $\leq 0,8$.

³ Validitas: "Valid" apabila $r_{pb} \geq 0,30$.

⁴ Fungsi Pengecoh: "Baik" apabila setiap alternatif jawaban butir dipilih oleh $\geq 1\%$ dari seluruh testee.

NO. BUTIR	Tingkat Kesulitan (p)	Korelasi butir-total (r_{pb})	Kesimpulan berdasarkan			Kesimpulan Akhir ¹
			Tingkat Kesulitan ²	Validitas ³	Fungsi Pengecoh ⁴	
31	0,09	0,15	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
32	0,23	0,57	Baik	Valid	Baik	Baik
33	0,44	0,59	Baik	Valid	Baik	Baik
34	0,37	0,49	Baik	Valid	Baik	Baik
35	0,18	0,32	Tdk. Baik	Valid	Baik	Tdk. Baik
36	0,26	0,51	Baik	Valid	Baik	Baik
37	0,35	0,51	Baik	Valid	Baik	Baik
38	0,21	0,54	Baik	Valid	Baik	Baik
39	0,36	0,48	Baik	Valid	Baik	Baik
40	0,49	0,41	Baik	Valid	Baik	Baik

Tabel tersebut memperlihatkan bahwa tingkat kesulitan butir merentang dari 0,06 (terendah, sangat sulit) sampai 0,68 (tertinggi, sedang). Tidak butir yang memiliki tingkat kesulitan tinggi atau sangat tinggi. Sedangkan rerata tingkat kesulitan adalah 0,287 (rendah). Hasil ini menunjukkan bahwa butir-butir tes masuk Bahasa Arab versi A termasuk kategori sulit bagi peserta tes pada umumnya, sehingga skor yang mereka capai pun sangat rendah.

Tabel tersebut juga memperlihatkan bahwa dayabeda atau validitas butir (diukur dengan korelasi antara masing-masing butir dengan skor total) merentang dari $r_{pb} = 0,00$ (terendah, tidak valid) sampai $r_{pb} = 0,62$ (tertinggi, tingkat validitas tinggi) dan tidak ada yang negatif. Sedangkan rerata tingkat validitasnya adalah $r_{pb} = 0,42$ (sedang). Hasil ini menunjukkan bahwa secara umum tingkat validitas butir-butir soal tes masuk Bahasa Arab versi A sedang, cukup menyakinkan untuk menghasilkan ukuran tingkat kemampuan berbahasa Arab.

Lebih lanjut, berdasarkan kriteria yang telah ditetapkan sebagaimana dalam bab sebelumnya, tabel tersebut juga memperlihatkan bahwa sebagian besar (26 [65 persen] dari 40) butir tes masuk Bahasa Arab versi A termasuk kategori baik, sedangkan 14 butir (35 persen) yang lain tidak baik. Ke-26 butir tersebut memenuhi kriteria tingkat kesulitan dan validitas/daya beda butir yang telah ditentukan. Dari 14 butir yang tidak baik, 6 butir tidak memenuhi kedua kriteria, 7 butir tidak memenuhi kriteria tingkat kesulitan saja, dan hanya 1 butir yang tidak memenuhi kriteria validitas saja. Sedangkan berdasarkan kriteria fungsi pengecoh, semua butir telah memenuhi kriteria.

Selanjutnya, data yang sama dari hasil tes masuk Bahasa Arab versi A dianalisis dengan menggunakan teori tes modern 1 parameter, atau *Rash model* untuk menghasilkan nilai indeks tingkat kesulitan butir dan probabilitas kesesuaiannya melalui uji fit/unfit pada taraf signifikansi 5 persen. Hasil analisis secara ringkas disajikan dalam tabel di bawah ini.

Tabel 6. Hasil Analisis Butir Tes dengan IRT 1 Parameter

NO. BUTIR	b^5	χ^2^6	$d.f.^7$	p^8	Kesimpulan ⁹
1	-0,253	86,841	19	0,00	Unfit
2	0,588	20,177	19	0,38	Fit

⁵ b : tingkat kesulitan butir, di mana testee dengan nilai $\theta = b$ memiliki peluang 50 persen untuk menjawab dengan benar pertanyaan butir tersebut.

⁶ χ^2 : nilai Kai kuadrat.

⁷ df : degrees of freedom (derajat kebebasan).

⁸ p : taraf signifikansi atau tingkat probabilitas kesesuaian model dengan data empiris.

⁹ Kesimpulan: "fit" bila $p > 0,05$ (model diterima), dan "unfit" bila $p \leq 0,05$ (model tidak diterima).

NO. BUTIR	b^5	χ^2^6	$d.f.^7$	p^8	Kesimpulan ⁹
3	0,549	31,290	19	0,04	Unfit
4	-1,244	28,017	19	0,08	Fit
5	-0,234	21,083	19	0,33	Fit
6	-1,294	25,794	19	0,14	Fit
7	-1,211	10,832	19	0,93	Fit
8	0,072	21,242	19	0,32	Fit
9	0,051	19,610	19	0,42	Fit
10	1,223	27,972	19	0,08	Fit
11	1,710	43,601	19	0,00	Unfit
12	0,762	26,602	19	0,11	Fit
13	-0,010	15,428	19	0,70	Fit
14	-1,437	13,633	19	0,80	Fit
15	-0,944	32,531	19	0,03	Unfit
16	0,487	26,486	19	0,12	Fit
17	0,082	24,711	19	0,17	Fit
18	0,863	63,046	19	0,00	Unfit
19	0,549	44,798	19	0,00	Unfit
20	-2,230	22,093	19	0,28	Fit
21	0,427	33,968	19	0,02	Unfit
22	0,356	16,069	19	0,65	Fit
23	0,666	31,880	19	0,03	Unfit
24	0,177	17,501	19	0,56	Fit
25	1,956	81,336	19	0,00	Unfit
26	0,693	44,037	19	0,00	Unfit
27	-0,994	42,858	19	0,00	Unfit
28	0,145	39,668	19	0,00	Unfit
29	0,135	39,239	19	0,00	Unfit
30	-0,708	35,412	19	0,01	Unfit
31	1,522	49,765	19	0,00	Unfit
32	0,243	32,960	19	0,02	Unfit
33	-0,953	31,620	19	0,03	Unfit
34	-0,604	12,670	19	0,86	Fit
35	0,614	30,348	19	0,05	Unfit
36	0,061	36,189	19	0,01	Unfit
37	-0,445	27,042	19	0,10	Fit
38	0,368	31,821	19	0,03	Unfit

NO. BUTIR	b^5	χ^2^6	$d.f.^7$	p^8	Kesimpulan ⁹
39	-0,543	11,600	19	0,90	Fit
40	-1,194	51,809	19	0,00	Unfit

Tabel tersebut memperlihatkan nilai tingkat kesulitan masing-masing butir, yang merentang dari $b = -2,230$ (terendah, sangat rendah/sulit) sampai $b = 1,956$ (tertinggi, tinggi/mudah). Tidak sebagaimana dalam analisis butir dengan pendekatan teori klasik, tingkat kesulitan butir dalam analisis butir dengan pendekatan modern ini distandarkan sehingga reratanya adalah $b = 0,00$ dan simpang baku 1,00.

Selanjutnya, tabel tersebut juga menunjukkan bahwa nilai Kai kuadrat merentang dari $\chi^2 = 10,832$ (terendah) sampai $\chi^2 = 86,841$ (tertinggi). Berdasarkan kriteria nilai probabilitas yang dapat diterima untuk menentukan apakah model didukung data empiris (*fit* [$p > 0,05$] atau *unfit* [$p \leq 0,05$]), tabel tersebut di atas juga memperlihatkan bahwa 19 butir soal tes masuk Bahasa Arab versi A memenuhi kriteria kesesuaian (*fit*) dan 21 butir yang lain tidak memenuhi kriteria kesesuaian (*unfit*). Hal ini menunjukkan lebih banyak butir yang tidak memenuhi kriteria kesesuaian dengan model dari pada yang memenuhi kriteria (sesuai model 1 parameter). Dengan kata lain, jumlah butir yang “baik” lebih sedikit dari pada jumlah butir yang “tidak baik.”

Berdasarkan hasil analisis butir dengan menggunakan kedua pendekatan teori tes, sebagaimana disajikan dalam tabel-tabel di atas, maka dapat dibandingkan kesimpulan dari keduanya, yang secara ringkas disajikan dalam tabel berikut ini.

Tabel 7. Perbandingan Hasil Analisis Butir Tes Bahasa Arab A Berdasarkan Pendekatan Teori Tes

Kesimpulan berdasarkan pendekatan		Teori Modern		Jumlah
		Baik	Tidak Baik	
Teori Klasik	Baik	15	11	26
	Tidak Baik	4	10	14
Jumlah		19	22	40

Tabel tersebut di atas memperlihatkan kesimpulan bahwa butir tes masuk Bahasa Arab versi A yang “baik” yang dihasilkan dari analisis dengan teori klasik lebih banyak dari pada yang dihasilkan oleh teori modern (masing-masing 26 butir [65 persen] dan 19 butir [47 persen]). Sebaliknya, butir yang “tidak baik” lebih banyak dihasilkan melalui analisis teori modern dari pada yang dihasilkan oleh analisis teori klasik. Hal ini menunjukkan bahwa teori tes modern memiliki kriteria yang lebih ketat dari pada teori klasik dalam menyeleksi butir tes yang baik.

Lebih lanjut, kedua pendekatan atau teori tes tidak selalu menghasilkan kesimpulan yang sama, meskipun lebih banyak kesimpulan yang sama (15 butir yang baik dan 10 butir yang tidak baik menurut keduanya) dari yang pada kesimpulan yang tidak sama (11 butir baik menurut teori klasik, tetapi tidak baik menurut teori modern; 4 butir tidak baik menurut teori klasik, tetapi baik menurut teori modern). Dari 26 butir tes masuk Bahasa Arab versi A yang dinyatakan baik berdasarkan analisis teori klasik, hanya 15 butir yang juga dinyatakan baik dan 11 butir yang lain dinyatakan tidak baik berdasarkan analisis teori modern. Sedangkan dari 14 butir yang dinyatakan tidak baik berdasarkan teori tes klasik, hanya 10 yang

dinyatakan tidak baik dan 4 butir yang lain dinyatakan baik berdasarkan hasil analisis dengan teori tes modern. Hal ini menunjukkan bahwa kedua teori tes tidak selalu sejalan dalam menentukan kualitas butir tes karena memang masing-masing menggunakan asumsi dan kriteria yang berbeda.

b. Reliabilitas Tes

Dalam rangka untuk mengetahui tingkat reliabilitas tes masuk Bahasa Arab versi A, data dari jawaban seluruh testee yang mengerjakan tes tersebut dianalisis dengan menggunakan teknik Cronbach alpha (α). Analisis dilakukan terhadap data dari seluruh (40) butir dan data dari butir-butir yang dinyatakan valid saja (33 butir) berdasarkan kriteria teori tes klasik. Berdasarkan data dari seluruh butir dihasilkan nilai alpha Cronbach, $\alpha = 0,889$. Sedangkan berdasarkan data dari butir-butir yang dinyatakan valid saja diperoleh nilai alpha Cronbach, $\alpha = 0,891$. Hasil ini menunjukkan bahwa tingkat reliabilitas butir-butir soal tes masuk Bahasa Arab versi A memiliki tingkat reliabilitas yang sangat tinggi, baik sebelum maupun sesudah dilakukan seleksi butir berdasarkan kriteria validitas butir.

Dengan nilai alpha Cronbach, $\alpha = 0,891$, dan nilai simpang baku, $SB = 7,438$, dapat dihitung nilai galat baku pengukuran atau *standard error of measurement*, $GBP/SEM = 2,482$. Hasil ini menunjukkan bahwa tes masuk Bahasa Arab versi A memiliki akurasi yang tinggi (tingkat kesalahan yang rendah) dalam mengukur kemampuan berbahasa Arab testee/calon mahasiswa baru angkatan

2012/2013. Dengan kata lain, skor tes yang diperoleh peserta tes cukup meyakinkan untuk mencerminkan/memperkirakan kemampuan peserta yang sebenarnya.

3. Kalibrasi Tes Masuk Bahasa Arab B

Kalibrasi terhadap respon testee pada masing-masing butir tes masuk Bahasa Arab versi B (untuk testee dengan nomor tes genap) telah dilakukan dengan menggunakan pendekatan teori tes klasik dan pendekatan teori tes modern. Secara lengkap, hasil analisis tersebut dapat dilihat dalam lampiran. Sedangkan secara ringkas hasil tersebut disajikan berikut ini.

a. Hasil Analisis Butir: Pendekatan Klasik dan Modern

Untuk menguji bagaimana kualitas butir-butir tes masuk Bahasa Arab versi B yang digunakan untuk menyeleksi calon mahasiswa baru tahun akademik 2012/2013, data respon peserta tes masuk Bahasa Arab versi B dianalisis dengan menggunakan pendekatan teori tes klasik. Analisis ini menghasilkan indeks tingkat kesulitan butir (p), koefisien korelasi antara skor butir dan skor seluruh butir (r_{pb}), fungsi pengecoh (kesimpulan dari hasil analisis). Hasil analisis tersebut secara ringkas disajikan dalam tabel berikut ini.

Tabel 8
Hasil Analisis Butir Tes Bahasa Arab B dengan Teori Klasik

NO. BUTIR	Tingkat Kesulitan (p)	r_{pb}	Kesimpulan berdasarkan			Kesimpulan Akhir ¹⁰
			Tingkat Kesulitan ¹¹	Validitas ¹²	Fungsi Pengecoh ¹³	
1	0,55	0,52	Baik	Valid	Baik	Baik
2	0,19	0,25	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
3	0,64	0,54	Baik	Valid	Baik	Baik
4	0,33	0,61	Baik	Valid	Baik	Baik
5	0,34	0,61	Baik	Valid	Baik	Baik
6	0,54	0,42	Baik	Valid	Baik	Baik
7	0,50	0,51	Baik	Valid	Baik	Baik
8	0,28	0,47	Baik	Valid	Baik	Baik
9	0,21	0,37	Baik	Valid	Baik	Baik
10	0,13	0,41	Tdk. Baik	Valid	Baik	Tdk. Baik
11	0,12	0,14	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
12	0,18	0,45	Tdk. Baik	Valid	Baik	Tdk. Baik
13	0,29	0,41	Baik	Valid	Baik	Baik
14	0,16	0,39	Tdk. Baik	Valid	Baik	Tdk. Baik
15	0,19	0,44	Tdk. Baik	Valid	Baik	Tdk. Baik
16	0,43	0,58	Baik	Valid	Baik	Baik
17	0,28	0,39	Baik	Valid	Baik	Baik
18	0,25	0,50	Baik	Valid	Baik	Baik
19	0,54	0,50	Baik	Valid	Baik	Baik
20	0,40	0,37	Baik	Valid	Baik	Baik
21	0,20	0,37	Tdk. Baik	Valid	Baik	Tdk. Baik
22	0,25	0,35	Baik	Valid	Baik	Baik
23	0,14	0,24	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
24	0,21	0,24	Baik	Tdk. Valid	Baik	Tdk. Baik
25	0,64	0,56	Baik	Valid	Baik	Baik
26	0,20	0,40	Baik	Valid	Baik	Baik
27	0,21	0,42	Baik	Valid	Baik	Baik
28	0,15	0,45	Tdk. Baik	Valid	Baik	Tdk. Baik
29	0,28	0,49	Baik	Valid	Baik	Baik

¹⁰ Kesimpulan: "Baik" apabila semua kriteria hasil analisis butir Baik/Valid.

¹¹ Tingkat Kesulitan: "Baik" apabila $p > 0,2$ dan $\leq 0,8$.

¹² Validitas: "Valid" apabila $r_{pb} \geq 0,30$.

¹³ Fungsi Pengecoh: "Baik" apabila setiap alternatif jawaban butir dipilih oleh $\geq 1\%$ dari seluruh testee.

NO. BUTIR	Tingkat Kesulitan (p)	r_{pb}	Kesimpulan berdasarkan			Kesimpulan Akhir ¹⁰
			Tingkat Kesulitan ¹¹	Validitas ¹²	Fungsi Pengecoh ¹³	
30	0,16	0,31	Tdk. Baik	Valid	Baik	Tdk. Baik
31	0,32	0,53	Baik	Valid	Baik	Baik
32	0,40	0,58	Baik	Valid	Baik	Baik
33	0,21	0,46	Baik	Valid	Baik	Baik
34	0,33	0,53	Baik	Valid	Baik	Baik
35	0,39	0,57	Baik	Valid	Baik	Baik
36	0,10	0,18	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
37	0,21	0,52	Baik	Valid	Baik	Baik
38	0,41	0,59	Baik	Valid	Baik	Baik
39	0,39	0,52	Baik	Valid	Baik	Baik
40	0,52	0,44	Baik	Valid	Baik	Baik

Dari tabel tersebut dapat dilihat bahwa tingkat kesulitan butir merentang dari 0,10 (terendah, sangat sulit) sampai 0,64 (tertinggi, sedang). Tabel tersebut juga menunjukkan tidak ada butir yang memiliki tingkat kesulitan tinggi atau sangat tinggi. Sedangkan rerata tingkat kesulitan adalah 0,307 (rendah). Hasil ini menunjukkan bahwa butir-butir tes masuk Bahasa Arab versi B termasuk kategori sulit bagi testee pada umumnya, sehingga skor yang mereka capai sangat rendah.

Dari tabel tersebut juga dapat dilihat bahwa dayabeda atau validitas butir (diukur dengan korelasi antara masing-masing butir dengan skor total) merentang dari $r_{pb} = 0,140$ (terendah, tingkat validitas sangat rendah) sampai $r_{pb} = 0,610$ (tertinggi, tingkat validitas tinggi) dan tidak ada yang bernilai negatif. Berdasarkan hasil tersebut, rerata tingkat validitas butir adalah $r_{pb} = 0,441$ (sedang). Hasil ini memberikan petunjuk bahwa secara umum tingkat validitas butir-butir soal tes masuk Bahasa Arab versi B termasuk

kategori sedang sehingga cukup memberikan keyakinan untuk membedakan tingkat kemampuan berbahasa Arab testee yang bervariasi.

Berdasarkan kriteria yang telah ditetapkan sebagaimana dalam bab sebelumnya, tabel tersebut di atas juga memperlihatkan bahwa sebagian besar (28 [70 persen] dari 40) butir tes masuk Bahasa Arab versi B termasuk kategori baik, sedangkan 12 butir (30 persen) yang lain termasuk kategori tidak baik. Ke-28 butir tersebut memenuhi kriteria tingkat kesulitan dan validitas/daya beda butir yang telah ditentukan. Dari 12 butir yang tidak baik, 4 butir tidak memenuhi kedua kriteria, 7 butir tidak memenuhi kriteria tingkat kesulitan saja, dan hanya 1 butir yang tidak memenuhi kriteria validitas saja. Sebagaimana versi A, semua butir soal versi B telah memenuhi kriteria fungsi pengecoh, yakni setiap pilhan jawaban telah dipilih oleh minimal 1 persen dari jumlah testee.

Lebih lanjut, data hasil tes masuk Bahasa Arab versi B juga dianalisis dengan menggunakan teori tes modern 1 parameter, atau *Rash model*. Analisis ini menghasilkan nilai indeks tingkat kesulitan butir dan probabilitas kesesuaiannya melalui uji fit/unfit pada taraf signifikansi 5 persen. Hasil analisis secara ringkas disajikan dalam tabel di bawah ini.

Tabel 9. Hasil Analisis Butir Tes dengan IRT 1 Parameter

NO. BUTIR	b	χ^2	df	p	Kesimpulan
1	-1,399	30,943	19	0,04	Unfit
2	0,726	94,445	19	0,00	Unfit
3	-1,890	31,991	19	0,03	Unfit
4	-0,218	30,762	19	0,04	Unfit
5	-0,281	31,225	19	0,04	Unfit

NO. BUTIR	<i>b</i>	χ^2	<i>df</i>	<i>p</i>	Kesimpulan
6	-1,331	50,776	19	0,00	Unfit
7	-1,139	19,540	19	0,42	Fit
8	0,068	19,189	19	0,44	Fit
9	0,518	20,425	19	0,37	Fit
10	1,204	20,490	19	0,37	Fit
11	1,306	70,190	19	0,00	Unfit
12	0,792	48,327	19	0,00	Unfit
13	0,018	26,754	19	0,11	Fit
14	0,958	15,587	19	0,68	Fit
15	0,675	27,103	19	0,10	Fit
16	-0,757	28,858	19	0,07	Fit
17	0,088	37,204	19	0,01	Unfit
18	0,251	12,132	19	0,88	Fit
19	-1,331	43,070	19	0,00	Unfit
20	-0,605	37,617	19	0,01	Unfit
21	0,613	21,895	19	0,29	Fit
22	0,230	31,880	19	0,03	Unfit
23	1,139	55,888	19	0,00	Unfit
24	0,506	60,043	19	0,00	Unfit
25	-1,890	32,096	19	0,03	Unfit
26	0,577	20,979	19	0,34	Fit
27	0,553	10,623	19	0,94	Fit
28	0,987	33,977	19	0,02	Unfit
29	0,098	11,570	19	0,90	Fit
30	0,929	30,812	19	0,04	Unfit
31	-0,144	22,678	19	0,25	Fit
32	-0,596	31,685	19	0,03	Unfit
33	0,518	23,005	19	0,24	Fit
34	-0,236	17,512	19	0,56	Fit
35	-0,562	26,851	19	0,11	Fit
36	1,572	50,429	19	0,00	Unfit
37	0,518	26,521	19	0,12	Fit
38	-0,647	42,264	19	0,00	Unfit
39	-0,579	16,819	19	0,60	Fit
40	-1,239	27,812	19	0,09	Fit

Dalam tabel tersebut dapat dilihat nilai tingkat kesulitan masing-masing butir tes masuk Bahasa Arab versi B, yang merentang dari $b = -1,890$ (terendah, rendah) sampai $b = 1,572$ (tertinggi, cukup tinggi). Tidak sebagaimana dalam pendekatan teori klasik, tingkat kesulitan butir dalam analisis butir dengan pendekatan modern ini distandarkan sehingga reratanya adalah $b = 0,00$.

Tabel diatas juga memperlihatkan nilai Kai kuadrat yang merentang dari $\chi^2 = 10,623$ (terendah) sampai $\chi^2 = 94,445$ (tertinggi). Berdasarkan kriteria nilai probabilitas/ taraf signifikansi yang dapat diterima untuk menentukan apakah model didukung data empiris (*fit* [$p > 0,05$] atau *unfit* [$p \leq 0,05$]), tabel tersebut di atas juga memperlihatkan bahwa 26 butir soal tes masuk Bahasa Arab versi B memenuhi kriteria (*fit*) dan 14 butir yang lain tidak memenuhi kriteria (*unfit*). Tidak seperti versi A, hal ini menunjukkan bahwa butir yang memenuhi kriteria kesesuaian dengan model lebih banyak dari pada yang tidak memenuhi kriteria (model 1 parameter).

Berdasarkan hasil analisis butir dengan menggunakan kedua pendekatan teori tes, sebagaimana disajikan dalam tabel-tabel di atas, maka dapat dibandingkan kesimpulan dari keduanya, yang secara ringkas disajikan dalam tabel berikut ini.

Tabel 10. Perbandingan Hasil Analisis Butir Tes masuk Bahasa Arab B Berdasarkan Pendekatan Teori Tes

Kesimpulan berdasarkan		Teori Modern		Jumlah
		Baik	Tidak Baik	
Teori Klasik	Baik	16	12	28
	Tidak Baik	4	8	12
Jumlah		20	20	40

Dari tabel tersebut di atas dapat diketahui bahwa butir tes masuk Bahasa Arab versi B yang “baik” yang dihasilkan dari analisis dengan teori klasik lebih banyak dari pada yang dihasilkan oleh teori modern (masing-masing 28 butir [70 persen] dan 20 butir [50 persen]). Jumlah butir yang “tidak baik” yang dihasilkan melalui analisis teori modern sama dengan jumlah butir yang baik, yakni 20 butir. Sementara jumlah butir yang tidak baik yang dihasilkan oleh analisis teori klasik lebih kecil, yakni 12 butir (30 persen). Hal ini juga menunjukkan bahwa teori tes modern memiliki kriteria yang lebih ketat dari pada teori klasik dalam menyeleksi butir tes yang baik.

Lebih lanjut, sebagaimana dalam tes Bahasa Arab versi A, dalam versi B kedua pendekatan atau teori tes juga tidak selalu menghasilkan kesimpulan yang sama, meskipun lebih banyak kesimpulan yang sama (16 butir yang baik dan 8 butir yang tidak baik menurut keduanya) dari yang pada kesimpulan yang tidak sama (12 butir baik menurut teori klasik, tetapi tidak baik menurut teori modern; 4 butir tidak baik menurut teori klasik, tetapi baik menurut teori modern). Dari 28 butir tes masuk Bahasa Arab versi B yang dinyatakan baik berdasarkan analisis teori klasik, hanya 16 butir yang juga dinyatakan baik dan 12 butir yang lain dinyatakan tidak baik berdasarkan analisis teori modern. Sedangkan dari 12 butir yang dinyatakan tidak baik berdasarkan teori tes klasik, hanya 8 yang dinyatakan tidak baik dan 4 butir yang lain dinyatakan baik berdasarkan hasil analisis dengan teori tes modern. Hal ini juga

menunjukkan bahwa kedua teori tes tidak selalu sejalan dalam menentukan kualitas butir tes karena memang masing-masing menggunakan asumsi dan kriteria yang berbeda.

b. Reliabilitas Tes

Sebagaimana untuk versi A, data dari jawaban seluruh testee yang mengerjakan tes masuk Bahasa Arab versi B (621 testee) dianalisis dengan menggunakan teknik Cronbach alpha (α) untuk mengetahui tingkat reliabilitasnya. Analisis dilakukan terhadap data dari seluruh (40) butir dan data dari butir-butir yang dinyatakan valid saja berdasarkan kriteria teori tes klasik (35 butir). Berdasarkan data dari seluruh butir maupun dari butir-butir yang dinyatakan valid saja dihasilkan nilai alpha Cronbach, $\alpha = 0,898$. Hasil ini menunjukkan bahwa tingkat reliabilitas butir-butir soal tes masuk Bahasa Arab versi B memiliki tingkat reliabilitas yang sama-sama sangat tinggi.

Dengan nilai $\alpha = 0,898$ dan nilai simpang baku $SB = 7,835$ dapat dihitung nilai galat baku pengukuran atau standard error of measurement, $GBP/SEM = 2,505$. Hal ini menunjukkan bahwa tes masuk Bahasa Arab versi B memiliki akurasi yang tinggi (tingkat kesalahan yang rendah) dalam mengukur kemampuan berbahasa Arab testee/calon mahasiswa baru angkatan 2012/2013. Dengan kata lain, berdasarkan teori tes klasik skor tes yang diperoleh peserta tes cukup meyakinkan untuk mencerminkan/memperkirakan kemampuan peserta yang mendekati sebenarnya karena tingkat kesalahan yang kecil.

4. Kalibrasi Tes Masuk Bahasa Inggris A

Sebagaimana tes masuk Bahasa Arab, data masing-masing butir tes masuk Bahasa Inggris versi A dianalisis dengan menggunakan pendekatan tes klasik dan modern untuk mengetahui kualitasnya. Penggunaan pendekatan klasik dalam analisis ini menghasilkan indeks tingkat kesulitan butir, validitas atau daya beda butir, dan fungsi pengecoh masing-masing pilihan jawaban butir. Sedangkan penggunaan pendekatan modern dimaksudkan untuk menghasilkan karakteristik laten butir yang berupa tingkat kesulitan butir. Selanjutnya, berdasarkan kriteria yang telah ditentukan sebagaimana telah disajikan dalam bab sebelumnya diambil kesimpulan apakah butir tersebut dinyatakan baik (bila semua kriteria terpenuhi) atau tidak baik (bila setidaknya salah satu kriteria tidak terpenuhi). Secara lengkap, hasil analisis tersebut dapat dilihat dalam lampiran. Sedangkan secara ringkas hasil tersebut disajikan berikut ini.

a. Hasil Analisis Butir: Pendekatan Klasik dan Modern

Untuk menguji apakah butir-butir tes masuk Bahasa Inggris A sebagai calon mahasiswa baru memiliki kualitas yang baik atau tidak, data dari respon peserta tes masuk untuk masing-masing pertamanya dianalisis dengan menggunakan pendekatan teori tes klasik untuk menghasilkan indeks tingkat kesulitan butir (p), koefisien korelasi point biserial (r_{pb}) antara skor masing-masing butir secara terpisah dan skor total (seluruh butir), fungsi pengecoh (dikemukakan kesimpulan dari hasil analisis). Tabel 11 berikut ini menyajikan ringkasan hasil analisis tersebut.

Tabel 11
 Hasil Analisis Butir Tes Bahasa Inggris A dengan Teori Klasik

NO. Butir	Tingkat Kesulitan (p)	r_{pb}	Kesimpulan berdasarkan			Kesimpulan Akhir ¹⁴
			Tingkat Kesulitan ¹⁵	Validitas ¹⁶	Fungsi Pengecoh ¹⁷	
1	0,073	0,114	Tdk Baik	Tdk Valid	Baik	Tdk Baik
2	0,145	0,186	Tdk Baik	Tdk Valid	Baik	Tdk Baik
3	0,287	0,268	Baik	Tdk Valid	Baik	Tdk Baik
4	0,125	0,172	Tdk Baik	Tdk Valid	Baik	Tdk Baik
5	0,239	0,177	Baik	Tdk Valid	Baik	Tdk Baik
6	0,402	0,247	Baik	Tdk Valid	Baik	Tdk Baik
7	0,520	0,310	Baik	Valid	Baik	Baik
8	0,156	0,240	Tdk Baik	Tdk Valid	Baik	Tdk Baik
9	0,366	0,223	Baik	Tdk Valid	Baik	Tdk Baik
10	0,205	0,253	Baik	Tdk Valid	Baik	Tdk Baik
11	0,114	0,126	Tdk Baik	Tdk Valid	Baik	Tdk Baik
12	0,112	0,140	Tdk Baik	Tdk Valid	Baik	Tdk Baik
13	0,524	0,396	Baik	Valid	Baik	Baik
14	0,301	0,327	Baik	Valid	Baik	Baik
15	0,189	0,239	Tdk Baik	Tdk Valid	Baik	Tdk Baik
16	0,153	0,145	Tdk Baik	Tdk Valid	Baik	Tdk Baik
17	0,158	0,180	Tdk Baik	Tdk Valid	Baik	Tdk Baik
18	0,116	0,267	Tdk Baik	Tdk Valid	Baik	Tdk Baik
19	0,179	0,208	Tdk Baik	Tdk Valid	Baik	Tdk Baik
20	0,466	0,241	Baik	Tdk Valid	Baik	Tdk Baik
21	0,495	0,429	Baik	Valid	Baik	Baik
22	0,101	0,155	Tdk Baik	Tdk Valid	Baik	Tdk Baik
23	0,054	0,130	Tdk Baik	Tdk Valid	Baik	Tdk Baik
24	0,267	0,350	Baik	Valid	Baik	Baik
25	0,259	0,332	Baik	Valid	Baik	Baik
26	0,153	0,213	Tdk Baik	Tdk Valid	Baik	Tdk Baik
27	0,572	0,497	Baik	Valid	Baik	Baik

¹⁴ Kesimpulan: "Baik" apabila semua kriteria hasil analisis butir Baik/Valid.

¹⁵ Tingkat Kesulitan: "Baik" apabila $p > 0,2$ dan $\leq 0,8$.

¹⁶ Validitas: "Valid" apabila $r_{pb} \geq 0,30$.

¹⁷ Fungsi Pengecoh: "Baik" apabila setiap alternatif jawaban butir dipilih oleh $\geq 1\%$ dari seluruh testee.

NO. Butir	Tingkat Kesulitan (p)	r_{pb}	Kesimpulan berdasarkan			Kesimpulan Akhir ¹⁴
			Tingkat Kesulitan ¹⁵	Validitas ¹⁶	Fungsi Pengecoh ¹⁷	
28	0,316	0,444	Baik	Valid	Baik	Baik
29	0,163	0,375	Tdk Baik	Valid	Baik	Tdk Baik
30	0,077	0,235	Tdk Baik	Tdk Valid	Baik	Tdk Baik
31	0,124	0,250	Tdk Baik	Tdk Valid	Baik	Tdk Baik
32	0,169	0,297	Tdk Baik	Tdk Valid	Baik	Tdk Baik
33	0,251	0,354	Baik	Valid	Baik	Baik
34	0,371	0,328	Baik	Valid	Baik	Baik
35	0,407	0,355	Baik	Valid	Baik	Baik
36	0,241	0,285	Baik	Tdk Valid	Baik	Tdk Baik
37	0,109	0,206	Tdk Baik	Tdk Valid	Baik	Tdk Baik
38	0,109	0,201	Tdk Baik	Tdk Valid	Baik	Tdk Baik
39	0,085	0,172	Tdk Baik	Tdk Valid	Baik	Tdk Baik
40	0,137	0,308	Tdk Baik	Valid	Baik	Tdk Baik

Dari tabel tersebut dapat dilihat bahwa tingkat kesulitan butir tes masuk Bahasa Inggris A merentang dari 0,05 (terendah, kategori sangat rendah) sampai 0,57 (tertinggi, kategori sedang). Tidak ada butir tes yang memiliki tingkat kesulitan sangat tinggi. Sedangkan rerata tingkat kesulitan adalah 0,23 (rendah). Hasil ini menunjukkan bahwa butir-butir tes masuk Bahasa Inggris versi A pada umumnya termasuk kategori sangat sulit bagi peserta tes masuk yang bernomor ganjil, sehingga skor yang mereka capai pun sangat rendah.

Tabel tersebut juga memperlihatkan bahwa dayabeda atau validitas butir (diukur dengan korelasi antara masing-masing butir dengan skor total) yang merentang dari $r_{pb} = 0,114$ (terendah, tingkat validitas sangat rendah) sampai $r_{pb} = 0,497$ (tertinggi, tingkat validitas sedang) dan tidak ada yang negatif. Sedangkan rerata tingkat validitasnya adalah $r_{pb} = 0,259$ (rendah). Hasil ini

menunjukkan bahwa secara umum tingkat validitas butir-butir soal tes masuk Bahasa Inggris versi A rendah, kurang menyakinkan untuk menghasilkan ukuran yang membedakan tingkat kemampuan berbahasa Inggris.

Lebih lanjut, tabel tersebut juga memperlihatkan bahwa hanya sebagian kecil (11 [27,5 persen] dari 40) butir tes masuk Bahasa Inggris versi A termasuk kategori baik, dan sebagian besar, 29 butir (80 persen), yang lain tidak baik. Ke-11 butir yang baik tersebut memenuhi kriteria tingkat kesulitan dan sekaligus validitas/daya beda butir yang telah ditentukan. Sedangkan dari 29 butir yang tidak baik, 20 butir tidak memenuhi kedua kriteria, 2 butir tidak memenuhi kriteria tingkat kesulitan saja, dan hanya 7 butir yang lain tidak memenuhi kriteria validitas saja. Sedangkan berdasarkan kriteria fungsi pengecoh, semua butir telah memenuhi kriteria.

Selanjutnya, data yang sama dari hasil tes masuk Bahasa Inggris versi A dianalisis dengan menggunakan teori tes modern 1 parameter atau *Rash model* untuk menghasilkan nilai indeks tingkat kesulitan butir dan probabilitas kesesuaiannya melalui uji fit/unfit pada taraf signifikansi 5 persen. Hasil analisis secara ringkas disajikan dalam tabel di bawah ini.

Tabel 12. Hasil Analisis Butir Tes dengan IRT 1 Parameter

NO. Butir	b	χ^2	df	p	Kesimpulan
1	1,172	21,604	17	0,200	Fit
2	0,429	16,216	17	0,509	Fit
3	-0,451	14,733	17	0,615	Fit
4	0,625	13,331	17	0,714	Fit
5	-0,203	25,697	17	0,080	Fit
6	-1,005	21,578	17	0,201	Fit

NO. Butir	b	χ^2	df	p	Kesimpulan
7	-1,525	29,866	17	0,027	Unfit
8	0,364	14,427	17	0,637	Fit
9	-0,848	26,756	17	0,062	Fit
10	0,022	10,183	17	0,896	Fit
11	0,750	30,299	17	0,024	Unfit
12	0,750	27,86	17	0,047	Unfit
13	-1,561	21,103	17	0,222	Fit
14	-0,509	19,639	17	0,293	Fit
15	0,129	14,999	17	0,596	Fit
16	0,390	17,125	17	0,446	Fit
17	0,339	23,25	17	0,141	Fit
18	0,734	15,731	17	0,543	Fit
19	0,196	22,607	17	0,162	Fit
20	-1,302	31,608	17	0,017	Unfit
21	-1,439	15,779	17	0,540	Fit
22	0,853	25,183	17	0,091	Fit
23	1,573	23,637	17	0,130	Fit
24	-0,339	14,593	17	0,625	Fit
25	-0,295	10,345	17	0,889	Fit
26	0,339	9,967	17	0,905	Fit
27	-1,759	26,241	17	0,070	Fit
28	-0,583	29,029	17	0,034	Unfit
29	0,302	35,736	17	0,005	Unfit
30	1,172	30,483	17	0,023	Unfit
31	0,655	22,607	17	0,162	Fit
32	0,266	14,049	17	0,664	Fit
33	-0,240	16,491	17	0,489	Fit
34	-0,863	20,98	17	0,227	Fit
35	-1,019	15,584	17	0,553	Fit
36	-0,203	12,122	17	0,793	Fit
37	0,767	17,375	17	0,429	Fit
38	0,750	15,905	17	0,531	Fit
39	1,043	9,482	17	0,924	Fit
40	0,524	12,024	17	0,799	Fit

Tabel tersebut memperlihatkan nilai tingkat kesulitan masing-masing butir, yang merentang dari $b = -1,759$ (terendah, sangat rendah/sulit) sampai $b = 1,573$ (tertinggi, tinggi/mudah). Tidak sebagaimana dalam analisis butir dengan pendekatan teori klasik, tingkat kesulitan butir dalam analisis butir dengan pendekatan modern ini distandarkan sehingga reratanya adalah $b = 0,00$ dan simpang baku 1,00.

Selanjutnya, tabel tersebut juga menunjukkan bahwa nilai Kai kuadrat merentang dari $\chi^2 = 9,482$ (terendah) sampai $\chi^2 = 35,736$ (tertinggi). Berdasarkan kriteria nilai probabilitas yang dapat diterima untuk menentukan apakah model didukung data empiris (*fit* [$p > 0,05$] atau *unfit* [$p \leq 0,05$]), tabel tersebut di atas juga memperlihatkan bahwa 33 butir soal tes masuk Bahasa Inggris versi A memenuhi kriteria kesesuaian (*fit*) dan 7 butir yang lain tidak memenuhi kriteria kesesuaian (*unfit*). Hal ini menunjukkan lebih banyak butir yang memenuhi kriteria kesesuaian dengan model dari pada yang tidak memenuhi kriteria (sesuai model 1 parameter). Dengan kata lain, jumlah butir yang “baik” lebih banyak dari pada jumlah butir yang “tidak baik.”

Berdasarkan hasil analisis butir dengan menggunakan kedua pendekatan teori tes, sebagaimana disajikan dalam tabel-tabel di atas, maka dapat dibandingkan kesimpulan dari keduanya, yang secara ringkas disajikan dalam tabel berikut ini.

Tabel 13
Perbandingan Hasil Analisis Butir Tes Bahasa Inggris A
Berdasarkan Pendekatan Teori Tes

Kesimpulan berdasarkan pendekatan		Teori Modern		Jumlah
		Baik	Tidak Baik	
Teori Klasik	Baik	9	3	12
	Tidak Baik	24	4	28
Jumlah		33	7	40

Tabel tersebut di atas memperlihatkan kesimpulan bahwa butir tes masuk Bahasa Inggris versi A yang “baik” yang dihasilkan dari analisis dengan teori klasik jauh lebih sedikit dari pada yang dihasilkan oleh teori modern (masing-masing 12 butir [30 persen] dan 33 butir [82,5 persen]). Sebaliknya, butir yang “tidak baik” lebih banyak dihasilkan melalui analisis dengan pendekatan teori tes klasik dari pada yang dihasilkan melalui analisis dengan pendekatan teori tes modern. Hal ini menunjukkan bahwa teori tes klasik memiliki kriteria yang lebih ketat dari pada teori modern dalam menyeleksi butir tes yang baik.

Lebih lanjut, kedua pendekatan atau teori tes tidak selalu menghasilkan kesimpulan yang sama, meskipun lebih sedikit kesimpulan yang sama (9 butir yang baik dan 4 butir yang tidak baik menurut keduanya) dari pada kesimpulan yang tidak sama (3 butir baik menurut teori klasik, tetapi tidak baik menurut teori modern; 24 butir tidak baik menurut teori klasik, tetapi baik menurut teori modern). Dari 12 butir tes masuk Bahasa Inggris versi A yang dinyatakan baik berdasarkan analisis teori klasik, 9 butir juga dinyatakan baik dan 3 butir yang lain dinyatakan tidak baik

berdasarkan analisis teori modern. Sedangkan dari 28 butir yang dinyatakan tidak baik berdasarkan teori tes klasik, 24 dinyatakan baik dan 4 butir yang lain dinyatakan tidak baik berdasarkan hasil analisis dengan teori tes modern. Hal ini menunjukkan bahwa kedua teori tes tidak selalu sejalan dalam menentukan kualitas butir tes karena memang masing-masing menggunakan asumsi dan kriteria yang berbeda.

b. Reliabilitas Tes

Dalam rangka untuk mengetahui tingkat reliabilitas tes masuk Bahasa Inggris versi A, data dari jawaban seluruh testee yang mengerjakan tes tersebut dianalisis dengan menggunakan teknik Cronbach alpha (α). Analisis dilakukan terhadap data dari seluruh (40) butir dan data dari butir-butir yang dinyatakan valid saja (12 butir) berdasarkan kriteria teori tes klasik. Berdasarkan data dari seluruh butir dihasilkan nilai alpha Cronbach, $\alpha = 0,666$. Sedangkan berdasarkan data dari butir-butir yang dinyatakan valid saja diperoleh nilai alpha Cronbach, $\alpha = 0,641$. Hasil ini menunjukkan bahwa tingkat reliabilitas butir-butir soal tes masuk Bahasa Inggris versi A memiliki tingkat reliabilitas yang sedang, baik sebelum maupun sesudah dilakukan seleksi butir berdasarkan kriteria validitas butir.

Dengan nilai $\alpha = 0,666$ dan nilai simpang baku $SB = 4,254$ dapat dihitung nilai galat baku pengukuran atau standard error of measurement, $GBP/SEM = 2,458$. Hal ini menunjukkan bahwa tes masuk Bahasa Inggris versi A memiliki akurasi yang rendah (tingkat kesalahan yang tinggi) dalam mengukur kemampuan berbahasa

Inggris testee/calon mahasiswa baru angkatan 2012/2013. Dengan kata lain, skor tes yang diperoleh peserta tes kurang meyakinkan untuk mencerminkan/memperkirakan kemampuan peserta yang sebenarnya.

5. Kalibrasi Tes Masuk Bahasa Inggris B

Sebagaimana terhadap respon pada tes yang lain, kalibrasi terhadap respon testee pada masing-masing butir tes masuk Bahasa Inggris versi B (untuk testee dengan nomor tes genap) telah dilakukan dengan menggunakan pendekatan teori tes klasik dan teori tes modern. Analisis dengan kedua pendekatan teori tes tersebut dilakukan terhadap data yang sama. Secara lengkap, hasil analisis tersebut dapat dilihat dalam lampiran. Sedangkan secara ringkas hasil tersebut disajikan berikut ini.

a. Hasil Analisis Butir: Pendekatan Klasik dan Modern

Sebagaimana untuk tes dan versi yang lain, untuk mengetahui bagaimana kualitas butir-butir tes masuk Bahasa Inggris versi B yang digunakan untuk menyeleksi calon mahasiswa baru tahun akademik 2012/2013, pertama-tama data respon peserta tes masuk Bahasa Inggris versi B (peserta dengan nomor tes genap) dinalisis dengan menggunakan pendekatan teori tes klasik. Hasil yang diperoleh dari analisis ini adalah indeks tingkat kesulitan butir (p), koefisien korelasi antara skor butir dan skor seluruh butir (r_{pb}), fungsi pengecoh (kesimpulan dari hasil analisis). Hasil analisis tersebut secara ringkas disajikan dalam tabel berikut

Tabel 14. Hasil Analisis Butir Tes Bahasa Inggris B dengan Teori Klasik

NO. Butir	Tingkat Kesulitan (p)	r_{pb}	Kesimpulan berdasarkan			Kesimpulan Akhir ¹⁸
			Tingkat Kesulitan ¹⁹	Validitas ²⁰	Fungsi Pengecoh ²¹	
1	0,28	0,28	Baik	Tdk. Valid	Baik	Baik
2	0,31	0,18	Baik	Tdk. Valid	Baik	Baik
3	0,10	0,11	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
4	0,17	0,29	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
5	0,17	0,19	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
6	0,25	0,09	Baik	Tdk. Valid	Baik	Tdk. Baik
7	0,10	0,09	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
8	0,17	0,21	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
9	0,13	0,11	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
10	0,48	0,30	Baik	Valid	Baik	Baik
11	0,16	0,29	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
12	0,69	0,42	Baik	Valid	Baik	Baik
13	0,61	0,48	Baik	Valid	Baik	Baik
14	0,15	0,32	Tdk. Baik	Valid	Baik	Tdk. Baik
15	0,22	0,42	Baik	Valid	Baik	Baik
16	0,26	0,39	Baik	Valid	Baik	Baik
17	0,21	0,25	Baik	Tdk. Valid	Baik	Tdk. Baik
18	0,05	0,07	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
19	0,33	0,27	Baik	Tdk. Valid	Baik	Tdk. Baik
20	0,37	0,43	Baik	Valid	Baik	Baik
21	0,81	0,53	Tdk. Baik	Valid	Tdk. Baik	Tdk. Baik
22	0,65	0,50	Baik	Valid	Baik	Baik
23	0,73	0,54	Baik	Valid	Baik	Baik
24	0,52	0,55	Baik	Valid	Baik	Baik
25	0,44	0,43	Baik	Valid	Baik	Baik
26	0,76	0,56	Baik	Valid	Baik	Baik
27	0,38	0,38	Baik	Valid	Baik	Baik
28	0,42	0,53	Baik	Valid	Baik	Baik
29	0,41	0,51	Baik	Valid	Baik	Baik
30	0,22	0,42	Baik	Valid	Baik	Baik

¹⁸ Kesimpulan: "Baik" apabila semua kriteria hasil analisis butir Baik/Valid.

¹⁹ Tingkat Kesulitan: "Baik" apabila $p > 0,2$ dan $\leq 0,8$.

²⁰ Validitas: "Valid" apabila $r_{pb} \geq 0,30$.

²¹ Fungsi Pengecoh: "Baik" apabila setiap alternatif jawaban butir dipilih oleh $\geq 1\%$ dari seluruh testee.

NO. Butir	Tingkat Kesulitan (p)	r_{pb}	Kesimpulan berdasarkan			Kesimpulan Akhir ¹⁸
			Tingkat Kesulitan ¹⁹	Validitas ²⁰	Fungsi Pengecoh ²¹	
31	0,24	0,39	Baik	Valid	Baik	Baik
32	0,22	0,45	Baik	Valid	Baik	Baik
33	0,33	0,36	Baik	Valid	Baik	Baik
34	0,33	0,36	Baik	Valid	Baik	Baik
35	0,53	0,54	Baik	Valid	Baik	Baik
36	0,11	0,13	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
37	0,08	0,15	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
38	0,18	0,29	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
39	0,07	0,05	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik
40	0,06	0,04	Tdk. Baik	Tdk. Valid	Baik	Tdk. Baik

Tabel tersebut memperlihatkan tingkat kesulitan butir-butir tes Bahasa Inggris versi B, yang merentang dari 0,05 (terendah, sangat rendah) sampai 0,81 (tertinggi, sangat tinggi). Tabel tersebut juga menunjukkan bahwa butir-butir tes Bahasa Inggris versi B memiliki tingkat kesulitan sangat variatif, merentang dari sangat sulit sampai sangat mudah. Sedangkan rerata tingkat kesulitan adalah 0,318 (rendah). Hasil ini menunjukkan bahwa butir-butir tes masuk Bahasa Inggris versi B termasuk kategori sulit bagi testee bernomor tes genap, sehingga skor yang mereka capai sangat rendah.

Dari tabel tersebut juga dapat dilihat bahwa dayabeda atau validitas butir (diukur dengan korelasi antara masing-masing butir dengan skor total) merentang dari $r_{pb} = 0,04$ (terendah, tingkat validitas sangat rendah) sampai $r_{pb} = 0,560$ (tertinggi, tingkat validitas sedang) dan tidak ada yang bernilai negatif. Berdasarkan hasil tersebut, rerata tingkat validitas butir adalah $r_{pb} = 0,323$ (rendah). Hasil ini memberikan petunjuk bahwa secara umum tingkat validitas butir-butir soal tes masuk Bahasa Inggris versi B termasuk

kategori rendah sehingga kurang memberikan keyakinan yang memadai untuk membedakan tingkat kemampuan berbahasa Inggris testee yang bervariasi.

Berdasarkan kriteria yang telah ditetapkan sebagaimana dalam bab sebelumnya, tabel tersebut di atas juga memperlihatkan bahwa sebagian besar (22 [55 persen] dari 40) butir tes masuk Bahasa Inggris versi B termasuk kategori baik, sedangkan 18 butir (45 persen) yang lain termasuk kategori tidak baik. Ke-22 butir yang baik tersebut memenuhi kriteria tingkat kesulitan dan validitas/daya beda butir yang telah ditentukan. Dari 18 butir yang tidak baik, 14 butir tidak memenuhi kedua kriteria, 2 butir tidak memenuhi kriteria tingkat kesulitan saja, dan hanya 2 butir yang tidak memenuhi kriteria validitas saja. Tidak sebagaimana versi A, ada 1 butir soal versi B yang tidak memenuhi kriteria fungsi pengecoh, yakni pilihan jawaban telah dipilih oleh kurang dari 1 persen dari jumlah testee.

Lebih lanjut, data hasil tes masuk Bahasa Inggris versi B juga dianalisis dengan menggunakan teori tes modern 1 parameter, atau *Rash model*. Analisis ini menghasilkan nilai indeks tingkat kesulitan butir dan probabilitas kesesuaiannya melalui uji fit/unfit pada taraf signifikansi 5 persen. Tabel di bawah ini menyajikan hasil analisis secara ringkas.

Tabel 15. Hasil Analisis Butir Tes Bahasa Inggris B dengan IRT 1 Parameter

NO. Butir	b	Chi Sq.	df	p	Kesimpulan
1	0,043	42,099	19	0,00	Unfit
2	-0,096	55,601	19	0,00	Unfit
3	1,382	53,314	19	0,00	Unfit
4	0,749	19,574	19	0,42	Fit

NO. Butir	b	Chi Sq.	df	p	Kesimpulan
5	0,761	42,136	19	0,00	Unfit
6	0,198	85,536	19	0,00	Unfit
7	1,364	93,776	19	0,00	Unfit
8	0,749	36,134	19	0,01	Unfit
9	1,100	43,048	19	0,00	Unfit
10	-0,936	34,887	19	0,01	Unfit
11	0,862	19,599	19	0,42	Fit
12	-1,994	15,203	19	0,71	Fit
13	-1,564	21,790	19	0,29	Fit
14	0,902	19,446	19	0,43	Fit
15	0,393	30,345	19	0,05	Unfit
16	0,142	17,810	19	0,54	Fit
17	0,466	16,815	19	0,60	Fit
18	2,131	68,523	19	0,00	Unfit
19	-0,188	31,694	19	0,03	Unfit
20	-0,415	18,358	19	0,50	Fit
21	-2,758	30,961	19	0,04	Unfit
22	-1,781	20,187	19	0,38	Fit
23	-2,245	30,562	19	0,05	Unfit
24	-1,133	32,496	19	0,03	Unfit
25	-0,762	18,869	19	0,47	Fit
26	-2,458	40,645	19	0,00	Unfit
27	-0,454	23,312	19	0,22	Fit
28	-0,647	39,434	19	0,00	Unfit
29	-0,624	38,693	19	0,00	Unfit
30	0,424	24,201	19	0,19	Fit
31	0,274	25,236	19	0,15	Fit
32	0,383	26,214	19	0,12	Fit
33	-0,213	24,359	19	0,18	Fit
34	-0,188	14,901	19	0,73	Fit
35	-1,163	28,812	19	0,07	Fit
36	1,276	44,728	19	0,00	Unfit
37	1,691	25,923	19	0,13	Fit
38	0,700	15,988	19	0,66	Fit
39	1,738	53,968	19	0,00	Unfit
40	1,891	36,810	19	0,01	Unfit

Berdasarkan tabel tersebut dapat dilihat nilai tingkat kesulitan masing-masing butir tes masuk Bahasa Inggris versi B, yang merentang dari $b = -2,758$ (terendah, rendah) sampai $b = 2,131$ (tertinggi, tinggi). Tidak sebagaimana dalam pendekatan teori klasik, tingkat kesulitan butir dalam analisis butir dengan pendekatan modern ini distandarkan sehingga reratanya adalah $b = 0,00$.

Tabel diatas juga memperlihatkan nilai Kai kuadrat yang merentang dari $\chi^2 = 14,901$ (terendah) sampai $\chi^2 = 93,776$ (tertinggi). Berdasarkan kriteria nilai probabilitas yang dapat diterima untuk menentukan apakah model didukung data empiris (*fit* [$p > 0,05$] atau *unfit* [$p \leq 0,05$]), tabel tersebut di atas juga memperlihatkan bahwa 19 butir soal tes masuk Bahasa Inggris versi B memenuhi kriteria (*fit*) dan 21 butir yang lain tidak memenuhi kriteria (*unfit*). Hal ini menunjukkan bahwa butir yang memenuhi kriteria kesesuaian dengan model lebih sedikit dari pada yang tidak memenuhi kriteria (sesuai model 1 parameter).

Berdasarkan hasil analisis butir dengan menggunakan kedua pendekatan teori tes, sebagaimana disajikan dalam tabel-tabel di atas, maka dapat dibanding kesimpulan dari keduanya, yang secara ringkas disajikan dalam tabel berikut ini.

Tabel 16. Perbandingan Hasil Analisis Butir Tes Bahasa Inggris B Berdasarkan Pendekatan Teori Tes

Kesimpulan berdasarkan pendekatan		Teori Modern		Jumlah
		Baik	Tidak Baik	
Teori Klasik	Baik	13	6	19
	Tidak Baik	9	12	21
Jumlah		22	18	40

Sebagaimana untuk versi A, data dari jawaban seluruh testee yang mengerjakan tes masuk Bahasa Inggris versi B (621 testee) dianalisis dengan menggunakan teknik Cronbach alpha (α) untuk mengetahui tingkat reliabilitasnya. Analisis dilakukan terhadap data dari seluruh (40) butir dan data dari butir-butir yang dinyatakan valid saja berdasarkan kriteria teori tes klasik (35 butir). Berdasarkan data dari seluruh butir dihasilkan nilai alpha Cronbach, $\alpha = 0,804$. Sedangkan untuk data dari butir-butir yang dinyatakan valid saja dihasilkan nilai alpha Cronbach, $\alpha = 0,815$. Hasil ini menunjukkan bahwa tingkat reliabilitas butir-butir soal tes masuk Bahasa Inggris versi B memiliki tingkat reliabilitas yang sama-sama sangat tinggi.

Dengan nilai $\alpha = 0,804$ dan nilai simpang baku SB = 5,673 dapat dihitung nilai galat baku pengukuran atau standard error of measurement, GBP/SEM = 2,512. Hal ini menunjukkan bahwa tes masuk Bahasa Arab versi B memiliki akurasi yang tinggi (tingkat kesalahan yang rendah) dalam mengukur kemampuan berbahasa Arab testee/calon mahasiswa baru angkatan 2012/2013. Dengan kata lain, berdasarkan teori tes klasik skor tes yang diperoleh peserta tes cukup meyakinkan untuk mencerminkan/memperkirakan kemampuan peserta yang sebenarnya.

B. Uji Hipotesis: Tes Masuk sebagai Prediktor Hasil Belajar

Salah satu tujuan penelitian ini adalah untuk menguji kemampuan prediksi skor tes masuk Bahasa Arab dan Bahasa Inggris dalam meramalkan hasil belajar mahasiswa IAIN Walisongo dalam menempuh matakuliah masing-masing untuk Bahasa Arab dan Bahasa Inggris.

Untuk menguji tujuan tersebut, data yang telah terkumpul dianalisis dengan menggunakan teknik regresi sederhana (*simple regression*).²² Sebagai variabel prediktor dalam analisis ini adalah skor tes masuk dan variabel kriterianya adalah hasil belajar mahasiswa. Setiap rangkaian skor tes masuk berdasarkan mata tes dan versinya dianalisis secara terpisah dengan skor hasil belajar bahasa yang relevan. Berdasarkan pengolahannya, skor tes masuk dibedakan menjadi dua: skor tes klasik dan skor tes modern (Rush). Analisis ini utamanya dimaksudkan untuk menghasilkan koefisien determinasi untuk menentukan besarnya sumbangan skor tes masuk (variabel prediktor) dalam menjelaskan varian hasil belajar mahasiswa (variabel kriteria), persamaan regresi untuk memperkirakan besaran nilai hasil belajar berdasarkan skor tes masuk, serta taraf signifikansinya untuk memungkinkan generalisasi dari sampel ke populasi. Analisis terpisah untuk masing-masing pendekatan teori tes untuk variabel prediktor tersebut memungkinkan melakukan perbandingan antara teori tes klasik dan teori klasik. Hasil analisis untuk masing-masing versi dan model tes masuk disajikan dalam bagian berikut ini.

1. Tes Masuk Bahasa Arab A terhadap Hasil Belajar Bahasa Arab

Adakah skor tes masuk Bahasa Arab versi A (baik klasik maupun modern) signifikan untuk memprediksi hasil belajar Bahasa Arab mahasiswa? Hasil analisis data skor tes masuk (masing-masing pendekatan teori tes) dan skor hasil belajar dengan menggunakan

²² Pedhazur, 1982, *Multiple Regression*,

teknik analisis regresi sederhana (satu prediktor) secara ringkas disajikan dalam tabel berikut ini.

Tabel 18. Prediksi hasil belajar Bahasa Arab berdasarkan skor tes masuk Bahasa Arab Versi A

Kriteria	Prediktor	R^2	a	b	β	t	p
Hasil Belajar B. Arab	TM Klasik A	,269	2,319	,045	,519	6,251	,000
	TM Modern A	,266	3,213	,300	,515	6,191	,000

Keterangan:

- R^2 : besarnya sumbangan skor tes masuk pada varian hasil belajar
- a : nilai intersep
- b : nilai slop/regresi hasil belajar pada skor tes masuk
- β : nilai beta/nilai baku regresi
- t : nilai t
- p : taraf signifikansi

Hasil analisis data sebagaimana disajikan dalam tabel tersebut di atas menunjukkan bahwa skor tes masuk Bahasa Arab versi A yang dihitung berdasarkan pendekatan klasik dan modern memiliki kontribusi yang relatif sama dalam menjelaskan varian hasil belajar Bahasa Arab mahasiswa. Keduanya menunjukkan kemampuan prediksi yang signifikan (masing-masing $t = 6,251$ dan $t = 6,191$; $p < 0,001$) dan masing-masing model memberikan sumbangan 26,9 persen dan 26,6 persen dari nilai hasil belajar mahasiswa. Hal ini menunjukkan bahwa skor kedua pendekatan tes untuk skor tes masuk Bahasa Arab versi A (yang dikerjakan oleh peserta tes dengan nomor tes gasal) mempunyai kemampuan prediktis yang signifikan untuk meramalkan hasil belajar Bahasa Arab mereka setelah diterima dan belajar di IAIN Walisongo Semarang. Namun demikian, nilai intersep dan slop (regresi) keduanya berbeda cukup jauh (masing-masing $a = 2,319$ dan $a = 3,213$; $b = 0,045$ dan $b = 0,300$). Dengan

demikian, hasil belajar mahasiswa tersebut dapat diperkirakan dengan persamaan regresi sebagai berikut:

- a. Persamaan regresi berdasarkan skor tes masuk dengan pendekatan teori tes klasik:

$$Y' = a + bX = 2,319 + 0,045X$$

- b. Persamaan regresi berdasarkan skor tes masuk dengan pendekatan teori tes modern:

$$Y' = a + bX = 3,213 + 0,300X$$

2. Tes Masuk Bahasa Arab B terhadap Hasil Belajar Bahasa Arab

Sebagaimana skor tes masuk Bahasa Arab versi A, skor tes masuk Bahasa Arab versi B juga digunakan untuk memprediksi keberhasilan belajar Bahasa Arab mahasiswa di IAIN Walisongo. Hasil analisis regresi sederhana untuk menguji prediksi tersebut secara singkat disajikan dalam tabel berikut ini.

Tabel 19. Prediksi hasil belajar Bahasa Arab berdasarkan skor tes masuk Bahasa Arab Versi B

Kriteria	Prediktor	R^2	a	b	β	t	p
Hasil Belajar	TM Klasik B	,295	2,219	,046	,543	5,958	,000
B. Arab	TM Modern B	,264	3,125	,290	,513	5,516	,000

Dari tabel tersebut di atas dapat diketahui bahwa skor tes masuk Bahasa Arab versi B memberikan kontribusi yang signifikan untuk menjelaskan varian hasil belajar Bahasa Arab mahasiswa, baik berdasarkan skor teori tes klasik maupun teori tes modern (masing-masing $t = 5,958$ dan $t = 5,5166$; $p < 0,001$). Namun demikian, skor dengan pendekatan teori klasik memberikan sumbangan yang lebih besar dari pada skor dengan pendekatan modern (masing-masing

model memberikan sumbangan 29,5 persen dan 26,4 persen dari varian hasil belajar mahasiswa. Hasil ini menunjukkan bahwa skor masuk dengan menggunakan kedua pendekatan tes untuk skor tes masuk Bahasa Arab versi B (yang dikerjakan oleh peserta tes dengan nomor tes genap) mempunyai kemampuan prediktis yang signifikan untuk meramalkan hasil belajar Bahasa Arab mereka setelah diterima dan belajar di IAIN Walisongo Semarang. Sebagaimana kontribusi terhadap varian, nilai intersep dan slop (regresi) keduanya juga berbeda cukup jauh (masing-masing $a = 2,219$ dan $a = 3,125$; $b = 0,046$ dan $b = 0,290$). Berdasarkan hasil konstan tersebut di atas, hasil belajar mahasiswa tersebut dapat diperkirakan dengan persamaan regresi sebagai berikut:

- a. Persamaan regresi berdasarkan skor tes masuk dengan pendekatan teori tes klasik:

$$Y' = a + bX = 2,219 + 0,046X$$

- b. Persamaan regresi berdasarkan skor tes masuk dengan pendekatan teori tes modern:

$$Y' = a + bX = 3,215 + 0,290X$$

3. Tes Masuk Bahasa Inggris A terhadap HB Bahasa Inggris

Sebagaimana matakuliah Bahasa Arab, hasil belajar matakuliah Bahasa Inggris yang dicapai oleh mahasiswa juga diprediksi dengan skor hasil tes masuk Bahasa Inggris, salah satunya versi A, baik yang dihitung berdasarkan teori klasik maupun

modern. Hasil analisis tersebut secara ringkas disajikan dalam tabel berikut ini.

Tabel 20. Prediksi hasil belajar Bahasa Inggris berdasarkan skor tes masuk Bahasa Inggris Versi A

Kriteria	Prediktor	R^2	a	b	β	T	p
Hasil Belajar	TM Klasik A	,022	2,570	,028	,147	1,946	,053
B. Inggris	TM Modern A	,020	3,025	,129	,143	1,887	,061

Hasil analisis data sebagaimana disajikan dalam tabel tersebut di atas menunjukkan bahwa skor tes masuk Bahasa Inggris versi A yang dihitung berdasarkan pendekatan klasik dan modern tidak memiliki kemampuan prediksi yang signifikan (masing-masing $t = 1,946$ dan $p = 0,053$; $t = 1,887$ dan $p = 0,061$). Hal ini menunjukkan bahwa skor kedua pendekatan tes untuk skor tes masuk Bahasa Inggris versi A (yang dikerjakan oleh peserta tes dengan nomor tes gasal) tidak mempunyai kemampuan prediktif yang signifikan untuk meramalkan hasil belajar Bahasa Inggris mereka setelah diterima dan belajar di IAIN Walisongo Semarang. Dengan kata lain, hasil belajar Bahasa Inggris mereka tidak dipengaruhi oleh kemampuan dasar Bahasa Inggris yang telah mereka miliki sebelum mereka masuk ke IAIN Walisongo sebagaimana tercerminkan dalam hasil tes masuk bahasa Inggris versi A. Karena tidak signifikan, maka kontribusi maupun persamaan regresi dari skor tes masuk Bahasa Inggris versi A tidak dapat digunakan untuk prediksi keberhasilan belajar mahasiswa. Karena itu, penghitungan persamaan regresi hasil belajar Bahasa Inggris pada skor tes masuk Bahasa Inggris versi A tidak perlu dilakukan.

4. Tes Masuk Bahasa Inggris B terhadap HB Bahasa Inggris

Sebagaimana skor tes masuk Bahasa Inggris versi A, skor tes masuk Bahasa Inggris versi B juga digunakan untuk memprediksi keberhasilan belajar Bahasa Inggris mahasiswa di IAIN Walisongo. Dalam uji prediktif ini digunakan teknik analisis regresi sederhana (dengan satu prediktor). Hasil analisis prediktif tersebut secara singkat disajikan dalam tabel berikut ini.

Tabel 21. Prediksi hasil belajar Bahasa Inggris berdasarkan skor tes masuk Bahasa Inggris Versi B

Kriteria	Prediktor	R^2	a	b	β	T	p
Hasil Belajar B. Inggris	TM Klasik B	,138	2,199	,052	,372	5,433	,000
	TM Modern B	,132	3,154	,277	,363	5,292	,000

Tabel tersebut di atas memperlihatkan bahwa skor tes masuk Bahasa Inggris versi B memberikan kontribusi yang signifikan untuk menjelaskan varian hasil belajar Bahasa Inggris yang dicapai oleh mahasiswa, baik berdasarkan skor tes masuk berdasarkan pendekatan teori tes klasik maupun pendekatan teori tes modern (masing-masing $t = 5,433$ dan $t = 5,292$; $p < 0,001$). Tabel tersebut juga memperlihatkan bahwa skor dengan pendekatan teori klasik memberikan sumbangan yang relatif sama skor dengan pendekatan modern (masing-masing model memberikan sumbangan 13,8 persen dan 13,2 persen dari varian hasil belajar Bahasa Inggris B yang diperoleh mahasiswa setelah belajar di IAIN Walisongo. Hasil ini menunjukkan bahwa skor masuk dengan menggunakan kedua pendekatan tes untuk skor tes masuk Bahasa Inggris versi B (yang dikerjakan oleh peserta tes dengan nomor tes genap) mempunyai kemampuan prediktif yang signifikan untuk meramalkan hasil belajar

Bahasa Inggris mereka setelah diterima dan belajar di IAIN Walisongo Semarang. Sebagaimana kontribusi terhadap varian, nilai intersep dan slop (regresi) keduanya juga berbeda cukup jauh (masing-masing $a = 2,199$ dan $a = 3,154$; $b = 0,052$ dan $b = 0,277$). Berdasarkan hasil konstan tersebut di atas, hasil belajar mahasiswa tersebut dapat diperkirakan dengan persamaan regresi sebagai berikut:

- a. Persamaan regresi berdasarkan skor tes masuk dengan pendekatan teori tes klasik:

$$Y' = a + bX = 2,199 + 0,052X$$

- b. Persamaan regresi berdasarkan skor tes masuk dengan pendekatan teori tes modern:

$$Y' = a + bX = 3,154 + 0,277X$$

C. Pembahasan

Sebagaimana telah dibahas sebelumnya, tujuan penelitian ini adalah untuk menguji kualitas tes masuk calon mahasiswa baru IAIN Walisongo tahun 2012/2013, baik secara internal maupun eksternal. Untuk tujuan tersebut, data telah dikumpulkan berdasarkan dokumen atau data base yang ada di bagian Akademik dan Kemahasiswaan serta Pusat Data dan Teknologi Informasi. Selanjutnya data tersebut dianalisis dengan menggunakan tekni yang sesuai dengan tujuannya.

Hasil analisis data mengungkap temuan bahwa semua versi tes masuk Bahasa Arab dan Bahasa Inggris yang digunakan untuk menyeleksi calon mahasiswa baru memiliki kualitas yang bervariasi, baik dari sisi kualitas butir maupun keandalan tes. Semua tes memiliki butir-butir yang baik dan tidak baik. Jumlah butir yang tidak baik

tersebut bervariasi antar satu tes dengan yang lainnya. Untuk tes masuk Bahasa Arab, jumlah butir yang tidak baik lebih banyak diperoleh dari analisis butir tes dengan menggunakan pendekatan teori tes modern/Rash model dari pada yang diperoleh dengan menggunakan pendekatan teori tes klasik. Meskipun ada perbedaan hasil antara tes versi A dan B, jumlah butir yang baik relatif tidak jauh berbeda (masing-masing 19 dan 20 butir tes modern dan 26 dan 28 butir untuk tes klasik). Hal ini kemungkinan karena kedua versi tes berisi butir-butir yang sama, tetapi hanya berbeda urutannya. Hasil tersebut menunjukkan bahwa tes masuk Bahasa Arab memiliki kualitas butir yang kurang baik karena banyaknya butir yang tidak memenuhi kriteria minimal yang telah ditentukan. Namun demikian, keandalan tes masuk tersebut secara keseluruhan cukup bagus untuk menghasilkan skor yang relatif stabil bila dilakukan pengukuran berulang karena reliabilitasnya tinggi, baik untuk butir-butir tes secara keseluruhan maupun butir-butir yang valid saja. Dengan demikian dapat dikatakan bahwa meskipun validitas butirnya hanya sedang saja, tes masuk Bahasa Arab versi A dan B memiliki keandalan untuk menghasilkan kestabilan skor hasil pengukuran yang relatif baik.

Sementara itu, analisis terhadap tes masuk Bahasa Inggris menghasilkan kesimpulan yang berbeda untuk versi A dan B. Untuk versi A, jumlah butir yang tidak baik dan baik berbanding terbalik untuk kedua versi. Untuk versi A, jumlah butir yang tidak baik yang dihitung dengan pendekatan teori tes klasik jauh lebih banyak dari pada dengan pendekatan teori tes modern (masing-masing 28 dan 8 butir). Sebaliknya, untuk versi B jumlah butir yang tidak baik yang dihitung

dengan pendekatan teori tes klasik lebih sedikit dari pada dengan pendekatan teori tes modern (masing-masing 22 dan 19 butir). Dengan demikian, meskipun berbeda, hasil ini menunjukkan bahwa kedua tes masuk Bahasa Inggris memiliki kualitas yang kurang baik karena banyaknya butirbutir soal yang tidak memenuhi kriteria yang telah ditetapkan. Namun demikian, keandalan tes versi B tergolong tinggi sedangkan untuk versi A tergolong sedang. Dengan kata lain, dari segi butir kedua versi tes masuk Bahasa Inggris kurang baik, meskipun khusus untuk versi B memiliki keandalan yang tinggi sehingga mampu menghasilkan skor yang relatif stabil bila dilakukan pengulangan pengukuran. Sedangkan untuk versi A keandalannya kurang bisa dipercaya karena hanya sedang saja. Hal ini kemungkinan karena butir-butirnya lebih banyak yang tidak baik sehingga keandalannya pun tidak tinggi.

Lebih lanjut, hasil analisis uji prediktif skor tes masuk untuk meramalkan hasil belajar mahasiswa menunjukkan hasil yang beragam. Hasil analisis dengan prediktor skor tes masuk Bahasa Arab signifikan, baik untuk versi A maupun B, baik skor dengan pendekatan teori tes klasik maupun modern. Untuk pendekatan teori tes modern, kedua versi tes masuk bahasa Arab memberikan sumbangan yang relatif sama (yakni 26,6 dan 26,4 persen) untuk menjelaskan varian hasil belajar Bahasa Arab mahasiswa. Sedangkan untuk pendekatan klasik, masing-masing versi memberikan sumbangan berbeda (meskipun hanya sedikit), yakni 26,9 dan 29,5 persen untuk menjelaskan hasil belajar Bahasa Arab mahasiswa. Dengan demikian dapat dikatakan bahwa tes masuk Bahasa

Arab memiliki kemampuan prediksi yang baik untuk meramalkan keberhasilan belajar mahasiswa.

Senentara itu, analisis prediktif untuk tes masuk Bahasa Inggris menunjukkan hasil yang berbeda antar versi. Hasil analisis data untuk tes masuk Bahasa Inggris versi A tidak signifikan sehingga tes tersebut tidak dapat digunakan untuk memprediksi keberhasilan belajar mahasiswa. Sedangkan untuk versi B hasil analisis menunjukkan signifikansinya. Tes Masuk Bahasa Inggris versi B ini memberikan sumbangan 13,8 persen (pendekatan teori tes klasik) dan 13,2 persen (pendekatan teori tes modern) untuk menjelaskan varian hasil belajar Bahasa Inggris mahasiswa. Dengan kata lain, tidak sebagaimana versi A, tes masuk Bahasa Inggris versi B memiliki kemampuan prediksi, meskipun sumbangannya kecil, untuk memperkirakan hasil belajar mahasiswa.

D. Keterbatasan Penelitian

Meskipun telah didisain sedemikian rupa dan dilakukan dengan sebaik-baiknya, penelitian ini tidak terlepas dari keterbatasan yang mungkin dapat mempengaruhi kualitas hasilnya. Di antara keterbatasan tersebut adalah:

1. Permasalahan yang diajukan dalam penelitian ini hanya dibatasi pada validitas empiris yang didasarkan pada data. Validitas konseptual yang terkait dengan validitas konstruk dan validitas isi tidak dimasukkan dalam fokus penelitian. Seperti diketahui, kedua validitas konseptual tersebut sangat penting untuk menentukan kualitas suatu tes karena menjadi pijakan dalam pengembangan suatu. Dalam penelitian ini keduanya tidak dianalisis karena tidak

dijumpai “*academic paper*” atau dokumen lain yang menjadi pedoman dalam pengembangan tes yang diteliti. Meskipun kerangka pemikiran dalam pengembangan tes tersebut kemungkinan dapat diperoleh dari pengembang/pembuat tes melalui wawancara, misalnya, peneliti tidak melakukannya karena hanya fokus pada validitas empiris. Penelitian yang akan datang hendaknya juga memasukkan validitas konseptual agar dapat diperoleh hasil yang lebih meyakinkan.

2. Pengujian validitas prediktif dalam penelitian ini hanya melibatkan skor tes masuk secara keseluruhan (total skor seluruh butir tes masuk yang berjumlah 40), baik dengan pendekatan teori tes klasik maupun teori tes modern. Dari hasil analisis terhadap butir-butir tes masuk dengan kedua pendekatan ditemukan banyak di antaranya yang tidak baik/valid. Untuk mendapatkan prediksi yang terbaik, seharusnya butir-butir tersebut dihilangkan sehingga hanya skor dari butir-butir yang baik saja yang digunakan sebagai prediktor keberhasilan belajar. Akan tetapi, dalam penelitian ini hal tersebut tidak dilakukan karena butir-butir ini juga dilibatkan dalam prediksi hasil belajar mahasiswa. Karena itu, dalam penelitian yang akan datang hendaknya juga dilakukan prediksi hasil belajar yang melibatkan prediktor skor butir-butir tes yang valid saja sehingga memungkinkan untuk membandingkan hasilnya dengan efektivitas prediksi skor hasil tes secara keseluruhan.
3. Karena penelitian ini hanya menggunakan skor hasil belajar mahasiswa yang telah jadi, kualitas tes yang digunakan untuk mengukur hasil belajar tersebut tidak diketahui. Hal ini dapat terjadi

karena sejauh informasi yang peneliti peroleh, tidak ada upaya dari Pusat Pengembangan Bahasa untuk melakukan analisis tes yang digunakan. Hal ini diduga juga dapat mempengaruhi hasil penelitian ini, terutama untuk uji prediktif kualitas tes masuk. Penelitian yang akan datang sebaiknya melibatkan juga analisis kualitas tes hasil belajar sehingga memungkinkan untuk memperoleh hasil prediktif yang maksimal.

Dengan adanya beberapa keterbatasan penelitian sebagaimana di atas, hasil penelitian ini harus dipahami secara hati-hati agar kesimpulan atau generalisasinya tepat guna dalam penggunaannya untuk perbaikan tes masuk yang akan datang. Dengan memahami keterbatasan tersebut, perbaikan tes masuk hendaknya lebih fokus pada aspek-aspek yang telah ditemukan kekurangannya. Dengan perbaikan aspek-aspek tersebut diharapkan tes masuk yang akan datang akan lebih baik, dengan memberikan kontribusi prediksi keberhasilan belajar mahasiswa secara maksimal. Dengan cara tersebut diharapkan IAIN Walisongo akan memperoleh calon mahasiswa yang memiliki potensi keberhasilan belajar yang tinggi.

BAB V PENUTUP

Sesuai dengan tujuan penelitian ini, Bab Penutup ini akan menyajikan kesimpulan dari seluruh pembahasan dalam bab-bab sebelumnya, yang sekaligus menjawab secara singkat masalah yang diajukan dalam Bab I. Kesimpulan tersebut ditarik berdasarkan hasil penelitian dan pembahasannya, sebagaimana yang disajikan dalam bab sebelumnya. Selanjutnya, berdasarkan kesimpulan tersebut, bab ini akan mengakhiri sajiannya dengan saran-saran, baik yang sifatnya teoritis (untuk pengembangan penelitian lebih lanjut), maupun praktis (untuk digunakan dalam memperbaiki tes masuk).

A. Kesimpulan

Dari hasil analisis yang dilakukan terhadap data yang terkumpul, sebagaimana disajikan dalam bab sebelumnya, dapat disimpulkan sebagai berikut:

1. Secara umum kualitas internal Tes Masuk IAIN Walisong tahun akademik 2012/2013 untuk mata ujian bahasa asing cenderung moderat, meskipun ada variasi antar bahasa, versi, maupun pendekatan teori tes yang digunakan, kecuali tes masuk Bahasa Inggris versi A yang tidak baik. Berdasarkan kualitas butirnya, butir yang baik berdasarkan teori klasik adalah 47,5 persen (Bahasa Inggris versi B), 65 persen (Bahasa Arab versi A), dan 70 persen (Bahasa Arab versi B). Sedangkan berdasarkan teori tes modern (model Rash, 1 parameter) lebih sedikit jumlah butir yang baik, yakni 55 persen (Bahasa Inggris B), 50 persen (Bahasa Arab B), dan 47,5

persen (Bahasa Arab A). Sedangkan untuk tes masuk persen Bahasa Inggris versi A) hasil analisis butir menunjukkan hasil yang kontrasi antar kedua pendekatan, yakni 82,5 persen (pendekatan teori tes modern dan 30 persen (pendekatan teori tes klasik). Tidak tingginya kualitas butir-butir tersebut terutama karena tes tersebut terlalu sulit untuk para testee (secara umum kemampuan testee sangat rendah).

Meskipun demikian, kecuali tes masuk Bahasa Inggris versi A, semua tes masuk memiliki tingkat reliabilitas yang tinggi, mendekati alpha 0,90. Hasil ini menunjukkan bahwa meskipun kualitas tes masuk tidak begitu baik, hasil pengukuran yang diperoleh dengan tes tersebut relatif memiliki tingkat kesalahan yang rendah sehingga hasilnya dapat mencerminkan realitas kemampuan testee (calon mahasiswa baru) yang sebenarnya. Sedangkan untuk tes masuk Bahasa Inggris versi A memiliki reliabilitas yang rendah sehingga hasilnya kurang mencerminkan keadaan testee yang sebenarnya karena kurang stabil.

2. Berdasarkan kualitas prediktifnya, kecuali tes masuk Bahasa Inggris versi A, semua tes masuk Bahasa Asing memiliki kemampuan prediktif yang meyakinkan (signifikan) untuk meramalkan hasil belajar mahasiswa setelah diterima dan kuliah di IAIN Walisongo. Kemampuan prediktif ini diperoleh baik melalui analisis dengan pendekatan teori tes klasik maupun modern. Namun demikian, skor tes masuk tersebut hanya memberikan sumbangan yang rendah untuk menjelaskan variansi thasil belajar mahasiswa. Sementara itu, tes masuk Bahasa Inggris versi A tidak dapat digunakan untuk

memprediksi hasil belajar Bahasa Inggris mahasiswa karena tidak signifikan.

B. Saran

Berdasarkan hasil uji kualitas tes masuk tersebut di atas, maka disarankan sebagai berikut:

1. Untuk mendapatkan calon mahasiswa baru yang memiliki potensi keberhasilan yang tinggi, IAIN Walisongo ke depan hendaknya mengembangkan tes seleksi calon mahasiswa baru yang dirancang berdasarkan prinsip-prinsip pengembangan tes yang terpercaya. Pengembangan tersebut dilakukan dengan menggunakan kaidah-kaidah dan prosedur yang jelas dan terukur. Sebaiknya tes tersebut memiliki butir-butir dengan tingkat kesulitan yang cenderung moderat dengan rentang yang luas (dari yang sangat sulit sampai yang sangat mudah). Dengan tingkat kesulitan yang seperti ini diharapkan tes tersebut dapat membedakan kemampuan testee yang memiliki latar belakang pendidikan yang beragam, terutama terkait dengan Bahasa Arab dan Inggris.
2. Untuk memberikan landasan yang lebih kuat untuk pengembangan tes masuk di masa mendatang, penelitian lebih lanjut sangat diperlukan, terutama yang terkait dengan butir-butir yang valid saja dalam kaitannya dengan prediksi hasil belajar. Di samping itu, skor tes maupun hasil belajar bahasa asing juga perlu diteliti secara rinci (berdasarkan variasi cakupan materinya [listening/istima', speaking/muhadatsah, reading/mutholaah, dan writing/kitabah]). Dengan penelitian ini diharapkan akan menghasilkan kesimpulan yang lebih akurat dalam memprediksi hasil belajar, yang sekaligus

juga akan memungkinkan untuk menghasilkan calon mahasiswa yang memiliki potensi yang baik untuk berhasil belajar di IAIN Walisongo.

DAFTAR PUSTAKA

- Anastasi, Anne & Susana Urbina, *Psychological Testing* Upper Saddle River: Prentice Hall, ed. Ke-7, 1997.
- Assessment Systems Corporation, *Rasch Model Item Calibration Program-RASCAL*, Version 3.00 1988.
- Baker, Frank, *The Basic of Item Response Theory*, Madison: ERIC Clearinghouse on Assessment and Evaluation, 2001.
- Bejar, Isaac I., "Introducing to Item Response Models and Their Assumption," dalam Ronald K. Hambleton (Ed.), *Application of Item Response Theory* Burnaby: Educational Institute of British Columbia, 1983: 1-23.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Jump, *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain*, New York: Longmans, 1956.
- Crocker, Linda & James Algina, *Introduction to Classical and Modern Test Theory* Fort Worth: Holt, Rinehart and Winston, 1986.
- Fulcher, G., & F. Davidson, *Language testing and assessment: An advanced resource book*, London: Routledge, 2007.
- Gagné, Robert M., *The Condition of Learning* New York: Holt Rinehart, & Winston, 1997.
- Gagné, Robert M., dan Marcy P. Driscoll, *Essential of Learning for Instruction* Englewood Cliffs, NJ: Prentice hall, 1988.
- Gray, D.E., *Doing research in the real world*, Los Angeles: Sage, 2009.
- Gronlund, N.E. & R.L. Linn, *Measurement and Evaluation in Teaching*, New York: Macmillan Pub. Co., 1985.
- Gustafsson, Ja-Eric, *An Introduction to Rasch's Measurement Model* Princeton: Educational Testing Service, 1981.
- Hadjar, Ibnu, "Pengembangan model seleksi calon mahasiswa baru IAIN Walisongo Semarang", *Jurnal Penelitian Walisongo*, 12(1), 2004: 47-59.
- Hadjar, Ibnu, "Uji kualitas ujian masuk IAIN Walisongo Semarang dengan pendekatan klasik dan item response theory," *Jurnal Penelitian Walisongo*, 9(19), 2001:84-92.
- Hadjar, Ibnu, *Dasar-dasar metodologi penelitian kuantitatif dalam pendidikan*, Jakarta: RajGrafindo Persada, 1999.
- Hadjar, Ibnu, *W-Stats, Program Aplikasi Statistik Walisongo Semarang: IAIN Walisongo*, 2013.

- Hambleton, R.K., H. Swaminathan, & H.J. Rogers, *Fundamentals of item response theory*, Newbury Park: Sage, 1991.
- Hambleton, Ronald K., H. Swaminathan, dan H. Jane Rogers, *Fundamentals of Item Response Theory* Newbury Park: Sage Publication, 1991.
- Hulin, Charles L., Fritz Drasgow, & Charles K. Parsons, *Item Response Theory: Application to Psychological Measurement* Homewood: Dow Jone-Irwin, 1983.
- IAIN Walisongo Semarang, *Pedoman Akademik IAIN Walisongo Semarang Semarang*: IAIN Walisongo, 2012-2013.
- Johnson, R.L., J.A. Penny, & B. Gordon *Assessing Performance: Designing, Scoring, and Validating Performance Tasks* New York: The Guilford Press, 2009.
- Krathwohl, David R., Benjamin S. Bloom, dan Bertram B. Masia, *Taxonomy of Educational Obyectivies, Book II: Affective Domain* London: Longman Group, 1973.
- Linn, Robert L., & Nourman E. Grounlund, *Measurement and Assessment in Teaching* Englewood Clift: Prentice Hall, 1995.
- Lodico, M.D., D.T. Spoulding & K.H. Voegtler, *Methods in educational research: From theory to practice*, San Francisco: Josey-Bass, 2006), h. 87.
- Lyman, Howard B., *Test Score and What They Mean*, Englewood Cliffs: Prentice-Hall, 1978.
- M.A.K. Muhammad, *Ikhtibārāt al-Lughoh*, Riyādh: Jāmi'ah al-Mālik Su'ūd, 1996.
- Magno, C., "Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data," *The International Journal of Educational and Psychological Assessment*, 1(1): 1-11, 2009.
- Martin, W.E., & K.D. Bridgmon, *Quantitative and statistical research methods: From hypothesis to results*, San Francisco, 2012.
- Mcdonald, Roderick P., *Test Theory: A Unified Treatmen* Mahwah: Lawrence Erlbaum Association, 1999.
- Murdapi, Djemari, *Analisis Butir Dengan Teori Klasik dan Teori Respon Butir* Yogyakarta: IKIP Yogyakarta, 1994.
- Nitko, Anthony J., *Educational Assesment of Students* Englewood Cliffs: Prentice Hall, 1996.
- Panitia SBMPTN 2013, *INFORMASI Seleksi Bersama Masuk Perguruan Tinggi Negeri SBMPTN) Tahun 2013* diakses dari:

- <http://www.sbmptn.or.id/utama.php?hmenu=3&hitem=1&dikirim=1>, pada 12 Oktober 2013.
- Panitia SNMPTN 2013, *Informasi Umum SNMPTN 2013* diakses dari: <http://www.snmptn.ac.id/informasi.html#pengantar>, pada 12 Oktober 2013.
- Partchev, Ivaelo, 2004, *A Visual Guide to Item Response Theory* Jena: Friedrick Scheiler Universitat.
- Pedhazur, E.J., *Multiple regression in behavioral research: Explanation and prediction*, New York: Holt, Rinehart, & Winston, 1982.
- Pratt, David, *Curriculum : Design and development* San Diego : Harcourt Brace Jovanovich, 1980.
- Purwanto, *Evaluasi Hasil Belajar Yogyakarta* : Pustaka Pelajar, 2009.
- Romiszowski, A. J., *Designing intruactional system* new York: Kognan Page, 1981.
- Soediyono, dkk., 1995, *Tes Masuk sebagai Prediktor Keberhasilan Bealajar*, Semarang: Fakultas Tarbiyah, 1995.
- Suryabrata, Sumadi, *Seleksi Calon Mahasiswa Baru Perguruan Tinggi: Yang Sekarang dan Kemungkinan untuk Masa Mendatang* Yogyakarta: Andi Offset, 1989.
- Traub, Rose E., "A Priori Consideration in Choosing an Item Response Model," dalam Roland K. Hambleton Ed.), *Aplication of Item Response Theory* Burnaby: Educational Research Institut of British Columbia, 1983:57-70.

LAMPIRAN 1A1: HASIL ANALISIS BUTIR TES MASUK BAHASA ARAB A DENGAN PENDEKATAN TEORI TES KLASIK

W-Stats

PROCESOR: SPSS 16.0.0
Date: 2011-08-18 10:00:00

HASIL ANALISIS BUTIR TES

Mata Tes: A Arab-A
Jenis Instrumen: Tesya Multichoice
Jumlah Butir: 40
Jumlah Pilihan Jawaban: 5
Jumlah Tesur: 604
Validitas/IT: 0,800
Reliabilitas Butir yang valid saja/MT: 0,882

Kriteria Tingkat Kesulitan antara p= 0,20 dan q = 0,30
Kriteria Validitas r2 0,3
Kriteria Pengaruh/ pe > 3%
Skor Terendah: 0
Nilai Rata-Rata: 13,629
Nilai Simpang Baku: 7,438
Galat Baku Pengukuran (Standard Error of Measurement/SEM): 2,482

BERANDA

DATA

No. Butir	Tingkat Kesulitan	Kondisi Butir-total/r	Proporsi Pilihan Jawaban:					Lain-lain	Tingkat Kesulitan	Validitas/IT	Pengaruh
			A	B	C	D	E				
1	0,313	0,288	0,20	0,13	0,06	0,23	0,38	0,27	0,8	0,000	0,00
2	0,188	0,362	0,34	0,30	0,10	0,20	0,16	0,54	0,84	0,00	0,00
3	0,188	0,336	0,36	0,40	0,10	0,10	0,10	0,54	0,84	0,00	0,00
4	0,488	0,570	0,67	0,40	0,06	0,07	0,15	0,24	0,8	0,00	0,00
5	0,388	0,464	0,36	0,30	0,05	0,08	0,20	0,20	0,8	0,00	0,00
6	0,588	0,481	0,64	0,30	0,10	0,00	0,10	0,10	0,8	0,00	0,00
7	0,482	0,525	0,33	0,40	0,00	0,40	0,30	0,25	0,8	0,00	0,00
8	0,252	0,458	0,31	0,40	0,17	0,00	0,10	0,25	0,8	0,00	0,00
9	0,282	0,587	0,17	0,10	0,17	0,07	0,1	0,16	0,8	0,00	0,00
10	0,116	0,354	0,36	0,17	0,17	0,30	0,18	0,44	0,84	0,00	0,00
11	0,078	0,346	0,30	0,13	0,06	0,07	0,17	0,42	0,84	0,00	0,00
12	0,162	0,485	0,40	0,17	0,05	0,30	0,10	0,14	0,84	0,00	0,00
13	0,278	0,487	0,17	0,10	0,00	0,30	0,17	0,10	0,8	0,00	0,00
14	0,538	0,466	0,34	0,30	0,11	0,10	0,10	0,14	0,8	0,00	0,00
15	0,448	0,493	0,30	0,30	0,00	0,00	0,15	0,11	0,8	0,00	0,00
16	0,182	0,517	0,41	0,10	0,11	0,14	0,10	0,10	0,84	0,00	0,00
17	0,256	0,432	0,16	0,17	0,16	0,16	0,10	0,16	0,8	0,00	0,00
18	0,152	0,441	0,12	0,11	0,11	0,10	0,10	0,10	0,84	0,00	0,00
19	0,188	0,221	0,10	0,10	0,10	0,10	0,10	0,10	0,84	0,00	0,00
20	0,178	0,533	0,10	0,10	0,07	0,10	0,10	0,10	0,8	0,00	0,00

No. Unit	Tingkat Kendaraan	Kategori Baru masuk	Properti PDRB					Revisi				
			A	B	C	D	E	Unit-unit	Tingkat Kendaraan	Nilai/Unit	Properti	
21	0205	0323	0,07	0,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
22	0215	0407	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
23	0214	0405	0,07	0,04	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
24	0211	0409	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
25	0014	0005	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
26	0211	0208	0,07	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
27	0410	0417	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
28	0216	0306	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
29	0218	0308	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
30	0214	0300	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
31	0001	0102	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
32	0211	0300	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
33	0411	0301	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
34	0315	0400	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
35	0111	0300	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
36	0219	0300	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
37	0315	0311	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
38	0213	0300	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
39	0313	0400	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
40	0419	0415	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

**LAMPIRAN 1A2: HASIL ANALISIS BUTIR TES
MASUK BAHASA ARAB A DENGAN
PENDEKATAN TEORI TES
MODERN (RASCH MODEL)**

MicroCAT (tm) Testing System
Copyright (c) 1982, 1984, 1986, 1988 by Assessment
Systems Corporation

Rasch Model Item Calibration Program -- RASCAL
(tm) Version 3.00

Final Parameter Estimates for Data from File ARABA.DAT

Item ----	Difficulty -----	Chi Sq. -----	df ----
1	-0,253	86,841	19
2	0,588	20,177	19
3	0,549	31,290	19
4	-1,244	28,017	19
5	-0,234	21,083	19
6	-1,294	25,794	19
7	-1,211	10,832	19
8	0,072	21,242	19
9	0,051	19,610	19
10	1,223	27,972	19
11	1,710	43,601	19
12	0,762	26,602	19
13	-0,010	15,428	19
14	-1,437	13,633	19
15	-0,944	32,531	19
16	0,487	26,486	19
17	0,082	24,711	19
18	0,863	63,046	19
19	0,549	44,798	19
20	-2,230	22,093	19
21	0,427	33,968	19
22	0,356	16,069	19
23	0,666	31,880	19
24	0,177	17,501	19
25	1,956	81,336	19

26	0,693	44,037	19
27	-0,994	42,858	19
28	0,145	39,668	19
29	0,135	39,239	19
30	-0,708	35,412	19
31	1,522	49,765	19
32	0,243	32,960	19
33	-0,953	31,620	19
34	-0,604	12,670	19
35	0,614	30,348	19
36	0,061	36,189	19
37	-0,445	27,042	19
38	0,368	31,821	19
39	-0,543	11,600	19
40	-1,194	51,809	19

LAMPIRAN 1B1: HASIL ANALISIS BUTIR TES
MASUK BAHASA ARAB B DENGAN
PENDEKATAN TEORI TES KLASIK



HASIL ANALISIS BUTIR TES

Nama Tes: A Arab-B
 Jenis Instrumen: Tesya Mekanis
 Jumlah butir: 40
 Jumlah Pilihan Jawaban: 5
 Jumlah Tesee: 621
 Reliabilitas/ICC: 0,838
 Reliabilitas butir yang Valid/ICC: 0,838

Kriteria Tingkat Kesulitan antara $p = 0,20$ dan $p = 0,80$
 Kriteria Validitas: $r \geq 0,3$
 Kriteria Pengantar/ $pe \geq 1\%$
 Skor Terendah: 0
 Nilai Rendah Mutlak: 52,205
 Nilai Sampang Bawah: 1,825

Salah Satu Pengantar (Standard Error of Measurement/SEM): 2,585

BERANDA
DATA

No. Butir	Tingkat Kesulitan	Korelasi Butir-total/r	Proporsi/Persentase Jawaban:					lain-lain	Tingkat Kesulitan	Validitas/ Pengantar	
			A	B	C	D	E				
1	0,352	0,517	0,62	0,82	0,84	0,82	0,84	0,25	baik	Valid	baik
2	0,065	0,252	0,29	0,46	0,53	0,59	0,69	0,28	Tidak Baik	Tidak Valid	baik
3	0,643	0,544	0,88	0,89	0,69	0,67	0,81	0,31	baik	Valid	baik
4	0,329	0,686	0,15	0,21	0,54	0,82	0,29	0,29	baik	Valid	baik
5	0,348	0,686	0,89	0,18	0,82	0,38	0,89	0,32	baik	Valid	baik
6	0,524	0,416	0,89	0,81	0,58	0,87	0,84	0,36	baik	Valid	baik
7	0,382	0,515	0,48	0,81	0,87	0,82	0,82	0,34	baik	Valid	baik
8	0,280	0,607	0,29	0,82	0,28	0,88	0,88	0,38	baik	Valid	baik
9	0,213	0,373	0,80	0,14	0,24	0,18	0,13	0,28	baik	Valid	baik
10	0,832	0,411	0,33	0,83	0,56	0,24	0,18	0,36	Tidak Baik	Valid	baik
11	0,822	0,344	0,25	0,48	0,84	0,32	0,84	0,41	Tidak Baik	Tidak Valid	baik
12	0,177	0,449	0,20	0,64	0,83	0,37	0,89	0,38	Tidak Baik	Valid	baik
13	0,288	0,432	0,60	0,48	0,87	0,28	0,84	0,36	baik	Valid	baik
14	0,858	0,388	0,88	0,81	0,58	0,32	0,32	0,38	Tidak Baik	Valid	baik
15	0,882	0,448	0,28	0,24	0,32	0,37	0,85	0,38	Tidak Baik	Valid	baik
16	0,428	0,585	0,86	0,86	0,48	0,88	0,88	0,33	baik	Valid	baik
17	0,277	0,394	0,46	0,23	0,83	0,81	0,86	0,38	baik	Valid	baik
18	0,251	0,363	0,86	0,18	0,82	0,23	0,23	0,21	baik	Valid	baik
19	0,524	0,885	0,88	0,80	0,21	0,87	0,81	0,32	baik	Valid	baik
20	0,229	0,371	0,48	0,46	0,29	0,88	0,88	0,27	baik	Valid	baik

No. Data	Tingkat Kesulitan	Gredasi Int-traffic	Proporsi Perforansi					Gredasi	Aspek		
			A	B	C	D	E		Isi-Isi	Tingkat Kesulitan	Validasi
2	0,200	0,371	0,10	0,20	0,34	0,04	0,11	0,30	Tinggi	Valid	Salah
22	0,214	0,355	0,10	0,19	0,34	0,04	0,11	0,30	Salah	Valid	Salah
23	0,218	0,348	0,10	0,18	0,34	0,04	0,11	0,30	Tinggi	Valid	Salah
24	0,214	0,342	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
25	0,213	0,335	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
26	0,205	0,328	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
27	0,200	0,322	0,10	0,17	0,34	0,04	0,11	0,30	Salah	Valid	Salah
28	0,215	0,338	0,10	0,19	0,34	0,04	0,11	0,30	Tinggi	Valid	Salah
29	0,215	0,338	0,10	0,19	0,34	0,04	0,11	0,30	Salah	Valid	Salah
30	0,214	0,332	0,10	0,18	0,34	0,04	0,11	0,30	Tinggi	Valid	Salah
31	0,216	0,327	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
32	0,200	0,325	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
33	0,213	0,335	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
34	0,212	0,328	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
35	0,201	0,319	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
36	0,200	0,312	0,10	0,18	0,34	0,04	0,11	0,30	Tinggi	Valid	Salah
37	0,213	0,329	0,10	0,19	0,34	0,04	0,11	0,30	Salah	Valid	Salah
38	0,197	0,310	0,10	0,17	0,34	0,04	0,11	0,30	Salah	Valid	Salah
39	0,205	0,321	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah
40	0,212	0,331	0,10	0,18	0,34	0,04	0,11	0,30	Salah	Valid	Salah

**LAMPIRAN 1B2: HASIL ANALISIS BUTIR TES
MASUK BAHASA ARAB B DENGAN
PENDEKATAN TEORI TES
MODERN (RASCH MODEL)**

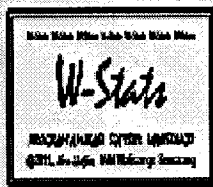
MicroCAT (tm) Testing System
Copyright (c) 1982, 1984, 1986, 1988 by Assessment
Systems Corporation
Rasch Model Item Calibration Program -- RASCAL (tm)
Version 3.00

Final Parameter Estimates for Data from File ARABA.DAT

Item	Difficulty	Chi Sq.	df
1	-0,253	86,841	19
2	0,588	20,177	19
3	0,549	31,290	19
4	-1,244	28,017	19
5	-0,234	21,083	19
6	-1,294	25,794	19
7	-1,211	10,832	19
8	0,072	21,242	19
9	0,051	19,610	19
10	1,223	27,972	19
11	1,710	43,601	19
12	0,762	26,602	19
13	-0,010	15,428	19
14	-1,437	13,633	19
15	-0,944	32,531	19
16	0,487	26,486	19
17	0,082	24,711	19
18	0,863	63,046	19
19	0,549	44,798	19
20	-2,230	22,093	19
21	0,427	33,968	19
22	0,356	16,069	19
23	0,666	31,880	19
24	0,177	17,501	19
25	1,956	81,336	19
26	0,693	44,037	19
27	-0,994	42,858	19
28	0,145	39,668	19

29	0,135	39,239	19
30	-0,708	35,412	19
31	1,522	49,765	19
32	0,243	32,960	19
33	-0,953	31,620	19
34	-0,604	12,670	19
35	0,614	30,348	19
36	0,061	36,189	19
37	-0,445	27,042	19
38	0,368	31,821	19
39	-0,543	11,600	19
40	-1,194	51,809	19

LAMPIRAN 1C1: HASIL ANALISIS BUTIR TES MASUK BAHASA INGGRIS A DENGAN PENDEKATAN TEORI TES KLASIK



HASIL ANALISIS BUTIR TES

Nama Tes: **K. Inggris A**
 Jenis Instrumen: **Cerita Mendengar**
 Jumlah Butir: **40**
 Jumlah Pilihan Jawaban: **5**
 Jumlah Tesrer: **634**
 Reliabilitas/rt: **0,666**
 Reliabilitas Butir yang Valid (ajp/rt): **0,441**

Kriteria Tingkat Kesulitan antara p: **0,20**
 Kriteria Validitas: r > **0,3**
 Kriteria Pengesahan/ pe > **1%**
 Skor Terendah: **0**
 Nilai Rerata Mula: **9,292**
 Nilai Simpang Baku: **4,254**
 Error Butir Pengukuran (Standard Error of Measurement/SEM): **2,458**

skor p = **0,00**
 Tingkat: **2A**

BERANDA
 DATA

No. Butir	Tingkat Kesulitan	Korelasi Butir-total/A	Pangarsu Pilihan Jawaban:					Lain-lain	Tingkat Kesulitan	Validitas/ Pengesahan	
			A	B	C	D	E				
1	0,073	0,134	0,130	0,089	0,208	0,073	0,256	0,276	Tinggi	Tinggi	Baik
2	0,115	0,106	0,146	0,142	0,148	0,153	0,129	0,263	Tinggi	Tinggi	Baik
3	0,250	0,206	0,164	0,157	0,157	0,123	0,121	0,263	Baik	Tinggi	Baik
4	0,125	0,172	0,181	0,135	0,084	0,125	0,101	0,194	Tinggi	Tinggi	Baik
5	0,228	0,177	0,230	0,185	0,229	0,091	0,152	0,176	Baik	Tinggi	Baik
6	0,402	0,247	0,120	0,142	0,097	0,123	0,142	0,131	Baik	Tinggi	Baik
7	0,529	0,100	0,109	0,120	0,140	0,079	0,139	0,134	Baik	Tinggi	Baik
8	0,156	0,200	0,154	0,102	0,134	0,140	0,103	0,188	Tinggi	Tinggi	Baik
9	0,356	0,223	0,164	0,145	0,080	0,119	0,173	0,226	Baik	Tinggi	Baik
10	0,225	0,253	0,093	0,140	0,093	0,075	0,123	0,140	Baik	Tinggi	Baik
11	0,118	0,126	0,164	0,114	0,221	0,122	0,111	0,140	Tinggi	Tinggi	Baik
12	0,112	0,140	0,200	0,110	0,112	0,040	0,100	0,130	Tinggi	Tinggi	Baik
13	0,520	0,106	0,105	0,124	0,084	0,040	0,125	0,151	Baik	Tinggi	Baik
14	0,301	0,177	0,151	0,181	0,080	0,100	0,150	0,226	Baik	Tinggi	Baik
15	0,110	0,220	0,141	0,109	0,089	0,120	0,142	0,194	Tinggi	Tinggi	Baik
16	0,153	0,145	0,134	0,155	0,152	0,110	0,142	0,150	Tinggi	Tinggi	Baik
17	0,150	0,180	0,150	0,155	0,129	0,125	0,119	0,180	Tinggi	Tinggi	Baik
18	0,116	0,267	0,100	0,100	0,150	0,116	0,142	0,167	Tinggi	Tinggi	Baik
19	0,120	0,220	0,140	0,120	0,120	0,040	0,130	0,120	Tinggi	Tinggi	Baik
20	0,466	0,241	0,141	0,125	0,073	0,036	0,137	0,170	Baik	Tinggi	Baik

No. Berir	Tingkat Kesulitan	Membaca Berir-totally	Proses Adiksin					Kesimpulan				
			A	B	C	D	E	akhir-hari	Tingkat Kesulitan	Validasi/Valid	Pengantar	
21	0,485	0,129	0,385	0,082	0,05	0,05	0,054	0,101	0,29	Yaah	Valid	Yaah
22	0,104	0,155	0,201	0,121	0,221	0,157	0,201	0,291	0,29	Yaah	Valid	Yaah
23	0,054	0,139	0,251	0,08	0,08	0,05	0,05	0,054	0,05	Yaah	Valid	Yaah
24	0,257	0,159	0,057	0,15	0,04	0,01	0,07	0,05	0,05	Yaah	Valid	Yaah
25	0,254	0,132	0,08	0,17	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
26	0,153	0,213	0,153	0,05	0,13	0,11	0,05	0,05	0,05	Yaah	Valid	Yaah
27	0,572	0,087	0,05	0,17	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
28	0,314	0,04	0,08	0,07	0,08	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
29	0,161	0,175	0,05	0,04	0,04	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
30	0,077	0,135	0,07	0,07	0,04	0,01	0,05	0,05	0,05	Yaah	Valid	Yaah
31	0,121	0,259	0,05	0,16	0,11	0,04	0,05	0,05	0,05	Yaah	Valid	Yaah
32	0,158	0,297	0,05	0,15	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
33	0,251	0,154	0,08	0,08	0,05	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
34	0,371	0,128	0,08	0,16	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
35	0,487	0,155	0,119	0,07	0,08	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
36	0,211	0,185	0,01	0,08	0,04	0,01	0,05	0,05	0,05	Yaah	Valid	Yaah
37	0,188	0,205	0,119	0,16	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
38	0,188	0,205	0,119	0,16	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
39	0,085	0,172	0,07	0,05	0,11	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah
40	0,117	0,308	0,15	0,17	0,08	0,05	0,05	0,05	0,05	Yaah	Valid	Yaah

**LAMPIRAN 1C2: HASIL ANALISIS BUTIR TES
MASUK BAHASA INGGRIS A
DENGAN PENDEKATAN TEORI TES
MODERN (RASCH MODEL)**

MicroCAT (tm) Testing System

Copyright (c) 1982, 1984, 1986, 1988 by Assessment
Systems Corporation

Rasch Model Item Calibration Program -- RASCAL (tm)
Version 3.00

Final Parameter Estimates for Data from File
INGGRISA.DAT

Item	Difficulty	Chi Sq.	df
1	1,172	21,604	17
2	0,429	16,216	17
3	-0,451	14,733	17
4	0,625	13,331	17
5	-0,203	25,697	17
6	-1,005	21,578	17
7	-1,525	29,866	17
8	0,364	14,427	17
9	-0,848	26,756	17
10	0,022	10,183	17
11	0,750	30,299	17
12	0,750	27,860	17
13	-1,561	21,103	17
14	-0,509	19,639	17
15	0,129	14,999	17
16	0,390	17,125	17
17	0,339	23,250	17
18	0,734	15,731	17
19	0,196	22,607	17
20	-1,302	31,608	17
21	-1,439	15,779	17
22	0,853	25,183	17
23	1,573	23,637	17
24	-0,339	14,593	17
25	-0,295	10,345	17
26	0,339	9,967	17

27	-1,759	26,241	17
28	-0,583	29,029	17
29	0,302	35,736	17
30	1,172	30,483	17
31	0,655	22,607	17
32	0,266	14,049	17
33	-0,240	16,491	17
34	-0,863	20,980	17
35	-1,019	15,584	17
36	-0,203	12,122	17
37	0,767	17,375	17
38	0,750	15,905	17
39	1,043	9,482	17
40	0,524	12,024	17

LAMPIRAN 1D1: HASIL ANALISIS BUTIR TES
 MASUK BAHASA INGGRIS B
 DENGAN PENDEKATAN TEORI TES
 KLASIK

W-Stats
 PROGRAM STUDI PENDIDIKAN
 GURU B. & S. PAJARAN

HASIL ANALISIS BUTIR TES

Forma Tes: B, Inggris-B
 Jenis Instrumen: Essay Multichoice
 Jumlah Butir: 40
 Jumlah Siswa Jawaban: 5
 Jumlah Tes: 621
 Reliabilitas/rtt: 0,894
 Reliabilitas butir yang
 Valid saja/rtt: 0,815

Kriteria Tingkat Kesulitan antara p: 0,20 dan q = 0,30
 Kriteria Validitas r2: 0,3
 Kriteria Pengesakan/pe2: 1/6
 Star/Tamabah: 0
 Nilai Mean Molek: 52,721
 Nilai Simpang Baku: 5,673
 Galat Baku Pengukuran (Standard
 Error of Measurement/SEM): 2,312

BERANDA
 DATA

No. Butir	Tingkat Kesulitan	Skorasi Butir-total/r	Proporsi Pilihan Jawaban					Lain-lain	Kesimpulan		
			A	B	C	D	E		Tingkat Kesulitan	Validitas/rtt	Pengesakan
1	0,282	0,277	0,102	0,382	0,484	0,138	0,248	0,108	Bah	Tidak Valid	Bah
2	0,308	0,182	0,388	0,172	0,382	0,138	0,465	0,308	Bah	Tidak Valid	Bah
3	0,581	0,113	0,388	0,382	0,382	0,138	0,465	0,155	Tidak Bah	Tidak Valid	Bah
4	0,171	0,286	0,371	0,382	0,388	0,121	0,468	0,188	Tidak Bah	Tidak Valid	Bah
5	0,189	0,187	0,385	0,388	0,386	0,377	0,189	0,378	Tidak Bah	Tidak Valid	Bah
6	0,254	0,087	0,118	0,342	0,338	0,132	0,258	0,227	Bah	Tidak Valid	Bah
7	0,188	0,087	0,388	0,488	0,378	0,132	0,482	0,237	Tidak Bah	Tidak Valid	Bah
8	0,171	0,215	0,371	0,488	0,388	0,148	0,138	0,358	Tidak Bah	Tidak Valid	Bah
9	0,129	0,188	0,118	0,377	0,352	0,148	0,148	0,245	Tidak Bah	Tidak Valid	Bah
10	0,489	0,285	0,386	0,487	0,352	0,488	0,488	0,148	Bah	Valid	Bah
11	0,156	0,289	0,177	0,488	0,388	0,138	0,138	0,388	Tidak Bah	Tidak Valid	Bah
12	0,681	0,423	0,481	0,388	0,488	0,488	0,488	0,138	Tidak Bah	Tidak Valid	Bah
13	0,518	0,482	0,118	0,488	0,482	0,485	0,485	0,138	Bah	Valid	Bah
14	0,151	0,322	0,381	0,388	0,378	0,138	0,488	0,388	Tidak Bah	Tidak Valid	Bah
15	0,222	0,422	0,118	0,322	0,488	0,487	0,158	0,388	Bah	Valid	Bah
16	0,264	0,383	0,381	0,377	0,378	0,487	0,158	0,388	Bah	Valid	Bah
17	0,211	0,255	0,388	0,388	0,372	0,138	0,138	0,188	Bah	Tidak Valid	Bah
18	0,092	0,088	0,484	0,288	0,482	0,138	0,138	0,188	Tidak Bah	Tidak Valid	Bah
19	0,125	0,288	0,375	0,371	0,377	0,488	0,488	0,188	Bah	Tidak Valid	Bah
20	0,178	0,435	0,412	0,378	0,388	0,485	0,138	0,158	Bah	Valid	Bah

No. Rata	Tempat Kondisi	Kondisi Rata- rata/1	Proporsi Pilihan					Kategori					
			A	B	C	D	E	Latihan	Kejadian Insidien	Validasi/1	Pengujian		
21	0,885	0,574	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
22	0,632	0,889	0,88	0,82	0,89	0,79	0,83	0,83	0,88	0,88	0,88	0,88	0,88
23	0,733	0,584	0,83	0,86	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
24	0,522	0,598	0,83	0,81	0,82	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83
25	0,443	0,428	0,83	0,86	0,89	0,85	0,88	0,88	0,88	0,88	0,88	0,88	0,88
26	0,785	0,582	0,85	0,86	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,87
27	0,528	0,381	0,88	0,85	0,88	0,85	0,88	0,88	0,88	0,88	0,88	0,88	0,88
28	0,489	0,526	0,88	0,89	0,89	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
29	0,484	0,583	0,88	0,89	0,89	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
30	0,277	0,828	0,88	0,84	0,88	0,87	0,88	0,88	0,88	0,88	0,88	0,88	0,88
31	0,242	0,384	0,82	0,85	0,88	0,82	0,87	0,87	0,88	0,88	0,88	0,88	0,88
32	0,224	0,486	0,82	0,84	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
33	0,330	0,382	0,87	0,82	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
34	0,325	0,384	0,88	0,82	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
35	0,528	0,581	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
36	0,111	0,128	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
37	0,877	0,448	0,87	0,83	0,87	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
38	0,177	0,298	0,87	0,84	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
39	0,884	0,351	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88
40	0,864	0,382	0,88	0,84	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88

**LAMPIRAN 1D2: HASIL ANALISIS BUTIR TES
MASUK BAHASA INGGRIS B
DENGAN PENDEKATAN TEORI TES
MODERN (RASCH MODEL)**

MicroCAT (tm) Testing System
Copyright (c) 1982, 1984, 1986, 1988 by Assessment
Systems Corporation

Rasch Model Item Calibration Program -- RASCAL (tm)
Version 3.00

Final Parameter Estimates for Data from File
INGGRISB.DAT

Item	Difficulty	Chi Sq.	df
1	0,043	42,099	19
2	-0,096	55,601	19
3	1,382	53,314	19
4	0,749	19,574	19
5	0,761	42,136	19
6	0,198	85,536	19
7	1,364	93,776	19
8	0,749	36,134	19
9	1,100	43,048	19
10	-0,936	34,887	19
11	0,862	19,599	19
12	-1,994	15,203	19
13	-1,564	21,790	19
14	0,902	19,446	19
15	0,393	30,345	19
16	0,142	17,810	19
17	0,466	16,815	19
18	2,131	68,523	19
19	-0,188	31,694	19
20	-0,415	18,358	19
21	-2,758	30,961	19
22	-1,781	20,187	19
23	-2,245	30,562	19
24	-1,133	32,496	19
25	-0,762	18,869	19
26	-2,458	40,645	19

27	-0,454	23,312	19
28	-0,647	39,434	19
29	-0,624	38,693	19
30	0,424	24,201	19
31	0,274	25,236	19
32	0,383	26,214	19
33	-0,213	24,359	19
34	-0,188	14,901	19
35	-1,163	28,812	19
36	1,276	44,728	19
37	1,691	25,923	19
38	0,700	15,988	19
39	1,738	53,968	19
40	1,891	36,810	19

**LAMPIRAN 2A: SKOR TES MASUK TESTEE UNTUK
TES MASUK BAHASA ARAB A
(KLASIK DAN MODERN [θ]) DAN
HASIL BELAJAR BAHASA ARAB**

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ¹
	Skor ²	θ ³	
1	14	-0,724	2,53
2	2	-3,323	3,37
3	19	-0,11	3,80
4	29	1,141	4,00
5	11	-1,135	0,60
6	30	1,289	3,73
7	37	2,844	3,93
8	26	0,736	3,23
9	32	1,615	4,00
10	28	1,001	4,00
11	30	1,289	3,20
12	32	1,615	4,00
13	30	1,289	3,70
14	2	-3,323	2,27
15	26	0,736	3,60
16	27	0,866	3,57
17	11	-1,135	3,73
18	10	-1,285	1,20
19	19	-0,11	3,67
20	27	0,866	3,90
21	19	-0,11	2,90

¹ **Nilai Bahasa Arab:** nilai rerata matakuliah Bahasa Arab I, II, III yang diperoleh mahasiswa pada semester Gasal 2012/2013.

² **Skor:** jumlah butir tes yang dijawab dengan benar oleh testee (0-40).

³ **θ:** theta: nilai testee yang diperoleh dari hasil analisis tes teori modern.

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ¹
	Skor ²	θ^3	
22	21	0,127	2,97
23	29	1,141	3,97
24	8	-1,617	1,80
25	12	-0,992	3,07
26	6	-2,011	2,40
27	10	-1,285	3,30
28	20	0,009	3,63
29	23	0,365	3,40
30	14	-0,724	3,60
31	7	-1,804	1,43
32	23	0,365	3,63
33	28	1,001	3,83
34	5	-2,246	3,40
35	6	-2,011	3,50
36	18	-0,229	3,67
37	29	1,141	3,63
38	4	-2,522	3,07
39	9	-1,445	3,03
40	29	1,141	3,53
41	20	0,009	2,67
42	9	-1,445	3,37
43	21	0,127	3,30
44	9	-1,445	3,77
45	32	1,615	2,87
46	27	0,866	3,77
47	8	-1,617	2,80
48	6	-2,011	2,60
49	5	-2,246	2,67
50	31	1,446	3,80
51	23	0,365	3,17

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ¹
	Skor ²	θ^3	
52	1	-4,07	2,00
53	21	0,127	3,77
54	27	0,866	3,63
55	17	-0,349	3,07
56	20	0,009	3,77
57	11	-1,135	3,00
58	22	0,246	3,47
59	2	-3,323	1,73
60	4	-2,522	1,70
61	22	0,246	2,50
62	20	0,009	3,17
63	13	-0,856	2,23
64	20	0,009	3,53
65	14	-0,724	3,67
66	22	0,246	3,87
67	10	-1,285	2,83
68	31	1,446	3,73
69	27	0,866	3,80
70	27	0,866	1,80
71	8	-1,617	2,27
72	9	-1,445	1,43
73	8	-1,617	2,00
74	9	-1,445	3,40
75	10	-1,285	3,47
76	3	-2,863	2,67
77	3	-2,863	1,90
78	8	-1,617	0,87
79	8	-1,617	2,70
80	14	-0,724	3,33
81	10	-1,285	1,77

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ¹
	Skor ²	θ^3	
82	16	-0,472	1,57
83	5	-2,246	2,47
84	35	2,235	4,00
85	19	-0,11	3,23
86	28	1,001	3,83
87	15	-0,596	3,73
88	16	-0,472	3,90
89	20	0,009	2,83
90	9	-1,445	3,77
91	31	1,446	2,27
92	9	-1,445	3,70
93	12	-0,992	1,00
94	28	1,001	3,90
95	12	-0,992	4,00
96	19	-0,11	2,73
97	7	-1,804	3,30
98	13	-0,856	2,50
99	13	-0,856	2,80
100	6	-2,011	2,47
101	24	0,487	3,10
102	3	-2,863	2,43
103	11	-1,135	3,40
104	15	-0,596	3,17
105	6	-2,011	3,40
106	14	-0,724	2,67
107	14	-0,724	2,57
108	15	-0,596	3,70

**LAMPIRAN 2B: SKOR TES MASUK TESTEE UNTUK
TES MASUK BAHASA ARAB B
(KLASIK DAN MODERN [θ]) DAN
HASIL BEAJAR BAHASA ARAB**

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ⁴
	Skor ⁵	θ ⁶	
1	15	-0,596	3,17
2	30	1,295	4,00
3	34	1,995	3,77
4	13	-0,85	2,90
5	18	-0,217	3,43
6	30	1,295	3,60
7	27	0,877	3,33
8	13	-0,85	1,60
9	35	2,222	4,00
10	37	2,819	3,93
11	29	1,149	2,47
12	27	0,877	2,07
13	1	-4,075	2,50
14	34	1,995	3,77
15	14	-0,717	2,57
16	27	0,877	3,80
17	20	0,022	2,57
18	21	0,14	3,13
19	22	0,259	3,17
20	10	-1,284	0,60
21	14	-0,717	2,90
22	33	1,795	3,40
23	24	0,5	3,03
24	5	-2,251	2,37
25	7	-1,806	1,70

⁴ *Nilai Bahasa Arab*: nilai rerata matakuliah Bahasa Arab I, II, III yang diperoleh mahasiswa pada semester Gasal 2012/2013.

⁵ *Skor*: jumlah butir tes yang dijawab dengan benar oleh testee (0-40).

⁶ *θ*: theta: nilai testee yang diperoleh dari hasil analisis tes teori modern.

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ⁴
	Skor ⁵	θ^6	
26	19	-0,098	3,20
27	23	0,379	4,00
28	12	-0,988	2,93
29	20	0,022	3,07
30	7	-1,806	2,80
31	12	-0,988	2,27
32	21	0,14	3,50
33	23	0,379	3,10
34	2	-3,328	2,80
35	16	-0,462	2,77
36	16	-0,462	3,67
37	13	-0,85	3,33
38	24	0,5	3,00
39	9	-1,445	2,23
40	18	-0,217	2,00
41	17	-0,339	2,57
42	13	-0,85	1,43
43	9	-1,445	3,07
44	18	-0,217	3,27
45	13	-0,85	2,53
46	8	-1,618	3,20
47	20	0,022	3,03
48	16	-0,462	4,00
49	29	1,149	4,00
50	9	-1,445	2,20
51	21	0,14	2,93
52	24	0,5	3,23
53	9	-1,445	2,17
54	30	1,295	4,00
55	15	-0,588	2,43
56	20	0,022	3,40
57	14	-0,717	2,07
58	22	0,259	3,57

No. Testee	Skor Tes Masuk		Nilai Bhs. Arab ⁴
	Skor ⁵	θ^6	
59	21	0,14	4,00
60	29	1,149	4,00
61	28	1,01	3,57
62	19	-0,098	3,40
63	25	0,623	3,97
64	14	-0,717	2,27
65	17	-0,339	3,47
66	17	-0,339	3,97
67	12	-0,988	2,30
68	26	0,748	3,87
69	18	-0,217	3,00
70	13	-0,85	3,90
71	6	-2,015	3,33
72	6	-2,015	2,73
73	15	-0,588	3,30
74	8	-1,618	2,73
75	16	-0,462	2,37
76	5	-2,251	3,40
77	9	-1,445	3,13
78	2	-3,328	3,23
79	12	-0,988	3,10
80	15	-0,588	3,67
81	15	-0,588	2,03
82	4	-2,528	2,13
83	14	-0,717	2,30
84	4	-2,528	1,17
85	3	-2,869	2,23
86	6	-2,015	3,23
87	15	-0,588	3,47

**LAMPIRAN 2C: SKOR TES MASUK TESTEE UNTUK
TES MASUK BAHASA INGGRIS A
(KLASIK DAN MODERN [θ]) DAN
HASIL BELAJAR BAHASA INGGRIS**

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ⁷
	Skor ⁸	θ ⁹	
1	10	-1,269	2,57
2	10	-1,269	3,13
3	3	-2,838	3,03
4	14	-0,709	4,00
5	12	-0,977	3,00
6	8	-1,599	3,07
7	1	-4,036	1,80
8	7	-1,785	3,20
9	4	-2,499	4,00
10	10	-1,269	2,83
11	13	-0,84	4,00
12	10	-1,269	2,13
13	6	-1,991	2,30
14	6	-1,991	3,60
15	10	-1,269	2,60
16	8	-1,599	3,90
17	14	-0,709	3,10
18	6	-1,991	3,37
19	6	-1,991	3,13
20	11	-1,119	0,57
21	5	-2,225	2,57

⁷ *Nilai Bahasa Arab*: nilai rerata matakuliah Bahasa Inggris I, II, III yang diperoleh mahasiswa pada semester Gasal 2012/2013.

⁸ *Skor*: jumlah butir tes yang dijawab dengan benar oleh testee (0-40).

⁹ *θ*: theta: nilai testee yang diperoleh dari hasil analisis tes teori modern.

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ⁷
	Skor ⁸	θ^9	
52	7	-1,785	3,33
53	1	-4,036	2,77
54	14	-0,709	3,80
55	7	-1,785	2,50
56	7	-1,785	2,67
57	6	-1,991	3,13
58	11	-1,119	0,00
59	11	-1,119	3,00
60	1	-4,036	3,03
61	10	-1,269	3,83
62	11	-1,119	2,73
63	8	-1,599	3,70
64	1	-4,036	2,47
65	13	-0,84	2,80
66	8	-1,599	3,70
67	7	-1,785	2,27
68	3	-2,838	0,00
69	9	-1,428	2,90
70	7	-1,785	2,70
71	9	-1,428	2,70
72	8	-1,599	3,13
73	11	-1,119	2,73
74	2	-3,293	2,67
75	6	-1,991	3,17
76	5	-2,225	3,60
77	1	-4,036	2,57
78	3	-2,838	2,63
79	12	-0,977	3,47
80	12	-0,977	3,90
81	3	-2,838	2,20

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ⁷
	Skor ⁸	θ^9	
82	8	-1,599	2,30
83	12	-0,977	3,50
84	2	-3,293	1,27
85	16	-0,458	3,63
86	9	-1,428	3,93
87	7	-1,785	3,43
88	8	-1,599	3,10
89	3	-2,838	1,77
90	8	-1,599	1,27
91	15	-0,582	1,53
92	7	-1,785	3,03
93	3	-2,838	4,00
94	10	-1,269	3,40
95	5	-2,225	3,53
96	15	-0,582	3,87
97	13	-0,84	3,50
98	6	-1,991	0,43
99	7	-1,785	2,27
100	15	-0,582	2,50
101	9	-1,428	2,97
102	3	-2,838	3,20
103	2	-3,293	2,63
104	5	-2,225	2,80
105	10	-1,269	3,13
106	8	-1,599	3,23
107	13	-0,84	3,50
108	13	-0,84	2,40
109	13	-0,84	3,30
110	9	-1,428	2,43
111	11	-1,119	2,53

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ⁷
	Skor ⁸	θ^9	
112	13	-0,84	3,13
113	16	-0,458	3,33
114	15	-0,582	2,37
115	7	-1,785	2,40
116	5	-2,225	1,00
117	6	-1,991	2,60
118	3	-2,838	3,47
119	10	-1,269	3,10
120	4	-2,499	2,30
121	11	-1,119	2,87
122	7	-1,785	2,63
123	8	-1,599	3,17
124	11	-1,119	2,40
125	7	-1,785	2,10
126	7	-1,785	2,97
127	11	-1,119	2,30
128	10	-1,269	2,80
129	1	-4,036	2,23
130	9	-1,428	2,90
131	2	-3,293	3,60
132	5	-2,225	3,00
133	5	-2,225	2,43
134	4	-2,499	2,93
135	11	-1,119	2,73
136	9	-1,428	3,13
137	9	-1,428	0,77
138	8	-1,599	2,73
139	12	-0,977	0,67
140	8	-1,599	2,37
141	8	-1,599	3,07

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ⁷
	Skor ⁸	θ^9	
142	4	-2,499	2,87
143	8	-1,599	3,30
144	4	-2,499	2,70
145	1	-4,036	3,53
146	3	-2,838	1,33
147	13	-0,84	0,93
148	3	-2,838	3,13
149	5	-2,225	3,03
150	15	-0,582	3,13
151	10	-1,269	3,37
152	5	-2,225	2,77
153	8	-1,599	2,40
154	11	-1,119	3,37
155	6	-1,991	3,53
156	8	-1,599	1,63
157	7	-1,785	2,67
158	7	-1,785	3,67
159	8	-1,599	2,97
160	11	-1,119	3,60
161	18	-0,216	2,30
162	5	-2,225	3,27
163	9	-1,428	3,20
164	13	-0,84	4,00
165	8	-1,599	2,33
166	6	-1,991	2,53
167	15	-0,582	3,47
168	13	-0,84	3,87
169	6	-1,991	3,07
170	16	-0,458	3,47
171	10	-1,269	2,87

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ⁷
	Skor ⁸	θ^9	
172	17	-0,336	2,70
173	6	-1,991	2,03

**LAMPIRAN 2D: SKOR TES MASUK TESTEE UNTUK
TES MASUK BAHASA INGGRIS B
(KLASIK DAN MODERN [θ]) DAN
HASIL BELAJAR BAHASA INGGRIS**

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ ¹²	
1	13	-0,923	3,53
2	15	-0,629	2,53
3	17	-0,352	3,73
4	3	-3,166	2,57
5	18	-0,217	3,20
6	21	0,178	3,10
7	18	-0,217	3,47
8	14	-0,774	3,03
9	12	-1,077	3,13
10	7	-1,997	2,97
11	4	-2,796	3,47
12	16	-0,489	3,70
13	12	-1,077	3,50
14	7	-1,997	3,73
15	15	-0,629	3,47
16	16	-0,489	2,83
17	12	-1,077	3,20
18	9	-1,591	1,07
19	15	-0,629	2,57
20	15	-0,629	2,93
21	14	-0,774	2,53

¹⁰ **Nilai Bahasa Arab:** nilai rerata matakuliah Bahasa Inggris I, II, III yang diperoleh mahasiswa pada semester Gasal 2012/2013.

¹¹ **Skor:** jumlah butir tes yang dijawab dengan benar oleh testee (0-40).

¹² **θ** theta: nilai testee yang diperoleh dari hasil analisis tes teori modern.

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ^{12}	
22	17	-0,352	3,07
23	12	-1,077	1,40
24	16	-0,489	3,03
25	15	-0,629	3,30
26	12	-1,077	2,37
27	11	-1,239	1,60
28	13	-0,923	2,67
29	22	0,309	3,00
30	8	-1,786	3,10
31	8	-1,786	3,33
32	5	-2,492	3,47
33	14	-0,774	3,33
34	10	-1,41	2,13
35	10	-1,41	3,27
36	3	-3,166	3,50
37	11	-1,239	3,10
38	5	-2,492	2,87
39	23	0,441	3,60
40	10	-1,41	2,53
41	10	-1,41	2,90
42	16	-0,489	3,10
43	6	-2,23	3,17
44	18	-0,217	2,93
45	8	-1,786	2,50
46	13	-0,923	2,77
47	22	0,309	3,23
48	19	-0,084	3,67
49	20	0,047	4,00
50	13	-0,923	2,87
51	22	0,309	3,67

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ^{12}	
52	13	-0,923	3,40
53	6	-2,23	2,80
54	14	-0,774	4,00
55	22	0,309	3,77
56	5	-2,492	3,43
57	1	-4,433	0,77
58	15	-0,629	3,83
59	21	0,178	4,00
60	15	-0,629	3,47
61	24	0,574	3,87
62	14	-0,774	3,00
63	23	0,441	3,33
64	13	-0,923	2,47
65	17	-0,352	3,67
66	11	-1,239	3,07
67	18	-0,217	3,20
68	15	-0,629	3,10
69	20	0,047	3,90
70	13	-0,923	3,67
71	20	0,047	3,57
72	18	-0,217	3,30
73	15	-0,629	2,70
74	21	0,178	3,93
75	17	-0,352	3,17
76	6	-2,23	0,00
77	5	-2,492	1,73
78	12	-1,077	2,73
79	22	0,309	3,23
80	7	-1,997	3,50
81	7	-1,997	2,73

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ^{12}	
82	9	-1,591	3,80
83	5	-2,492	2,57
84	8	-1,786	3,30
85	17	-0,352	3,53
86	2	-3,655	0,67
87	15	-0,629	2,53
88	12	-1,077	2,63
89	2	-3,655	2,87
90	14	-0,774	3,20
91	11	-1,239	3,83
92	15	-0,629	2,90
93	8	-1,786	3,13
94	6	-2,23	2,13
95	20	0,047	3,30
96	1	-4,433	3,33
97	14	-0,774	2,93
98	10	-1,41	2,80
99	10	-1,41	2,43
100	11	-1,239	3,60
101	5	-2,492	2,23
102	9	-1,591	2,07
103	11	-1,239	3,07
104	14	-0,774	3,57
105	14	-0,774	2,60
106	14	-0,774	2,23
107	12	-1,077	3,03
108	12	-1,077	2,23
109	23	0,441	3,13
110	13	-0,923	3,10
111	6	-2,23	2,63

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ ¹²	
112	6	-2,23	2,10
113	13	-0,923	3,83
114	15	-0,629	2,87
115	9	-1,591	2,60
116	17	-0,352	2,33
117	12	-1,077	3,10
118	12	-1,077	2,77
119	14	-0,774	3,20
120	16	-0,489	3,20
121	13	-0,923	3,00
122	6	-2,23	1,70
123	15	-0,629	2,80
124	7	-1,997	1,47
125	4	-2,796	2,80
126	11	-1,239	1,80
127	13	-0,923	3,00
128	6	-2,23	1,73
129	16	-0,489	2,73
130	3	-3,166	1,80
131	14	-0,774	3,30
132	18	-0,217	2,50
133	21	0,178	2,67
134	11	-1,239	2,97
135	6	-2,23	3,27
136	19	-0,084	3,27
137	11	-1,239	2,23
138	6	-2,23	3,27
139	10	-1,41	1,97
140	20	0,047	3,20
141	11	-1,239	3,60

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ ¹²	
142	11	-1,239	0,80
143	14	-0,774	3,07
144	10	-1,41	1,10
145	16	-0,489	3,00
146	12	-1,077	2,70
147	23	0,441	3,50
148	9	-1,591	2,27
149	3	-3,166	2,43
150	15	-0,629	3,10
151	14	-0,774	2,47
152	12	-1,077	3,00
153	8	-1,786	1,50
154	12	-1,077	2,87
155	11	-1,239	2,17
156	12	-1,077	3,53
157	7	-1,997	2,50
158	13	-0,923	2,93
159	16	-0,489	3,00
160	11	-1,239	2,50
161	5	-2,492	3,23
162	17	-0,352	2,83
163	18	-0,217	2,27
164	18	-0,217	1,37
165	24	0,574	1,47
166	14	-0,774	2,73
167	10	-1,41	2,53
168	12	-1,077	3,43
169	11	-1,239	1,97
170	2	-3,655	2,17
171	10	-1,41	2,83

No. Testee	Skor Tes Masuk		Nilai Bhs. Inggris ¹⁰
	Skor ¹¹	θ ¹²	
172	11	-1,239	2,30
173	14	-0,774	0,73
174	17	-0,352	3,10
175	11	-1,239	2,43
176	14	-0,774	2,87
177	14	-0,774	2,40
178	17	-0,352	3,87
179	8	-1,786	3,00
180	12	-1,077	3,80
181	11	-1,239	2,37
182	10	-1,41	2,30
183	8	-1,786	2,60
184	14	-0,774	2,93
185	18	-0,217	3,77
186	13	-0,923	2,20
187			
188			

**LAMPIRAN 3A1: HASIL ANALISIS REGRESI NILAI
BAHASA ARAB A PADA SKOR TES
MASUK BHS ARAB A DENGAN
PENDEKATAN TEORI TES KLASIK**

Regression

[DataSet4] D:\LAIN-LAIN\PENELITIAN\DATA\B Arab A.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK Bahasa Arab	3.0565	.79802	108
Skor Tes B Arab A Klasik	16.56	9.302	108

Correlations

		Nilai MK Bahasa Arab	Skor Tes B Arab A Klasik
Pearson Correlation	Nilai MK Bahasa Arab	1.000	.519
	Skor Tes B Arab A Klasik	.519	1.000
Sig. (1-tailed)	Nilai MK Bahasa Arab	.	.000
	Skor Tes B Arab A Klasik	.000	.
N	Nilai MK Bahasa Arab	108	108
	Skor Tes B Arab A Klasik	108	108

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.519 ^a	.269	.262	.68534	.269	39.076	1	106	.000

a. Predictors: (Constant), Skor Tes B Arab A Klasik

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.354	1	18.354	39.076	.000 ^a
	Residual	49.788	106	.470		
	Total	68.141	107			

a. Predictors: (Constant), Skor Tes B Arab A Klasik

b. Dependent Variable: Nilai MK Bahasa Arab

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.319	.135		17.157	.000
	Skor Tes B Arab A Klasik	.045	.007	.519	6.251	.000

a. Dependent Variable: Nilai MK Bahasa Arab

**LAMPIRAN 3A2: HASIL ANALISIS REGRESI NILAI
BAHASA ARAB PADA SKOR TES
MASUK BHS ARAB A DENGAN
PENDEKATAN TEORI TES
MODERN**

Regression

[DataSet4] D:\LAIN-LAIN\PENELITIAN\DATA\B Arab A.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK Bahasa Arab	3.0565	.79802	108
Skor Tes B Arab A Modern	-.5230	1.37029	108

Correlations

		Nilai MK Bahasa Arab	Skor Tes B Arab A Modern
Pearson Correlation	Nilai MK Bahasa Arab	1.000	.515
	Skor Tes B Arab A Modern	.515	1.000
Sig. (1-tailed)	Nilai MK Bahasa Arab	.	.000
	Skor Tes B Arab A Modern	.000	.
N	Nilai MK Bahasa Arab	108	108
	Skor Tes B Arab A Modern	108	108

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.515 ^a	.266	.259	.68712	.266	38.324	1	106	.000

a. Predictors: (Constant), Skor Tes B Arab A Modern

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.094	1	18.094	38.324	.000 ^a
	Residual	50.047	106	.472		
	Total	68.141	107			

a. Predictors: (Constant), Skor Tes B Arab A Modern

b. Dependent Variable: Nilai MK Bahasa Arab

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.213	.071		45.379	.000
	Skor Tes B Arab A Modern	.300	.048	.515	6.191	.000

a. Dependent Variable: Nilai MK Bahasa Arab

**LAMPIRAN 3B1: HASIL ANALISIS REGRESI NILAI
BAHASA ARAB A PADA SKOR TES
MASUK BHS ARAB B DENGAN
PENDEKATAN TEORI TES KLASIK**

Regression

[DataSet3] D:\LAIN-LAIN\PENELITIAN\DATA\B Arab B.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK B Arab	2.9979	.72626	87
Skor Tes B Arab B Klasik	16.97	8.583	87

Correlations

		Nilai MK B Arab	Skor Tes B Arab B Klasik
Pearson	Nilai MK B Arab	1.000	.543
Correlation	Skor Tes B Arab B Klasik	.543	1.000
Sig. (1-tailed)	Nilai MK B Arab		.000
	Skor Tes B Arab B Klasik	.000	
N	Nilai MK B Arab	87	87
	Skor Tes B Arab B Klasik	87	87

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.543 ^a	.295	.286	.61356	.295	35.492	1	85	.000

a. Predictors: (Constant), Skor Tes B Arab B Klasik

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13.361	1	13.361	35.492	.000 ^a
	Residual	31.999	85	.376		
	Total	45.360	86			

a. Predictors: (Constant), Skor Tes B Arab B Klasik

b. Dependent Variable: Nilai MK B Arab

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.219	.146		15.156	.000
	Skor Tes B Arab B Klasik	.046	.008	.543	5.958	.000

a. Dependent Variable: Nilai MK B Arab

**LAMPIRAN 3B2: HASIL ANALISIS REGRESI NILAI
BAHASA ARAB PADA SKOR TES
MASUK BHS ARAB B DENGAN
PENDEKATAN TEORI TES
MODERN**

Regression

[DataSet3] D:\LAIN-LAIN\PENELITIAN\DATA\B Arab B.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK B Arab	2.9979	.72626	87
Skor Tes B Arab B Modern	-.4376	1.28570	87

Correlations

		Nilai MK B Arab	Skor Tes B Arab B Modern
Pearson	Nilai MK B Arab	1.000	.513
Correlation	Skor Tes B Arab B Modern	.513	1.000
Sig. (1-tailed)	Nilai MK B Arab		.000
	Skor Tes B Arab B Modern	.000	
N	Nilai MK B Arab	87	87
	Skor Tes B Arab B Modern	87	87

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.513 ^a	.264	.255	.62687	.264	30.430	1	85	.000

a. Predictors: (Constant), Skor Tes B Arab B Modern

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11.958	1	11.958	30.430	.000 ^a
	Residual	33.402	85	.393		
	Total	45.360	86			

a. Predictors: (Constant), Skor Tes B Arab B Modern

b. Dependent Variable: Nilai MK B Arab

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.125	.071		43.989	.000
	Skor Tes B Arab B Modern	.290	.053	.513	5.516	.000

a. Dependent Variable: Nilai MK B Arab

**LAMPIRAN 4C1: HASIL ANALISIS REGRESI NILAI
BAHASA INGGRIS PADA SKOR TES
MASUK BHS INGGRIS A DENGAN
PENDEKATAN TEORI TES KLASIK**

Regression

[DataSet1] D:\LAIN-LAIN\PENELITIAN\DATA\B Inggris A.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK B Inggris	2.8049	.76253	173
Skor Tes B Inggris Klasik A	8.27	3.950	173

Correlations

		Nilai MK B Inggris	Skor Tes B Inggris Klasik A
Pearson Correlation	Nilai MK B Inggris	1.000	.147
	Skor Tes B Inggris Klasik A	.147	1.000
Sig. (1-tailed)	Nilai MK B Inggris		.027
	Skor Tes B Inggris Klasik A	.027	
N	Nilai MK B Inggris	173	173
	Skor Tes B Inggris Klasik A	173	173

Model Summary

Model	Change Statistics				
	R Square Change	F Change	df1	df2	Sig. F Change
1	.022 ^a	3.789	1	171	.053

a. Predictors: (Constant), Skor Tes B Inggris Klasik A

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.168	1	2.168	3.789	.053 ^a
	Residual	97.842	171	.572		
	Total	100.010	172			

a. Predictors: (Constant), Skor Tes B Inggris Klasik A

b. Dependent Variable: Nilai MK B Inggris

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.570	.134		19.212	.000
	Skor Tes B Inggris Klasik A	.028	.015	.147	1.946	.053

a. Dependent Variable: Nilai MK B Inggris

**LAMPIRAN 3C2: HASIL ANALISIS REGRESI NILAI
BAHASA INGGRIS PADA SKOR TES
MASUK BHS INGGRIS A DENGAN
PENDEKATAN TEORI MODERN**

REGRESI

[DataSet1] D:\LAIN-LAIN\PENELITIAN\DATA\B Inggris A.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK B Inggris	2.8049	.76253	173
Skor Tes B Inggris Modern A	-1.7069	.84430	173

Correlations

		Nilai MK B Inggris	Skor Tes B Inggris Modern A
Pearson Correlation	Nilai MK B Inggris	1.000	.143
	Skor Tes B Inggris Modern A	.143	1.000
Sig. (1-tailed)	Nilai MK B Inggris		.030
	Skor Tes B Inggris Modern A	.030	
N	Nilai MK B Inggris	173	173
	Skor Tes B Inggris Modern A	173	173

Model Summary

Model	Change Statistics				
	R Square Change	F Change	df1	df2	Sig. F Change
1	.020 ^a	3.562	1	171	.061

a. Predictors: (Constant), Skor Tes B Inggris Modern A

ANOVA^b

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	2.041	1	2.041	3.562	.061 ^a
Residual	97.969	171	.573		
Total	100.010	172			

a. Predictors: (Constant), Skor Tes B Inggris Modern A

b. Dependent Variable: Nilai MK B Inggris

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	3.025	.130		23.252	.000
Skor Tes B Inggris Modern A	.129	.068	.143	1.887	.061

a. Dependent Variable: Nilai MK B Inggris

**LAMPIRAN 3D1: HASIL ANALISIS REGRESI NILAI
BAHASA INGGRIS PADA SKOR TES
MASUK BHS INGGRIS B DENGAN
PENDEKATAN TEORI TES KLASIK**

Regression

[DataSet2] D:\LAIN-LAIN\PENELITIAN\DATA\B Inggris B.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK B Inggris	2.8466	.71608	186
Skor Tes B Inggris B Klasik	12.55	5.162	186

Correlations

		Nilai MK B Inggris	Skor Tes B Inggris B Klasik
Pearson Correlation	Nilai MK B Inggris	1.000	.372
	Skor Tes B Inggris B Klasik	.372	1.000
Sig. (1- tailed)	Nilai MK B Inggris		.000
	Skor Tes B Inggris B Klasik	.000	
N	Nilai MK B Inggris	186	186
	Skor Tes B Inggris B Klasik	186	186

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.372 ^a	.138	.134	.66655	.138	29.513	1	184	.000

a. Predictors: (Constant), Skor Tes B Inggris B Klasik

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13.112	1	13.112	29.513	.000 ^a
	Residual	81.749	184	.444		
	Total	94.862	185			

a. Predictors: (Constant), Skor Tes B Inggris B Klasik

b. Dependent Variable: Nilai MK B Inggris

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.199	.129		17.073	.000
	Skor Tes B Inggris B Klasik	.052	.009	.372	5.433	.000

a. Dependent Variable: Nilai MK B Inggris

**LAMPIRAN 3D2: HASIL ANALISIS REGRESI NILAI
BAHASA INGGRIS PADA SKOR TES
MASUK BHS INGGRIS B DENGAN
PENDEKATAN TEORI MODERN**

Regression

[DataSet2] D:\LAIN-LAIN\PENELITIAN\DATA\B Inggris B.sav

Descriptive Statistics

	Mean	Std. Deviation	N
Nilai MK B Inggris	2.8466	.71608	186
Skor Tes B Inggris B Modern	-1.1088	.93842	186

Correlations

		Nilai MK B Inggris	Skor Tes B Inggris B Modern
Pearson	Nilai MK B Inggris	1.000	.363
Correlation	Skor Tes B Inggris B Modern	.363	1.000
Sig. (1-tailed)	Nilai MK B Inggris	.	.000
	Skor Tes B Inggris B Modern	.000	.
N	Nilai MK B Inggris	186	186
	Skor Tes B Inggris B Modern	186	186

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.363 ^a	.132	.127	.66892	.132	28.000	1	184	.000

a. Predictors: (Constant), Skor Tes B Inggris B Modern

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12.529	1	12.529	28.000	.000 ^a
	Residual	82.333	184	.447		
	Total	94.862	185			

a. Predictors: (Constant), Skor Tes B Inggris B Modern

b. Dependent Variable: Nilai MK B Inggris

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.154	.076		41.477	.000
	Skor Tes B Inggris B Modern	.277	.052	.363	5.292	.000

a. Dependent Variable: Nilai MK B Inggris



LEMBAGA PENELITIAN DAN PENGABDIAN
KEPADA MASYARAKAT