

PENGEMBANGAN INSTRUMEN TES TERTULIS SELEKSI MASUK STAIN PONOROGO TAHUN AKADEMI 2010/2011

Oleh: Athok Fu'adi, M.Pd
Dosen STAIN Ponorogo
E-mail: athokfuadi@yahoo.com

Abstract

This research for knowing quality of written test instruments for STAIN Ponorogo entrance selection in academic year 2010/2011, they are : 1) how validity test is; 2) how far level of difficulty is; 3) how the different power test is; 4) how the validity result analysis, the level of difficulty, different power the test are.

This research is quantitaf with the reseach subjects are student applicants and test question. Data is collected by interview and documentation. Sample is taken 10% from all of answer. Validity analysis uses SPSS, while the level of difficulty and different power use microsoft excel programe. The reseach results demonstrate that validity of written test selection STAIN Ponorogo in academic year 2010/2011 for islamic education is 17 items, for Arabic language is 53 items, for English language is 15 items, and General Knoledge is 25 items, so the validity of the test is very low.

Whereas inthe level of difficulty of test instruments for islamic education, Too Difficult (TD) is 16 items, Middle (M) is 39 items, and Too Esasy (TE) is 55 items, for Arabic language Too Difficult (TD) is 25 items, Middle (M) is 69 items, Too Easy (TE) is 6 items, for English language Too Difficult (TD) is 45 items, Middle (M) is 51 items, Too Easy (TE) is 4 items. For general knowledge, Too Difficult (TD) is 18 items, Middle (M) is 46 items, Too Easy (TE) is 36 items. The different power for islamic education, exellent is 0 item, good is 16 items, enough is 41 items, bad is 44 items, very bad is 4 items, so the items were worth at the preserve is on the criteria of excellent, good and enough that are 52 items, for 44 bad items is replaced, and 4 very bad items of this matter should not be issued again the future. 70 items is maintained, while 23 items is replaced , which is 7 items should not be used in the future.

Different power for English exellent is 0 item, good is 7 items, enough is 29 items, bad is 43 items, very bad is 20 items, so the items are worth at the preserved there are 36 items, in exchange for 43 points, while 20 points in the future should not be used. Different power general knowledge, exellent is 0 items, 13 items well, 36 items is sufficient, 46 items is bad, 8 items is very bad, then the matter should be maintained 49 items, 46 items are replaced while the 8 items in the future should not be used. Analysis of the validity of a written test on Islamic education is 17%, Arabic is 53%, English is 15%, whereas 25% for general knowledge, it demonstrates the validity of the test is very low for the entrance test STAIN Ponorogo in academic year 2010/2011. Whereas the difficulty level of Islamic Educationis average of 39% is good, 69% for Arabic, 51% for English, 46% for general knowledge, while the good

different power for Islamic Education is 52%, 70% for Arabic, 37% for English language, and 49% for general knowledge.

Keywords: Test development , test instrument , the University entrance selection

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas instrumen tes tertulis seleksi masuk STAIN Ponorogo tahun akademi 2010/2011 yang meliputi: 1) bagaimanakah validitas tes, 2) bagaimana tingkat kesukaran tes, 3) bagaimana daya beda tes, 4)bagaimana analisis hasil validitas, tingkat kesukaran, daya beda tes..

Penelitian ini merupakan penelitian kuantitatif dengan subjek penelitian adalah jawaban calon mahasiswa dan soal tes. Data dikumpulkan melalui wawancara, dokumentasi.Pengambilan sampel dengan mengambil jumlah jawaban 10% dari jawaban seluruhnya. Dengan analisis validitas menggunakan SPSS sedangkan tingkat kesukaran dan daya beda menggunakan program Excel

Hasil penelitian menunjukkan bahwa Bahwa validitas dari tes tertulis ujian masuk STAIN Ponorogo tahun akademik 2010/2011 pada pendidikan agama Islam yang valid 17 butir tes, untuk bahasa Arab yang valid 53 butir tes, untuk bahasa Inggris yang valid 15 butir tes, sedangkan pengetahuan umum yang valid 25 butir tes, sehingga validitas tesnya sangat rendah. Sedangkan pada tingkat kesukaran instrumen tes pada pendidikan agama Islam, terlalu sukar (TS) 16 butir, baik (S) 39 butir, dan terlalu mudah (TM) 55 butir, pada bahasa arab, terlalu sukar (TS) 25 butir, baik (S) 69 butir, telalu mudah (TM) 6 butir, bahasa Inggris, terlalu sukar (TS) 45 butir, baik (S) 51 butir, terlalu mudah (TM) 4 butir, pada pengetahuan umum, terlalu sukar (TS) 18 butir, baik (S) 46 butir, terlalu mudah (TM) 36 butir. Daya beda untuk pendidikan agama Islam baik sekali 0 butir, baik 11 butir, cukup 41 butir, jelek 44 butir, jelek sekali 4 butir, maka butir soal yang patut di pertahankan ada di kriteria baik sekali, baik dan cukup 52 butir, untuk 44 jelek diganti, dan 4 butir jelek sekali kedepannya soal ini tidak boleh di keluarkan lagi. Daya beda bahasa arab, baik sekali 6 butir, baik 40 butir, cukup 24 butir, jelek 23 butir dan jelek sekali 7 butir, maka soal yang patut di pertahankan 70 butir, sedangkan di ganti 23 butir, yang 7 butir kedepannya tidak boleh dipakai. Daya beda bahasa Inggris baik sekalo 0 butir, baik 7 butir, cukup 29 butir, jelek 43 butir, 20 butir jelek sekali, maka butir soal yang patut di pertahankan ada 36 butir, 43 butir di ganti, sedangkan 20 butir kedepannya tidak boleh dipakai. Daya beda pengetahuan umum, baik sekali 0 butir, baik 13 butir, cukup 36 butir, jelek 46 butir, jelek sekali 8 butir, maka soal yang patut dipertahankan 49 butir, sadangkan yang diganti 46 butir sedangkan 8 butir kedepannya tidak boleh dipakai. Analisis validitas tes tertulis pada pendidikan agama Islam 17%, bahasa arab 53%, bahasa Inggris 15%, sedangkan pengetahuan umum 25%, hal ini menunjukkan validitas tes sangat rendah untuk tes masuk STAIN

Ponorogo tahun akademik 2010/2011. Sedangkan Tingkat kesukaran PAI rata-rata 39% yang baik, bahasa arab 69%, bahasa inggris 51%, pengetahuan umum 46%, sedangkan daya beda yang baik untuk PAI 52%, bahasa arab 70%, bahasa inggris 37% , pengetahuan umum 49%.

Kata Kunci: Uji Pembangunan, Alat Tes, Seleksi Masuk Universitas

A. PENDAHULUAN

Pendidikan merupakan has asasi manusia yang paling mendasar dan bersifat universal. Maju mundurnya suatu bangsa akan tergantung dari kualitas sumberdaya manusia, hal ini akan berpengaruh pada tingkat kualitas pendidikan. Era globalisasi menuntut perubahan di segala bidang dengan cepat. Berbagai rumpun keilmuan mengalami kemajuan dengan pesat, termasuk dalam bidang pendidikan. Peningkatan mutu pendidikan di Indonesia telah menjadi kebutuhan yang mendesak saat ini. Persaingan lulusan pendidikan semakin ketat, terutama dalam segi kualitas. Tuntutan masyarakat bahwa mutu pendidikan bisa meningkatkan kesejahteraan hidup masyarakat, perlu dijawab dengan berbagai inovasi dalam dunia pendidikan.

Peningkatan kualitas pendidikan merupakan tujuan bersama yang harus dilakukan demi meningkatkan mutu pendidikan di Indonesia. Pihak pemerintah telah berusaha untuk merumuskan berbagai macam strategi untuk meningkatkan kualitas pendidikan dan melakukan inovasi pendidikan sebagai bagian dari reformasi pendidikan. Undang-Undang No 20 tahun 2003 tentang sistem pendidikan nasional, yang berisi penjelasan tentang tujuan pendidikan serta upaya-upaya peningkatan pendidikan, merupakan salah satu produk reformasi pendidikan.

Fungsi dan tujuan pendidikan nasional seperti yang tercantum pada Undang-undang sistem pendidikan pada Bab II pasal 3 tahun 2003, yaitu:

Mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri, dan menjadi warga negara yang demokratis serta bertanggungjawab.

Perwujudan masyarakat Indonesia yang berkualitas menjadi tanggung jawab dunia pendidikan, terutama dalam mempersiapkan peserta didik menjadi subjek yang makin tangguh, kreatif, mandiri dan profesional menurut bidang keilmuannya masing-masing. Hal ini merupakan sesuatu yang perlu diperhatikan demi mengantisipasi era kesejagatan yang mulai kita rasakan.

Menurut PP No 19 tahun 2005 tentang Standar Nasional Pendidikan, pada bab IV tentang standar proses, diungkapkan bahwa proses pembelajaran pada satuan pendidikan diselenggarakan secara interaktif, inspiratif, menyenangkan, menantang, memotivasi peserta didik untuk berpartisipasi aktif, serta memberikan ruang yang cukup bagi prakarsa, kreativitas, dan kemandirian sesuai dengan bakat, minat, dan perkembangan fisik serta psikologis peserta didik. Ini berarti bahwa dalam kegiatan pembelajaran juga dilakukan secara interaktif, inspiratif, menyenangkan, menantang, memotivasi peserta didik untuk berpartisipasi aktif, serta memberikan ruang yang cukup bagi prakarsa, kreativitas, dan kemandirian sesuai dengan bakat, minat, dan perkembangan fisik serta psikologis peserta didik. Oleh karena proses pendidikan yang baik dan mutunya baik maka harus ditunjang juga dengan sistem seleksi yang baik.

Hampir semua perguruan tinggi di setiap negara bersikap selektif ketika melaksanakan program penerimaan mahasiswa baru. Sikap selektif ini sekurang-kurangnya didasari oleh alasan-alasan berikut. Pertama, perguruan tinggi merupakan tempat untuk mempersiapkan calon-calon pemimpin masyarakat yang akan datang. Oleh karena itu, para pembuat kebijakan dan pimpinan perguruan tinggi pun menginginkan "kepastian" bahwa para mahasiswa yang disiapkan untuk menduduki posisi-posisi penting di masa yang akan datang itu adalah individu-individu yang benar-benar bermutu sesuai dengan yang diharapkan. Kedua, kesempatan untuk belajar di perguruan tinggi merupakan semacam kesempatan yang "langka", sehingga hanya tersedia bagi mereka yang benar-benar berhak mendapatkannya, yaitu mereka yang tergolong "bibit unggul" dari angkatan muda suatu bangsa. Ketiga, "*human talent*", yaitu potensi yang dimiliki oleh para calon mahasiswa, adalah sesuatu yang sangat berharga, yang tidak boleh disia-siakan oleh siapapun. Keempat, pendidikan tinggi adalah suatu upaya yang sangat mahal, sehingga harus dimanfaatkan sebaik mungkin. Oleh sebab hal-hal yang diuraikan di

atas itulah, seleksi mahasiswa baru selalu menjadi hal yang penting di setiap negara. Keputusan mengenai penerimaan atau penolakan calon-calon mahasiswa merupakan keputusan yang sangat besar maknanya bagi generasi mendatang suatu bangsa karena pendidikan tinggi memegang peranan yang sangat penting dalam upaya pengembangan sumber daya manusia.

Sebuah sistem seleksi mahasiswa baru yang baik harus mempunyai landasan yang cukup kuat bila di pandang dari segi akademik, ekonomi, pendidikan, maupun psiko-sosial. Karena itu sebuah sistem penerimaan mahasiswa baru harus mempertimbangkan sekaligus sekurang-kurangnya empat hal, yaitu (a) kecermatan atau akurasi prediksi (*prediction effectiveness*), (b) efisiensi ekonomi (*economic efficiency*), (c) insentif belajar-mengajar (*teaching-learning incentive*), dan (d) keadilan (*equity*). Dengan melihat berbagai hal yang menjadi dasar suatu seleksi penerimaan mahasiswa baru perlu diperhatikan, selain itu yang sangat bberperan dalam seleksi adalah soal ujian seleksi itu sendiri, karena jika suatu perguruan tinggi ingin mendapatkan mahasiswa yang berkualitas maka alat tes sebagai alat seleksi harus baik.

Baik disini alat uji ataupun tes ini harus dapat dipertanggungjawabkan secara validitas maupun secara indek kesukaran dan daya beda sehingga akan dapat diketahui bahwa mahasiswa yang memang terjaring dalam program seleksi penerimaan mahasiswa baru ini merupakan mahasiswa pilihan sesuai dengan tujuan dari perguruan tinggi tersebut. Sekolah Tinggi Agama Islam (STAIN) Ponorogo merupakan salah satu perguruan tinggi agama Islam yang menerapkan seleksi langsung mahasiswanya, sehingga untuk penyedia alat uji soal tes dibuat oleh pihak pengembang soal tes, karena dari setiap tahun diadakan tes, maka perlu diadakan analisis butir soal yang diteskan untuk seleksi masuk STAIN Ponorogo, sehingga akuntabiliatas tes dapat dipertanggungjawabkan dalam rangka meningkatkan kualitas mahasiswa baru.

Ujian yang diadakan di STAIN Ponorogo tidak hanya tes tertulis tetapi ada non tes yaitu tes baca tulis al Quran, dengan adanya tes seleksi yang lengkap maka diharapkan mahasiswa yang di terima di perguruan tinggi merupakan betul-betul mahasiswa pilihan, sehingga diharapkan kedepan dapat menjadi manusia yang berguna bagi bangsa dan agama.

B. RUMUSAN MASALAH

Dengan melihat latar belakang masalah maka ada beberapa rumusan masalah, yaitu:

1. Bagaimanakah validitas soal tes tertulis seleksi masuk STAIN Ponorogo tahun akademik 2010/2011?
2. Bagaimanakah tingkat kesukaran soal tes tertulis seleksi masuk STAIN Ponorogo tahun akademik 2010/2011?
3. Bagaimanakah daya beda soal tes tertulis seleksi masuk STAIN Ponorogo tahun akademik 2010/2011?
4. Bagaimanakah analisis hasil dari validitas, tingkat kesukaran dan daya beda dari soal tes tertulis masuk STAIN Ponorogo tahun akademik 2010/2011?

C. MANFAAT PENELITIAN

Manfaat penelitian ini antara lain:

1. Untuk lembaga STAIN Ponorogo, hasil penelitian ini dapat dijadikan dasar untuk pengembangan tes tertulis seleksi masuk dalam rangka meningkatkan kualitas mahasiswa baru.
2. Untuk pengembang tes tertulis soal, hasil penelitian ini dapat dijadikan alat untuk mengganti soal-soal yang kurang baik, dan mengembangkan yang sudah baik.
3. Untuk peneliti, bahwa penelitian kebijakan ini untuk mengembangkan keilmuan dibidang evaluasi pendidikan yang dapat membuat kebijakan yang membangun bagi lembaga.
4. Untuk masyarakat, bahwa hasil penelitian dapat menjadikan khasanah keilmuan.

D. METODE PENELITIAN

1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di STAIN Ponorogo pada tahun akademik 2010/2011

2. Subjek dan Objek Penelitian

Subjek dan objek penelitian ini adalah calon mahasiswa baru, dan pengembang tes tertulis seleksi masuk STAIN Ponorogo tahun akademik 2010/2011.

3. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini adalah dokumentasi.

Dokumentasi, digunakan untuk mendapatkan data dari soal tes seleksi masuk STAIN Ponorogo Tahun akademik 2010/2011 serta kunci jawaban, selain itu dibutuhkan juga hasil ujian dari calon mahasiswa yang akan masuk STAIN Ponorogo. Sedangkan untuk membatasi data maka peneliti mengambil sampel sesuai dengan buku Suharsimi Arikunto, 10% dari jumlah keseluruhan.

4. Instrumen Penelitian

Penelitian ini bersifat kuantitatif, karena instrumen penelitian adalah soal tertulis seleksi masuk STAIN Ponorogo yang terdiri dari Tes Pengetahuan Agama Islam, Bahasa Arab, Bahasa Inggris, dan Pengetahuan Umum, serta kunci jawaban dari soal tersebut.

5. Teknik Analisis Data

Setelah data terkumpul dari hasil ujian tes seleksi masuk STAIN Ponorogo maka akan dilakukan analisis menggunakan SPSS untuk Validitas. Sedangkan untuk mencari Tingkat Kesukaran memakai rumus

$$TK = \frac{B}{N \times skor.maks}$$

Sedangkan untuk mencari daya beda memakai rumus

$$DB = \frac{U - L}{Nup \times skor.max}$$

Sedangkan untuk menganalisis tes tertulis seleksi masuk STAIN Ponorogo secara keseluruhan untuk validitas menggunakan SPSS sedangkan tingkat kesukaran dan daya beda menggunakan program *EXCEL*.

E. Tinjauan Pustaka

Hasil Penelitian Ali Ridho dan Saifuddin Azwar Universitas Gajah Mada tentang keberfungsian aitem tes UAN Matematika SMA di Propinsi DIY tahun pelajaran 2003/2004.

Penelitian ini dilakukan untuk mempelajari secara empiris karakteristik aitem-aitem tes UAN SMA di Daerah Istimewa Yogyakarta khusus mata ujian matematika tahun pelajaran 2003/2004 yang berjumlah 40. Fokus penelitian adalah keberfungsian aitem diferensial atau *diferensial item functioning* (DIF) berdasarkan kelompok gender.

Hasil penelitian menunjukkan bahwa, dari 38 aitem yang dianalisis: (1) dilihat dari parameter-parameter aitemnya, terdapat 33 (86,84%) aitem diterima dan 5 (13,16%) aitem ditolak; (2) dilihat dari informasi tes, tes UAN matematika SMA 2004 paling efektif dalam membedakan antara *testee* dengan kemampuan θ di atas 1 dan di bawah 1, (3) terdapat 12 aitem yang terdeteksi DIF di mana 5 aitem menguntungkan laki-laki dan 7 aitem menguntungkan perempuan.

F. KAJIAN TEORI

1. Pengertian Tes

Tes merupakan sejumlah pertanyaan yang memiliki jawaban yang benar atau salah. Sedangkan menurut Slameto tes menurut materi yang akan dinilai, bentuknya, dan caranya membuat. Tes Tertulis merupakan tes dimana soal dan jawaban yang diberikan kepada peserta didik dalam bentuk tulisan. Dalam menjawab soal peserta didik tidak selalu merespon dalam bentuk menulis jawaban tetapi dapat juga dalam bentuk yang lain seperti memberi tanda, mewarnai, menggambar dan lain sebagainya.

Ada dua bentuk soal tes tertulis, yaitu:

- a. Soal dengan memilih jawaban
 - 1) pilihan ganda
 - 2) dua pilihan (benar-salah, ya-tidak)
 - 3) menjodohkan
- b. Soal dengan mensuplai-jawaban.
 - 1) isian singkat atau melengkapi
 - 2) uraian terbatas
 - 3) uraian obyektif / non obyektif
 - 4) uraian terstruktur / nonterstruktur .

Dari berbagai alat penilaian tertulis, tes memilih jawaban *benar-salah, isian singkat, dan menjodohkan* merupakan alat yang hanya menilai *kemampuan berpikir rendah*, yaitu kemampuan mengingat (pengetahuan). Tes pilihan ganda dapat digunakan untuk menilai kemampuan mengingat dan memahami. Pilihan ganda mempunyai kelemahan, yaitu peserta didik tidak mengembangkan sendiri jawabannya tetapi cenderung hanya memilih jawaban yang benar dan jika peserta didik tidak mengetahui jawaban yang benar, maka peserta didik akan menerka. Hal ini menimbulkan kecenderungan peserta didik tidak belajar untuk memahami pelajaran tetapi menghafalkan soal dan jawabannya. Selain itu pilihan ganda kurang mampu memberikan informasi yang cukup untuk dijadikan umpan balik guna mendiagnosis atau memodifikasi pengalaman belajar.

Tes tertulis bentuk uraian adalah alat penilaian yang menuntut peserta didik untuk mengingat, memahami, dan mengorganisasikan gagasannya atau hal-hal yang sudah dipelajari. Peserta didik mengemukakan atau mengekspresikan gagasan tersebut dalam bentuk uraian tertulis dengan menggunakan kata-katanya sendiri. Alat ini dapat menilai berbagai jenis kompetensi, misalnya mengemukakan pendapat, berpikir logis, dan menyimpulkan. Kelemahan alat ini antara lain cakupan materi yang ditanyakan terbatas.

2. Soal yang Bermutu

Bahan ujian atau soal yang bermutu dapat membantu pendidik meningkatkan pembelajaran dan memberikan informasi dengan tepat tentang peserta didik mana yang belum atau sudah mencapai kompetensi. Salah satu ciri soal yang bermutu adalah bahwa soal itu dapat membedakan setiap kemampuan peserta didik. Semakin tinggi kemampuan peserta didik dalam memahami materi pembelajaran, semakin tinggi pula peluang menjawab benar soal atau mencapai kompetensi yang ditetapkan. Makin rendah kemampuan peserta didik dalam memahami materi pembelajaran, makin kecil pula peluang menjawab benar soal untuk mengukur pencapaian kompetensi yang ditetapkan.

Syarat soal yang bermutu adalah bahwa soal harus sah (*valid*), dan handal. Sah maksudnya bahwa setiap alat ukur hanya mengukur satu dimensi/aspek saja. Mistar hanya mengukur panjang, timbangan hanya mengukur berat Handal

maksudnya bahwa setiap alat ukur harus dapat memberikan hasil pengukuran yang tepat, cermat, dan ajeg. Untuk dapat menghasilkan soal yang sah dan handal, penulis soal harus merumuskan kisi-kisi dan menulis soal berdasarkan kaidah penulisan soal yang baik (kaidah penulisan soal bentuk objektif/pilihan ganda, uraian, atau praktik).

Linn dan Gronlund menyatakan bahwa tes yang baik harus memenuhi tiga karakteristik, yaitu: validitas, reliabilitas, dan usabilitas. Validitas artinya ketepatan interpretasi hasil prosedur pengukuran, reliabilitas artinya konsistensi hasil pengukuran, dan usabilitas artinya praktis prosedurnya. Di samping itu, Djemari Mardapi juga menyatakan bahwa tes yang baik adalah tes yang valid artinya mengukur apa yang hendak diukur, validitas berhubungan dengan interpretasi atau makna dan penggunaan hasil pengukuran peserta didik. Validitas tes merupakan suatu integrasi pertimbangan evaluatif derajat keterangan empiris yang mendasarkan pemikiran teoritis yang mendukung ketepatan dan kesimpulan berdasarkan pada skor tes.

3. Validitas

Didalam buku "*Encyclopedia of Educational Evaluation*" yang ditulis oleh Scarvia B. Anderson dan kawan-kawan disebutkan : A test is valid if it measures what it purpose to measure. Maksudnya bahwa sebuah tes dikatakan valid apabila tes tersebut mengukur apa yang hendak diukur. Dalam bahasa Indonesia "valid" disebut dengan istilah "sahih".

Validitas bukanlah suatu ciri atau sifat yang mutlak dari suatu teknik evaluasi, ia merupakan suatu ciri yang relatif terhadap tujuan yang hendak dicapai oleh pembuat tes. Teknik yang sama dapat digunakan untuk beberapa tujuan yang berbeda, dan validitasnya dapat berbeda-beda dari yang tinggi kepada yang rendah, bergantung pada tujuan. Suatu tes pengerjaan berhitung misalnya, dapat mempunyai validitas yang tinggi untuk menentukan siswa-siswa dalam kecakapannya mengerjakan berhitung. Validitas itu mungkin sedang atau cukup untuk mengukur kecakapan murid-murid dalam hitung dagang (*business arithmetic*). Dan mungkin juga tes tersebut mempunyai validitas rendah dalam mengukur atau meramalkan keberhasilan dalam aspek-aspek matematis dari suatu pelajaran ilmu alam yang akan

datang. Oleh karena itu, validitas harus ditentukan dengan tujuan yang akan dicapai dengan alat evaluasi itu.

Validitas sebuah tes dibedakan menjadi dua macam yaitu validitas logis dan validitas empiris. Validitas logis sama dengan analisis kualitatif terhadap sebuah soal, yaitu untuk menentukan berfungsi tidaknya suatu soal berdasarkan kriteria yang telah ditentukan, yang dalam hal ini adalah kriteria materi, konstruksi, dan bahasa. Untuk dapat menentukan apakah tes hasil belajar sudah memiliki validitas logis atau belum, dapat dilakukan penelusuran dari dua segi, yaitu dari segi isinya (*content*) dan dari segi susunan atau konstruksinya (*construct*).

Validitas empiris adalah ketepatan mengukur yang didasarkan pada hasil analisis yang bersifat empirik, maksudnya adalah validitas yang bersumber pada atau diperoleh atas dasar pengamatan lapangan. Untuk dapat menentukan apakah tes hasil belajar sudah memiliki validitas empiris atau belum, dapat dilakukan penelusuran dari dua segi, yaitu dari segi daya ketepatan meramalnya (*predictive validity*) dan daya ketepatan bandingannya (*concurrent validity*).

Mengukur Validitas, salah satu cara untuk menentukan validitas alat ukur adalah dengan menggunakan korelasi *product moment* dengan simpangan yang dikemukakan oleh pearson seperti berikut:

$$r_{xy} = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2] \cdot [N\sum Y^2 - (\sum Y)^2]}}$$

Dengan rumus tersebut dapat dihitung validitas suatu tes dengan membandingkan atau mencari korelasi antara dua kelompok skor, kemudian dihitung berdasarkan deviasi setiap skor dari mean.

Validitas suatu tes dinyatakan dengan angka korelasi koefisien (r). Kriteria korelasi koefisien menurut Purwanto adalah sebagai berikut:

Angka Korelasi	Makna
0,00 – 0,20	Sangat rendah
0,20 – 0,40	Rendah
0,40 – 0,70	Cukup
0,70 – 0,90	Tinggi
0,90 – 1,00	Sangat tinggi

4. Tingkat Kesukaran

Tingkat kesukaran merupakan salah satu ciri dari tes yang perlu diperhatikan, karena tingkat kesukaran tes menunjukkan seberapa sukar atau mudahnya butir-butir tes atau tes secara keseluruhan yang telah diselenggarakan. Butir tes yang baik adalah butir yang memiliki tingkat kesukaran yang sedang, yaitu yang dapat dijawab dengan benar oleh sekitar 40 sampai 80 persen peserta tes. Sebab butir tes yang hanya dijawab oleh 10 persen atau bahkan 90 persen, akan sulit dibedakan manakah kelompok yang benar-benar mampu dan kelompok yang benar-benar kurang mampu dalam menjawab soal. Uji tingkat kesukaran dilakukan dengan per item.

Butir tes harus diketahui tingkat kesukarannya, karena setiap pembuat tes perlu mengetahui apakah soal itu sukar, sedang atau mudah. Tingkat kesukaran tes adalah pernyataan tentang seberapa mudah atau seberapa sukar sebuah butir tes itu bagi testee atau peserta terkait. Tingkat kesukaran itu dapat dilihat dari jawaban peserta ujian. Semakin sedikit jumlah peserta ujian yang dapat menjawab soal itu dengan benar, berarti soal itu termasuk sukar dan sebaliknya semakin banyak peserta ujian yang dapat menjawab soal itu dengan benar berarti itu mengindikasikan soal itu tidak sukar atau mudah.

Tingkat kesukaran butir tes dinyatakan dengan indeks berkisar antara 0,00 sampai dengan 1,00.

0 ————— 1

Indeks 0,00 berarti butir soal sangat sukar karena tidak seorangpun dapat menjawab dengan benar butir tes tersebut. Sebaliknya jika indeksnya 1,00 berarti butir soal tersebut sangat mudah karena semua siswa dapat menjawabnya dengan benar.

Rumus yang digunakan untuk menghitung tingkat kesukaran butir tes adalah

$$TK = \frac{B}{N \times skor.maks}$$

TK = Tingkat Kesukaran

B = Jumlah skor siswa yang menjawab dengan benar

N = Jumlah siswa

Secara lebih terperinci tentang penafsiran tingkat kesukaran dapat diperhatikan sebagai berikut :

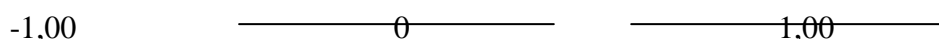
= 0,30	>	Terlalu Sukar
= 0,70 - 0,30		Sedang (baik)
< 0,70		Terlalu Mudah

Untuk sebuah butir tes yang ideal, tingkat kesukaran butir berkisar antara 0,31 hingga 0,70.

5. Daya Beda

Salah satu ciri butir yang baik adalah yang mampu membedakan antara kelompok atas (yang mampu) dan kelompok bawah(kurang mampu). Karena itu butir tes harus diketahui daya bedanya. Daya beda adalah analisis yang mengungkapkan seberapa besar suatu butir tes dapat membedakan anatara siswa kelompok tinggi dengan siswa kelompok rendah. Siswa yang termasuk kelompok tinggi adalah siswa yang mempunyai rata-rata skor paling baik. Siswa yang termasuk kelompok terendah adalah siswa yang mempunyai rata-rata skor yang rendah.

Tingkat daya pembeda butir-butir tes dinyatakan dalam skala indeks -1,00 sampai dengan 1,00.



Penjelasan :

- 1)Indeks -1,00 berarti butir tes terbalik, siswa kurang pandai dalam kelompok Lower dapat menjawab butir tes dengan sempurna, dan kelompok yang paling pandai dalam Upper tidak ada satupun yang mampu menjawab dengan benar.
- 2)Indeks 0,00 berarti butir tes tidak dapat membedakan siswa yang pandai dengan yang kurang pandai. Atau kemampuan kelompok pandai (Upper) sama dengan kemampuan kelompok kurang pandai (Lower)
- 3)Indeks 1,00 berarti butir tes secara sempurna dapat membedakan siswa berdasarkan tingkat kempuannya.

Rumus yang digunakan untuk menghitung daya pembeda butir tes adalah

$$DB = \frac{U - L}{Nup \times skor.max}$$

DB = Daya Beda
 U = Kelompok tinggi
 L = Kelompok rendah
 Nup = Jumlah siswa Upper atau Lower

Secara lebih terperinci tentang penafsiran Daya Beda butir soal dapat diperhatikan sebagai berikut :

0,70 - 1,00 Baik Sekali
 0,40 - 0,69 Baik
 0,20 - 0,39 Cukup
 0,00 - 0,19 Jelek
 -1,00 - 0,00 Jelek Sekali

G. Hasil Penelitian dan Analisis Validitas, Tingkat Kesukaran, Daya Beda Tes

1. Validitas

a. Pendidikan agama Islam

Butir	Ket	Butir	Ket	Butir	Ket	Butir	Ket
1	novalid	26	Novalid	51	valid	76	novalid
2	Novalid	27	Novalid	52	novalid	77	Novalid
3	Novalid	28	Novalid	53	Novalid	78	Novalid
4	Novalid	29	Novalid	54	Novalid	79	Novalid
5	Novalid	30	Novalid	55	Novalid	80	Novalid
6	Valid	31	valid	56	Novalid	81	Novalid
7	Novalid	32	Valid	57	Valid	82	Valid
8	Novalid	33	Valid	58	Novalid	83	Novalid
9	Valid	34	novalid	59	Valid	84	Novalid
10	Valid	35	Valid	60	Valid	85	Novalid
11	Novalid	36	Novalid	61	Novalid	86	Novalid
12	Novalid	37	Novalid	62	Valid	87	Novalid
13	Novalid	38	valid	63	Novalid	88	Novalid
14	Novalid	39	Novalid	64	Novalid	89	novalid
15	Novalid	40	Novalid	65	Novalid	90	valid
16	novalid	41	Novalid	66	Novalid	91	Novalid
17	Novalid	42	Novalid	67	Novalid	92	Novalid
18	Novalid	43	Novalid	68	Valid	93	Novalid
19	Novalid	44	novalid	69	Novalid	94	Novalid

20	Novalid	45	Novalid	70	Novalid	95	Novalid
21	Novalid	46	Novalid	71	Novalid	96	novalid
22	novalid	47	Novalid	72	Novalid	97	Novalid
23	Novalid	48	Novalid	73	Novalid	98	Novalid
24	Novalid	49	Novalid	74	Novalid	99	Novalid
25	Novalid	50	novalid	75	Novalid	100	Novalid

Untuk Validitas tes pendidikan agama Islam hanya 17% butir soal yang valid, sisanya tidak valid.

b. Bahasa Arab

Butir	ket	Butir	Ket	Butir	Ket	Butir	Ket
1	Valid	26	Novalid	51	Novalid	76	Novalid
2	Novalid	27	Novalid	52	Novalid	77	Novalid
3	Novalid	28	valid	53	Novalid	78	Novalid
4	Valid	29	valid	54	Novalid	79	Novalid
5	Valid	30	valid	55	valid	80	Novalid
6	Valid	31	novalid	56	valid	81	novalid
7	Novalid	32	Novalid	57	valid	82	Novalid
8	Valid	33	valid	58	Novalid	83	valid
9	Novalid	34	valid	59	Novalid	84	valid
10	Valid	35	Novalid	60	valid	85	valid
11	Valid	36	valid	61	Novalid	86	valid
12	Valid	37	valid	62	novalid	87	valid
13	Valid	38	valid	63	valid	88	valid
14	Valid	39	valid	64	valid	89	valid
15	Valid	40	valid	65	valid	90	valid
16	Valid	41	valid	66	valid	91	valid
17	Valid	42	Novalid	67	Novalid	92	valid
18	novalid	43	novalid	68	novalid	93	valid
19	Valid	44	Novalid	69	valid	94	Novalid
20	Novalid	45	Novalid	70	valid	95	Novalid
21	Valid	46	Novalid	71	Novalid	96	Novalid
22	Valid	47	Novalid	72	Novalid	97	Novalid
23	Valid	48	Novalid	73	Novalid	98	Novalid
24	Valid	49	novalid	74	novalid	99	novalid
25	Valid	50	Novalid	75	valid	100	valid

Untuk validitas tes bahasa arab yang valid 53% sedangkan sisanya tidak valid.

c. Bahasa Inggris

Butir	ket	Butir	Ket	Butir	Ket	Butir	Ket
1	Novalid	26	Novalid	51	Novalid	76	Novalid
2	Novalid	27	Novalid	52	Novalid	77	Novalid
3	Novalid	28	Novalid	53	Novalid	78	Novalid
4	Novalid	29	Novalid	54	Novalid	79	Novalid
5	Novalid	30	Novalid	55	Novalid	80	Novalid
6	novalid	31	novalid	56	novalid	81	novalid
7	Novalid	32	Novalid	57	Novalid	82	Novalid
8	Novalid	33	Novalid	58	Novalid	83	Novalid
9	Novalid	34	Novalid	59	Novalid	84	Novalid
10	Novalid	35	Novalid	60	Novalid	85	valid
11	Valid	36	Novalid	61	Novalid	86	Novalid
12	novalid	37	novalid	62	novalid	87	novalid
13	Novalid	38	Novalid	63	Novalid	88	Novalid
14	Novalid	39	Novalid	64	Novalid	89	valid
15	Novalid	40	Novalid	65	Novalid	90	valid
16	Novalid	41	Novalid	66	Novalid	91	Novalid
17	Novalid	42	Novalid	67	valid	92	Novalid
18	Valid	43	novalid	68	novalid	93	novalid
19	Novalid	44	Novalid	69	Novalid	94	valid
20	Novalid	45	Novalid	70	Novalid	95	valid
21	Novalid	46	Novalid	71	Novalid	96	valid
22	Novalid	47	Novalid	72	valid	97	valid
23	Novalid	48	Novalid	73	Novalid	98	Novalid
24	novalid	49	novalid	74	novalid	99	valid
25	Valid	50	Novalid	75	Novalid	100	valid

Untuk validasi tes bahasa Inggris yang valid 15% sedangkan sisanya tidak valid

d. Pengetahuan Umum

Butir	ket	Butir	Ket	Butir	Ket	Butir	Ket
1	Novalid	26	Novalid	51	Novalid	76	Novalid
2	Novalid	27	Novalid	52	Novalid	77	Novalid
3	Novalid	28	Novalid	53	Novalid	78	Novalid
4	Valid	29	Novalid	54	Novalid	79	valid
5	Novalid	30	Novalid	55	Novalid	80	valid
6	novalid	31	novalid	56	valid	81	novalid
7	Novalid	32	valid	57	Novalid	82	Novalid
8	Novalid	33	valid	58	valid	83	Novalid

9	Novalid	34	valid	59	Novalid	84	valid
10	Novalid	35	Novalid	60	Novalid	85	Novalid
11	Novalid	36	Novalid	61	Novalid	86	Novalid
12	novalid	37	valid	62	novalid	87	novalid
13	Novalid	38	Novalid	63	Novalid	88	Novalid
14	Novalid	39	Novalid	64	valid	89	Novalid
15	Valid	40	valid	65	valid	90	Novalid
16	Novalid	41	valid	66	Novalid	91	Novalid
17	Novalid	42	Novalid	67	Novalid	92	Novalid
18	novalid	43	novalid	68	novalid	93	novalid
19	Novalid	44	Novalid	69	Novalid	94	Novalid
20	Novalid	45	valid	70	Novalid	95	Novalid
21	Valid	46	valid	71	ovalid	96	Novalid
22	Novalid	47	Novalid	72	Novalid	97	Novalid
23	Valid	48	Novalid	73	Novalid	98	Novalid
24	novalid	49	novalid	74	novalid	99	novalid
25	Valid	50	valid	75	Novalid	100	Novalid

Untuk Validitas tes pengetahuan umum yang valid 25 % sedangkan sisanya tidak valid.

2. Tingkat Kesukaran Tes Masuk STAIN Ponorogo

a. Pendidikan Agama Islam

Distribusi jawaban dan hasil penelitian diatas dapat dianalisa sebagai berikut, untuk analisis ini menggunakan kriteria yang ada pada bagian teori, adapun hasil analisisnya adalah:

Hasil penelitian tingkat kesukaran (TK) pada tes tertulis pendidikan agama Islam. Jumlah soal tes 100 dengan di bedakan kriteria menjadi 3, yaitu terlalu sukar (TS), sedang (S), terlalu mudah (TM). Adapun kriteria hasil tes untuk TS adalah $=0,30$ sedangkan untuk S $=0,70-0,30$ sedangkan untuk TM adalah $0,70$.

Hasilnya;

Kriteria terlalu sukar (TS) ada pada butir soal nomor:

21, 37, 38, 39, 41, 56, 60, 62, 71, 81, 86, 88, 92, 93, 95, 96 dengan jumlah 16 butir soal.

Kriteria sedang (S) ada pada butir soal nomor:

5, 7, 13, 17, 19, 25, 27, 32, 33, 34, 35, 36, 42, 47, 48, 50, 58, 59, 61, 63, 64, 68, 69, 70, 72, 74, 75, 78, 80, 82, 85, 87, 89, 90, 94, 97, 98, 99, 100 dengan jumlah 39 butir soal.

Kriteria terlalu mudah (TM) ada pada butir soal nomor:

91, 2, 3, 4, 6, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 22, 23, 24, 28, 29, 30, 31, 40, 43, 44, 45, 46, 49, 51, 52, 53, 54, 55, 57, 65, 66, 67, 73, 76, 77, 79, 83, 84, 91 berjumlah 55 butir soal.

Hasil intepretasi dan analisisnya adalah

Untuk butir soal tes yang terlalu sukar (TS) sebaiknya di drop atau dibuang, atau diteliti ulang untuk digunakan pada tes seleksi yang ketat, item ini berjumlah 16 butir soal tes.

Untuk butir soal tes yang sedang (S) sebaiknya di catat pada kumpulan soal tes masuk pada STAIN Ponorogo, karena soal ini baik. Item ini berjumlah 39 butir tes.

Untuk butir soal yang terlalu mudah (TM) sebaiknya soal tesnya di drop atau dibuang, diteliti ulang untuk digunakan pada tes seleksi yang longgar, item ini berjumlah 55 butir soal.

b. Bahasa Arab

Distribusi jawaban dan hasil penelitian diatas dapat dianalisa sebagai berikut, untuk analisis ini menggunakan kriteria yang ada pada bagian teori, adapun hasil analisisnya adalah:

Hasil penelitian tingkat kesukaran (TK) pada tes tertulis Bahasa Arab. Jumlah soal tes 100 dengan di bedakan kriteria menjadi 3, yaitu terlalu sukar (TS), sedang (S), terlalu mudah (TM). Adapun kriteria hasil tes untuk TS adalah $=0,30$ sedangkan untuk S $=0,70-0,30$ sedangkan untuk TM adalah $0,70$.

Hasilnya;

Kriteria terlalu sukar (TS) ada pada butir soal nomor:

2, 16, 18, 19, 20, 44, 52, 53, 54, 56, 58, 59, 62, 68, 69, 71, 74, 76, 78, 79, 81, 86, 95, 96, 99 dengan jumlah 25 butir soal.

Kriteria sedang (S) ada pada butir soal nomor:

1, 3, 4, 5, 6, 7, 8, 9,10, 11, 12, 13, 14, 15, 17, 23, 24, 25, 26, 27, 28, 29,30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 55, 57, 60, 61, 63, 64, 65, 66, 67, 70, 72, 73, 75, 77, 80, 82, 83, 84, 85, 87, 88, 89, 90, 91, 92, 93, 100 dengan jumlah 69 butir soal.

Kriteria terlalu mudah (TM) ada pada butir soal nomor:

21, 22, 39, 94, 97, 98 berjumlah 6 butir soal.

Hasil intrepetasi dan analisisnya adalah

Untuk butir soal tes yang terlalu sukar (TS) sebaiknya di drop atau dibuang, atau diteliti ulang untuk digunakan pada tes seleksi yang ketat, item ini berjumlah 25 butir soal tes.

Untuk butir soal tes yang sedang (S) sebaiknya di catat pada kumpulan soal tes masuk pada STAIN Ponorogo, karena soal ini baik. Item ini berjumlah 69 butir tes.

Untuk butir soal yang terlalu mudah (TM) sebaiknya soal tesnya di drop atau dibuang, diteliti ulang untuk digunakan pada tes seleksi yang longgar, item ini berjumlah 6 butir soal.

c. Bahasa Inggris

Distribusi jawaban dan hasil penelitian diatas dapat dianalisa sebagai berikut, untuk analisis ini menggunakan kriteria yang ada pada bagian teori, adapun hasil analisisnya adalah:

Hasil penelitian tingkat kesukaran (TK) pada tes tertulis Bahasa Inggris. Jumlah soal tes 100 dengan di bedakan kriteria menjadi 3, yaitu terlalu sukar (TS), sedang (S), terlalu mudah (TM). Adapun kriteria hasil tes untuk TS adalah $=0,30$ sedangkan untuk S $=0,70-0,30$ sedangkan untuk TM adalah $0,70$.

Hasilnya;

Kriteria terlalu sukar (TS) ada pada butir soal nomor:

2, 3, 5, 8, 13, 15, 16, 17, 19, 21, 23, 26, 27, 28, 29, 30,31, 32, 33, 34, 36, 38, 40, 42, 43, 44, 45, 47, 48, 49, 53, 57, 62, 63, 64, 65, 69, 70, 73, 75, 77, 78, 82, 91, 92 dengan jumlah 45 butir soal.

Kriteria sedang (S) ada pada butir soal nomor:

1, 4, 6, 7, 9, 11, 12, 14, 18, 20, 22, 24, 35, 37, 39, 41, 46, 50, 51, 52, 55, 56, 58, 59, 60, 61, 66, 67, 68, 71, 72, 76, 79, 80, 83, 84, 85, 86, 88, 89, 90, 93, 94, 95, 96, 97, 98, 99, 100 dengan jumlah 51 butir soal.

Kriteria terlalu mudah (TM) ada pada butir soal nomor:

25, 54, 74, 81 berjumlah 4 butir soal.

Hasil intrepetasi dan analisisnya adalah

Untuk butir soal tes yang terlalu sukar (TS) sebaiknya di drop atau dibuang, atau diteliti ulang untuk digunakan pada tes seleksi yang ketat, item ini berjumlah 45 butir soal tes.

Untuk butir soal tes yang sedang (S) sebaiknya di catat pada kumpulan soal tes masuk pada STAIN Ponorogo, karena soal ini baik. Item ini berjumlah 51 butir tes.

Untuk butir soal yang terlalu mudah (TM) sebaiknya soal tesnya di drop atau dibuang, diteliti ulang untuk digunakan pada tes seleksi yang longgar, item ini berjumlah 4 butir soal.

d. Pengetahuan Umum

Distribusi jawaban dan hasil penelitian diatas dapat dianalisa sebagai berikut, untuk analisis ini menggunakan kriteria yang ada pada bagian teori, adapun hasil analisisnya adalah:

Hasil penelitian tingkat kesukaran (TK) pada tes tertulis pengetahuan Umum. Jumlah soal tes 100 dengan di bedakan kriteria menjadi 3, yaitu terlalu sukar (TS), sedang (S), terlalu mudah (TM). Adapun kriteria hasil tes untuk TS adalah =0,30 sedangkan untuk S =0,70-0,30 sedangkan untuk TM adalah 0,70.

Hasilnya;

Kriteria terlalu sukar (TS) ada pada butir soal nomor:

5, 6, 18, 20, 30, 38, 42, 51, 53, 54, 60, 61, 70, 76, 85, 87, 96, 99 dengan jumlah 18 butir soal.

Kriteria sedang (S) ada pada butir soal nomor:

1, 9, 14, 16, 19, 21, 22, 24, 25, 26, 28, 29, 31, 35, 39, 40, 43, 44, 45, 47, 50, 53, 54, 56, 57, 60, 61, 63, 64, 65, 67, 68, 70, 75, 77, 79, 80, 81, 83, 84, 86, 88, 89, 90, 94, 95, 98, 100 dengan jumlah 46 butir soal.

Kriteria terlalu mudah (TM) ada pada butir soal nomor:

2, 4, 7, 8, 10, 11, 12, 13, 15, 17, 23, 27, 32, 33, 34, 37, 41, 46, 48, 49, 52, 55, 58, 59, 62, 66, 69, 71, 72, 73, 74, 78, 91, 92, 93, 97 berjumlah 36 butir soal.

Hasil intrepetasi dan analisisnya adalah

Untuk butir soal tes yang terlalu sukar (TS) sebaiknya di drop atau dibuang, atau diteliti ulang untuk digunakan pada tes seleksi yang ketat, item ini berjumlah 18 butir soal tes.

Untuk butir soal tes yang sedang (S) sebaiknya di catat pada kumpulan soal tes masuk pada STAIN Ponorogo, karena soal ini baik. Item ini berjumlah 46 butir tes.

Untuk butir soal yang terlalu mudah (TM) sebaiknya soal tesnya di drop atau dibuang, diteliti ulang untuk digunakan pada tes seleksi yang longgar, item ini berjumlah 36 butir soal.

3. Daya Beda

a. Pendidikan Agama Islam

Hasil intrepetasi dan analisis data item dari daya beda memakai kriteria sebagai berikut:

0,70 - 1,00	Baik Sekali
0,40 - 0,69	Baik
0,20 - 0,39	Cukup
0,00 - 0,19	Jelek
-1,00 - 0,00	Jelek Sekali

Maka untuk hasil dari ujian Pendidikan Agama Islam adalah:

Untuk kriteria Baik Sekali pada butir item no. 0 berjumlah 0 butir soal.

Untuk kriteria Baik pada butir item no. 7, 17, 32, 36, 45, 50, 60, 61, 63, 69, 83 berjumlah 11 butir soal.

Untuk kriteria Cukup pada butir no. 3, 5, 8, 10, 11, 15, 19, 20, 21, 25, 26, 27, 31, 33, 34, 37, 39, 40, 42, 47, 48, 49, 52, 58, 62, 68, 70, 72, 73, 74, 78, 82, 85, 86, 89, 90, 91, 96, 97, 98, 99 berjumlah 41 butir soal

Untuk kriteria Jelek pada butir no. 1, 2, 4, 6, 9, 12, 13, 14, 16, 18, 22, 23, 24, 28, 29, 30, 35, 38, 44, 46, 51, 53, 54, 55, 56, 57, 59, 64, 65, 66, 67, 71, 75, 76, 77, 79, 80, 84, 87, 88, 92, 93, 94, 100 berjumlah 44 butir soal.

Untuk kriteria Jelek sekali pada butir no. 41, 43, 81, 95 berjumlah 4 butir soal.

Maka hasil analisisnya adalah;

Butir item soal yang mempunyai daya beda cukup, baik dan baik sekali maka perlu di simpan dalam kumpulan soal tes ujian, untuk Pendidikan agama Islam berjumlah 52 buah soal.

Butir item soal yang mempunyai daya beda Jelek maka perlu ditelusuri kemudian diperbaiki atau di drop. Untuk pendidikan agama Islam berjumlah 44 butir soal

Butir item soal yang mempunyai daya beda negatif atau jelek sekali maka untuk ke depannya soal ini tidak perlu di pakai lagi. Untuk pendidikan agama Islam berjumlah 4 butir soal.

b. Bahasa Arab

Hasil intrepetasi dan analisis data item dari daya beda memakai kriteria sebagai berikut:

0,70 - 1,00	Baik Sekali
0,40 - 0,69	Baik
0,20 - 0,39	Cukup
0,00 - 0,19	Jelek
-1,00 - 0,00	Jelek Sekali

Maka untuk hasil dari ujian Bahasa Arab adalah:

Untuk kriteria Baik Sekali pada butir item no. 5, 8, 10, 25, 65, 90 berjumlah 6 butir soal.

Untuk kriteria Baik pada butir item no. 4, 6, 9, 11, 12, 13, 15, 17, 19, 21, 23, 24, 28, 29, 30, 33, 34, 36, 37, 38, 40, 41, 50, 55, 56, 57, 63, 64, 66, 67, 70, 83, 85, 87, 88, 89, 91, 92, 93, 100 berjumlah 40 butir soal.

Untuk kriteria Cukup pada butir no. 1, 3, 7, 14, 16, 22, 27, 31, 32, 39, 43, 44, 45, 46, 48, 49, 60, 69, 73, 75, 80, 84, 86, 97 berjumlah 24 butir soal

Untuk kriteria Jelek pada butir no.2, 18,20, 26, 35, 42, 47, 51, 52, 54, 58, 59, 62, 68, 72, 78, 79, 82, 94, 95, 96, 98, 99 berjumlah 23 butir soal.

Untuk kriteria Jelek sekali pada butir no. 53, 61, 71, 74, 76, 77, 81 berjumlah 7 butir soal.

Maka hasil analisisnya adalah;

Butir item soal yang mempunyai daya beda cukup, baik dan baik sekali maka perlu di simpan dalam kumpulan soal tes ujian, untuk Bahasa Arab berjumlah 70 buah soal.

Butir item soal yang mempunyai daya beda Jelek maka perlu ditelusuri kemudian diperbaiki atau di drop. Untuk Bahasa Arab berjumlah 23 butir soal

Butir item soal yang mempunyai daya beda negatif atau jelek sekali maka untuk ke depannya soal ini tidak perlu di pakai lagi. Untuk Bahasa Arab berjumlah 7 butir soal.

c. Bahasa Inggris

Hasil intrepetasi dan analisis data item dari daya beda memakai kriteria sebagai berikut:

0,70 - 1,00	Baik Sekali
0,40 - 0,69	Baik
0,20 - 0,39	Cukup
0,00 - 0,19	Jelek
-1,00 - 0,00	Jelek Sekali

Maka untuk hasil dari ujian Bahasa Inggris adalah:

Untuk kriteria Baik Sekali pada butir item no. 85 berjumlah 1 butir soal.

Untuk kriteria Baik pada butir item no. 1, 4, 11, 67, 68, 72, 80 berjumlah 7 butir soal.

Untuk kriteria Cukup pada butir no. 9, 12, 14, 15, 17, 18, 20, 22, 24, 25, 33, 35, 39, 42, 46, 50, 52, 55, 59, 66, 71, 76, 84, 89, 90, 94, 95, 97, 98 berjumlah 29 butir soal

Untuk kriteria Jelek pada butir no. 5, 6, 7, 10, 13, 16, 21, 26, 27, 28, 29, 30, 31, 32, 34, 37, 40, 41, 45, 47, 48, 49, 53, 54, 57, 58, 60, 61, 62, 63, 65, 69, 73, 74, 79, 82, 83, 86, 87, 88, 91, 92, 93 berjumlah 43 butir soal.

Untuk kriteria Jelek sekali pada butir no.2, 3, 8, 19, 23, 36, 38, 43, 44, 51, 56, 64, 70, 75, 77, 78, 81, 96, 99, 100 berjumlah 20 butir soal.

Maka hasil analisisnya adalah;

Butir item soal yang mempunyai daya beda cukup, baik dan baik sekali maka perlu di simpan dalam kumpulan soal tes ujian, untuk bahasa inggris berjumlah 37 buah soal.

Butir item soal yang mempunyai daya beda Jelek maka perlu ditelusuri kemudian diperbaiki atau di drop. Untuk bahasa inggris berjumlah 43 butir soal

Butir item soal yang mempunyai daya beda negatif atau jelek sekali maka untuk ke depannya soal ini tidak perlu di pakai lagi. Untuk bahasa inggris berjumlah 20 butir soal.

d. Pengetahuan Umum

Hasil intrepetasi dan analisis data item dari daya beda memakai kriteria sebagai berikut:

0,70 - 1,00	Baik Sekali
0,40 - 0,69	Baik
0,20 - 0,39	Cukup
0,00 - 0,19	Jelek
-1,00 - 0,00	Jelek Sekali

Maka untuk hasil dari ujian Pengetahuan Umum adalah:

Untuk kriteria Baik Sekali pada butir item no. 0 berjumlah 0 butir soal.

Untuk kriteria Baik pada butir item no. 21, 25, 26, 32, 33, 45, 50, 52, 58, 64, 65, 71, 80 berjumlah 13 butir soal.

Untuk kriteria Cukup pada butir no. 1, 2, 4, 5, 15, 23, 27, 31, 36, 37, 40, 41, 44, 47, 49, 56, 57, 61, 62, 67, 68, 72, 73, 75, 79, 81, 84, 85, 88, 89, 97, 98, 99 berjumlah 33 butir soal

Untuk kriteria Jelek pada butir no. 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 22, 24, 28, 34, 35, 38, 39, 42, 43, 48, 53, 54, 55, 59, 60, 66, 69, 70, 74, 76, 78, 82, 83, 86, 90, 91, 92, 93, 94, 95, 100 berjumlah 46 butir soal.

Untuk kriteria Jelek sekali pada butir no. 3, 29, 30, 51, 63, 77, 87, 96 berjumlah 8 butir soal.

Maka hasil analisisnya adalah;

Butir item soal yang mempunyai daya beda cukup, baik dan baik sekali maka perlu di simpan dalam kumpulan soal tes ujian, untuk pengetahuan umum berjumlah 46 buah soal.

Butir item soal yang mempunyai daya beda Jelek maka perlu ditelusuri kemudian diperbaiki atau di drop. Untuk pengetahuan umum berjumlah 46 butir soal

Butir item soal yang mempunyai daya beda negatif maka untuk ke depannya soal ini tidak perlu di pakai lagi. Untuk pengetahuan umum berjumlah 8 butir soal..

H. Kesimpulan

1. Bahwa validitas dari tes tertulis ujian masuk STAIN Ponorogo tahun akademik 2010/2011 pada pendidikan agama Islam yang valid 17 butir tes, untuk bahasa Arab yang valid 53 butir tes, untuk bahasa Inggris yang valid 15 butir tes, sedangkan pengetahuan umum yang valid 25 butir tes, sehingga validitas tesnya sangat rendah..
2. Sedangkan pada tingkat kesukaran pada instrumen tes tertulis masuk STAIN ponorogo pada pendidikan agama Islam, terlalu sukar (TS) 16 butir, baik (S) 39 butir, dan terlalu mudah (TM) 55 butir, maka soal yang patut di pertahankan hanya berjumlah 39 butir sisanya didrop atau diganti. Sedangkan tingkat kesukaran pada bahasa arab, terlalu sukar (TS) 25 butir, baik (S) 69 butir, telalu mudah (TM) 6 butir, maka soal yang patut di pertahankan berjumlah 69 butir dan sisanya di ganti. Untuk tingkat kesukaran bahasa Inggris, terlalu sukar (TS) 45 butir, baik (S) 51 butir, terlalu mudah (TM) 4 butir, maka soal yang ppatut dipertahankan ada 51 butir, sisanya di ganti, demikian juga dengan tingkat kesukaran tes pada pengetahuan umum, terlalu sukar (TS) 18 butir, baik (S) 46 butir, terlalu mudah (TM) 36 butir, maka soal yang patut dipertahankan berjumlah 46 butir, sisanya diganti atau didrop.
3. Daya beda untuk pendidikan agama Islam baik sekali 0 butir, baik 11 butir, cukup 41 butir, jelek 44 butir, jelek sekali 4 butir, maka butir soal yang patut di pertahankan ada di kriteria baik sekali, baik dan cukup 52 butir, untuk 44 jelek diganti, dan 4 butir jelek sekali kedepannya soal ini tidak boleh di keluarkan lagi. Daya beda bahasa arab, baik sekali 6 butir, baik 40 butir, cukup 24 butir,

jelek 23 butir dan jelek sekali 7 butir, maka soal yang patut di pertahankan 70 butir, sedangkan di ganti 23 butir, yang 7 butir kedepannya tidak boleh dikeluarkan lagi. Daya beda bahasa Inggris baik sekalo 0 butir, baik 7 butir, cukup 29 butir, jelek 43 butir, 20 butir jelek sekali, maka butir soal yang patut di pertahankan ada 36 butir, 43 butir di ganti, sedangkan 20 butir kedepannya tidak boleh dimunculkan lagi. Daya beda pengetahuan umum, baik sekali 0 butir, baik 13 butir, cukup 36 butir, jelek 46 butir, jelek sekali 8 butir, maka soal yang patut dipertahankan 49 butir, sadangkan yang diganti 46 butir sedangkan 8 butir kedepannya tidak boleh dimunculkan lagi.

4. Analisis dari validitas tes tertulis pada pendidikan agama Islam 17%, bahasa arab 53%, bahasa Inggris 15%, sedangkan pengetahuan umum 25%, hal ini menunjukkan validitas tes sangat rendah untuk tes masuk STAIN Ponorogo tahun akademik 2010/2011. Sedangkan Tingkat kesukaran PAI rata-rata 39% yang baik, bahasa arab 69%, bahasa inggris 51%, pengetahuan umum 46%, sedangkan daya beda yang baik untuk PAI 52%, bahasa arab 70%, bahasa inggris 37% , pengetahuan umum 49% .

DAFTAR PUSTAKA

- Ali Ridho dan Saifuddin Azwar. 2005. Makalah Seminar Nasional
- Arikunto, Suharsimi. 2000. *Dasar-dasar Evaluasi*. Bumi Aksara: Jakarta
- Azwar, Saepuddin. 2000. *Tes Prestasi*. Pustaka Pelajar. Yogyakarta
- Badrun K.2003. *Penilaian Berbasis Kompetensi*. Jakarta: Depdiknas.
- Depdiknas.1999.*Pengelolaan Pengujian*. Jakarta:Dirjend Dikdasmen
- Herminarto.2001.*Pengembangan Instrumen Penelitian*.Jakarta; Delima Press.
- Linn, Robert L. and Gronlund, Norman E. 1995. *Measurement and Assessment in Teaching*. (Seventh Edition). Ohio: Prentice-Hall, Inc.
- Mardapi, Djemari, 2008. *Teknik Peyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekia
- Pusat Pengujian Depdiknas.2001.*Panduan Pembuatan Butir Soal*. Jakarta: Depdiknas
- Sujana.2007.*Pengantar Evaluasi Pendidikan*.Jakarta: Grafindo
- Slameto.2001. *Evaluasi Pendidikan*. Bumi Aksara: Jakarta
- Sumadi Suryabrata.2005. *Rekayasa Sistem Penilaian Dalam Rangka Meningkatkan Kualitas Pendidikan*. Yogyakarta: HEPI
- Tim Lapis.*Evaluasi Pembelajaran*.(Jakarta: LAPIS-DEPAG RI, 2008), hal 56.
- UUSPN tahun 2003