

Analisis Kualitas Naskah Ujian Akhir Semester Mata Kuliah Bahasa Inggris II Politeknik Negeri Lampung Tahun 2006

The Analysis of Quality of Semester II English Test for Students of the State Polytechnics of Lampung in the Year 2006

Refdi Akmal¹⁾

1) Dosen Bahasa Inggris Politeknik Negeri Lampung
Jl. Soekarno-Hatta Rajabasa Bandar Lampung

Abstract

It is believed that without the evaluation it would be almost impossible to teach. As students need to know their abilities, teachers use some kind of evaluations. Evaluation of teaching learning process can be administrated in the form of tests. In other words teaching must be followed by testing. The information gained can be useful for both students and teacher. Hence there are two points interrelated. Firstly, the test is concerned with the teaching that has taken place. It is that the test administered should give the students feeling that the teacher's evaluation matches what they have taught and sense of accomplishment. Secondly, the teaching is concerned with the test. The teacher can diagnose his effort in his teaching whether or not his teaching has been effective. Madsen (1987) pointed out that the criteria of a good test comprise: (1) validity, (2) reliability, (3) difficulty level, and (4) discrimination power. This research was to know and describe quality of semester II English test for the students of The State Polytechnic of Lampung (Polinela) in the year 2006. The study showed that the Semester II English Test administered for the students did not meet the expected level of validity, reliability, level of difficulty nor discrimination power of a good test.

Key words: evaluation, validity, reliability, difficulty level, discrimination power.

Pendahuluan

Evaluasi merupakan satu istilah yang difahami dengan banyak makna. Secara umum evaluasi difahami dalam kerangka fikir ujian dan penilaian. Ten Brink (1974), menyatakan bahwa evaluasi tidak hanya berkenaan dengan pengujian dan pengambilan nilai, melainkan lebih dari itu, evaluasi merupakan suatu proses penjaringan informasi yang dibutuhkan untuk digunakan sebagai bahan pertimbangan pengambilan keputusan dalam hal proses belajar mengajar secara umum. Hal serupa juga ditegaskan oleh Danfey dan Wakistron (1976) yang menyatakan bahwa evaluasi merupakan suatu proses untuk mendapatkan informasi tertentu yang akan digunakan dalam mengambil keputusan, seperti mengenai pelaksanaan dan perkembangan suatu kegiatan, efektivitas materi pengajaran, perkembangan anak, dan lain-lain.

Dengan demikian, proses melaksanakan evaluasi pembelajaran tidak hanya sebatas mendistribusikan soal dan mendapatkan nilai. Lebih jauh Ten Brink menyebutkan bahwa sedikitnya ada tiga langkah yang harus dilakukan dalam melaksanakan evaluasi ini. Pertama, kesiapan untuk melaksanakan suatu evaluasi. Hal ini secara sederhana berarti mempertimbangkan apakah suatu proses pembelajaran telah perlu dievaluasi. Kedua, pengumpulan informasi. Hal ini dilakukan dengan pelaksanaan tes dan penilaian. Faktor terpenting dalam hal ini adalah penyediaan alat evaluasi yang memiliki standar yang baik agar benar-benar dapat menjangkau informasi yang akurat. Selanjutnya adalah penggunaan informasi yang didapat dalam pengambilan keputusan mengenai hal-hal yang berkaitan dengan proses belajar.

Evaluasi belajar merupakan salah satu bagian yang sangat penting dalam proses belajar mengajar, baik bagi pengajar maupun bagi peserta belajar. Pengajar yang baik akan senantiasa mencari cara terbaik untuk dapat menyajikan materi dengan cara yang efektif bagi siswa-siswanya. Guna mewujudkan hal tersebut, maka seorang guru akan sangat membutuhkan informasi mengenai para siswanya terutama mengenai kemampuan, pencapaian, ketertarikan, kelebihan dan kekurangan serta permasalahan dari para siswa tersebut.

Guna mendapatkan data yang akurat mengenai hal tersebut maka seorang pengajar akan membutuhkan suatu proses evaluasi. Di sisi lain, evaluasi juga akan memberikan informasi bagi para peserta belajar untuk mengetahui perkembangan pencapaian proses belajar yang dilakukan, sehingga mereka dapat mengambil keputusan dalam hal kegiatan belajar selanjutnya.

Evaluasi belajar biasanya dilaksanakan dalam bentuk ujian (tes). Ujian atau tes dapat didefinisikan sebagai suatu cara atau prosedur yang digunakan untuk mengukur sampel dari suatu hasil dari proses belajar. Shohamy (1985) menyatakan bahwa tes hanyalah sampel atas pengetahuan tertentu dan bukan merupakan keseluruhan pengetahuan atau kemampuan dari apa yang telah diajarkan oleh guru atau dipelajari oleh siswa. Oleh karena itu, sampel yang diperoleh

harus sebaik mungkin mewakili bahwa apa yang telah diujikan mencakup apa yang telah diajarkan. Heaton dalam Rahman (1999) menyatakan bahwa pengajaran dan evaluasi (tes) merupakan hal yang sangat terkait, sehingga sangat sulit untuk dapat melaksanakan dengan baik salah satu dari kedua hal tersebut dan mengabaikan yang lain. Keterkaitan tersebut dapat dijelaskan dalam dua bentuk. Pertama, evaluasi berkaitan dengan pengukuran hasil belajar pada saat sebuah tes dilaksanakan untuk menguji proses pembelajaran yang telah dilakukan. Tes harus dapat mengukur pencapaian siswa sesuai dengan materi yang telah disajikan. Kedua, evaluasi berkaitan dengan proses pengajaran. Dalam hal ini sebuah evaluasi akan menjadi tolak ukur bagi seorang guru guna mengetahui efektivitas perlakuan dalam pengajaran yang telah dilakukannya.

Lebih lanjut, Hopkins dalam Rahman (1999) menyatakan bahwa suatu tes juga dapat dijadikan bahan untuk menentukan pencapaian prestasi siswa yang mengikuti tes. Hal ini dilakukan dengan menganalisis keberhasilan siswa dalam menyelesaikan ujian yang diberikan. Bertitik tolak pada hal-hal tersebut di atas, jelas bahwa suatu tes tidak hanya menyediakan informasi tentang nilai siswa, namun lebih jauh adalah mengenai pengembangan pendidikan dan pengajaran itu sendiri, yang meliputi guru, siswa, sumber-sumber belajar, metode belajar hingga teknik mengajar seorang guru. Dengan demikian sudah selayaknya suatu alat uji dipastikan memenuhi kriteria tes yang baik.

Sebagai sumber informasi penting dalam pengukuran keberhasilan belajar, sebuah tes tentunya harus disajikan dalam bentuk yang baik dan benar. Bila tes yang disajikan tidak benar, baik bentuk dan substansi isinya maka informasi yang dihasilkan tentu akan meragukan. Suatu tes diharapkan memenuhi beberapa kriteria, agar dapat menyajikan informasi yang akurat. Madsen (1987) menyatakan bahwa tes yang baik memiliki tingkat (1) validitas (*validity*), (2) reliabilitas (*reliability*), (3) tingkat kesulitan (*level of difficulty*), dan (4) daya beda (*discrimination power*) yang baik.

Penelitian ini mendeskripsikan informasi mengenai validitas, reliabilitas, tingkat kesulitan dan daya beda dari soal-soal dalam naskah ujian mata kuliah Bahasa Inggris semester II di Politeknik Negeri Lampung tahun 2006. Soal-soal ujian yang menjadi objek penelitian ini adalah soal-soal ujian yang disusun tim dosen Bahasa Inggris Politeknik Negeri Lampung untuk kebutuhan pelaksanaan ujian akhir semester. Untuk dapat mengetahui dengan baik kualitas soal ujian tersebut, maka dalam penelitian ini dilakukan analisis tes dan analisis soal-per-soal. Dengan demikian, penelitian ini diharapkan dapat menjelaskan dengan baik tingkat validitas, reliabilitas, tingkat kesulitan dan daya beda alat evaluasi tersebut.

Metode Penelitian

Penelitian ini dilakukan di Kampus Politeknik Negeri Lampung (Polinela) Kota Bandar Lampung. Penentuan lokasi dilakukan secara sengaja (*purposive*) dengan pertimbangan bahwa naskah soal yang menjadi objek penelitian adalah naskah soal yang disusun oleh tim dosen Bahasa Inggris pada kampus ini. Penelitian ini berlangsung selama empat bulan dan dilaksanakan mulai bulan Desember 2006 sampai dengan Maret 2007. Bahan penelitian yang diperlukan dalam penelitian ini adalah naskah ujian dari Ujian Akhir Semester (UAS) mahasiswa semester II di Politeknik Negeri Lampung (Polinela), lembar jawaban mahasiswa yang dihasilkan dari pelaksanaan Ujian Akhir Semester (UAS) mahasiswa, silabus pengajaran mata kuliah Bahasa Inggris mahasiswa semester II di Politeknik Negeri Lampung (Polinela), buku teks mata kuliah Bahasa Inggris semester II di Politeknik Negeri Lampung (Polinela).

Data yang dibutuhkan dalam penelitian ini didapat dengan cara pendokumentasian dan pengambilan sampel dari data yang dibutuhkan. Data tersebut meliputi naskah soal ujian semester, silabus pelajaran dan buku ajar mata kuliah Bahasa Inggris semester II di lingkungan Polinela dan hasil ujian semester mahasiswa Polinela yang mengikuti ujian tersebut yang diambil dari 200 lembar jawaban mahasiswa dari 10 program studi pada lima jurusan di Polinela. Data yang berupa lembar jawaban mahasiswa ini kemudian dianalisis dan

dikelompokkan menjadi dua kelompok yaitu kelompok kelas atas (*upper group*) dan kelas bawah (*lower group*). Masing-masing kelompok diwakili dengan 28 lembar jawaban, sehingga total sampel yang didapat adalah 56 lembar jawaban atau 27,5%.

Data yang diperoleh dianalisis melalui analisis naskah dan analisis butir soal. Analisis naskah dilakukan untuk mendapatkan informasi mengenai validitas (validitas isi dan validitas soal) dan reliabilitas soal ujian semester tersebut. Analisis butir per butir soal dilakukan untuk mendapatkan data mengenai tingkat kesulitan dan daya beda dari setiap soal yang telah diujikan tersebut.

Validitas isi dianalisis dengan cara membandingkan butir soal secara keseluruhan dengan silabus pelajaran yang menggambarkan tujuan pengajaran yang diharapkan. Hasilnya kemudian dijelaskan dengan tabel penghitungan validitas sebagai berikut:

Tabel 1. Rancangan analisis validitas

| No | Tujuan (1) | Topik/sub topik (2) | Metode (3) | Media (4) | Waktu/jam (5) | Butir soal (6) | jumlah (7) | Persentase (4) |
|-------|---------------|---------------------------|---------------|--------------|------------------|----------------------|---------------|-------------------|
| | | | | | | | | |
| Total | | | | | | | | |

Validitas soal akan dilakukan dengan menggunakan Korelasi *Product Moment* dengan rumus sbb:

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

Selanjutnya, reliabilitas soal akan dikalkulasi dengan rumusan KR 21. Hal ini dilakukan dengan cara menyusun data dalam tabel penghitungan sebagai berikut:

Tabel 2. Rancangan analisis reliabilitas

| Variabel | R hitung x | R table x - \bar{x} | Hasil (x - \bar{x}) ² |
|----------|--------------------|-------------------------------|--|
| | | | |
| N = | $\Sigma x = \dots$ | $\Sigma(x - \bar{x}) = \dots$ | $\Sigma(x - \bar{x})^2 = \dots$ |

Standar deviasi ditentukan dengan menggunakan perhitungan sebagai berikut:

$$S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}}$$

Selanjutnya reliabilitas ini dikalkulasi dengan rumusan KR 21 sebagai berikut:

$$r(K21) = \frac{K}{K-1} \left\{ 1 - \frac{x(x - \bar{x})}{KS^2} \right\}$$

Guna mendapatkan informasi mengenai dua poin ini dilakukan langkah-langkah sebagai berikut:

1. Menyusun hasil jawaban mahasiswa dari nilai tertinggi hingga terendah.
2. Mengidentifikasi kelompok atas (*Upper group* = U) dan kelompok bawah (*Lower Group* = L) secara terpisah dengan memilih 27,5% dari masing masing jumlah kelompok atas dan bawah.
3. Penghitungan daya beda (*Discrimination Power* = DP) dilakukan dengan rumus sebagai berikut

$$DP = \frac{U - L}{1/2 T}$$

Hasil perhitungan tersebut kemudian diinterpretasikan dengan kriteria sebagai berikut:

- DP = 0.00--0.20 = buruk
- DP = 0.21--0.40 = memuaskan
- DP = 0.41--0.70 = baik
- DP = 0.71--1.00 = sangat baik

4. Penghitungan tingkat kesulitan dilakukan dengan rumusan sebagai berikut:

$$LD = \frac{U - L}{T}$$

Kalkulasinya kemudian diinterpretasikan dengan kriteria:

$$LD = 0.00 - 0.14 = \text{sulit}$$

$$LD = 0.15 - 0.85 = \text{rata-rata}$$

$$LD = 0.86 - 1.00 = \text{mudah}$$

Hasil dari analisis tingkat kesulitan dan daya beda soal tersebut, kemudian ditampilkan dalam bentuk tabel penghitungan sebagai berikut:

Tabel 3. Rancangan analisis tingkat kesulitan dan daya beda soal

| Soal | KA | KB | A - B | A + B | Daya Beda | | Tingkat Kesulitan | | |
|------|----|----|-------|-------|-----------|--------------|-------------------|--------------|--|
| | | | | | Hasil | Interpretasi | Hasil | Interpretasi | |
| | | | | | | | | | |

Keterangan:

KA = Kelompok Atas

KB = Kelompok Bawah

A = Nilai Kelompok Atas

B = Nilai Kelompok Bawah

Hasil dan Pembahasan

Validitas sampling/logis

Analisis data mengenai validitas logis menggambarkan ketidakseimbangan sebaran jumlah soal yang digunakan sebagai naskah ujian akhir (UAS) semester II mata kuliah Bahasa Inggris tahun 2006 di Polinela. Gambaran ini menunjukkan jumlah pokok bahasan dan jumlah soal yang dapat merepresentasikan topik/sub topik yang dibahas ternyata tidak terwakili dengan baik.

Dengan pokok bahasan seperti yang dibebankan dalam rencana pengajaran sebanyak delapan (8) unit yang masing masing dibahas dalam rata-rata waktu pembahasan yang relatif sama, maka asumsi ideal sebaran soal per pokok bahasan adalah 10 hingga 14 persen atau lima (5) hingga tujuh (7) butir soal untuk setiap pokok bahasan. Namun tabel analisis validitas logis isi menunjukkan bahwa jumlah soal yang mewakili pokok bahasan pada unit satu adalah 14

butir soal atau mencapai angka 28% dari total keseluruhan jumlah soal. Sebaliknya, soal-soal yang mewakili pokok bahasan pada unit 4 hanya berjumlah 3 soal atau 6% dari total jumlah soal, dan pokok bahasan pada unit 5 hanya terwakili oleh 2 butir soal atau 4% saja. Namun demikian, pokok-pokok bahasan pada unit-unit yang lain sudah cukup terwakili dengan sebaran jumlah soal yang cukup memadai dalam naskah ujian tersebut.

Validitas soal

Berdasarkan analisis yang dilakukan di atas dapat disimpulkan bahwa sebagian besar soal ujian yang dipakai dalam ujian akhir semester II mata kuliah Bahasa Inggris tidak memenuhi standar validitas yang baik. Ada fenomena tersendiri yang nampak dalam tabel di atas yaitu bahwa sejumlah soal cukup valid di kelas atas namun tidak valid di kelas bawah. Contohnya pada soal-soal nomor 7, 8, 10, 28, dan 35 valid di kelompok atas tetapi tidak valid di kelompok bawah. Sebaliknya soal-soal nomor 23, 27, 31, 41, 42, 46, 47, 48, 49 valid untuk dipakai di kelompok bawah namun tidak valid bagi kelompok atas.

Secara sederhana dapat difahami bahwa hanya 12 % dari soal-soal ujian yang diujikan memenuhi standar validitas untuk dapat diterapkan di kelompok bawah, dan hanya 20% yang valid bagi kelompok atas. Fenomena inkonsistensi ini menggambarkan bahwa soal-soal ujian yang telah dipakai dapat dikategorikan secara umum tidak valid.

Reliabilitas Tes

Hasil penghitungan reliabilitas soal-soal dalam naskah ujian yang telah digunakan dalam Ujian Akhir Semester II tersebut menunjukkan bahwa pada kelompok atas (*Upper Group*) tingkat reliabilitasnya adalah -0,063, sedangkan pada kelompok bawah (*Lower Group*) hasil penghitungan menunjukkan angka -0,057. Angka-angka tersebut mengindikasikan bahwa soal-soal ujian yang dipergunakan tidak memenuhi kriteria reliabilitas soal yang baik yaitu 0,1848.

Berdasarkan hasil perhitungan di atas dapat disimpulkan bahwa soal ujian semester II Bahasa Inggris yang telah diterapkan tidak reliabel.

Tingkat Kesulitan dan Daya Beda

Data yang diproses untuk mengetahui tingkat kesulitan dan daya beda soal, menghasilkan informasi bahwa soal-soal yang dipergunakan sebagai alat uji tersebut memiliki daya beda yang buruk. Hal ini berarti bahwa tes tersebut tidak dapat membedakan dengan baik antara kelompok siswa yang berkemampuan baik dan kelompok siswa yang tidak berkemampuan baik.

Penghitungan tersebut juga menggambarkan bahwa keseluruhan soal ujian tersebut tergolong sulit bagi seluruh siswa. Soal yang baik selayaknya tidak terlalu sukar dan tidak terlalu mudah bagi seluruh kelompok siswa.

Tabel 1. Hasil Analisis Validitas Isi

| No | Tujuan (1) | Topik/sub topik (2) | Metode (3) | Media (4) | Waktu/jam (5) | Butir soal (6) | Jumlah (7) | Persentase (4) |
|-------|---|--|-------------------------------|----------------|---------------|---|------------|----------------|
| 1 | Mampu memperkenalkan diri dengan menggunakan ungkapan yang benar | Greeting and Introduction | Latihan tertulis Role Play | BPP Lab Set | 2 X 170 menit | 1, 2, 3, 4, 16, 17, 18, 19, 20, 41, 42, 43, 44, 45 | 14 | 28% |
| 2 | Mampu menggunakan ungkapan untuk menunjuk arah ke suatu tempat | Giving Direction Modal Auxiliary (can, may, will) | Latihan tertulis Role Play | BPP Lab Set | 2 X 170 menit | 38, 39, 46, 47, 48, 49, 50 | 7 | 14% |
| 3 | Mampu menceritakan dan menuliskan pengalaman masa lalu | Past Tense Perfect Tense | Latihan tertulis Role Play | BPP Lab Set | 2 X 170 menit | 9, 21, 22, 23, 25, 26 | 6 | 12% |
| 4 | Mampu berbicara menggunakan ungkapan meminta pertolongan, menolak, menyetujui | Request and Imperatives | Latihan tertulis Role Play | BPP Lab Set | 2 X 170 menit | 13, 14, 28, | 3 | 6% |
| 5 | Mampu menjelaskan ciri-ciri fisik dan sifat seseorang | Describing Someone | Latihan tertulis Role Play | BPP Lab Set | 2 X 170 menit | 5, 6, 7, 24, 29, 40 | 6 | 12% |
| 6 | Mampu menggunakan kata sifat dalam Bahasa Inggris dengan benar | Adjectives | Latihan tertulis Role Play | BPP Lab Set | 2 X 170 menit | 15, 35 | 2 | 4% |
| 7 | Mampu menggunakan ungkapan untuk memberi saran dengan benar | Giving Suggestion Modal Auxiliaries (should, must, have to) | Role Play Latihan tertulis | BPP Lab Set | 2 X 170 menit | 27, 30, 31, 32, 33, 34 | 6 | 12% |
| 8 | Mampu mengungkapkan rencana kegiatan yang akan dilakukan | Future Tense | Role Play Latihan tertulis | BPP Lab Set | 2 X 170 menit | 8, 10, 11, 12, 36, 37 | 6 | 12% |
| Total | | | | | | | 50 | 100 % |

Tabel 2. Analisis Validitas Konstrak

Kelompok Bawah (Lower Group)

| Variabel | Butir dikoreksi-korelasi Total (R –hitung) | R tabel | Hasil |
|-----------------|---|----------------|--------------|
| item 1 | 0.009 | 0.1848 | Tidak valid |
| item 2 | -0.145 | 0.1848 | Tidak valid |
| item 3 | -0.033 | 0.1848 | Tidak valid |
| item 4 | -0.134 | 0.1848 | Tidak valid |
| item 5 | -0.200 | 0.1848 | Tidak valid |
| item 6 | -0.253 | 0.1848 | Tidak valid |
| item 7 | -0.270 | 0.1848 | Tidak valid |
| item 8 | -0.123 | 0.1848 | Tidak valid |
| item 9 | -0.140 | 0.1848 | Tidak valid |
| item 10 | -0.259 | 0.1848 | Tidak valid |
| item 11 | -0.287 | 0.1848 | Tidak valid |
| item 12 | 0.046 | 0.1848 | Tidak valid |
| item 13 | 0.024 | 0.1848 | Tidak valid |
| item 14 | 0.068 | 0.1848 | Tidak valid |
| item 15 | -0.297 | 0.1848 | Tidak valid |
| item 16 | 0.059 | 0.1848 | Tidak valid |
| item 17 | 0.005 | 0.1848 | Tidak valid |
| item 18 | -0.033 | 0.1848 | Tidak valid |
| item 19 | -0.183 | 0.1848 | Tidak valid |
| item 20 | -0.198 | 0.1848 | Tidak valid |
| item 21 | -0.067 | 0.1848 | Tidak valid |
| item 22 | -0.081 | 0.1848 | Tidak valid |
| item 23 | 0.207 | 0.1848 | Valid |
| item 24 | 0.111 | 0.1848 | Tidak valid |
| item 25 | 0.027 | 0.1848 | Tidak valid |
| item 26 | 0.029 | 0.1848 | Tidak valid |
| item 27 | 0.280 | 0.1848 | Valid |
| item 28 | 0.000 | 0.1848 | Tidak valid |
| item 29 | -0.024 | 0.1848 | Tidak valid |
| item 30 | -0.002 | 0.1848 | Tidak valid |
| item 31 | 0.250 | 0.1848 | Valid |
| item 32 | 0.236 | 0.1848 | Valid |
| item 33 | -0.031 | 0.1848 | Tidak valid |
| item 34 | 0.175 | 0.1848 | Tidak valid |
| item 35 | 0.083 | 0.1848 | Tidak valid |
| item 36 | -0.140 | 0.1848 | Tidak valid |
| item 37 | -0.002 | 0.1848 | Tidak valid |
| item 38 | -0.059 | 0.1848 | Tidak valid |
| item 39 | 0.057 | 0.1848 | Tidak valid |
| item 40 | -0.066 | 0.1848 | Tidak valid |
| item 41 | 0.383 | 0.1848 | Valid |
| item 42 | 0.229 | 0.1848 | Valid |
| item 43 | -0.059 | 0.1848 | Tidak valid |
| item 44 | -0.002 | 0.1848 | Tidak valid |
| item 45 | -0.164 | 0.1848 | Tidak valid |
| item 46 | 0.187 | 0.1848 | Valid |
| item 47 | 0.226 | 0.1848 | Valid |
| item 48 | 0.378 | 0.1848 | Valid |
| item 49 | 0.364 | 0.1848 | Valid |
| item 50 | 0.330 | 0.1848 | Valid |

Kelompok Atas

| Variabel | Butir dikoreksi-korelasi Total (R –hitung) | R tabel | Hasil |
|----------------|---|---------------|--------------|
| Item 1 | -0.192 | 0.1848 | Tidak valid |
| Item 2 | 0.000 | 0.1848 | Tidak valid |
| Item 3 | -0.084 | 0.1848 | Tidak valid |
| Item 4 | 0.095 | 0.1848 | Tidak valid |
| Item 5 | -0.356 | 0.1848 | Tidak valid |
| Item 6 | 0.163 | 0.1848 | Tidak valid |
| Item 7 | 0.193 | 0.1848 | Valid |
| Item 8 | 0.317 | 0.1848 | Valid |
| Item 9 | 0.044 | 0.1848 | Tidak valid |
| item 10 | 0.296 | 0.1848 | valid |
| item 11 | -0.146 | 0.1848 | Tidak valid |
| item 12 | 0.106 | 0.1848 | Tidak valid |
| item 13 | -0.076 | 0.1848 | Tidak valid |
| item 14 | 0.054 | 0.1848 | Tidak valid |
| item 15 | -0.004 | 0.1848 | Tidak valid |
| item 16 | -0.149 | 0.1848 | Tidak valid |
| item 17 | 0.000 | 0.1848 | Tidak valid |
| item 18 | 0.072 | 0.1848 | Tidak valid |
| item 19 | -0.192 | 0.1848 | Tidak valid |
| item 20 | 0.163 | 0.1848 | Tidak valid |
| item 21 | -0.003 | 0.1848 | Tidak valid |
| item 22 | 0.050 | 0.1848 | Tidak valid |
| item 23 | -0.024 | 0.1848 | Tidak valid |
| item 24 | 0.011 | 0.1848 | Tidak valid |
| item 25 | -0.297 | 0.1848 | Tidak valid |
| item 26 | 0.038 | 0.1848 | Tidak valid |
| item 27 | -0.105 | 0.1848 | Tidak valid |
| item 28 | 0.234 | 0.1848 | valid |
| item 29 | 0.000 | 0.1848 | Tidak valid |
| item 30 | -0.053 | 0.1848 | Tidak valid |
| item 31 | 0.098 | 0.1848 | Tidak valid |
| item 32 | 0.258 | 0.1848 | valid |
| item 33 | -0.127 | 0.1848 | Tidak valid |
| item 34 | 0.000 | 0.1848 | Tidak valid |
| item 35 | 0.188 | 0.1848 | valid |
| item 36 | 0.072 | 0.1848 | Tidak valid |
| item 37 | 0.000 | 0.1848 | Tidak valid |
| item 38 | 0.000 | 0.1848 | Tidak valid |
| item 39 | 0.149 | 0.1848 | Tidak valid |
| item 40 | 0.133 | 0.1848 | Tidak valid |
| item 41 | 0.000 | 0.1848 | Tidak valid |
| item 42 | 0.000 | 0.1848 | Tidak valid |
| item 43 | 0.000 | 0.1848 | Tidak valid |
| item 44 | 0.000 | 0.1848 | Tidak valid |
| item 45 | -0.169 | 0.1848 | Tidak valid |
| item 46 | 0.040 | 0.1848 | Tidak valid |
| item 47 | 0.054 | 0.1848 | Tidak valid |
| item 48 | 0.072 | 0.1848 | Tidak valid |
| item 49 | 0.092 | 0.1848 | Tidak valid |
| item 50 | 0.051 | 0.1848 | Tidak valid |

Tabel 3. Analisis Reliabilitas

| No | Subjek | X | |
|----|--------|-------------|------------|
| | | Kelas Bawah | Kelas Atas |
| 1 | A | 6 | 38 |
| 2 | B | 8 | 38 |
| 3 | C | 11 | 38 |
| 4 | D | 11 | 38 |
| 5 | E | 11 | 39 |
| 6 | F | 11 | 39 |
| 7 | G | 12 | 39 |
| 8 | H | 13 | 39 |
| 9 | I | 13 | 39 |
| 10 | J | 13 | 39 |
| 11 | K | 14 | 40 |
| 12 | L | 15 | 40 |
| 13 | M | 15 | 40 |
| 14 | N | 15 | 40 |
| 15 | O | 16 | 40 |
| 16 | P | 16 | 41 |
| 17 | Q | 16 | 41 |
| 18 | R | 16 | 41 |
| 19 | S | 16 | 41 |
| 20 | T | 16 | 41 |
| 21 | U | 16 | 41 |
| 22 | V | 17 | 42 |
| 23 | W | 17 | 44 |
| 24 | X | 17 | 44 |
| 25 | Y | 18 | 44 |
| 26 | Z | 18 | 46 |
| 27 | AA | 18 | 47 |
| 28 | BB | 18 | 47 |

$$r_t(KR21) = \frac{K}{K-1} \left\{ 1 - \frac{x(K-x)}{KS^2} \right\}$$

Kelas bawah

Kelas atas

K = 50
 X = 14.39
 S² = 9.65
 KR = -0.063
 R tabel = 0.1848

K = 50
 x = 40.93
 s² = 7.03
 KR = -0.057
 r tabel = 0.1848

$$r_t(KR21) = -0,063$$

Hasil = **Titak Reliabel**

$$r_t(KR21) = -0,057$$

Hasil = **Tidak Reliabel**

Tabel 4. Analisis Daya Beda dan Tingkat Kesulitan

| Soal | KA | KB | A-B | A+B | Daya Beda | | Tingkat Kesulitan | |
|------|----|----|-----|-----|-----------|--------------|-------------------|--------------|
| | | | | | Hasil | Interpretasi | Hasil | Interpretasi |
| 1 | 76 | 12 | 64 | 88 | 0.05 | Buruk | 0.03 | Sulit |
| 2 | 76 | 16 | 60 | 92 | 0.04 | Buruk | 0.03 | Sulit |
| 3 | 76 | 22 | 54 | 98 | 0.04 | Buruk | 0.04 | Sulit |
| 4 | 76 | 22 | 54 | 98 | 0.04 | buruk | 0.04 | Sulit |
| 5 | 78 | 22 | 56 | 100 | 0.04 | buruk | 0.04 | Sulit |
| 6 | 78 | 22 | 56 | 100 | 0.04 | buruk | 0.04 | Sulit |
| 7 | 78 | 24 | 54 | 102 | 0.04 | buruk | 0.04 | Sulit |
| 8 | 78 | 26 | 52 | 104 | 0.04 | buruk | 0.04 | Sulit |
| 9 | 78 | 26 | 52 | 104 | 0.04 | buruk | 0.04 | Sulit |
| 10 | 78 | 26 | 52 | 104 | 0.04 | buruk | 0.04 | Sulit |
| 11 | 80 | 28 | 52 | 108 | 0.04 | buruk | 0.04 | Sulit |
| 12 | 80 | 30 | 50 | 110 | 0.04 | buruk | 0.04 | Sulit |
| 13 | 80 | 30 | 50 | 110 | 0.04 | buruk | 0.04 | Sulit |
| 14 | 80 | 30 | 50 | 110 | 0.04 | buruk | 0.04 | Sulit |
| 15 | 80 | 32 | 48 | 112 | 0.03 | buruk | 0.04 | Sulit |
| 16 | 82 | 32 | 50 | 114 | 0.04 | buruk | 0.04 | Sulit |
| 17 | 82 | 32 | 50 | 114 | 0.04 | buruk | 0.04 | Sulit |
| 18 | 82 | 32 | 50 | 114 | 0.04 | buruk | 0.04 | Sulit |
| 19 | 82 | 32 | 50 | 114 | 0.04 | buruk | 0.04 | Sulit |
| 20 | 82 | 32 | 50 | 114 | 0.04 | buruk | 0.04 | Sulit |
| 21 | 82 | 32 | 50 | 114 | 0.04 | buruk | 0.04 | Sulit |
| 22 | 84 | 34 | 50 | 118 | 0.04 | buruk | 0.04 | Sulit |
| 23 | 88 | 34 | 54 | 122 | 0.04 | buruk | 0.04 | Sulit |
| 24 | 88 | 34 | 54 | 122 | 0.04 | buruk | 0.04 | Sulit |
| 25 | 88 | 36 | 52 | 124 | 0.04 | buruk | 0.04 | Sulit |
| 26 | 92 | 36 | 56 | 128 | 0.04 | buruk | 0.05 | Sulit |
| 27 | 94 | 36 | 58 | 130 | 0.04 | buruk | 0.05 | Sulit |
| 28 | 94 | 36 | 58 | 130 | 0.04 | buruk | 0.05 | Sulit |

Kesimpulan

Berdasarkan informasi dari data yang telah diproses dapat diambil kesimpulan sebagai berikut:

- 1 Validitas logis tidak menunjukkan sebaran keterwakilan isi materi yang dibahas dalam kegiatan belajar mengajar. Validitas konstruk juga tidak cukup memenuhi kriteria validitas tes yang baik dengan hanya sebagian kecil soal yang dapat dianggap valid.
- 2 Soal-soal dalam naskah ujian Bahasa Inggris semester II tahun 2006 di Polinela tidak reliabel. Tes tersebut menunjukkan angka negatif, yaitu -0,063 pada kelas bawah dan -0,057 pada kelas atas. sedangkan kriteria reliabilitas soal yang baik adalah 0,1848
- 3 Seluruh soal ujian tidak memiliki kriteria daya beda soal yang baik. Soal ujian tersebut terlalu sukar bagi mahasiswa yang mengikuti ujian semester II mata kuliah Bahasa Inggris tahun 2006 di Polinela.

Daftar Pustaka

- Arikunto, Suharsimi. 2003. *Dasar-dasar Evaluasi Pendidikan (edisi revisi)*. Bumi Aksara. Jakarta.
- Azwar, Saifudin. 1997. *Reliabilitas dan Validitas*. Pustaka Pelajar. Jogjakarta.
- Brown, H. Douglas. 1980. *Principle of Language Learning and Teaching*. Prentice-Hall, Inc. New Jersey.
- Henning, Grant. 1987. *A Guide to Language Testing Development Evaluation Research*. U.S.A. Peacock Publisher, Inc.
- Rahman, Anwar. 1999. *The Quality of The First Catur Wulan English Tes for The Second Year Students of SMUN 2 Bandar Lampung in The Year 1998/1999 (a script)*. Lampung University.
- Shohamy, Elena. 1985. *A Practical Handbook in Language Testing for the Second Language Teacher*. Tel-Aviv University. Israel.
- Ten Brink, Terry D. 1974. *Evaluation a Principle Guide for Teachers*. U.S.A.: McGraw-Hill, Inc.