KESESATAN DALAM PENGEMBANGAN TES UNTUK PENGUKURAN PENCAPAIAN HASIL BELAJAR PADA KURIKULUM BERBASIS KOMPETENSI

Oleh: Bambang Subali
Staf Pengajar FMIPA UNY

Abstract

The validity and reliability of measuring instruments determine their quality. The implementation of the competency-based curriculum at schools has caused basic problems to solve concerning requirements to be fulfilled related to the validity and reliability of such instruments. This article attempts to discuss the fulfillment of the requirements concerning the validity and reliability of tests used to measure learning achievement in educational research, or, for the interest of educational practice in the field, in relation to the implementation of the competency-based curriculum.

It can be concluded that the requirements for the validity of a test measuring such learning achievement can be fulfilled by making a test grid. The item validity is empirically determined; another test, a standardized one, is needed for comparison. In testing reliability of tests by finding out a correlation coefficient, a coefficient of homogeneity, or a standard error of measurement, one refers to a normality of distribution. It can be misleading if applied when testing the reliability of a criterion-referenced measuring instrument. Investigation on the reliability of such an instrument is based on the percentage of its consistency. Item analysis in norm-referenced tests is for investigating item effectiveness in discriminating testers or detecting their division into two groups of achievers, higher and lower, measured on the basis of the values of the point biserial coefficient for item discrimination, or, for the discriminating power of items, and also on the proportion of correct answers for item difficulty. Item analysis in criterion-referenced tests is for investigating effectiveness of learning processes, measured on the

basis of the values of the sensitivity index. So a researcher or teacher measuring learning achievement related to the competency-based curriculum must use a criterion-referenced test.

Keywords: validity, reliability, achievement test, assessment

Pendahuluan

Dalam penelitian pendidikan, banyak data yang dihimpun menggunakan instrumen yang harus dikembangkan sendiri oleh peneliti. Hal yang sama juga dialami oleh guru untuk kepentingan praktis di lapangan saat ia harus mengukur pencapaian hasil belajar, baik untuk kepentingan penilaian (asesmen) maupun untuk mengevaluasi program pembelajaran yang dirancangnya. Hal tersebut sangat berbeda dengan penelitian dalam ilmu natural yang pada umumnya peneliti tinggal menggunakan instrumen yang sudah tersedia.

Kualitas instrumen pengukuran, baik untuk kepentingan penelitian pendidikan maupun untuk kepentingan praktis selalu dilihat dari dua aspek. *Pertama*, persyaratan kesahihan (validitas) yang berkaitan dengan kemampuan alat ukur untuk mengukur apa yang seharusnya diukur. *Kedua*, persyaratan keandalan (reliabilitas) yang berkaitan dengan keajegan/konsistensi hasil pengukuran jika dilakukan pengulangan pengukuran. Dengan demikian, instrumen yang baik juga harus memiliki bukti dari aspek kesahihan dan keandalan.

Kurikulum 2004 sudah mulai diterapkan di sekolah dalam skala terbatas dalam bentuk *mini-piloting*. Namun demikian, banyak sekolah yang secara swadaya sudah ikut menerapkan. Kurikulum 2004 merupakan kurikulum berbasis kompetensi, sehingga keberhasilan belajar siswa harus berbasis standar. Oleh karena itu, kurikulum berbasis kompetensi juga disebut kurikulum berbasis standar. Sebagai konsekuensinya, keberhasilan peserta didik dalam pencapaian hasil belajar harus dinilai/diases dengan cara

dibandingkan dengan kriteria/standar (Direktorat PLP, 2004, Direktorat PMU, 2004). Pertanyaan yang mendasar adalah bagaimanakah pemenuhan kesahihan dan keandalan instrumen, khususnya tes pengukuran pencapaian hasil belajar agar hasilnya dapat dibandingkan dengan kriteria/standar.

Hal tersebut perlu dikaji dan dipaparkan secara tuntas mengingat dengan bergulirnya kurikulum baru akan memberi peluang dilakukannya penelitian, baik dalam konteks untuk mengevaluasi keberhasilan implementasi maupun dalam konteks untuk mengembangkan berbagai model ataupun strategi pembelajaran yang dapat diimplementasikan yang berkaitan dengan ketercapaian penguasaan kompetensi oleh peserta didik.

Selama ini, para guru di lapangan dalam mengukur pencapaian hasil belajar siswa terbiasa dengan pengembangan tes yang sebenarnya mengacu pada acuan norma. Dalam banyak pelatihan tanpa disadari masih ada instruktor yang mengenalkan kesahihan dan keandalan tes pengukuran pencapaian hasil belajar yang lebih mengacu pada acuan norma. Tulisan ini mencoba memaparkan karakteristik pemenuhan kesahihan dan keandalan tes pengukuran pencapaian hasil belajar yang beracuan pada kriteria/standar yang berkaitan dengan implementasi Kurikulum 2004 di sekolah.

Dasar Pemilihan Instrumen Penilaian/Asesmen

Dalam buku pedoman penilaian yang dikeluarkan oleh Direktorat PLP (2004 dan 2005), Direktorat PMU (2004), maupun draf buku pedoman asesmen berbasis kompetensi yang dikeluarkan oleh Dikti (2005), pemilihan instrumen untuk mengukur pencapaian hasil belajar tidak dapat dipisahkan dari pemilihan strategi penilaian/asesmen karena strategi penilaian/asesmen memuat metode penilaian dan bentuk instrumen. Sejalan dengan karakteristik kurikulum yang tidak hanya mengandalkan pada tes tulis, maka dalam pengembangan kisi-kisi penilaian terdapat berbagai bentuk instrumen yang dapat dipilih sesuai dengan

karakteristik metode/teknik penilaian. Berikut ini disajikan ragam metode dan bentuk instrumen penilaian dari Direktorat PLP, Direktorat PMU, dan dari Dikti.

Tabel 1. Jenis Tagihan dan Bentuk Instrumen Penilaian dalam Sistem Asesmen Berbasis Kompetensi menurut Buku Pedoman Penilaian dari Direktorat PLP dan Direktorat PMU Tahun 2004

Jenis Tagihan	Bentuk Instrumen
a. Kuis pada awal pelajaran	• Isian singkat • Pertanyaan singkat
b. Pertanyaan lisan pada akhir pelajaran	• Pertanyaan singkat
c. Ulangan harian	• Tes tertulis (tes pilihan ganda, isian singkat, menjodohkan, uraian, dan lainnya) • Tes kinerja/tes unjuk kerja
d. Ulangan semester	• Tes tertulis (tes pilihan ganda, isian singkat, menjodohkan, uraian, dan lainnya) • Tes kinerja/tes unjuk kerja
e. Ulangan kenaikan kelas	• Tes tertulis (tes pilihan ganda, isian singkat, menjodohkan, uraian, dan lainnya) • Tes kinerja/tes unjuk kerja
f. Tugas individu	• Proyek dan portofolio
g. Tugas kelompok	• Proyek

Tabel 2. Jenis Tagihan, Teknik Penilaian, Bentuk, dan Contoh Instrumen dalam Sistem Asesmen Berbasis Kompetensi menurut Buku Pedoman Penilaian dari Direktorat PLP (2005)*

Jenis Tagihan	Teknik Penilaian	Bentuk Instrumen	Contoh Instrumen
Tes	Kuis	• Pertanyaan lisan • Pertanyaan tertulis • Isian singkat • Pilihan ganda • Dsb.	Soal dan atau perintah
		• Unjuk kerja singkat • Dsb	
	Tes Harian (Ulangan Harian)	• Pertanyaan lisan • Pertanyaan tertulis • Isian singkat • Pilihan ganda • Uraian (disertai rubrik) • dsb.	Soal dan atau perintah
		• Unjuk kerja (disertai rubrik) • dsb.	
Non-Tes	Observasi	• Panduan (lembar) observasi	
	Wawancara	• Panduan (pedoman) wawancara	
	Pemberian angket	• Angket (kuesioner)	
	Pemberian tugas	• Perintah/arahan (dengan rubrik)	
	Proyek	• Perintah/arahan (dengan rubrik)	
	Portofolio	• Perintah/arahan (dengan rubrik)	

*) Pedoman Khusus Pengembangan Sistem Penilaian Mata Pelajaran IPA, Edisi April 2005

Tabel 3. Klasifikasi Metode dan Bentuk Instrumen Asesmen dalam Pedoman Sistem Asesmen Berbasis Kompetensi dari Ditjen Dikti*

No.	Metode Asesmen	Teknik Penilaian	Bentuk Instrumen Asesmen
1	Tes (gradasi benar-salah)		
A	Tes formal (ujian midsemester, ujian akhir, ujian responsi, dan sejenisnya)	• Tes tulis	• Tes isian • Tes uraian • Tes pilihan ganda • Dll.
		• Tes lisan	• Daftar pertanyaan
		• Tes kinerja	• Tes tulis keterampilan • Tes identifikasi • Tes simulasi • Uji petik kerja
B	Tes non formal (menyatu dengan proses pembelajaran)	• Penugasan	• Tugas proyek • Tugas portofolio • Tugas rumah
		• Observasi	• Lembar observasi
2	Nontes (gradasi positif-negatif, setuju-tidak setuju, suka-tidak suka)	• Observasi	• Lembar observasi
		• Wawancara	• Pedoman wawancara
		• Inventori	• Skala inventori
		• Self report	• Kuesioner

* Draf Pedoman Umum Pengembangan Instrumen Berbasis Kompetensi, Edisi Desember 2005

Berdasarkan contoh strategi penilaian yang diterbitkan oleh tiga lembaga di atas, tampak terdapat perbedaan mengenai dasar yang dipakai dalam klasifikasi, terutama antara Direktorat PLP Edisi Tahun 2005, dan Ditjen Dikti Edisi Tahun 2005. Direktorat PLP membedakan tes dan nontes atas dasar formal dan tidaknya diselenggarakannya suatu tes. Dengan demikian, tes diartikan sebagai pengukuran yang dilakukan dalam situasi ujian, sedangkan nontes diartikan pengukuran/nonpengukuran yang dilakukan selama pembelajaran. Ditjen Dikti membedakan tes dan nontes atas dasar gradasi hasil yang diperoleh. Disebut tes bila hasil pengukurannya dapat digradasikan benar-salah, dan disebut nontes jika hasilnya

tidak dapat digradasikan benar-salah, dan hanya digradasi positif-negatif, suka-tidak suka, atau setuju-tidak setuju. Terlepas dari perbedaan yang dijadikan sebagai dasar klasifikasi tagihan/metode penilaian, tampak bahwa ada perbedaan yang mendasar yang berkaitan dengan kisi-kisi untuk kepentingan tes tertulis dan kisi-kisi untuk penilaian penguasan kompetensi dasar.

Kesahihan Instrumen

Menurut Peter Hagul (1980:95-101), terdapat beberapa macam kesahihan/validitas suatu instrumen pengukuran dalam bidang pendidikan maupun dalam bidang ilmu sosial. Macam-macam validitas tersebut di antaranya kesahihan konstruks, kesahihan antarbudaya, kesahihan internal dan eksternal, kesahihan isi, kesahihan prediktif, dan kesahihan muka. Tidak setiap instrumen harus memenuhi seluruh persyaratan kesahihan. Hal tersebut bergantung kepada karakteristik variabel yang diukur dan tujuan pengukurannya. Contoh, instrumen untuk mengukur pencapaian hasil belajar akan berbeda persyaratannya dengan instrumen untuk mengukur minat ataupun motivasi belajar.

Kesahihan konstruks berkaitan dengan kemampuan alat ukur dalam mengukur aspek yang diukur. Tes untuk mengukur kecerdasan harus menjawab pertanyaan apakah kecerdasan itu berupa kemampuan mengingat fakta, membuat abstraksi, membuat aplikasi, menganalisis, membuat sintesis, atau melakukan evaluasi. Jika berkaitan dengan seluruh dimensi tersebut, setiap dimensi harus dicari indikator/unsurnya, apa saja yang termasuk indikator kemampuan mengingat fakta, membuat abstraksi, dan seterusnya. Kecermatan mendefinisikan dan menentukan dimensi variabel beserta indikatornya menjadi kunci pemenuhan kesahihan konstruks instrumen yang dikembangkan.

Kesahihan antarbudaya menuntut kenetralan suatu instrumen pengukuran jika subjek penelitiannya bersifat multikultural. Jika instrumen tidak nentral, berarti terdapat kelompok subjek yang diuntungkan dan kelompok subjek yang dirugikan.

Kesahihan internal berkaitan dengan kejelasan kedudukan suatu variabel yang diukur. Setiap variabel yang diukur harus jelas hubungannya dengan variabel lain, apakah hubungannya bersifat kausal ataukah korelasional, atau benar-benar independen. Kesahihan eksternal berkaitan dengan generalisasi yang akan diambil dari hasil penelitiannya.

Kesahihan isi berkaitan dengan pertanyaan seberapa jauh instrumen yang dipakai untuk mengukur sudah mencakup seluruh cakupan dari variabel yang akan diukur. Jika yang diukur berupa pencapaian hasil belajar, maka kesahihan isi selalu dikaitkan dengan cakupan keberhasilan menurut kurikulum, sehingga dalam konteks pengukuran prestasi kesahihan isi berupa kesahihan kurikuler.

Kesahihan prediktif adalah kemampuan alat ukur untuk memprediksi keberhasilan subjek pada waktu mendatang jika ia menempuh program lanjutan. Dengan demikian, pengujian validitas prediktif tidak dapat segera dipenuhi karena butuh tenggang waktu yang lama berkaitan dengan subjek dalam menempuh program lanjutannya.

Kesahihan muka meliputi dua aspek, yakni aspek substansi dan aspek bahasa. Kesahihan aspek substansi dapat dipenuhi melalui kaji ulang/*review* oleh pakar sebidang. Kesahihan aspek bahasa ditinjau dalam konteks kekomunikatifannya. Oleh karena itu, untuk pemenuhannya memerlukan kaji ulang/*review* dari pakar bahasa.

Ary (1980:216) mengemukakan bahwa kesahihan instrumen juga dapat dikaitkan dengan kriteria dalam bentuk kesahihan konkurens, yaitu dengan melihat kesejajaran suatu instrumen dengan instrumen lain yang baku/standar. Instrumen standar untuk tes buatan guru tidak ada yang dapat dijadikan sebagai pembanding, sehingga menjadikan kesulitan dalam langkah mengembangkan instrumen yang standar yang berhubungan dengan kriteria.

Berkaitan dengan pengukuran pencapaian hasil belajar peserta didik dalam penguasai kompetensi dasar yang menjadi target Kurikulum 2004, pemenuhan kesahihan kurikuler dapat dipenuhi dengan cara menyusun kisi-kisi yang memuat indikator pencapaian

dan strategi sistem penilaian/asesmennya. Dalam hal ini, terdapat pergeseran mendasar dalam penyusunan kisi-kisi sistem penilaian. Dalam kurikulum 1994, kisi-kisi tes hanya difokuskan pada pengembangan kisi-kisi untuk tes tertulis. Bahkan, karena ujian nasional hanya menggunakan tes pilihan ganda, praktis bentuk tes yang dipakai hanya bentuk pilihan ganda. Dari sisi pengembangan tes kinerja yang pernah dilakukan pun, untuk SD, SMP, dan SMA boleh dikata tidak ada pengembangan kisi-kisi karena tes kinerja jarang diuji, kecuali pada SMK.

Keandalan Instrumen

Sejalan dengan karakteristik penilaian berbasis kompetensi yang menggunakan acuan kriteria, maka keandalan/reliabilitas tes untuk mengukur pencapaian hasil belajar juga dalam konteks untuk memenuhi standar kualitas tes beracuan kriteria/standar. Dengan demikian, sudah semestinya bukan atas dasar standar kualitas tes beracuan norma. Keandalan tes beracuan norma dilakukan dengan alasan bahwa populasi siswa yang belajar memiliki kemampuan yang mengikuti distribusi normal. Oleh karena itu, hasil penelitian diharapkan dapat digeneralisasi pada tingkat populasi, jika penelitian sampling yang dilakukan memenuhi persyaratan keparametrikan.

Dalam rangka implementasi Kurikulum 2004, pada buku pedoman umum penilaian yang diterbitkan oleh Direktorat PMU (2004) dan Direktorat PLP (2004) dinyatakan bahwa persyaratan kualitas instrumen dari sisi keandalan harus dikaitkan dengan konsistensi suatu instrumen saat dipakai untuk mengukur. Instrumen yang andal dapat mengukur secara konsisten apa yang diukur dan menghasilkan ukuran yang tetap. Ibarat sebuah timbangan, ia dapat mengukur dan menunjukkan berat yang sama terhadap benda tertentu meskipun digunakan di daerah yang berbeda pada waktu yang berbeda. Keandalan suatu instrumen dapat dilihat dari konsistensi internal, stabilitas, dan konsistensi antarpenilai. Instrumen yang baik memberikan nilai yang sama meskipun dilakukan di tempat atau waktu yang berbeda sepanjang objek yang

dinilai belum berubah. Besarnya indeks keandalan digunakan untuk menghitung kesalahan pengukuran. Semakin andal suatu instrumen, semakin kecil kesalahan pengukuran. Kesalahan pengukuran tersebut dapat bersifat acak akibat kondisi yang diukur dan yang mengukur bervariasi, dapat pula karena pemilihan bahan yang diujikan tidak tepat, sedangkan kesalahan sistematik terjadi karena instrumennya atau cara penskorannya cenderung murah atau mahal untuk semua peserta.

Menurut Ary (1985: 231-234) dan Gronlund (1990: 77-87), pengujian keandalan tes pengukuran pencapaian hasil belajar melalui pendekatan klasik dapat dilakukan dengan tiga cara. *Pertama*, mencari nilai koefisien korelasi, baik korelasi tes-retes, korelasi tes yang setara, maupun korelasi internal dengan teknik belah dua. Kedua, yaitu dengan mencari koefisien homogenitas untuk mengukur konsistensi internal suatu instrumen, misalnya dengan menggunakan formula 20 dan formula 21 dari Kuder-Richardson. Koefisien homogenitas juga dapat dicari dengan menggunakan koefisien alfa dari Cronbach. *Ketiga*, mencari *standard error of measurement* yang menunjukkan besarnya kesalahan pengukuran.

Berikut ini disajikan ilustrasi perhitungan keandalan tes secara internal yang dihitung berdasarkan indeks Alfa Cronbach dan *standar error of measurement* (SEM) dengan menggunakan program ITEMAN. Misalnya, dari 12 testi yang mengerjakan 10 item tes, 6 orang berhasil sepenuhnya dan 6 orang gagal total.

Tabel 4. Hasil Tes dari 12 siswa/testi yang Mengerjakan 10 Item Tes Pencapaian Hasil Belajar untuk Materi Pokok YY dengan Hasil yang Berimbang

testi	i1	i2	i3	i4	I5	i6	i7	i8	i9	i10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1

testi	i1	i2	i3	i4	I5	i6	i7	i8	i9	i10
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0

Hasil analisis Program ITEMAN (tm) Version 3.00

Seq. No. Key	Scale Item	Prop. Correct.	Biser.	Point Biser.
1	0-1	0.500	1.000	1.000
2	0-2	0.500	1.000	1.000
3	0-3	0.500	1.000	1.000
4	0-4	0.500	1.000	1.000
5	0-5	0.500	1.000	1.000
6	0-6	0.500	1.000	1.000
7	0-7	0.500	1.000	1.000
8	0-8	0.500	1.000	1.000
9	0-9	0.500	1.000	1.000
10	0-10	0.500	1.000	1.000

Scale Statistics:
N of Items: 10; N of Examinees: 12; Mean: 5.000; Variance: 25.000; Std. Dev.: 5.000; Skew: 0.000; Kurtosis: -2.000; Minimum: 0.000; Maximum: 10.000; Median: 0.000; Alpha: 1.000; SEM: 0.000; Mean P: 0.500; Mean Item-Tot.: 1.000; Mean Biserial: 1.000

Hasil analisis menunjukkan nilai koefisien Alfa Cronbach sebesar 1,0 dan SEM 0.0 yang berarti instrumen sangat andal dan tidak ada kesalahan pengukuran. Bagaimana jika 11 orang berhasil sepenuhnya dan hanya seorang yang gagal total?

Tabel 5. Hasil Tes dari 12 Siswa/Testi yang Mengerjakan 10 Item Tes Pencapaian Hasil Belajar untuk Materi Pokok YY dengan 11 Orang Berhasil Sepenuhnya dan Seorang Gagal Total

testi	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	0	0

Hasil analisis Program ITEMAN (tm) Version 3.00

Seq. No. Key	Scale Item	Prop. Correct.	Biser.	Point Biser.
1	0-1	0.917	1.000	1.000
2	0-2	0.917	1.000	1.000
3	0-3	0.917	1.000	1.000
4	0-4	0.917	1.000	1.000
5	0-5	0.917	1.000	1.000
6	0-6	0.917	1.000	1.000
7	0-7	0.917	1.000	1.000
8	0-8	0.917	1.000	1.000
9	0-9	0.917	1.000	1.000
10	0-10	0.917	1.000	1.000

Scale Statistics:

N of Items: 10; N of Examinees: 12; Mean: 9.167; Variance: 7.639; Std. Dev.: 2.764; Skew: -3.015; Kurtosis: 7.091; Minimum: 0.000; Maximum: 10.000; Median: 10.000; Alpha: 1.000; SEM: 0.001; Mean P: 0.917; Mean Item-Tot.: 1.000; Mean Biserial: 1.000

Hasil analisis menunjukkan nilai koefisien Alfa Cronbach juga sebesar 1,0 dan SEM 0.001 yang berarti instrumen sangat andal dan kesalahan pengukuran sangat kecil. Bagaimana jika sebaliknya, yakni 11 orang gagal total dan hanya seorang yang berhasil sepenuhnya?

Tabel 6. Hasil Tes dari 12 Siswa/Testi yang Mengerjakan 10 Item Tes Pencapaian Hasil Belajar untuk Materi Pokok YY dengan 11 Orang Gagal Total dan Seorang Berhasil Sepenuhnya

Testi	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0

Hasil analisis Program ITEMAN (tm) Version 3.00

Seq. No. Key	Scale Item	Prop. Correct.	Biser.	Point Biser.
1	0-1	0.083	1.000	1.000
2	0-1	0.083	1.000	1.000
3	0-1	0.083	1.000	1.000

Seq. No. Key	Scale Item	Prop. Correct.	Biser.	Point Biser.
4	0-1	0.083	1.000	1.000
5	0-1	0.083	1.000	1.000
6	0-1	0.083	1.000	1.000
7	0-1	0.083	1.000	1.000
8	0-1	0.083	1.000	1.000
9	0-1	0.083	1.000	1.000
10	0-1	0.083	1.000	1.000

Scale Statistics:
N of Items: 10; N of Examinees: 12; Mean: 0.833; Variance: 7.639; Std. Dev.: 2.764; Skew: 3.015; Kurtosis: 7.091; Minimum: 0.000; Maximum: 10.000; Median: 0.000; Alpha: 1.000; SEM: 0.000; Mean P: 0.083; Mean Item-Tot.: 1.000; Mean Biserial: 1.000

Hasil analisis menunjukkan bahwa nilai koefisien Alfa Cronbach juga sebesar 1,0 dan SEM 0.000 yang berarti instrumen sangat andal dan tidak ada kesalahan pengukuran. Bagaimana jika hasil tes bervariasi dan ada satu item yang berhasil dikerjakan seluruh testi dan ada pula satu item yang gagal dikerjakan seluruh testi?

Tabel 7. Hasil Tes dari 12 Siswa/Testi yang Mengerjakan 10 Item Tes Pencapaian Hasil Belajar untuk Materi Pokok YY dengan 1 Item Berhasil Dikerjakan Seluruh Testi dan Satu Item Gagal Dikerjakan Seluruh Testi

Testi	i1	i2	I3	i4	i5	i6	i7	i8	i9	i10
1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1	1	0
3	1	1	1	1	1	1	1	1	1	0
4	1	1	1	1	1	1	1	1	0	0
5	1	1	1	1	1	1	1	0	0	0
6	1	1	1	1	1	1	0	0	0	0
7	1	1	1	1	1	0	0	0	0	0
8	1	1	1	1	0	0	0	0	0	0
9	1	1	1	0	0	0	0	0	0	0

Testi	i1	i2	I3	i4	i5	i6	i7	i8	i9	i10
10	1	1	0	0	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0

Hasil analisis Program ITEMAN (tm) Version 3.00

Seq. No. Key	Scale Item	Prop. Correct.	Biser.	Point Biser.
1	0-1	1.000	-9.000	-9.000
2	0-2	0.833	0.970	0.650
3	0-3	0.750	1.000	0.775
4	0-4	0.667	1.000	0.850
5	0-5	0.583	1.000	0.888
6	0-6	0.500	1.000	0.894
7	0-7	0.417	1.000	0.869
8	0-8	0.333	1.000	0.810
9	0-9	0.250	0.968	0.710
10	0-10	0.000	-9.000	-9.000

Scale Statistics:
N of Items: 10; N of Examinees: 12; Mean: 5.333; Variance: 8.889; Std. Dev.: 2.981; Skew: -0.136; Kurtosis: -1.458; Minimum: 1.000; Maximum: 9.000; Median: 5.000; Alpha: 0.899; SEM: 0.946; Mean P: 0.533; Mean Item-Tot.: 0.806; Mean Biserial: 0.992

Jika keberhasilan testi bervariasi, namun masih nyata dapat dipisahkan kelompok atas dan kelompok bawah, maka hasil analisis menunjukkan nilai koefisien Alfa Cronbach juga masih tinggi yakni sebesar 0,899 dan SEM 0.946. Hal tersebut berarti instrumen masih tergolong sangat andal dan tetapi terdapat kesalahan pengukuran yang tinggi.

Berdasarkan contoh di atas, pengujian keandalan instrumen tes yang hanya melihat pada besarnya koefisien Alfa Cronbach hanya sekedar untuk menunjukkan keandalan ditinjau dari homogenitas

itemnya, sedangkan bila dilihat dari SEM akan dapat diketahui besarnya penyimpangan yang terjadi pada antaritem.

Bagaimanakah agar tes pengukuran pencapaian hasil belajar dapat memenuhi persyaratan tes beracuan kriteria? Menurut Ary, dkk. (1985: 238-239) sukar untuk menetapkannya. Alasannya, apabila siswa belajar terus secara efektif, maka semua siswa akan menguasai kompetensi yang ditargetkan. Akibatnya, variabilitas antarsiswa semakin kecil, bahkan boleh jadi tidak ada atau sama dengan 0. Demikian pula jika peserta didik memiliki potensi yang tinggi, sehingga dengan mudah menguasai kompetensi yang dipelajari, maka dengan sendirinya variasi kemampuan antarsubjek juga kecil atau tidak ada.

Prosedur untuk menganalisis keandalan tes beracuan kriteria menurut Gronlund & Linn (1990: 98-100) dan Gronlund (1998: 215-217) berdasarkan persen konsistensi. Hal yang sama juga dikemukakan oleh Frisbie (2005: 26). Persen konsistensi diperoleh dengan cara sekelompok testi yang dites dengan dua set tes yang setara. Persen konsistensi suatu pasangan tes dapat dihitung atas dasar banyaknya testi yang konsisten menjawab benar ditambah dengan banyaknya testi yang konsisten menjawab salah dari pasangan tes yang bersangkutan dibagi dengan jumlah testi peserta tes.

Misalnya, ada dua perangkat tes yang setara, yakni tes A dan B yang terdiri dari 25 item, kemudian diujikan pada 40 peserta. Seandainya batas penguasaan jika siswa berhasil mengerjakan 80% atau 20 item. Hasil tes menunjukkan 30 siswa yang memiliki skor 20 atau lebih pada kedua pasangan tes. Sebanyak 6 siswa memiliki skor di bawah 20 pada kedua pasangan tes. Sebanyak 4 siswa yang menguasai salah satu tes, yakni 2 siswa berhasil mengerjakan 20 item atau lebih tes A, tetapi tidak berhasil mengerjakan 20 item tes B, demikian pula sebaliknya, 2 siswa berhasil mengerjakan 20 item atau lebih tes B, tetapi gagal mengerjakan 20 item tes A. Dengan demikian, ada 36 dari 40 siswa yang dikategorikan konsisten dalam

mengerjakan pasangan tes tersebut, sehingga persen konsistensi tes pengukuran hasil belajar tersebut sebesar (36/40)x100% atau 90%. Jika seluruh testi gagal semua ataupun berhasil semua dalam mengerjakan pasangan tes, persen konsistensinya tetap akan tinggi, yaitu sebesar 1.0. Implikasinya adalah, jika persen konsistensi suatu tes awal/*pretest* sebesar 1.0 dan angka tersebut berasal dari semua testi yang gagal mengerjakan pasangan tes, maka diartikan memang tes tersebut mampu menunjukkan bahwa testi belum menguasai kompetensi karena ia belum belajar. Sebaliknya, jika hal tersebut terjadi pada *posttest* dan angka tersebut berasal dari semua testi yang berhasil mengerjakan pasangan tes yang bersangkutan, maka harus diartikan bahwa tes tersebut mampu menunjukkan siswa telah berhasil dalam melakukan proses belajar. Oleh karena itu, setiap item tes harus sudah memenuhi persyaratan materi, konstruksi, dan bahasa. Persoalannya, membuat sepasang tes yang setara bukan pekerjaan yang mudah.

Analisis Item

Analisis item ditujukan untuk melihat kesahihan dan keandalan setiap item penyusun tes secara empiris. Menurut Dali S. Naga (2004: 105), kesahihan item adalah daya beda item yang dihitung berdasarkan korelasi antara skor satuan item dengan skor total atau skor testi. Menurut Gulliksen (1950: 375-377), menganalisis korelasi antara skor satuan item dengan skor total menghasilkan indeks keandalan item (*reliability index*). Namun demikian, menurut Kumaidi (1994:108) cara tersebut lebih tepat untuk menentukan daya beda item (*item discrimination*), yakni keefektifan item dalam membedakan kelompok atas dan kelompok bawah. Dalam program ITEMAN, koefisien korelasi tersebut dinyatakan sebagai koefisien biserial atau koefisien poin biserial (Ditjen PMU, 1999: 115-116). Jika analisis item hanya didasarkan pada nilai koefisien korelasi biserial, keefektifannya belum tentu baik karena item akan efektif memisahkan kelompok atas dan kelompok bawah jika memiliki

tingkat kesukaran yang berkualifikasi sedang.

Menurut Kumaidi (2004: 110-111), untuk menentukan validitas item dengan mencari korelasi antara skor satuan item dengan skor total tidak tepat. Kesahihan item tidak dapat dilihat secara internal, tetapi harus dibandingkan dengan tes lain sebagai kriteria, baik dalam konteks pemenuhan validitas konkuren maupun validitas prediktif. Misalnya, validitas item suatu tes potensi akademik yang sedang dikembangkan seorang peneliti harus menggunakan angka prestasi seperti UAN sebagai kriteria untuk memenuhi validitas konkuren atau menggunakan prestasi semester di universitas sebagai kriteria untuk memenuhi validitas prediktif.

Berdasarkan contoh hasil analisis yang sudah dipaparkan di atas, yakni data Tabel 4, 5, dan 6, instrumen yang terbaik adalah yang hasil pengukurannya tersaji pada Tabel 4. Hal ini ditunjukkan dengan nilai koefisien korelasi poin biserial 1,0 untuk semua item dan proporsi siswa menjawab benar untuk tiap item juga 0,5, yang berarti bahwa instrumen dapat memisahkan secara tegas kelompok atas (kelompok yang sepenuhnya berhasil) dan kelompok bawah (kelompok yang gagal total) dengan banyak siswa yang berimbang.

Analisis item juga dapat dilakukan menggunakan teori respons item (*item response theory*) yang sering dinyatakan dengan pendekatan yang modern. Kelebihan prinsip teori respons item adalah hasil analisis dapat memisahkan antara karakteristik testi dan karakteristik tes sebagai alat ukur. Hal ini tidak dapat dipenuhi dalam pendekatan klasik. Jika kedua hal tersebut tidak dapat dipisahkan maka tidak dapat diketahui antara kemampuan testi dengan tingkat kesukaran tes karena testi akan kelihatan berkemampuan tinggi bila item-item tesnya mudah dan sebaliknya kemampuan testi akan terlihat rendah jika item-item tesnya sukar (Hambilton et al., 1991:2). Oleh karena itu, item tes yang baik jika ia benar-benar dapat mengukur kemampuan testi. Sebagai contoh, kemampuan seseorang menyelesaikan soal aljabar karena memang pengetahuannya tentang teori matematika memadai, sehingga semakin tinggi pemahaman tentang teori matematika semakin besar

peluangnya untuk dapat memecahkan soal-soal aljabar (Gronlund, 1990: 467-468). Namun demikian, untuk menguji keandalan berdasar teori respons item diperlukan sampel yang sangat besar, misalnya dalam program ASCAL dari MicroCAT (tm) Testing System (1988) dipersyaratkan banyaknya testi 500. Hal ini sulit dipenuhi dalam pekerjaan sehari-hari seorang guru. Dalam pendekatan klasik pun untuk kestabilan informasi menurut Numally analisis untuk 50 item memerlukan 500 testi, menurut Davis 400 testi, sedangkan menurut Croker & Algina 200 testi (Dali S. Naga, 2004: 107-108).

Pada buku pedoman umum sistem penilaian dari Direktorat PMU (2004) dan Direktorat PLP (2004), keandalan item instrumen pengukur pencapaian hasil belajar dapat dinyatakan sebagai indeks keandalan yang berkisar antara 0 sampai 1. Item instrumen yang diterima (yang baik) memiliki indeks keandalan minimal 0,7. Item instrumen yang memiliki indeks keandalan antara 0,31 - 0,69 dapat diperbaiki, jika kurang dari 0,3 sebaiknya diganti. Sementara menurut Frisbie (2005: 26), daya pembeda item (*item discrimination*) masih dapat dipakai untuk mencirikan item beracuan kriteria sepanjang nilainya tidak negatif, sementara indeks kesukarannya boleh bervariasi dari rendah sampai tinggi.

Dengan demikian, apabila mengikuti pedoman dari Direktorat PMU dan PLP, data Tabel 4, 5, 6, dan 7 dengan tanpa memperhatikan proporsi siswa yang menjawab benar, item yang harus diganti adalah nomor 1 dan 10 yang menghasilkan data Tabel 7, dan item yang harus diperbaiki adalah item nomor 2 juga yang menghasilkan data pada Tabel 7. Jika dalam perbaikan atau penggantian item juga harus memperhatikan proporsi siswa yang menjawab benar, maka seluruh item yang menghasilkan data pada Tabel 5 juga harus diganti karena item-itemnya sangat mudah. Demikian pula seluruh item yang menghasilkan data pada Tabel 6 karena tergolong sangat sukar. Akan tetapi, sekali lagi, batasan tersebut hanya untuk item instrumen beracuan norma agar dapat membedakan kelompok atas dan kelompok bawah.

Keandalan item instrumen untuk mengukur pencapaian hasil belajar dapat pula dilakukan melalui analisis faktor (Imam Ghozali, 2001: 132-140; Crocker & Algina, 1986: 295-296; Harman, 1976: 20-21; Fruchter, 1967: 47-50). Melalui analisis faktor, akan diketahui homogenitas dari seluruh item yang digunakan. Analisis faktor menggunakan prinsip reduksi. Jika *specific variance* dari tes i diberi simbol s_i^2 dan *error variance* diberi simbol e_i^2 *yang diasumsikan sama dengan 0,* maka *total variance* dapat dituangkan dalam rumus $h_i^2 + s_i^2 + e_i^2 = 1$ di mana *reliable variance* adalah $h_i^2 + s_i^2$ dengan rentang harga 0 sampai 1. Nilai ini tidak lain adalah reliabilitas dari skor tes, sehingga h_i^2 dapat dipertimbangkan sebagai *lower bond* dari keandalan/reliabilitas skor pada tes *i*. Dengan demikian, item mana yang memenuhi persyaratan dan yang tidak, dilihat berdasarkan kesamaan besarnya nilai komponen utama/*principle component* (KU/PC) yang diperoleh. Semakin jauh selisih nilai KU suatu item dengan item-item lainnya maka item tersebut semakin tidak andal.

Berikut disajikan hasil analisis ITEMAN, hasil analisis korelasi menggunakan program SPSS, juga hasil analisis faktor melalui program SPSS terhadap data pada Tabel 8.

Tabel 8. Hasil Tes dari 15 Testi/Peserta Didik yang Mengerjakan 10 Item Tes Prestasi untuk Materi Pokok YY.

Subjek	Item 1	item 2	item 3	Item 4	item 5	item 6	item 7	item 8	item 9	Item 10	Total
A	0	0	0	0	0	0	0	0	0	1	1
B	0	0	0	0	0	0	0	0	0	1	1
C	0	0	0	0	0	0	0	0	1	1	2
D	0	0	0	0	0	0	0	0	1	1	2
E	0	0	0	0	0	0	0	1	1	1	3
F	0	0	0	0	0	0	1	1	1	1	4
G	0	0	0	0	0	0	1	1	1	1	4
H	0	0	0	0	0	1	1	1	1	1	5
I	0	0	0	0	0	1	1	1	1	1	5
J	0	0	0	0	1	1	1	1	1	1	6

Subjek	Item 1	item 2	item 3	Item 4	item 5	item 6	item 7	item 8	item 9	Item 10	Total
K	0	0	0	1	1	1	1	1	1	1	7
L	0	0	1	1	1	1	1	1	1	1	8
M	0	0	1	1	1	1	1	1	1	1	8
N	0	1	1	1	1	1	1	1	1	1	9
O	1	1	1	1	1	1	1	1	1	0	9

Hasil analisis Program ITEMAN (tm) Version 3.00 dan pada kolom terakhir adalah hasil analisis korelasi Pearson menggunakan program SPSS adalah sebagai berikut.

Seq. No. Key	Scale Item	Prop. Correct.	Biser.	Point Biser.	Pearson Correlation^
1	0-1	0.067	0.771	0.400	.400
2	0-2	0.133	0.926	0.587	.587(*)
3	0-3	0.267	1.000	0.791	.791(**)
4	0-4	0.333	1.000	0.849	.849(**)
5	0-5	0.400	1.000	0.871	.871(**)
6	0-6	0.533	1.000	0.862	.862(**)
7	0-7	0.667	1.000	0.815	.815(**)
8	0-8	0.733	1.000	0.761	.761(**)
9	0-9	0.867	0.896	0.567	.567(*)
10	0-10	0.933	-0.771	-0.400	-.400

Scale Statistics:
N of Items: 10; N of Examinees: 15; Mean: 4.933; Variance: 7.396; Std. Dev.: 2.719; Skew: 0.070; Kurtosis: -1.311; Minimum: 1.000; Maximum: 9.000; Median: 5.000; Alpha: 0.859; SEM: 1.022: Mean P: 0.493; Mean Item-Tot.: 0.610; Mean Biserial: 0.782.

Dengan batas minimal koefisien korelasi 0,7, item 3, 4, 5, 6, 7, dan 8 memiliki daya pembeda yang baik. Jika data di atas dianalisis dengan analisis faktor berdasar varians maksimum hasilnya sebagai berikut.

Factor Analysis

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.251	52.514	52.514	5.251	52.514	52.514
2	2.106	21.058	73.572			
3	1.110	11.104	84.676			
4	.520	5.198	89.874			
5	.361	3.611	93.485			
6	.265	2.647	96.132			
7	.174	1.737	97.869			
8	.120	1.201	99.069			
9	.093	.931	100.000			
10	.000	.000	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix(a)

	Component
	1
item1	.598
item2	.699
item3	.819
item4	.850
item5	.850
item6	.810
item7	.744
item8	.687
item9	.502
item10	-.598

Extraction Method: Principal Component Analysis.
a 1 components extracted.

Rotated Component Matrix(a)

a Only one component was extracted. The solution cannot be rotated.

Hasil analisis faktor menunjukkan bahwa keragaman jawaban dapat diterangkan oleh item-item tes hanya sebesar 52.514% (% of variance 52.514), selebihnya tidak dapat dijelaskan oleh item-item tes tersebut. Dilihat dari besarnya nilai KU, hanya item 3, 4, 5, dan 6, yang memenuhi syarat karena yang paling homogen dan konsisten, kemudian diikuti oleh item 7, selanjutnya item 2 dan 8.

Secara sederhana, keefektifan suatu item tes beracuan norma untuk mengukur pencapaian hasil belajar juga dihitung berdasarkan indeks daya beda. Indeks daya beda sama dengan selisih antara banyaknya testi kelompok atas dan kelompok bawah dibagi dengan setengah jumlah testi kelompok atas dan kelompok bawah. Suatu item dinyatakan efektif membedakan kelompok atas dan kelompok bawah bila memiliki indeks daya beda ≥ 0.3 dan memiliki indeks kesukaran antara 0.3 sampai 0.7. Bahkan, khusus untuk item bentuk pilihan ganda setiap pengecoh (distraktor) pun harus ada yang memilih (terkecoh) minimal sebesar 5%. Jika suatu item bentuk pilihan ganda memiliki empat pilihan alternatif jawaban, maka paling sedikit ada 15% siswa yang terkecoh. Siswa yang terkecoh tersebut adalah siswa dari kelompok bawah. Dengan demikian, mengacu pada kurve normal dalam suatu kelas/populasi siswa yang telah belajar harus ada yang dinyatakan gagal.

Keefektifan suatu item tes untuk mengukur pencapaian hasil belajar beracuan kriteria didasarkan pada prinsip bahwa siswa dinyatakan benar-benar berhasil dalam belajar bila mencapai suatu kriteria tertentu. Dengan demikian, jika seluruh siswa dalam suatu kelas/sekolah semuanya benar-benar berhasil, maka ia dapat mengerjakan item tes yang diujikan. Oleh karena itu, Gronlund (1977: 115-116) mengajukan suatu prosedur analisis untuk menghitung keadalan item tes beracuan kriteria dengan menggunakan indeks sensitivitas item, yang menunjukkan keefektifan proses pembelajaran. Hal itu dapat diketahui jika dilakukan tes awal/*pretest* dan tes akhir/*posttest*.

Indeks sensitivitas item memiliki interval –1 sampai dengan 1. Indeks sensitivitas sebesar 1 menunjukkan bahwa suatu item gagal dikerjakan seluruh testi pada saat *pretest* dan berhasil dikerjakan seluruh testi pada saat *posttest*. Kalau daya beda menunjukkan perbedaan kemampuan antara kelompok atas dan kelompok bawah yang berhasil mengerjakan suatu tes, maka indeks sensitivitas menunjukkan perbedaan kemampuan saat testi sebagai peserta *posttest* dan saat testi sebagai peserta *pretest*.

Penutup

Berdasarkan hasil analisis di atas, dapat ditarik kesimpulan bahwa pemenuhan persyaratan kesahihan dan keandalan tes untuk mengukur pencapaian hasil belajar yang berkaitan dengan implementasi kurikulum berbasis kompetensi harus memenuhi kaidah tes beracuan kriteria. Pengujian keandalan tes dengan mencari koefisien korelasi, koefisien homogenitas, ataupun dengan mencari *standard error of measurement* perhitungannya mengacu pada normalitas distribusi, sehingga dapat menimbulkan kesesatan untuk memenuhi keandalan tes beracuan kriteria. Oleh karena itu, pengujiannya harus didasarkan pada persen konsistensi. Efektivitas item tes beracuan norma untuk memisahkan kelompok atas dan kelompok bawah, bukan untuk menunjukkan efektivitas pembelajaran. Dengan demikian, perhitungannya bukan mengandalkan pada besarnya proporsi jawaban benar sebagai indeks kesukaran, indeks daya beda atau nilai point biserial, melainkan pada besarnya indeks sensitivitas butir. Para peneliti maupun praktisi di lapangan yang ingin mengukur pencapaian hasil belajar yang berkaitan dengan kurikulum berbasis kompetensi hendaknya tunduk pada persyaratan instrumen beracuan kriteria.

Daftar Pustaka

Apache Software Foundation. 2003. *SPSS 12.0 for Window*.

Ary, D., Jacobs, L.Ch. & Razavieh, A. 1985. *Introduction to Research in Education*, 3-rd ed. New York: Holt, Rinehart, and Winston.

Assessment Systems Corporation. 1988. *MicroCAT (tm) Testing System: Item Parameter Estimation Program -- ASCAL (tm) Version 3.20.*

-------. 1988. *MicroCAT (tm) Testing System: Item and Test Analysis Program -- ITEMAN (tm) Version 3.00*

Croker, L. & Algina, J. 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.

Direktorat P2TK dan KPT. 2005. *Pedoman Sistem Asesmen Berbasis Kompetensi*. Jakarta: Direktorat Jenderal Pendididian Tinggi, Departemen Pendidikan Nasional.

Direktorat PLP. 2004. *Pedoman Umum Sistem Penilaian Kurikulum Berbasis Kompetensi*. Jakarta: Direktorat PLP, Ditjen Dikdasmen, Depdiknas.

Direktorat PMU. 1999. *Pengelolaan Pengujian Bagi Guru Mata pelajaran*. Jakarta: Direktorat PMU, Ditjen Dikdasmen, Depdiknas.

-------. 2004. *Pedoman Umum Sistem Penilaian Kurikulum Berbasis Kompetensi*. Jakarta: Direktorat PMU, Ditjen Dikdasmen, Depdiknas.

Frisbie, D.A. 2005. "Measurement 101: Some Fundamentals Revisited'. *Educational Measurement Issues and Practice. Vol. 24. No. 3, pp. 21-28.*

Fruchter, B. 1967. *Introduction Factor Analysis*. East-West Student Edition. Princeton: Affiliated East-West Press P, Ltd.

Ghozali, Iman. 2001. *Aplikasi Analisis Multivariate dengan Program SPSS*. Semarang: Badan Penerbit Universitas Diponegoro.

Gronlund, N.E. & Linn. R.L. 1990. *Measurement and Evaluation in Teaching. 6-th ed.* New York: Macmillan Publishing Company.

-------. 1977. *Constructing Achievement Test.* Englewood Clifft. N.J.: Prentice-Hall. Inc.

Gronlund, N.E. 1998. *Assessment of Student Achievement.* Boston: Allyn and Bacon.

Hagul, Peter. 1982. "Reliabilitas dan Validitas". Dalam: Masri Singarimbun. 1982. *Metode Penelitian Survei.* Jakarta: LP3ES.

Harman, H.H. 1976. *Modern Factor Analysis.* 3rd ed. Chicago: The University of Chicago.

Kumaidi. 2004. "Interpretasi Koefisien Korelasi Skor-Butir dengan Skor Total Uji Kebermaknaan Koefisien KR-20 dalam Penelitian Pendidikan dan Psikologi". *Jurnal Ilmu Pendidikan, Juni 2004: Jilid II, Nomor 2.* h.107-114.

Naga, Dali S. 2004. "Ketidaktepatan Penggunaan Validitas Butir dan Koefisien Reliabilitas dalam Penelitian Pendidikan dan Psikologi. *Jurnal Ilmu Pendidikan, Juni 2004: Jilid II, Nomor 2.* h. 99-106.