



EFEK *TEST LENGTH* DAN *NUMBER OF EXAMINEES* TERHADAP TEST RELIABILITY

Samritin

Universitas Muhammadiyah Buton

samritin75@gmail.com

Abstrak

Reliabilitas merupakan salah satu properti tes yang berkaitan dengan kesalahan atau konsistensi pengukuran. Kesalahan pengukuran dalam teori tes klasik dihasilkan perbedaan antara skor tampak dengan suatu skor yang sesungguhnya-tidak tampak (skor murni). Reliabilitas suatu tes terlihat dari koefisien reliabilitasnya. Koefisien reliabilitas dapat diestimasi dengan beberapa cara, salah satunya tergantung pada jenis tes. Hasil estimasi koefisien reliabilitas dapat dipengaruhi oleh beberapa faktor, antara lain panjang tes dan heterogenitas kemampuan peserta tes. Hasil penelitian simulasi ini menunjukkan bahwa semakin panjang suatu tes, koefisien reliabilitasnya semakin tinggi. Sebaliknya semakin pendek suatu tes, koefisien reliabilitasnya semakin rendah. Ditemukan juga bahwa tes yang memiliki tingkat kesukaran tinggi, koefisien reliabilitasnya lebih tinggi dibandingkan dengan koefisien reliabilitas tes yang lebih pendek. Dalam penelitian ini tidak ditemukan pola perubahan koefisien reliabilitas akibat peningkatan atau pengurangan banyak peserta tes.

Kata kunci: panjang tes, banyak peserta tes, tingkat kesukaran, reliabilitas.

Abstract

Reliability is one of the test properties that related to the error and measurement consistency. The measurement error in the test of classical theory is produced by the difference between visible score with the invisible score (pure score). The reliability of a test can be seen from its coefficient reliability, It can be estimated by several ways, one of them is depend on the kind of the test. The result of coefficient reliability estimation can be influenced by some factors, such as the length of the test and the heterogeneity of participants' ability in having test (examinees). The result of this simulation study showed that the longer of a test, the higher its reliability coefficient. Conversely, the shorter of a test, the lower its reliability coefficient. In this study, there was not founded the change pattern of reliability coefficient as the result of the increasing or the reduction of the number of examinees.

Key words: *the test length, the number of examinees, the level of difficulties, reliability.*

Pendahuluan

Dalam pengukuran di bidang pendidikan, tes merupakan instrumen penting yang sering digunakan. Salah satu kriteria tes yang baik dapat dilihat dari reliabilitasnya. Reliabilitas berkaitan dengan kesalahan pengukuran atau konsistensi pengukuran. Reliabilitas

merupakan properti dari sekumpulan skor yang menggambarkan konsistensi atau bebasnya hasil pengukuran dari kesalahan (Ebel & Frisbie, 1991, p.76). Reliabilitas menunjuk pada derajat kebebasan hasil pengukuran dari kesalahan (AERA, APA, & NCME, 1985, p.19). Reliabilitas mencerminkan perbedaan dalam skor amatan responden konsisten dengan perbedaan dalam skor murninya (Furr & Bacharach, 2008, p. 82).

Dalam teori tes, reliabilitas merupakan properti tes yang berasal dari skor amatan (*observed score*), skor murni (*true score*), dan kesalahan pengukuran (*measurement error*). Tingkat reliabilitas tes tergantung pada dua hal yaitu (1) perbedaan skor tes dikaitkan dengan perbedaan nyata antar-atau intra individual dan (b) sejauh mana perbedaan tersebut sebagai fungsi dari kesalahan pengukuran (Furr & Bacharach, 2008, p.84). Hal ini sesuai dengan salah satu asumsi dalam teori tes klasik, yaitu skor amatan seseorang adalah fungsi dari skor murni ditambah skor kesalahan.

Allen & Yen (1979, p.72) mengemukakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor murninya. Selanjutnya dijelaskan pula bahwa koefisien reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Pendapat ini menunjukkan bahwa suatu tes memiliki reliabilitas tinggi jika hasil pengukuran mendekati keadaan yang sebenarnya dari peserta tes.

Kesalahan pengukuran dalam teori tes klasik dihasilkan perbedaan antara skor manapun dengan suatu skor yang disebut skor murni (*true score*). Skor murni adalah hasil hasil yang diperoleh jika semua item dalam domain itu dikelola dengan benar. Reliabilitas secara statistik dicerminkan oleh rasio varians skor murni dengan varians skor amatan. Reliabilitas secara teoretis dinyatakan sebagai korelasi kuadrat antara skor murni dengan skor amatan. Menurut Roid & Haladyna (1982, p.43-47) dalam praktiknya, indeks atau koefisien reliabilitas akan menjadi maksimal apabila sekumpulan item yang sama diberikan kepada sekelompok peserta tes yang luas cakupannya. Hasil estimasi reliabilitas bergantung pada banyaknya item karena semakin banyak item yang diberikan akan semakin tinggi pula estimasi reliabilitasnya.

Reliabilitas suatu tes dapat dilihat dari koefisien reliabilitas (*reliability coefficient*) tes tersebut. Koefisien paling rendah yang harus dimiliki suatu tes adalah 0,7 (Nunnally, 1981, p.245; Urbina, 2004, p.137). Mardapi (2008, p.38) reliabilitas instrumen baik tes maupun

nontes dikategorikan menjadi tiga kelompok yaitu konsistensi internal, stabilitas, dan antarpemilai. Sesuai dengan klasifikasinya, cara mengestimasi reliabilitas dapat dikategorikan dalam tiga metode yaitu konsistensi internal, stabilitas, dan antarpemilai.

Lebih lanjut Mardapi menjelaskan bahwa metode estimasi reliabilitas tes yang hanya memerlukan satu kali penyajian tes dikenal dengan nama *single-trial administration*. Metode ini dapat menghindari masalah-masalah yang timbul akibat penyajian tes berulang. Estimasi reliabilitas berdasarkan konsistensi internal dapat dilakukan dengan menggunakan pendekatan paralel klasik, tau-equivalen, dan kongenerik. Pada pendekatan paralel klasik tes dibagi dua, bisa gasal dan genap atau bagian awal dan bagian akhir. Syarat yang harus dipenuhi penggunaan metode ini adalah rerata, varians, dan materi harus sama untuk kedua belahan. Untuk menghitung besarnya koefisien reliabilitas dapat digunakan formula Spearman-Brown. Untuk menghitung koefisien reliabilitas dengan pendekatan tau-equivalent menggunakan formula: (1) Rulon—tanpa asumsi bahwa kedua belahan mempunyai varians sama, yang perlu diperhitungkan adalah varians kesalahan, (2) koefisien Alpha untuk tes yang memiliki varians skor dari kedua belahan tes berbeda atau tidak cukup alasan untuk menganggap bahwa kedua belahan paralel, (3) KR-20 dan KR-21 untuk tes-tes yang diberi skor dikotomis, (4) koefisien reliabilitas menurut Feldt untuk tes yang bila dibagi dua menghasilkan belahan tidak sama panjang tetapi masih menghasilkan bagian yang isinya homogen. Pendekatan kongenerik digunakan bila varians kedua belahan tidak sama, demikian pula reratanya. Salah satu formula yang dapat digunakan adalah formula Kristoff untuk tes belah tiga.

Estimasi reliabilitas tes juga dapat dilakukan dengan metode stabilitas yang didasarkan pada teknik tes-retes. Metode ini menggunakan perhitungan korelasi *product moment* atau perhitungan analisis varians menurut Hoyt. Namun harus diingat bahwa pelaksanaan tes-retes akan melahirkan efek *testing (carry-over effect)*. Pelaksanaan tes dan retes yang terlalu singkat memungkinkan peserta tes mengingat jawaban pada saat tes sehingga jawaban pada tes awal akan diulangnya pada saat tes ulang. Pelaksanaan tes dan retes yang relatif lama dapat melahirkan perubahan suasana hati peserta tes atau bertambahnya informasi yang diperoleh peserta tes dalam hal yang diukur. Apalagi bila peserta tes mengingat butir-butir tes yang pernah ditempuhnya. Ia akan mencari jawaban tes tersebut sebelum pelaksanaan tes ulang.

Untuk mengestimasi reliabilitas tes yang penilainya lebih dari satu penilai dapat dilakukan menggunakan reliabilitas antarpenilai. Estimasi reliabilitas ini dilakukan pada tes-tes yang berbentuk uraian. Reliabilitas antarpenilai dapat dihitung menggunakan formula Kappa dari Cohen. Suatu tes dikatakan baik apabila memiliki koefisien reliabilitas paling rendah 0,7 (Nunnally, 1981, p.245; Urbina, 2004, p.137).

Koefisien reliabilitas dapat dipengaruhi oleh beberapa faktor, antara lain panjang tes dan heterogenitas kemampuan peserta tes. Crocker dan Algina (1976) menjelaskan bila banyak instrumen ditingkatkan dua kali lipat dari banyak semula, maka reliabilitasnya diharapkan meningkat sebesar 0,15 dari reliabilitas semula, dan bila ingin meningkatkan lebih dari itu, peneliti perlu memperbanyak jumlah butir tiga kali jumlah semula.

Berdasarkan uraian diatas, dapat dikatakan bahwa secara teoritis banyaknya butir dapat mempengaruhi peningkatan reliabilitas tes. Bagaimana halnya jika *number of examinees* berubah? Apakah perubahan *number of examinees* mempengaruhi reliabilitas? Penelitian simulasi ini bertujuan (1) untuk mendeskripsikan efek *test length* (panjang tes) terhadap reliabilitas tes, dan (2) pengaruh *number of examinees* (banyak peserta tes) terhadap reliabilitas.

Metode Penelitian

Jenis Penelitian

Penelitian ini merupakan penelitian simulasi. Penelitian ini menggunakan data rekaan yang dibangkitkan. Pembangkitan data dilakukan menggunakan program aplikasi komputer TAP. Penelitian simulasi ini mengkaji tentang pengaruh panjang tes (*test length*) terhadap reliabilitas tes (*test reliability*), dan pengaruh banyak peserta tes (*number of examinee*) terhadap reliabilitas tes. Dalam simulasi ini digunakan tiga jenis panjang tes yaitu 10 butir, 20 butir, dan 40 butir. Pemilihan ketiga ukuran tersebut dianggap sudah mewakili dua ukuran tes yakni tes pendek dan tes panjang. Tes dengan panjang 10 butir dan 20 butir mewakili tes tes pendek, dan tes dengan panjang 40 butir mewakili tes panjang.

Jenis tes yang dikaji dalam penelitian ini berbentuk pilihan ganda (*multiple choice*) dengan pilihan jawaban sebanyak 4 alternatif. Pembangkitan data digunakan 3 (tiga) tingkatan kesukaran, (1) Sangat Mudah-*Very Easy Exam* (VEE), (2) Mudah-*Moderately Difícult Exam* (MDE) dan (3) Sangat Sukar - *Very Difícult Exam* (VDE). Setiap tingkatan kesukaran (kasus), data dibangkitkan sebanyak 50 replikasi. Untuk melakukan replikasi setiap kasus,

Seed for Data Generation, digunakan kelompok angka random: 346323, 346324, 346355, 346375, dan 34677.

Pada penelitian ini, untuk melihat pengaruh ukuran sampel terhadap reliabilitas tes digunakan 3 (tiga) ukuran (*number of examinee*), yaitu: 32 orang, 212 orang dan 8480 orang. Penentuan ukuran sampel minimal didasarkan pada rata-rata jumlah siswa setiap kelas Sekolah Dasar (SD). Ukuran sampel minimal tersebut dianggap sudah memadai untuk dilakukan estimasi efek.

Langkah-langkah kegiatan yang dilaku-kan dalam penelitian ini adalah (1) menentukan variabel yang menjadi fokus kajian yaitu panjang tes (*test length-TL*) dan banyak peserta tes (*number of examinee-NE*), (2) mendesain penelitian dengan menentukan banyak banyaknya alternatif jawaban, dan replikasi, (3) memilih program aplikasi pembangkit data, (4) membangkitkan data, (5) menganalisis data. Analisis data dalam penelitian ini dilakukan dengan menghitung reliabilitas tes setiap replikasi kemudian dihitung rata-ratanya. Perhitungan reliabilitas digunakan formula KR-20..

Hasil Penelitian dan Pembahasan

Secara keseluruhan data yang dihasilkan melalui pembangkitan sebanyak 1350 replikasi. Replikasi data dilakukan sebanyak 50 kali pada setiap kasus yang telah ditentukan. Hasil Estimasi koefisien reliabilitas menggunakan formula KR-20 beserta *standard error measurement* (SEM) pada setiap kasus secara berturut-turut disajikan dalam tabel 1 dan Tabel 2.

Tabel 1. Reliabilitas Tes

NE	TK	TL		
		10	20	40
32	VEE	0,524	0,544	0,699
	MDE	0,615	0,686	0,832
	VDE	0,640	0,769	0,863
212	VEE	0,401	0,585	0,727
	MDE	0,545	0,725	0,832
	VDE	0,608	0,751	0,851
8480	VEE	0,425	0,577	0,719
	MDE	0,567	0,715	0,832
	VDE	0,589	0,738	0,852

Tabel 2. SEM Tes

NE	DL	TL		
		10	20	40
32	VEE	0.942	1.354	1.852
	MDE	1.343	1.866	2.637
	VDE	0.638	1.900	2.784
212	VEE	0.895	1.355	1.868
	MDE	1.265	1.860	2.545
	VDE	1.367	1.937	2.726
8480	VEE	0.992	1.307	1.768
	MDE	1.328	1.843	2.593
	VDE	1.381	1.957	2.840

Pada tabel 1 dan tabel 2 terlihat bahwa reliabilitas tertinggi yakni 0,863 terjadi pada tes terpanjang dengan tingkat kesukaran VDE dan banyak peserta tes rendah dengan SEM sebesar 2,784. Sementara itu, reliabilitas terendah yakni 0,401 terjadi pada tes terpendek dengan tingkat kesukaran VEE dan banyak peserta tes 212 orang dengan SEM sebesar 0,895. Tes yang panjangnya 10 butir memiliki koefisien reliabilitas yang sangat rendah. Hal ini terjadi pada hampir semua kasus.

Pada tingkat kesukaran VEE perubahan TR terendah terjadi ketika TL ditingkatkan dari 10 menjadi 20 yaitu sebesar 0,020 yang mana NE sebanyak 32. Pada tingkat kesukaran MDE perubahan TR terendah terjadi ketika TL ditingkatkan dari 10 menjadi 20 yaitu sebesar 0,071 yang mana NE sebanyak 32. Pada tingkat kesukaran MDE perubahan TR terendah terjadi ketika TL ditingkatkan dari 20 menjadi 40 yaitu sebesar 0,094 yang mana NE sebanyak 32.

Pada tingkat kesukaran VEE perubahan TR tertinggi terjadi ketika TL ditingkatkan dari 10 menjadi 20 yaitu sebesar 0,184 yang mana NE sebanyak 212. Pada tingkat kesukaran MDE perubahan TR tertinggi terjadi ketika TL ditingkatkan dari 10 menjadi 20 yaitu sebesar 0,180 yang mana NE sebanyak 32. Pada tingkat kesukaran VDE perubahan TR tertinggi terjadi ketika TL ditingkatkan dari 10 menjadi 20 yaitu sebesar 0,149 yang mana NE sebanyak 8480.

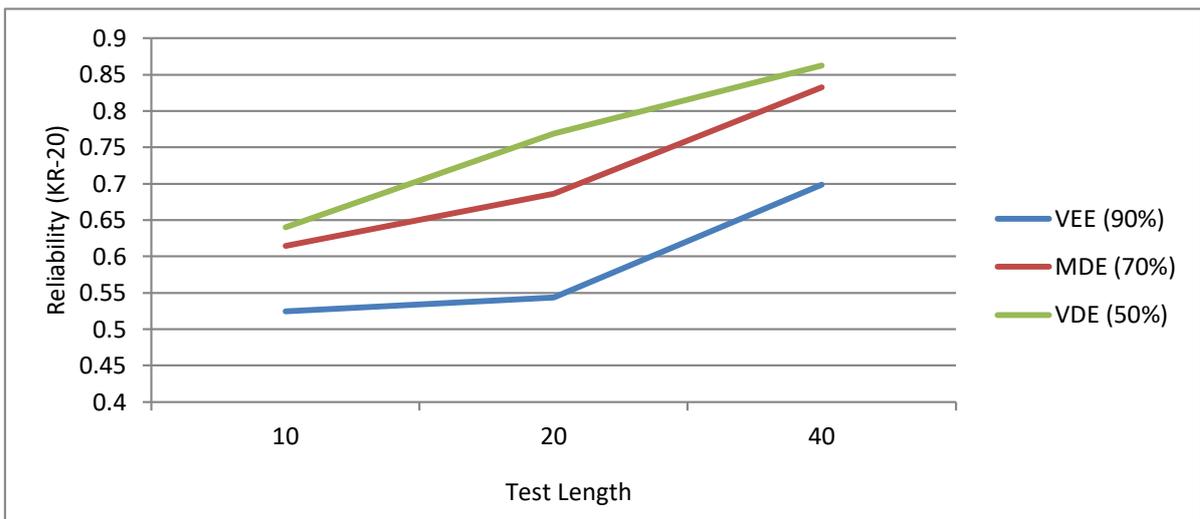
Rerata koefisien reliabilitas tes terpendek dengan tingkat kesukaran sangat mudah hanya 0,450. Rerata koefisien reliabilitas tes terpendek dengan tingkat kesukaran sedang 0,576. Rerata koefisien reliabilitas tes terpendek dengan tingkat kesukaran yang sangat tinggi 0,612. Rerata koefisien reliabilitas tes terpanjang pada tingkat tingkat kesukaran sangat mudah

mencapai 0,715. Rerata koefisien reliabilitas tes terpanjang dengan tingkat kesukaran sedang sebesar 0,832. Rerata koefisien reliabilitas tes terpanjang dengan tingkat kesukaran yang sangat tinggi mencapai 0,855. Rerata koefisien reliabilitas tes yang panjangnya 20 butir pada semua dengan tingkat kesukaran berada di antara koefisien reliabilitas tes terpanjang dan tes terpendek.

Secara keseluruhan rerata koefisien reliabilitas tes yang panjangnya 10 butir lebih kecil dibandingkan dengan koefisien reliabilitas tes yang panjangnya 20 butir. Rerata koefisien reliabilitas tes yang panjangnya 40 butir lebih tinggi dibandingkan koefisien reliabilitas tes yang panjangnya 20 butir. Hal ini berarti bahwa semakin panjang butir tes koefisien reliabilitas semakin tinggi.

Pengaruh panjang tes terhadap reliabilitas tes

Pengaruh panjang tes terhadap reliabilitas dengan tingkat kesukaran dan banyaknya peserta yang berbeda-beda disajikan pada gambar 1, gambar 2, dan gambar 3.



Gambar 1. Efek TL terhadap TR, NE = 32

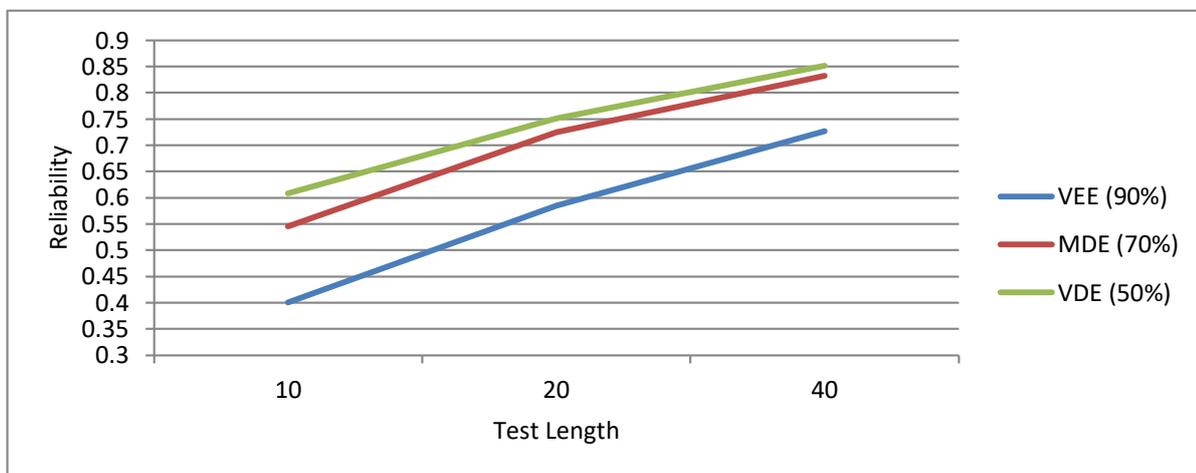
Grafik di atas menunjukkan bahwa jika NE = 32, tes yang panjangnya 10 butir memiliki reliabilitas kurang dari 0,70. Kondisi ini terjadi di semua tingkat kesukaran. Selain itu gambar di atas juga menunjukkan bahwa di setiap tingkat kesukaran tes dengan TL = 10 memiliki koefisien yang lebih rendah dibandingkan koefisien reliabilitas tes dengan TL = 20. Selanjutnya, di setiap tingkat kesukaran koefisien reliabilitas tes dengan TL = 20 lebih rendah dibandingkan dengan koefisien reliabilitas tes dengan TL = 40.

Tes yang panjangnya 20 butir koefisien reliabilitasnya hampir mencapai 0,8 atau berada di atas 0,70 pada tingkat kesukaran VDE, sedangkan pada tingkat kesukaran VEE dan

MDE koefisien reliabilitas tidak mencapai 0,70. Koefisien reliabilitas terendah pada tes yang panjangnya 20 butir terjadi pada tingkat kesukaran VEE yang tidak mencapai 0,55 atau koefisiennya hampir sama dengan tes yang panjangnya 10 butir pada tingkat kesukaran yang sama.

Tes yang panjangnya 40 butir (TL = 40) pada NE = 32 memiliki koefisien reliabilitas sebesar 0,70 pada tingkat kesukaran VEE, sedangkan pada tingkat kesukaran MDE dan VDE, koefisien reliabilitas berada di antara 0,8 dan 0,9. Koefisien reliabilitas tertinggi mencapai 0,86 yang terjadi pada tingkat VDE, sedangkan pada tingkat MDE mencapai 0,83. Koefisien reliabilitas terendah tingkat VEE pada TL = 40 lebih tinggi dibandingkan dengan koefisien reliabilitas tingkat MDE pada TL = 20. Keduanya lebih tinggi dibandingkan koefisien reliabilitas tes pada TL = 10 di semua tingkat kesukaran. Gambar 2 menunjukkan bahwa jika NE = 212, tes yang panjangnya 10 butir (TL = 10) memiliki reliabilitas kurang dari 0,70. Kondisi ini terjadi di semua tingkat kesukaran (VEE, MDE, dan VDE).

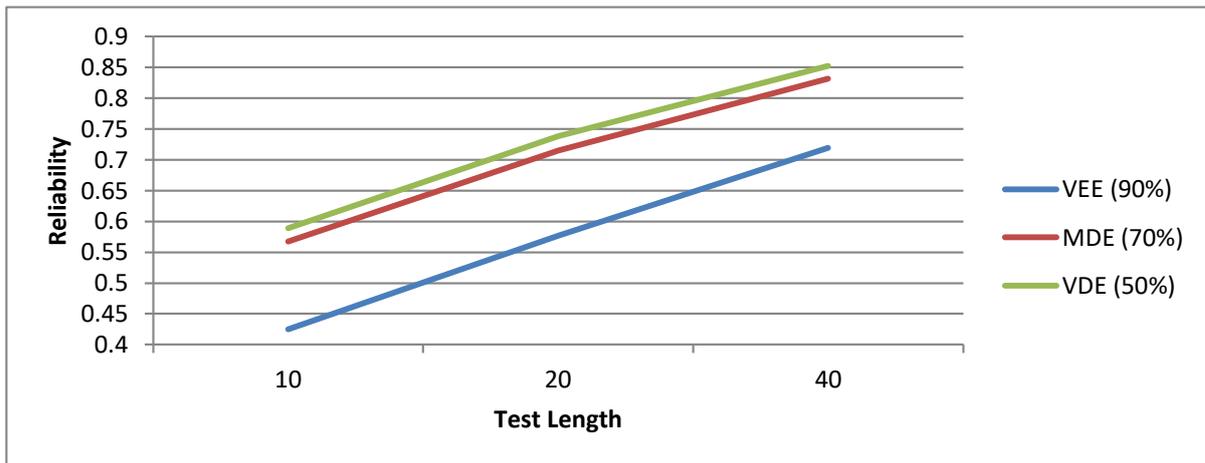
Gambar di atas juga menunjukkan bahwa di setiap tingkat kesukaran tes dengan TL = 10 memiliki koefisien yang lebih rendah dibandingkan koefisien reliabilitas tes dengan TL = 20. Selanjutnya, di setiap tingkat kesukaran koefisien reliabilitas tes dengan TL = 20 lebih rendah dibandingkan dengan koefisien reliabilitas tes dengan TL = 40.



Gambar 2. Grafik Efek TL terhadap TR, NE = 212

Tes yang panjangnya 20 butir (TL = 20) pada NE = 212 koefisien reliabilitasnya berada di atas 0,70 pada tingkat kesukaran MDE dan VDE, sedangkan pada tingkat kesukaran VEE koefisien reliabilitas hanya mendekati 0,60 atau tidak mencapai 0,70. Koefisien reliabilitas tertinggi pada tes yang panjangnya 20 butir terjadi pada tingkat kesukaran VDE yang mendekati 0,75.

Tes yang panjangnya 40 butir (TL = 40) pada NE = 212 memiliki koefisien reliabilitas di atas 0,70 pada semua level tingkat kesukaran. Koefisien reliabilitas tertinggi mencapai 0,85 yang terjadi pada tingkat VDE, sedangkan koefisien terendah mencapai 0,73 yang terjadi pada tingkat VEE. Koefisien reliabilitas terendah tingkat VEE pada TL = 40 sama dengan koefisien reliabilitas tingkat MDE pada TL = 20. Keduanya lebih tinggi dibandingkan koefisien reliabilitas tes pada TL = 10 di semua tingkat kesukaran.



Gambar 3. Efek TL terhadap TR, NE = 8480

Gambar 3 menunjukkan bahwa jika NE = 212, tes yang panjangnya 10 butir (TL = 10) memiliki reliabilitas kurang dari 0,70. Kondisi ini terjadi di semua tingkat kesukaran (VEE, MDE, dan VDE). Gambar di atas juga menunjukkan bahwa di setiap tingkat kesukaran tes dengan TL = 10 memiliki koefisien yang lebih rendah dibandingkan koefisien reliabilitas tes dengan TL = 20. Selanjutnya, di setiap tingkat kesukaran koefisien reliabilitas tes dengan TL = 20 lebih rendah dibandingkan dengan koefisien reliabilitas tes dengan TL = 40.

Tes yang panjangnya 20 butir (TL = 20) pada NE = 212 koefisien reliabilitasnya berada di atas 0,70 pada tingkat kesukaran MDE dan VDE, sedangkan pada tingkat kesukaran VEE koefisien reliabilitas hanya mendekati 0,60 atau tidak mencapai 0,70. Koefisien reliabilitas tertinggi pada tes yang panjangnya 20 butir terjadi pada tingkat kesukaran VDE yang mendekati 0,75.

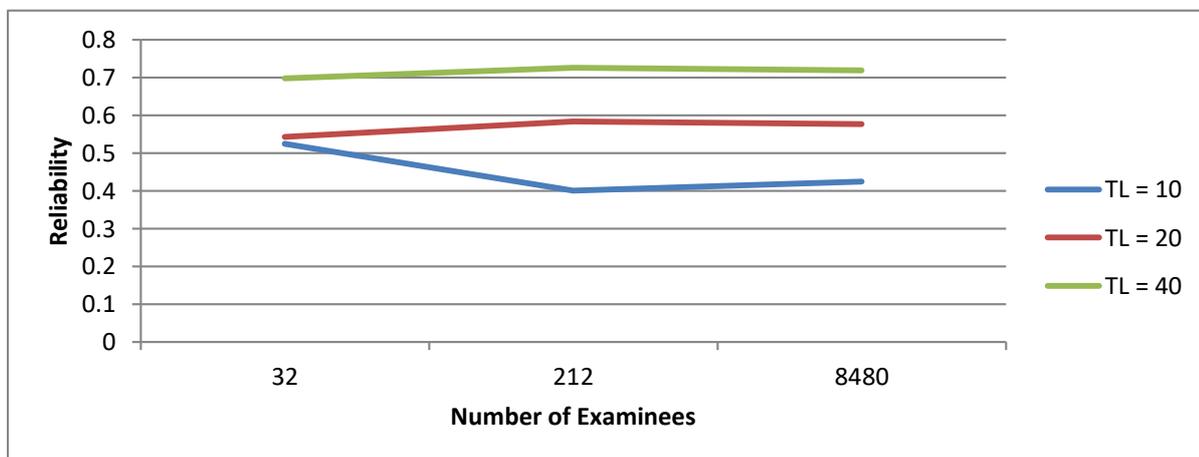
Tes yang panjangnya 40 butir (TL = 40) pada NE = 212 memiliki koefisien reliabilitas di atas 0,70 pada semua level tingkat kesukaran. Koefisien reliabilitas tertinggi mencapai 0,85 yang terjadi pada tingkat VDE, sedangkan koefisien terendah mencapai 0,73 yang terjadi pada tingkat VEE. Koefisien reliabilitas terendah tingkat VEE pada TL = 40 sama dengan

koefisien reliabilitas tingkat MDE pada TL = 20. Keduanya lebih tinggi dibandingkan koefisien reliabilitas tes pada TL = 10 di semua tingkat kesukaran.

Uraian di atas menunjukkan bahwa *test length* (panjang) tes berpengaruh pada hasil estimasi reliabilitas. Pada gambar 1 - gambar 3 memperlihatkan dengan jelas bahwa tes semakin panjang tes, hasil estimasi reliabilitas juga semakin tinggi. Sebaliknya semakin pendek tes, hasil estimasi reliabilitas juga semakin rendah. Kondisi ini terjadi pada semua tingkat kesukaran.

Gambar 1 – 3 menunjukkan bahwa koefisien reliabilitas juga dipengaruhi oleh tingkat kesukaran. Pada semua ukuran panjang tes, semakin sukar tes, reliabilitas tes juga semakin tinggi. Tes yang memiliki tingkat kesukaran tinggi menghasilkan variasi skor yang sangat tinggi. Hal ini berbeda dengan tes yang memiliki tingkat kesukaran rendah. Semakin mudah butir-butir tes, maka variasi skor semakin rendah. Dalam penelitian ini, koefisien reliabilitas paling tinggi ditemukan pada tes yang panjangnya 40 butir dengan tingkat kesukaran tinggi. Koefisien reliabilitas yang sangat rendah ditemukan pada tes yang panjangnya 10 butir dengan tingkat kesukaran rendah.

Efek *Number of Examinees* terhadap *Test Reliability*



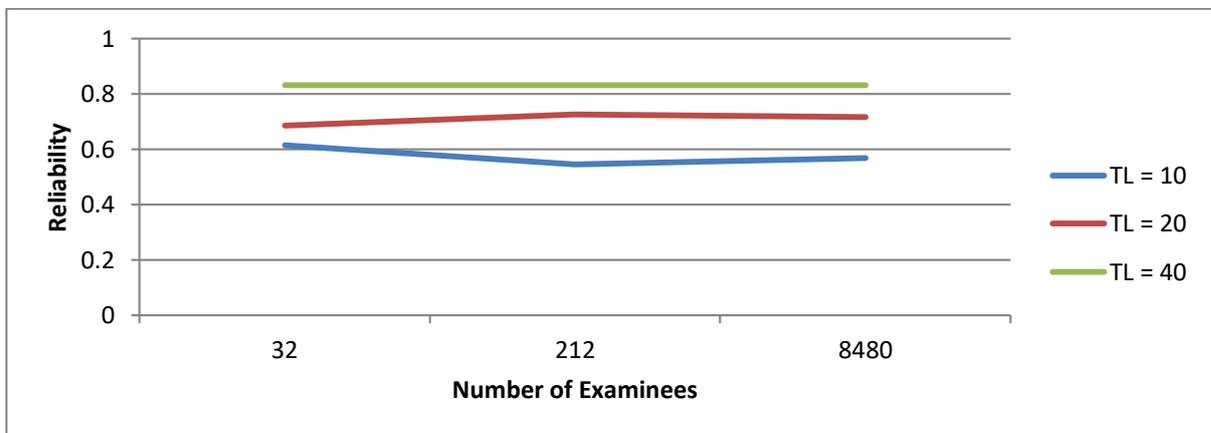
Gambar 4. Pengaruh NE terhadap TR, Tingkat Kesukaran VEE

Pengaruh peningkatan *number of examinees* (NE) terhadap reliabilitas (TR) digambarkan pada grafik 4. Grafik tersebut menunjukkan variasi koefisien reliabilitas tes yang memiliki tingkat kesukaran VEE pada tiga ukuran peserta tes (*number of examinee*). Pada TL = 10, koefisien KR tertinggi terjadi NE = 32. Ketika NE ditingkatkan, TR mengalami penurunan

pada NE = 212. Ketika NE ditingkatkan lagi menjadi 8480, TR meningkat kembali tetapi nilainya lebih rendah dibandingkan dengan TR pada NE = 32. Hal ini berarti bahwa pada tes yang panjangnya 10 butir, hasil estimasi TR akan tinggi jika menggunakan peserta tes 32.

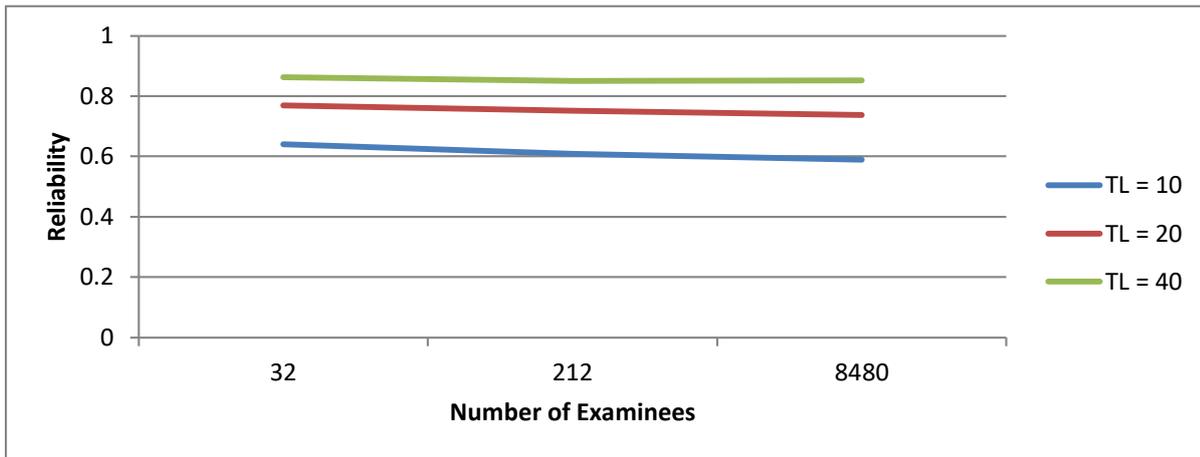
Efek peningkatan banyaknya peserta tes pada tes yang memiliki TL = 20 dan TL = 40, berbeda dengan kasus pada tes yang memiliki TL = 10. Pada tes dengan TL = 20 dan TL = 40, peningkatan NE mengakibatkan koefisien meningkat kemudian menurun kembali. Peningkatan NE menjadi 212 mengakibatkan meningkatnya TR. Peningkatan NE menjadi 8480 mengakibatkan berkurangnya TR mendekati TR pada NE = 32 tetapi nilainya masih lebih besar dibandingkan dengan TR pada NE = 32.

Gambar 5 menunjukkan variasi koefisien reliabilitas tes yang memiliki tingkat kesukaran MDE (sedang). Tes dengan TL = 10 memiliki koefisien reliabilitas sangat rendah pada semua NE. Grafik perubahan TR menunjukkan bahwa peningkatan NE menjadi 212 pada TL = 10 mengakibatkan TR berkurang dan TR meningkat kembali setelah NE ditingkatkan menjadi 8480. Kecenderungan perubahan TR tersebut sama dengan kecenderungan perubahan TR pada gambar 4.



Gambar 5. Efek NE terhadap TR, Tingkat Kesukaran MDE

Gambar 5 juga menunjukkan bahwa pada tes yang memiliki TL = 20 peningkatan NE menjadi 212, mengakibatkan peningkatan TR. Peningkatan NE menjadi 8480 mengakibatkan berkurangnya TR mendekati TR pada NE = 32 tetapi nilainya masih lebih besar dibandingkan dengan TR pada NE = 32. Sementara itu, gambar lima tidak memperlihatkan adanya pengaruh peningkatan NE terhadap TR pada tes yang memiliki tingkat kesukaran VEE.



Gambar 6. Pengaruh NE terhadap TR, Tingkat Kesukaran VDE

Gambar 6 memperlihatkan reliabilitas pada tes yang memiliki tingkat kesukaran VDE (tinggi). Pada gambar tampak bahwa NE = 32 memiliki TR yang lebih tinggi dibandingkan dengan TR pada NE yang lebih banyak. Pada tes yang memiliki TL = 10, peningkatan NE menjadi 212 maka TR berkurang. Peningkatan NE menjadi 8480 reliabilitas tes meningkat tetapi nilainya masih lebih kecil dibandingkan dengan reliabilitas tes pada NE = 32, bahkan masih lebih kecil dibandingkan pada NE = 212.

Uraian di atas menunjukkan bahwa perubahan koefisien reliabilitas tes akibat peningkatan jumlah peserta tes tidak teratur. Ketidakteraturan tersebut terjadi pada semua panjang tes. Hal ini dapat dicermati pada gambar 4 – gambar 6. Hasil penelitian ini belum memberikan informasi tentang banyaknya peserta tes baik untuk menghasilkan estimasi reliabilitas tinggi. Dengan demikian hasil penelitian ini belum memberikan konfirmasi yang sesuai dengan pernyataan Roid & Haladyna (1982: 43-47). Haladyna mengemukakan bahwa dalam praktiknya, indeks atau koefisien reliabilitas akan menjadi maksimal apabila sekumpulan item yang sama diberikan kepada sekelompok peserta tes yang luas cakupannya.

Simpulan

Berdasarkan uraian di atas dapat disimpulkan bahwa semakin panjang tes maka reliabilitas tes semakin tinggi. Hal ini ditunjukkan oleh grafik reliabilitas tes pada setiap ukuran sampel dan tingkat kesukaran. Efek peningkatan *number of examinees* terhadap reliabilitas tes tampak tidak teratur, bahkan *number of examinees* yang kecil memberikan

hasil estimasi *tes reliability* yang lebih tinggi dibandingkan dengan hasil estimasi menggunakan *number of examinees* lainnya.

Daftar Pustaka

- AERA, APA, & NCME. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association Inc.
- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Ebel, R.L. & Frisbie, D.A. (1991). *Essential of educational measurement* (5th ed.). New Delhi: Prentice-Hall of of India.
- Furr, R.M. & Bacharach, V.R. (2008). *Psychometrics*. Los Angeles: Sage Publication.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia Press.
- Nunnally, J.C. (1981). *Psychometric theory*. New Delhi: McGraw Hill.
- Roid, G.H. & Haladyna, T.M. (1982). *A Technology for test-item writing*. Orlando: Academic Press Inc.
- Urbina, S. (2004). *Essential of psychologycal testing*. New Jersey: Johnn Wiley & Sons, Inc.