

MENINGKATKAN ESTIMASI RELIABILITAS INSTRUMEN PILIHAN GANDA DENGAN MEMPERHATIKAN PROBABILITAS *GUESSING*

DIAH OGA NUSANTARI

Program Studi Pendidikan Matematika, Fakultas Teknik, Matematika dan IPA
Universitas Indraprasta PGRI

Abstrak. Koefisien reliabilitas adalah angka yang menunjukkan kestabilan dalam mengukur kemampuan siswa. Semakin tinggi angka hasil estimasi perhitungan koefisien reliabilitas instrumen menandakan semakin tinggi pula kepercayaan hasil tes siswa yang diukur dengan menggunakan instrumen tersebut. Derajat kepercayaan atas kehandalan sebuah instrumen dinyatakan dalam bentuk koefisien reliabilitas. Pada soal pilihan ganda, formula estimasi koefisien reliabilitas yang biasa digunakan adalah formula KR-20. Dengan menggunakan KR 20, estimasi reliabilitas dihitung berdasarkan jumlah skor siswa yang menjawab benar atau metoda *correct score*. Penelitian ini bertujuan memberikan alternatif formula estimasi reliabilitas dari penskoran *correct score*, selain menggunakan KR-20, yaitu dengan estimasi reliabilitas yang mempertimbangkan probabilitas *guessing*. Dengan mempertimbangkan faktor *guessing* maka koefisien reliabilitas instrumen akan meningkat.

Kata kunci: *Instrumen, pilihan ganda, Reliabilitas, probabilitas Guessing*

Abstract. Reliability coefficient is a number which refer to stability result of an instrument measurement. The higher coefficient of reliability, the higher trusted result of the instrument will be. The degree of confidence of an instrument, is expressed by reliability coefficient. For multiple choice instrument, estimation of reliability mostly calculated by KR-20. KR 20 formula calculate estimation of reliability instrument by using correct score method. This study aims to compare estimation of reliability coefficient by using KR-20 formula and by considering guessing probability. Hopefully by considering guessing factor, estimation of reliability coefficient will increased

Key words: Instrument, Multiple choice, Reliability, Guessing probability

PENDAHULUAN

Instrumen adalah alat ukur yang digunakan untuk mengukur dalam rangka pengumpulan data. Dalam pendidikan, instrumen alat ukur yang digunakan untuk mengumpulkan data dapat berupa tes atau nontes. Tes merupakan alat ukur pengumpulan data yang mendorong peserta memberikan penampilan maksimal (Djaali dan Muljono, 2008: 59). Sukardi berpendapat bahwa tes adalah prosedur sistematis guna membandingkan perilaku dua orang siswa atau lebih dengan menggunakan sekumpulan item (Sukardi, 2009: 27). Menurut Dali S. Naga, Sampel tingkah laku juga diukur oleh keberhasilan belajar atau kompetensi. Sampel yang terukur berfungsi sebagai hasil pengukuran suatu tes (Naga, 1992: 3). Hal ini berarti suatu tes disusun, dilaksanakan, dan dipakai untuk memberikan skor kepada seseorang berdasarkan aturan-aturan yang telah ditetapkan sebelum tes dilaksanakan. Butir-butir tes dipilih secara sistematis, dan diberikan kepada semua siswa sebagai peserta tes dalam bentuk, situasi, dan kondisi yang sama.

Salah satu bentuk tes yang digunakan untuk mengukur kemampuan siswa adalah bentuk tes pilihan ganda. Tes pilihan ganda disebut juga *multiple respon* atau pengenalan

kembali (*recognition test*), dimana dalam soal ini tersedia beberapa jawaban, tetapi hanya satu jawaban yang benar (Stamboel, 1991: 39). Tes pilihan ganda merupakan jenis tes yang paling fleksibel, karena banyak dipakai untuk menilai pencapaian tujuan pendidikan. Menurut Sukardi, tes ini dapat mengukur pengetahuan yang luas dengan tingkat domain yang bervariasi, tes pilihan ganda mempunyai semua persyaratan sebagai tes yang baik, yakni dilihat dari segi obyektifitas, reliabilitas dan daya beda antara siswa yang berhasil dengan siswa yang gagal atau bodoh (Sukardi, 2009: 125). Pilihan ganda juga sering digunakan untuk pengukuran kemampuan jika peserta tes berjumlah besar, karena pilihan ganda memiliki keunggulan mudah dalam pengkoreksian hasil tes. Akan tetapi suatu bentuk testidaklah lebih baik dari bentuk tes yang lainnya, hal tersebut dapat dipahami karena masing-masing bentuk tes memiliki beberapa kelebihan dan kelemahan. Soal berbentuk pilihan ganda memilikisalah satu kelemahan yaitu sangat sensitif terhadap menerka atau perilaku menebak (Surapranata, 2005: 178-179). *Guessing* atau menebak terjadi karena kurangnya kemampuan atau pengetahuan siswa atas materi yang ditanyakan karena kurang belajar atau kurang penguasaan materi. Karena itu salah satu faktor yang harus dipertimbangkan dalam pengukuran *maximum performed* dengan tes pilihan ganda adalah perilaku *guessing*. Perilaku *guessing* akan berkontribusi terhadap varians kesalahan pengukuran dan mengurangi reliabilitas tes (Nunnally dalam Kusaeri, 2012: 209).

Pada pengukuran menggunakan formula KR-20 penskoran yang digunakan adalah penskoran *corret score*, yaitu penskoran yang hanya berdasarkan jumlah butir yang dijawab benar tanpa memperhatikan faktor kelemahan pilihan ganda. Tebakan (*Guessing*) merupakan kondisi yang tidak wajar berkenaan dengan kemampuan peserta tes. Karena tebakan bersumber dari peserta tes. Alat ukur tes yang penyekorannya hanya menghitung respon betul (model *correct skor*) dimungkinkan memberikan peluang kepada peserta tes untuk melakukan tebakan pada butir-butir tesnya.

Didefinisikan oleh Gulliksen (dalam Zimmerman, 2009) bahwa reliabilitas merupakan perbandingan antara varians skor sebenarnya (σ_T^2) dengan varians skor teramati (σ_X^2). Selanjutnya dengan banyak memanipulasi formula Guliksen tersebut dan memanfaatkan persamaan yang diberikan oleh Hopkin dan Antes akan dilakukan penurunan rumus sehingga diperoleh formula reliabilitas yang dipengaruhi probabilitas *guessing*. Dengan demikian formula koefisien reliabilitas KR-20 dan koefisien reliabilitas yang dipengaruhi probabilitas *guessing* diturunkan dari definisi perbandingan antara variansi skor sebenarnya dengan skor amatan atau

$$\rho_{A1A2} = \frac{\sigma_T^2}{\sigma_A^2} \dots\dots\dots(1)$$

Jika suatu alat ukur dan setiap butirnya dianggap pilahan-pilahan yang setara sebanyak *L* pilahan dan ada responden sejumlah *g*, maka:

$$\begin{aligned} A_{g1} &= T_{g1} + K_{g1} \\ A_{g2} &= T_{g2} + K_{g2} \\ &\vdots \\ A_{gt} &= T_{gt} + K_{gt} \end{aligned}$$

Korelasi diantaradupilahanberurutanterjadi antara pilahan

$$\begin{aligned} A_1 &\text{ dan } A_2 \\ A_2 &\text{ dan } A_3 \\ &\vdots \end{aligned}$$

$$A_{L-1} \text{ dan } A_L$$

Atau pada umumnya diantarapilahan

$$A_r \text{ dan } A_s$$

Dengan $r = 1, 2, 3, \dots, L-1$

$$s = 2, 3, \dots, L$$

karena semua pilahan adalah setara dan memiliki variansi yang sama, maka

$$\sigma_{T1}^2 = \sigma_{T2}^2 = \dots = \sigma_{TL}^2$$

$$\sigma_{A1}^2 = \sigma_{A2}^2 = \dots = \sigma_{AL}^2$$

$$\rho_{TrTs} = 1$$

Dan dari (1.12) diketahui $\rho_{AA} = \frac{\sigma_T^2}{\sigma_A^2}$

Diperoleh $\rho_{ArAs} = \frac{\sigma_{Tr}^2}{\sigma_{Ar}^2} = \frac{\sigma_{Ts}^2}{\sigma_{As}^2}$

Sehingga

$$\sigma_{Tr}^2 = \rho_{ArAs} \sigma_{Ar}^2 \dots \dots \dots (1.13)$$

Perhatikan σ_T^2

$$\begin{aligned} \sigma_T^2 &= (\sigma_{T1} + \sigma_{T2} + \dots + \sigma_{TL})^2 \\ &= \sigma_{T1}^2 + \sigma_{T2}^2 + \dots + \sigma_{TL}^2 + \sum_r \sum_s \sigma_{TrTs} \\ &= L\sigma_{Tr}^2 + \sum_r \sum_s \rho_{TrTs} \sigma_{Tr} \sigma_{Ts} \\ &= L\sigma_{Tr}^2 + \sum_r \sum_s \sigma_{Tr} \sigma_{Ts} \end{aligned}$$

Untuk $r = 1, 2, 3, \dots, L-1$ dan $s = 2, 3, \dots, L$ maka

$$\begin{aligned} \sigma_T^2 &= L\sigma_{Tr}^2 + L(L-1)\sigma_{Tr}^2 \\ &= L\sigma_{Tr}^2 + L^2\sigma_{Tr}^2 - L\sigma_{Tr}^2 \\ &= L^2\sigma_{Tr}^2 \end{aligned}$$

Dari persamaan (1.13) maka

$$= L^2 \rho_{ArAs} \sigma_{Ar}^2$$

Selanjutnya perhatikan σ_A^2

$$\begin{aligned} \sigma_A^2 &= (\sigma_{A1} + \sigma_{A2} + \dots + \sigma_{AL})^2 \\ &= \sigma_{A1}^2 + \sigma_{A2}^2 + \dots + \sigma_{AL}^2 + \sum_r \sum_s \sigma_{ArAs} \\ &= L\sigma_{Ar}^2 + \sum_r \sum_s \rho_{ArAs} \sigma_{Ar} \sigma_{As} \\ &= L\sigma_{Ar}^2 + L(L-1)\rho_{ArAs} \sigma_{Ar}^2 \\ &= L\sigma_{Ar}^2 [1 + (L-1)\rho_{ArAs}] \end{aligned}$$

Sehingga dari

$$\rho_{AA} = \frac{\sigma_T^2}{\sigma_A^2}$$

$$= \frac{L^2 \rho_{ArAs} \sigma_{Ar}^2}{L \sigma_{Ar}^2 [1 + (L-1) \rho_{ArAs}]}$$

$$= \frac{L \rho_{ArAs}}{1 + (L-1) \rho_{ArAs}}$$

Selanjutnya kembali ke rumus σ_A^2 dengan mensubstitusikan $L \sigma_{Ar}^2 = \Sigma \sigma_{Ar}^2$ diperoleh

$$\sigma_A^2 = L \sigma_{Ar}^2 + L(L-1) \rho_{ArAs} \sigma_{Ar}^2$$

$$= \Sigma \sigma_{Ar}^2 + (L-1) \rho_{ArAs} \Sigma \sigma_{Ar}^2$$

$$\sigma_A^2 - \Sigma \sigma_{Ar}^2 = (L-1) \rho_{ArAs} \Sigma \sigma_{Ar}^2$$

Sehingga koefisien korelasi setiap pasangan pilahan menjadi

$$\rho_{ArAs} = \frac{\sigma_A^2 - \Sigma \sigma_{Ar}^2}{(L-1) \Sigma \sigma_{Ar}^2}$$

Karena ada L katakanlah pilahan setara dengan variansi yang sama maka koefisien reliabilitas seluruh skor adalah

$$\rho_{AA} = \frac{L \rho_{ArAs}}{(L-1) \rho_{ArAs} + 1}$$

$$= \frac{L \rho_{ArAs}}{(L-1) \rho_{ArAs} + \frac{\rho_{ArAs}}{\rho_{ArAs}}}$$

$$= \frac{L \rho_{ArAs}^2}{(L-1) \rho_{ArAs}^2 + \rho_{ArAs}}$$

$$= \frac{L}{(L-1) + \frac{1}{\rho_{ArAs}}}$$

Dengan substitusi ρ_{ArAs} maka

$$\rho_{AA} = \frac{L}{(L-1) + \frac{(L-1) \Sigma \sigma_{Ar}^2}{\sigma_A^2 - \Sigma \sigma_{Ar}^2}}$$

$$= \frac{L}{L-1} \cdot \frac{1}{\frac{\Sigma \sigma_{Ar}^2}{\sigma_A^2 - \Sigma \sigma_{Ar}^2}}$$

$$= \frac{L}{L-1} \cdot \frac{1}{\frac{\sigma_A^2 - \Sigma \sigma_{Ar}^2 + \Sigma \sigma_{Ar}^2}{\sigma_A^2 - \Sigma \sigma_{Ar}^2}}$$

$$= \frac{L}{L-1} \cdot \frac{\sigma_A^2 - \Sigma \sigma_{Ar}^2}{\sigma_A^2}$$

Jika setiap pilahan berisikan satu butir soal saja yakni butir ke-I sehingga variansi

$$\sigma_{Ar}^2 = \sigma_i^2$$

dan selanjutnya alat ukur mengandung N butir sehingga jumlah pilahan sama dengan jumlah butir

$$L = N$$

Dengan demikian semua butir adalah setara dan reliabilitas dikenal dengan Alpha Cronbach menjadi

$$\rho_{AA} = \frac{N}{N-1} \cdot \frac{\sigma_A^2 - \sum \sigma_{Ar}^2}{\sigma_A^2}$$

Dalam hal skor dikotomi maka variansi butir dapat disederhanakan menjadi

$$\sigma_i^2 = p_i q_i \quad \text{atau} \quad \sum \sigma_i^2 = \sum p_i q_i$$

Dengan ketentuan bahwa semua butir adalah setara maka diperoleh formula reliabilitas Kuder-Richardson 20

$$\rho_{KR-20} = \left(\frac{N}{N-1} \right) \left(\frac{\sigma_A^2 - \sum p_i q_i}{\sigma_A^2} \right)$$

Dimana

$$\rho_{KR-20} = \text{reliabilitas KR 20}$$

$$N = \text{Jumlah butir}$$

$$\sigma_A^2 = \text{Variansi skor total}$$

$$p_i = \text{Proporsi yang benar pada item ke-}i$$

$$q_i = \text{Proporsi yang salah pada item ke-}i$$

Kajian tentang tebakan buta yang bersumber dari peserta tes pada butir-butir tes objektif, khususnya bentuk pilihan ganda, berkenaan dengan pengaruhnya pada pelaksanaan pengukuran (Plake dan Melican, 1985: 3). Tebakan dalam hal ini merupakan faktor utama dalam penentuan skor peserta tes untuk tes objektif (Mehrens dan Lehmann, 1987: 142) Manakala peserta tes merespon butir-butir tes objektif pilihan ganda, ia akan melakukan tebakan buta dengan probabilitas untuk memperoleh respon betul sebesar $\frac{1}{K}$ dengan K sebagai ukuran opsi/pilihan (Fray, 1980: 80).

Dengan demikian dari uraian di sebelumnya, bahwa probabilitas *guessing* adalah peluang menebak benar pada tes objektif yang dilakukan oleh peserta tes yang tidak memiliki kemampuan apapun tentang jawaban/respon yang betul, dimana proses menebaknya dilakukan secara acak. Jika banyaknya ukuran pilihan/opsi adalah K , maka peserta tes akan melakukan tebakan dengan probabilitas untuk memperoleh respon betul sebesar $\frac{1}{K}$, dengan demikian jika ukuran pilihannya 3 opsi maka probabilitas *guessing*nya 0.33 dan jika ukuran pilihannya 4 opsi maka probabilitas *guessing*nya 0.25. Dengan demikian dapat disimpulkan bahwa probabilitas *guessing* adalah peluang menebak benar pada tes objektif yang dilakukan oleh peserta tes yang tidak memiliki kemampuan apapun tentang jawaban/respon yang betul, dimana proses menebaknya dilakukan secara acak.

Pengertian reliabilitas sendiri menurut teori klasik adalah sejauh mana hasil dari suatu pengukuran dapat dipercaya. Implikasi hasil pengukuran dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subjek yang sama

diperoleh hasil yang relatif sama, selama aspek dari subjek yang diukur memang belum berubah. Pengertian relatif sama menunjukkan bahwa ada toleransi terhadap perbedaan-perbedaan kecil di antara hasil skor beberapa kali pengukuran. Pengertian reliabilitas menurut Gulliksen dalam Zimmerman merupakan perbandingan antara varians skor sebenarnya (σ_T^2) dengan varians skor teramati (σ_X^2) (Gulliksen dan William, 2003: 366-

$$368). \text{ Atau } \rho_{A1A2} = \frac{\sigma_T^2}{\sigma_A^2} \dots\dots\dots(1.12)$$

Bila dinyatakan dalam bentuk persamaan, maka akan bentuk berikut seperti dalam persamaan

$$\sigma_{Tr}^2 = \rho_{ArAs} \sigma_{Ar}^2 \dots\dots\dots(2.1)$$

Dari beberapa pengertian reliabilitas yang dikemukakan oleh para ahli, selanjutnya akan difokuskan pada pengertian reliabilitas menurut Gulliksen, untuk itu pada penurunan rumus reliabilitas ini dilakukan dengan memanipulasi formula Gulliksen di atas dengan memanfaatkan persamaan yang diberikan Hopkin dan Antes. Hopkin dan Antes dalam Kusaeri menjelaskan bahwa di dalam skor tes terdapat komponen skor sebenarnya dan kekeliruan/kesalahan (Kusaeri, 2012: 210). Hubungan di antara ketiganya dinyatakan melalui persamaan matematika, sebagai berikut:

$$X = T + E \dots\dots\dots(2.2)$$

Dimana X sebagai skor teramati (*observed score*), T sebagai skor sebenarnya (*true score*) dan E sebagai kekeliruan (*error*). Ini berarti skor teramati dari peserta tes merupakan hasil pengukuran skor sebenarnya dengan kekeliruan. Jika diasumsikan kekeliruan adalah nol berarti rata-rata skor yang diperoleh sama dengan rata-rata skor sebenarnya. Bila kesalahan (*error*, E) dipartisi menjadi kesalahan yang disumbangkan oleh *guessing* (E_1) dan kesalahan karena sumber-sumber lain (E_2) maka

$$E = E_1 + E_2$$

Sehingga skor teramati menjadi

$$X = T + E_1 + E_2 \dots\dots\dots(2.3)$$

Dengan demikian untuk peserta tes individu ke-i diperoleh skor teramati

$$X_i = T_i + E_{1i} + E_{2i} \dots\dots\dots(2.4)$$

Dari rumus (2.2), diperoleh persamaa kesalahan (*error*), yaitu

$$E_{1i} + E_{2i} = X_i - T_i$$

Ekspektasi kesalahan (harapan) dari kesalahan

$$\begin{aligned} E(E_{1i} + E_{2i}) &= E(X_i - T_i) = 0 \\ &= E(X_i) - T_i = 0 \dots\dots\dots(2.5) \end{aligned}$$

Sehingga $E(X_i) - T_i = 0$

$$\text{Atau } E(X_i) = T_i \dots\dots\dots(2.6)$$

Dari persamaan (1.2) T_i = adalah banyaknya item yang pasti benarditambah harapan banyaknya item yang benar hasil tebakan.

Misalkan K_i adalah banyaknya item yang pasti benar untuk individu ke-i,

Maka akan ada $(N - K_i)$ item yang ditebak, dengan probabilitas melakukan tebakan adalah p maka banyaknya item yang benar hasil tebakan akan berdistribusi binomial $B(N - K_i, p)$

harapan banyaknya item benar hasil tebakan adalah sama dengan rata-rata distribusi binomial tersebut yaitu

$$\mu = p(N - K_i) \dots \dots \dots (2.7)$$

dari persamaa (2.6) dan (2.7) diperoleh

$$E(X_i) = T_i = K_i + p(N - K_i) \dots \dots \dots (2.8)$$

Berhubungan dengan varians maka

$$Var(X) = Var(T + E)$$

Karena T dan E kejadian yang saling bebas maka

$$Var(X) = Var(T) + Var(E)$$

Atau $\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \dots \dots \dots (2.9)$

Diketahui bahwa $E = E_1 + E_2$

Maka

$$\begin{aligned} Var(E) &= Var(E_1 + E_2) \\ &= Var(E_1) + Var(E_2) + 2Cov(E_1, E_2) \end{aligned}$$

Atau $\sigma_E^2 = \sigma_{E1}^2 + \sigma_{E2}^2 + 2Cov(E_1, E_2) \dots \dots \dots (2.10)$

Dari persamaan (2.9) dan (2.10)

$$\sigma_X^2 = \sigma_T^2 + \sigma_{E1}^2 + \sigma_{E2}^2 + 2Cov(E_1, E_2)$$

Jika dikembalikan kepada pengertian reliabilitas menurut Gulliksen yaitu bahwa reliabilitas adalah perbandingan dari skor sebenarnya (skor tulen) dengan skor amatan atau

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_X^2}$$

maka

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{E1}^2 + \sigma_{E2}^2 + 2Cov(E_1, E_2)} \dots \dots \dots (2.11)$$

Variansi kesalahan peserta tes yang disumbangkan oleh *guessing* adalah variansi dari distribusi Binomial $B(N - K_i, p)$ yaitu:

$$\sigma_{Ei}^2 = p(1 - p)(N - K_i) \dots \dots \dots (2.12)$$

Perhatikan kembali persamaan (2.8)

$$\begin{aligned} E(X_i) = T_i &= K_i + p(N - K_i) \\ T_i &= pN + (1 - p)K_i \end{aligned}$$

Sehingga varians dari T adalah

$$\begin{aligned} Var(T) &= Var[pN + (1 - p)K] \\ &= Var[(1 - p)K] \\ \sigma_T^2 &= (1 - p)^2 \sigma_K^2 \dots \dots \dots (2.13) \end{aligned}$$

Variansi skor kesalahan total terkait dengan *guessing* adalah:
Perhatikan kembali persamaan (2.12)

$$\sigma_{E_i}^2 = p(1-p)(N - K_i)$$

Harapan variansi skor kesalahan total terkait *guessing* adalah:

$$\begin{aligned} E(\sigma_{E_i}^2) &= E[p(1-p)(N - K_i)] \\ &= p(1-p)[N - E(K_i)] \end{aligned}$$

Sehingga $\sigma_{E_i}^2 = p(1-p)(N - \mu_K)$(2.14)

Dengan asumsi bahwa $Cov(E_1, E_2) = 0$ maka dari persamaan (2.11) dengan disubstitusikan persamaan (2.13) dan (2.14) diperoleh

$$\rho_{xx} = \frac{(1-p)^2 \sigma_K^2}{(1-p)^2 \sigma_K^2 + p(1-p)(N - \mu_K) + \sigma_{E_2}^2}$$

Dengan asumsi bahwa kesalahan yang disebabkan faktor lain adalah nol maka $\sigma_{E_2}^2 = 0$
 Dengan demikian diperoleh formula estimasi koefisien reliabilitas yang dipengaruhi probabilitas *guessing*, yaitu:

$$\rho_{xx} = \frac{(1-p)^2 \sigma_K^2}{(1-p)^2 \sigma_K^2 + p(1-p)(N - \mu_K)} \dots\dots\dots(2.15)$$

Sebagai ilustrasi untuk memperlihatkan meningkatnya hasil perhitungan estimasi reliabilitas yang memperhitungkan probabilitas *guessing* dibandingkan dengan KR-20 berikut ini digunakan data nilai Ujian Nasional (UN) mata pelajaran matematika kelas IX pada soal paket D. Skor yang diperoleh pada data ini adalah menggunakan metoda *correct score*. Dari hasil 10 kali replikasi dengan jumlah responden 100 untuk setiap replikasi maka diperoleh data sebagai berikut:

Tabel 1. Data Replikasi

	Dipengaruhi Prob Guessing	KR-20
1	0.9137	0.8680
2	0.9227	0.8641
3	0.9202	0.8553
4	0.9306	0.8383
5	0.9177	0.8589
6	0.9076	0.8179
7	0.9144	0.8097
8	0.9027	0.7949
9	0.9220	0.8663
10	0.9015	0.8276
Jumlah	9,1532	8,402
Mean	0,9153	0,8402
Median	0,9161	0,8468
SD	0,0093	0,0263
Var	0,0001	0,0007
Max	0,9306	0,8680
Min	0,9015	0,7949
Skewness	-0,2401	-0,7609

Setelah data dianggap memenuhi uji persyaratan analisis yaitu uji kenormalan data dengan uji Liliefors dan uji homogenitas varians dengan Bartlet, selanjutnya dilakukan analisis uji kesamaan rata-rata dengan uji t.

Hipotesis yang akan diuji adalah untuk mengetahui: Perbedaan reliabilitas yang dipengaruhi probabilitas *guessing* dan reliabilitas KR-20.

Tabel 2. Hasil Perhitungan Statistik Uji t Pada $\alpha = 0.05$ Dan $\alpha = 0.01$

Uji Perbedaan Rata-rata	t_{hitung}	$t_{tabel\alpha=0,01}$	$t_{tabel\alpha=0,05}$
Reliabilitas yang dipengaruhi probabilitas <i>guessing</i> dan KR-20	8,5151	2,88	2,10

Hasil analisa data dengan menggunakan menggunakan uji t pada taraf signifikansi $\alpha = 0,01$, didapat nilai $t_{hitung} = 8,5151$ berada di luar $t_{tabel} (t_{(0,01,1,18)}) = (\pm)2,88$ dan $(t_{(0,05,1,18)}) = (\pm) 2,10$ maka H_0 ditolak, sebagai konsekuensinya maka H_1 diterima, artinya terbukti bahwa rata-rata reliabilitas yang dipengaruhi probabilitas *guessing* lebih besar dari reliabilitas KR-20. nilai rata-rata reliabilitas yang dipengaruhi probabilitas *guessing* sebesar 0,9153 lebih tinggi dari nilai rata-rata reliabilitas KR-20 sebesar 0,8401. Dengan demikian dapat disimpulkan bahwa reliabilitas yang dipengaruhi probabilitas *guessing* lebih tinggi dari reliabilitas KR-20.

PENUTUP

Simpulan

Perhitungan estimasi reliabilitas instrumen dengan memperhitungkan reliabilitas *guessing* dapat meningkatkan kehandalan sebuah instrumen. Melalui formula yang disajikan dalam penelitian ini maka estimasi reliabilitas sebuah instrumen yang memperhitungkan probabilitas *guessing* tetap dapat dilakukan meskipun penskorannya menggunakan metoda *correct score*.

Melalui ilustrasi perhitungan dalam penelitian ini diketahui bahwa hasil estimasi reliabilitas instrumen pilihan ganda dengan memperhatikan faktor probabilitas *guessing* pada jawaban soal UN mata pelajaran matematika maka diperoleh bahwa perhitungan estimasi reliabilitas instrumen yang memperhitungkan probabilitas *guessing* memiliki angka estimasi 7,5% lebih tinggi lebih jika dibandingkan estimasi dengan formula KR-20

Saran

Semakin tingginya kualitas hasil pendidikan maka diperlukan instrumen pengukuran kemampuan siswa yang juga lebih berkualitas. Analisis atas instrumen sebaiknya dilakukan dengan tidak hanya mengacu kepada pedoman yang sudah lama berlaku tetapi ada baiknya dikembangkan dengan metoda-metoda lain yang dapat lebih meningkatkan kualitas instrumen. Kelemahan bentuk soal tidak hanya berpengaruh kepada reliabilitas tetapi juga kepada validitas setiap butir soal. Sebaiknya juga dilakukan penelitian tentang efek dari kelemahan butir soal terhadap validitas butir soal.

DAFTAR PUSTAKA

- Djaali dan Puji Muljono. 2008. **Pengukuran dalam Bidang Pendidikan**. Jakarta: PT Grasindo.
- Ebel, Robert L. 1986. **Essentials of Educational Measurement**. New Jersey: Prentice-Hall Inc.

- Kusaeri. **Probabilitas Guessing dan Pengaruhnya Terhadap Reliabilitas Tes Pilihan Ganda**. <http://jurnal.pdii.lipi.go.id/admin/jurnal/8209208222.pdf>, diakses. (tanggal 17 Desember 2010).
- Mehrens, William A., dan Irvin J. Lehmann. 1987. **Using Standardized Test in Education**. New York: Longman Inc.
- Naga, Dali S. 1992. **Pengantar Teori Skor pada Pengukuran Pendidikan**. Jakarta: Gunadarma.
- Naga, Dali S. 2009. **Bahan Perkuliahan Teori Tes**. Jakarta: Pasca Sarjana UNJ.
- Plake, Barbara S., dan Gerald J. Melican. 1985. **Prediction of Item Performance by Expert Judges: A Methodology for Examining the Impact of Correction for Guessing Instructions on Test Taking Behavior**. Research Report, ED 298-171 (National Institute of Education, US Departement of Education: Educational Resources Information Center (ERIC).
- Stamboel, Conny S. 1991. **Prinsip dan Teknik Pengukuran dan Penilaian Didalam Dunia Pendidikan**. Jakarta: Mutiara Sumber Widya.
- Sukardi. 2009. **Evaluasi Pendidikan Prinsip dan Operasionalnya**. Yogyakarta: Bumi Aksara.
- Surapranata, Sumarna. 2005. **Panduan Penulisan Tes Tertulis, Implementasi Kurikulum 2004**. Bandung: PT Remaja Rosdakarya.
- Zimmerman, Donald W., dan Richard H. Williams. 2003. **A new look at the influence of guessing on the reliability of multiple-choice tests**. *Journal Applied Psychological Measurement*, Vol. 27 No. 5, September 2003, pp. 366-368.