

Analisis Validitas Reabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika

Hery Susanto¹, Achi Rinaldi², Novalia³

^{1,2,3}Universitas Islam Negeri Raden Intan Lampung: hersus739@yahoo.co.id

Submitted : 15-07-2015, Revised : 23-09-2015, Accepted : 16-12-2015

Abstract

This study aims to (1) determine the items quality of odd semester final exam in social class XII mathematics subject academic year 2014/2015 in terms of validity, reliability, level of difficulty, and distinguishing power, (2) create a program shape for calculating validity, reliability, level of difficulty and distinguishing power. The research used quantitative descriptive method and documentation technique. The data consisted of 128 students' answer sheets in social class XII academic year 2014/2015. The data were analyzed by using several formulas of validity, reliability, level of difficulty, distinguishing power and the use of IBM SPSS Statistics 22 software and Anates V4 software. The results showed that 22 questions (55%) were valid and 18 questions (45%) were invalid. The reliability of items was quite good where the reliability coefficient value $r_{11} > r_{tabel}$ was 0.558 based on three formulas which were Anova Hoyt, Cronbach alpha and IBM SPSS Statistics 22. The difficulty level of items was not good because the items balances including easy, medium and difficult were not proportional, the proportional balance was 3-5-2 or 3-4-3. The distinguishing power of items showed that 1 item (2.50%) was very good, 9 items (22.50%) were good, 10 items (25%) were moderate, 14 items (35%) were less good and 6 items (15%) were bad. In addition to the above results, this study also resulted a program shape to facilitate the calculation of item analysis. The program calculations are the validity index analysis, reliability, level of difficulty and distinguishing power of test instrument.

Keywords : *distinguishing power, item analysis, level of difficulty, reliability, validity*

Abstrak

Penelitian ini bertujuan untuk (1) menentukan item kualitas ujian semester ganjil kelas sosial matematika kelas XII tahun akademik 2014/2015 dalam hal validitas, reliabilitas, tingkat kesulitan, dan daya pembeda, (2) membuat bentuk program untuk menghitung validitas, reliabilitas, Tingkat kesulitan dan daya pembeda. Metode penelitian yang digunakan adalah teknik deskriptif kuantitatif dan teknik dokumentasi. Datanya terdiri dari 128 lembar jawaban siswa pada kelas sosial XII tahun ajaran 2014/2015. Data-data tersebut dianalisis dengan menggunakan beberapa rumus validitas, reliabilitas, tingkat kesulitan, pembedaan daya dan penggunaan perangkat lunak IBM SPSS Statistics 22 dan perangkat lunak V4. Hasil penelitian menunjukkan bahwa 22 pertanyaan (55%) valid dan 18 pertanyaan (45%) tidak valid.

Keandalan item cukup bagus dimana nilai koefisien reliabilitas $r_{11} > r_{tabel}$ adalah 0,558 berdasarkan tiga rumus yaitu Anova Hoyt, Cronbach alpha dan IBM SPSS Statistics 22. Tingkat kesulitan item tidak baik karena saldo item termasuk media yang mudah, sedang Dan sulit tidak proporsional, keseimbangan proporsinya 3-5-2 atau 3-4-3. Daya pembeda item menunjukkan bahwa 1 item (2,50%) sangat baik, 9 item (22,50%) baik, 10 item (25%) sedang, 14 item (35%) kurang baik dan 6 item (15%) Yang buruk Selain hasil di atas, penelitian ini juga menghasilkan bentuk program untuk memudahkan perhitungan analisis item. Perhitungan program adalah analisis indeks validitas, reliabilitas, tingkat kesulitan dan daya pembeda instrumen uji.

Kata Kunci : daya beda, tingkat kesulitan, reliabilitas, validitas

PENDAHULUAN

Pendidikan merupakan sarana yang berperan untuk menciptakan manusia yang berkualitas dan berpotensi. Mutu pendidikan yang tinggi menjadi cermin dari keberhasilan pelaksanaan pendidikan formal pada suatu negara. Melalui proses pendidikan seseorang akan memperoleh pengetahuan, pemahaman, dan cara bertingkah Syah (2006). Pendidikan yang berkualitas berhubungan erat dengan proses kegiatan pembelajaran yang dilaksanakan di sekolah. Tiga unsur dalam proses pembelajaran adalah tujuan pengajaran (tujuan instruksional), pengalaman belajar (proses belajar mengajar), dan hasil belajar. Kegiatan penilaian dalam dunia pendidikan dikenal dengan istilah kegiatan evaluasi. Evaluasi merupakan bagian dari proses belajar mengajar. Evaluasi hasil belajar dilakukan oleh pendidik untuk memantau proses, kemajuan, pencapaian dan perbaikan hasil belajar peserta didik secara berkesinambungan, Sukardi (2011).

Evaluasi hasil belajar siswa diujikan melalui ujian akhir semester (UAS). Nilai UAS merupakan gambaran penguasaan kompetensi yang dipelajari siswa dalam menempuh proses pembelajaran di sekolah selama satu semester, sehingga diperlukan soal yang berkualitas baik. Evaluasi pembelajaran dilakukan oleh pendidik dalam bentuk tes. Fungsi tes secara umum yaitu (1) sebagai alat pengukur terhadap perkembangan atas kemajuan peserta didik; dan (2) sebagai alat ukur pengukur keberhasilan program pengajaran. Anas Sudijono (2013).

Kualitas tes sebagai salah satu alat evaluasi yang penting untuk diperhatikan. Kualitas tes dapat menjadi pedoman untuk penilaian ketepatan hasil belajar peserta didik. Pendidik perlu melakukan analisis terlebih dahulu sebelum melakukan penilaian hasil belajar. Analisis merupakan kemampuan seseorang untuk merinci atau menguraikan suatu bahan atau keadaan menurut bagian-bagian yang lebih kecil dan mampu memahami hubungan di antara bagian-bagian atau faktor-faktor yang satu dengan faktor-faktor lainnya. Anas Sudijono (2013).

Analisis yang perlu dilakukan pendidik adalah analisis butir soal. Kegiatan analisis butir soal merupakan kegiatan penting dalam penyusunan soal agar diperoleh butir soal yang bermutu. Tujuan kegiatan analisis butir soal menurut Aiken dalam kusaeri adalah mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum digunakan, meningkatkan kualitas butir tes melalui revisi atau membuang soal yang tidak efektif, serta mengetahui informasi diagnostik pada siswa apakah mereka telah memahami materi yang telah diajarkan. Soal yang bermutu adalah soal yang dapat memberikan informasi setepat-tepatnya tentang siswa yang telah menguasai materi dan siswa yang belum menguasai materi.

Kusaeri Suprananto (2012)

Fakta di lapangan menunjukkan bahwa analisis butir soal masih jarang dilakukan oleh pendidik. Beberapa alasan yang menyebabkan pendidik tidak melakukan analisis butir soal antara lain: (1) pendidik merasa terbebani dalam proses analisis butir soal sehingga tidak melakukannya, (2) pendidik meyakini bahwa kualitas soal tes yang dibuat sudah baik sehingga tidak melakukan penelaahan lebih lanjut. Menurut ketua MGMP mata pelajaran Matematika kota Bandar Lampung, soal yang digunakan untuk ujian semester dibuat oleh guru yang memang ahli di bidangnya dan tidak dibuat berdasarkan analisis yang telah dilakukan. Analisis butir soal dilakukan oleh sekolah atau masing-masing guru, karena menganalisis butir soal secara keseluruhan sekota Bandar Lampung membutuhkan biaya yang besar dan waktu yang cukup banyak.

Berdasarkan uraian di atas, maka perlu dilakukan analisis butir soal secara kuantitatif yang disusun oleh MGMP Matematika tahun 2014/2015 untuk mengetahui kualitas soal tersebut. Kualitas soal UAS dapat dilihat dari hasil uji validitas, reliabilitas, tingkat kesukaran dan daya beda. Peneliti tertarik untuk melakukan penelitian dengan judul “Analisis Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda Pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika Kelas XII IPS di SMA Negeri 12 Bandar Lampung Tahun Ajaran 2014/2015”.

Penelitian ini bertujuan untuk mengetahui validitas reliabilitas tingkat kesukaran dan daya beda butir soal Ujian Akhir Semester Ganjil mata pelajaran matematika kelas XII IPS di SMA Negeri 12 tahun ajaran 2014/2015. Penelitian ini bertujuan untuk (1) mengetahui kualitas butir soal ujian akhir semester ganjil kelas XII IPS mata pelajaran matematika tahun ajaran 2014/2015 ditinjau dari validitas, reliabilitas, tingkat kesukaran, dan daya beda, (2) membuat bentuk sebuah program untuk menghitung validitas, reliabilitas, tingkat kesukaran dan daya beda.

METODE PENELITIAN

Penelitian menggunakan metode deskriptif kuantitatif dan teknik dokumentasi. Data penelitian berupa 128 lembar jawaban siswa kelas XII IPS tahun ajaran 2014/2015. Data dianalisis menggunakan beberapa rumus validitas, reliabilitas, tingkat kesukaran dan daya beda serta penggunaan *software* IBM SPSS Statistics 22 dan *software* anates V4. Bentuk tes yang dianalisis adalah tes pilihan ganda atau multiple choice. Validitas adalah indeks yang menunjukkan sejauh mana suatu alat pengukur itu mengukur apa yang ingin diukur. Masri Singarimbun, Sodian Effendi (2011). Analisis validitas dilakukan dengan menggunakan teknik korelasi point biserial yaitu:

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Rumus lainnya dengan menggunakan korelasi *Product moment*

$$r_{xy} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \cdot \sum_{i=1}^n Y_i}{\sqrt{[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2][n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2]}}$$

Nilai r_{xy} atau r_{pbi} akan dibandingkan dengan koefisien korelasi *table* nilai “r” *product moment* pada taraf signifikan 5%. Apabila nilai r_{xy} atau r_{pbi} hasil koefisien korelasi lebih besar

(>) dari nilai r_{tabel} , maka hasil yang diperoleh adalah signifikan, artinya butir soal tes dinyatakan valid.

Nilai r_{xy} adalah nilai koefisien korelasi dari setiap butir/ item soal sebelum dikoreksi, kemudian dicari *corrected item-total correlation coefficient* dengan rumus sebagai berikut:

$$r_{x(y-1)} = \frac{r_{xy}S_y - S_x}{\sqrt{S_y^2 + S_x^2 - 2r_{xy}(S_y)(S_x)}}$$

Nilai $r_{x(y-1)}$ akan dibandingkan dengan koefisien korelasi tabel $r_{tabel} = r_{(\alpha, n-2)}$. Jika $r_{x(y-1)} \geq r_{tabel}$, maka instrumen valid. Pada output SPSS, *corrected item-total correlation coefficient* $t \geq r_{tabel}$, maka instrumen valid. Reliabilitas adalah indeks yang menunjukkan sejauh mana suatu alat pengukur dapat dipercaya atau dapat diandalkan. Suatu instrumen pengukuran dikatakan reliabel jika pengukurannya konsisten, cermat dan akurat. Masri Singarimbun, Sodian Effendi (2011). Rumus-rumus yang digunakan pada pengujian reliabilitas sebagai berikut

1) Rumus Spearman Brown:

$$r_i = \frac{2rb}{1 + rb}$$

2) Rumus KR 20 (Kuder Richardson)

$$r_i = \frac{k}{k-1} \left\{ \frac{st^2 - \sum p_i \cdot q_i}{st^2} \right\}$$

3) Rumus KR 21 Kuder Richardson

$$r_i = \frac{k}{(k-1)} \left\{ 1 - \frac{M(k-M)}{k \cdot st^2} \right\}$$

4) Rumus Rulon

$$r_{11} = 1 - \frac{S_d^2}{S_t^2}$$

5) Flanagan

$$r_{11} = 2 \left(1 - \frac{S_1^2 + S_2^2}{S_t^2} \right)$$

6) Analisis Varians Hoyt (Anova Hoyt)

$$r_i = 1 - \frac{MKe}{MKs}$$

7) Alfa Cronbach

$$r_i = \frac{k}{(k-1)} \left\{ 1 - \frac{\sum s_i^2}{st^2} \right\}$$

Nilai koefisien (r_i) akan dibandingkan dengan koefisien korelasi tabel $r_{tabel} = r_{(\alpha, n-2)}$. Jika $r_{11} > r_{tabel}$, maka instrumen reliabel. Pada Output SPSS, jika *Cronbach's Alpha* $> r_{tabel}$, maka instrumen reliabel. Analisis tingkat kesukaran soal artinya mengkaji soal-soal tes dari segi kesulitannya sehingga dapat diperoleh soal-soal mana yang termasuk mudah, sedang, dan sukar. Nana Sudjana (2009). Asumsi yang digunakan untuk memperoleh kualitas soal yang baik, di samping memenuhi validitas dan reliabilitas adalah adanya keseimbangan dari tingkat kesulitan soal tersebut. Keseimbangan yang dimaksudkan adalah adanya soal-

soal yang termasuk mudah, sedang dan sukar secara proporsional. Perbandingan antara soal mudah-sedang-sukar bisa dibuat 3-4-3, artinya 30% soal kategori mudah, 40% soal kategori sedang, dan 30% soal kategori sukar. Perbandingan lain yang termasuk sejenis dengan proporsi di atas adalah 3-5-2. Artinya 30% soal kategori mudah, 50% soal kategori sedang, dan 20% soal kategori sukar. Novalia, Muhammad Syazali (2014). Rumus yang digunakan untuk mengetahui tingkat/indeks kesukaran butir soal yaitu:

$$I = \frac{B}{J}$$

Penafsiran tingkat kesukaran butir tes digunakan kriteria menurut *Witherington* dalam Anas Sudijono sebagai berikut.

Tabel 1. Interpretasi tingkat kesukaran butir tes

Interval	Interpretasi
0,00 – 0,30	SoalSukar
0,31 - 0,70	SoalSedang
0,71 – 1,00	SoalMudah

Cara lain yang dapat ditempuh untuk menafsirkan analisis tingkat kesukaran butir soal dalam anates adalah:

Tabel 2. Interpretasi tingkat kesukaran anates

Interval	Interpretasi
0% - 15%	Sangat Sukar
16% - 30%	Sukar
31% - 70%	Sedang
71% - 85%	Mudah
86% - 100%	Sangat Mudah

Menganalisis daya pembeda artinya mengkaji soal-soal tes dari segi kesanggupan tes tersebut dalam membedakan siswa yang termasuk ke dalam kategori lemah/ rendah dan kategori kuat/ tinggi prestasinya. Novalia, Muhammad Syazali (2014). Penentuan daya beda butir soal pada Anates dapat diketahui dalam tabel daya pembeda pada kolom DP persen. Butir soal yang memiliki indeks daya beda $\geq 0,30$ dinyatakan baik dan butir soal yang indeks daya beda $< 0,30$ dinyatakan tidak baik. Daya pembeda butir soal memiliki manfaat yaitu untuk meningkatkan mutu setiap butir soal melalui data empiriknya dan untuk mengetahui seberapa jauh masing-masing butir soal dapat membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan pendidik. Kusaeri Suprananto (2012). Adapun untuk menghitung daya beda digunakan rumus sebagai berikut:

$$D = P_A - P_B$$

Sebagai catatan, dalam keadaan di mana jumlah *testee* adalah cukup besar (100 orang atau lebih), daya pembeda item cukup dihitung berdasar 27% *testee* kelompok atas dan 27% dari *testee* kelompok bawah, sedangkan *testee* yang terletak diantara kedua ujung ekstrem itu tidak perlu diikutsertakan dalam perhitungan analisis. Anas Sudijono (2013)

Tabel 3. Interpretasi Indeks Daya Pembeda Butir

Daya Pembeda	Klasifikasi	Interpretasi
0,70-1,00	Excellent	Baik Sekali
0,40-0,69	Good (baik)	Baik
0,20-0,39	Satisfactory (memuaskan)	Cukup
0,00-0,19	Poor (lemah)	Kurang baik
Bertanda negatif	-	Jelek Sekali

HASIL PENELITIAN DAN PEMBAHASAN

Hasil data yang peneliti peroleh berupa soal dan hasil jawaban tes pada mata pelajaran matematika kelas XII tahun 2014/2015 di SMA Negeri 12 Bandar Lampung

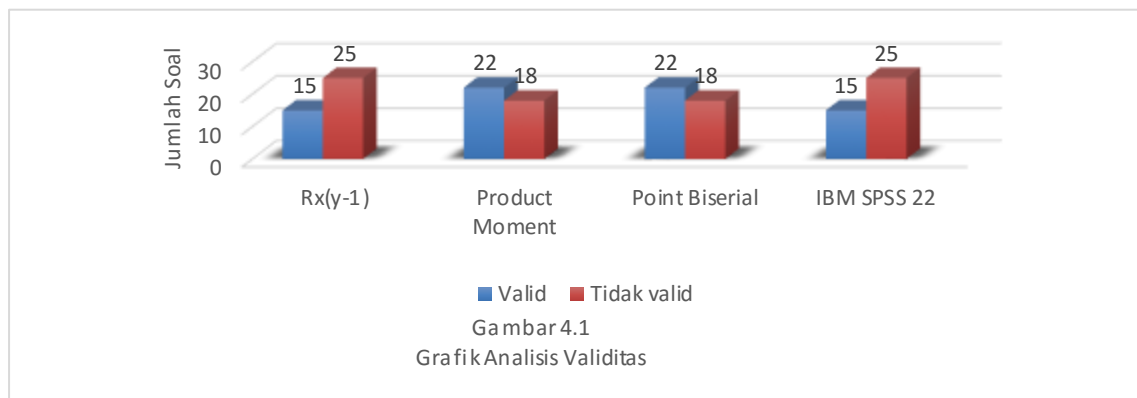
Tabel 4. Data Nilai Peserta Didik Kelas XII IPS

Jumlah Jawaban Benar	Nilai	Banyak siswa
4	10	1
5	12,5	2
6	15	1
7	17,5	4
8	20	1
9	22,5	2
10	25	1
11	27,5	6
12	30	9
13	32,5	9
14	35	13
15	37,5	6
16	40	11
17	42,5	26
18	45	11
19	47,5	9
20	50	9
21	52,5	4
22	55	3

Berdasarkan tabel tersebut, nilai maksimum atau nilai tertinggi didapatkan oleh 3 siswa yang mendapatkan nilai 55. Nilai minimum atau nilai terendah didapatkan oleh 1 orang siswa yang mendapatkan nilai 10. *Mean* atau rata-rata skor yang diperoleh untuk mata pelajaran matematika adalah 38,357. *Median* atau nilai tengah dari data hasil ujian akhir

semester adalah 40. *Mode* atau modus atau nilai yang sering muncul didapatkan oleh 26 siswa yang mendapatkan nilai 42,5. Jumlah siswa yang memperoleh nilai lebih besar dari *mean* adalah 73 orang.

Validitas adalah derajat yang menunjukkan suatu tes mengukur apa yang hendak diukur. Sebuah data atau informasi dapat dikatakan valid apabila sesuai dengan keadaan sebenarnya. Sugiyono (2012). Analisis validitas butir soal dilakukan pada soal matematika kelas XII IPS yang berjumlah 40 butir soal. Analisis validitas pada penelitian ini menggunakan tiga rumus meliputi $R_{x(y-1)}$, *Product moment*, *Point biserial* dan satu *software* IBM SPSS Statistics 22. Hasil analisis validitas yaitu:



Untuk menyatakan tinggi rendahnya validitas, beberapa rumus yang digunakan antara lain rumus v_1 atau rumus $R_{x(y-1)}$, rumus v_2 atau rumus *Product Moment*, rumus v_3 atau rumus *Point Biserial* dan *Software* IBM SPSS 22. Perhitungan validitas item tes dengan rumus v_2 dan v_3 menghasilkan 22 butir soal atau sekitar 55% butir soal pada tes tersebut yang dinyatakan valid, sedangkan 18 butir dinyatakan tidak valid atau sekitar 45%. Hasil berbeda ditunjukkan pada perhitungan menggunakan rumus v_1 dan IBM SPSS Statistics 22. Berdasarkan rumus v_1 dan IBM SPSS 22, jumlah soal yang memiliki validitas pada tes yaitu 15 butir atau sekitar 37,5%, sedangkan jumlah soal yang tidak valid yaitu 25 butir atau sekitar 62,5%.

Butir soal yang dinyatakan valid dan tidak ada perbedaan pada rumus v_1 , v_2 , v_3 dan SPSS adalah soal nomor 1, 8, 9, 10, 16, 18, 19, 26, 27, 28, 30, 33, 34, 35, 37, dan 38. Contoh butir soal yang valid adalah butir soal nomor 1, soal tersebut dinyatakan valid karena rata-rata siswa dapat menjawab dengan benar soal tersebut sehingga soal tersebut dapat mengukur kemampuan yang diharapkan. Butir soal yang tidak valid menurut rumus v_1 , v_2 , v_3 dan IBM SPSS 22 adalah butir soal nomor 2, 3, 4, 5, 11, 12, 13, 14, 15, 20, 21, 22, 23, 25, 31, 32, 36, 40. Contoh butir soal yang tidak valid adalah butir soal nomor 32 dan 33, hal ini terjadi karena pada butir soal tersebut tidak ditemukan jawaban yang tepat karena terjadi kesalahan penulisan pada pilihan jawaban. Butir soal yang memiliki hasil berbeda antara rumus v_1 dan IBM SPSS 22 dengan rumus v_2 dan v_3 adalah soal nomor 6, 7, 17, 24, 29, 35, dan 39. Soal nomor 6, 7, 17, 24, 29, 35, dan 39 dinyatakan valid dengan rumus v_2 dan v_3 , namun dinyatakan tidak valid dengan rumus v_1 dan IBM SPSS 22, pada IBM SPSS 22 nilai untuk memperoleh koefisien reliabilitas lebih teliti dan lebih cermat karena merupakan rumus lanjutan dari *product moment*. Soal nomor 6, 7, 17, 24, 29, 35, 39 secara berurutan memperoleh nilai korelasi pada

rumus v2 dan v3 adalah 0,192; 0,211; 0,191; 0,252; 0,184; 0,271; 0,188. Soal nomor 6, 7, 17, 24, 29, 35, 39 secara berurutan memperoleh nilai korelasi pada v1 dan IBM SPSS 22 adalah 0,143; 0,088; 0,073; 0,128; 0,063; 0,145; 0,072. Nilai validitas pada rumus v2 dan v3 memiliki hasil berbeda dengan rumus v1 dan IBM SPSS 22 dimana rumus yang diterapkan dalam IBM SPSS 22 adalah rumus v1. Rumus v2 dan v3 lebih dianjurkan karena memperoleh hasil perhitungan yang sama, serta jumlah validitas yang lebih banyak dibandingkan *software* IBM SPSS 22. Namun rumus v3 mempunyai kelemahan yaitu tidak bisa digunakan dalam tes subyektif, sehingga rumus v2 lebih disarankan untuk digunakan.

Reliabilitas tes adalah ketepatan atau keajegan alat dalam menilai apa yang dinilainya. Reliabilitas adalah indeks yang menunjukkan sejauh mana suatu alat pengukur dapat dipercaya. Sugiyono (2012). Analisis reliabilitas pada penelitian ini menggunakan tujuh rumus meliputi *Spearman Brown*, *KR-20*, *KR-21*, *Rulon*, *Flanagan*, *Anova Hoyt*, *Alpha Cronbach*, dan dua *software* yaitu *Anates* dan *software* IBM SPSS 22.

Berdasarkan hasil perhitungan, koefisien reliabilitas pada rumus r1 yaitu 0,483. Pada rumus r2, indeks reliabilitasnya adalah 0,562. Indeks reliabilitas pada rumus r3 adalah 0,373. Indeks reliabilitas pada rumus r4 adalah 0,612. Indeks reliabilitas pada rumus r5 adalah 0,482. Indeks reliabilitas pada rumus r6 adalah 0,558), sedangkan indeks reliabilitas pada rumus r7 adalah 0,558. *Software* *anates* memiliki indeks reliabilitas yaitu 0,460, sedangkan *software* IBM SPSS 22 mempunyai indeks reliabilitas yaitu 0,558. Hasil analisis menunjukkan bahwa seluruh rumus menyatakan soal reliabel meskipun angka reliabilitas berbeda.

Formula *Spearman-Brown* dilakukan untuk menentukan reliabilitas tes hasil belajar bentuk obyektif dengan cara membelah tes menjadi dua bagian, dengan menggunakan dua model yaitu *Model Gasal Genap* dan *Model Kiri-Kanan*, sedangkan formula *Flanagan* dan *Rulon* sama-sama menggunakan teknik belah dua, namun sasaran yang dijadikan landasan berpijak dalam penentuan reliabilitas tes tersebut berbeda. Dengan Formula *Spearman-Brown*, pengujian dan penentuan reliabilitas tes hasil belajar bentuk obyektif didasarkan pada korelasi antara belahan pertama dan belahan kedua. Formula *Flanagan* menentukan reliabilitas tes hasil belajar tidak berdasarkan korelasi belahan pertama dan belahan kedua, melainkan pada deviasi (selisih skor dengan rata-rata totalnya). Pada formula *Rulon*, penentuan reliabilitas tes hasil belajar bentuk obyektif didasarkan pada selisih skor yang dimiliki oleh belahan pertama dengan belahan kedua tes tersebut.

Berbeda dengan formula sebelumnya, formula *Kuder-Richardson* dan formula *C. Hoyt* tidak menggunakan teknik belah dua. Penentuan reliabilitas dengan formula *Kuder-Richardson* dilakukan dengan menganalisis skor-skor item tes hasil belajar yang bersangkutan secara langsung. Formula *C. Hoyt* menentukan reliabilitas tes hendaknya kita menganggap bahwa data yang berupa skor-skor hasil tes itu kita anggap sebagai data hasil eksperimen, faktor pertama adalah subyek dan faktor kedua adalah item. Selanjutnya kita cari interaksi antara testee (subyek yang menjawab item) dengan item tes hasil belajar itu sendiri (yang dijawab oleh subyek), teknik ini yang lebih dikenal dengan istilah Teknik Analisis Varian (ANAVA). Namun kelemahannya yaitu perhitungan harus lebih teliti dan memerlukan waktu yang lebih lama dibandingkan dengan rumus *alpha cronbach*.

Kelebihan rumus pertama (*KR₂₀*) adalah perhitungan lebih teliti dari pada rumus kedua (*KR₂₁*), namun kelemahannya adalah proses perhitungan lebih sulit. Adapun kelebihan rumus

kedua (KR₂₁) adalah perhitungan relatif lebih sederhana atau lebih mudah, sedangkan kelemahannya adalah hasil perhitungan kurang teliti.

Peneliti menyimpulkan bahwa ketujuh rumus di atas mempunyai kecermatan masing-masing dalam penentuan reliabilitas butir soal. Hal ini disebabkan oleh proses pengerjaan dengan menggunakan rumus di atas tidak terlalu sulit, waktu pengerjaan yang relatif singkat, dan semua rumus mempunyai derajat reliabilitas yang berbeda-beda, selain dari rumus r6 dan r7. Akan tetapi rumus r7 lebih dianjurkan untuk digunakan karena proses perhitungan lebih sederhana, lebih mudah dipahami, dan hasil yang sama dengan rumus r6. Selain itu, kemungkinan kesalahan perhitungan pada rumus r7 lebih kecil dibandingkan rumus lainnya. Rumus r7 dapat digunakan untuk soal tes obyektif maupun subyektif dan hasilnya sama dengan penggunaan *software* IBM SPSS 22.

Analisis tingkat kesukaran artinya mengkaji soal-soal tes dari segi kesulitannya sehingga dapat diperoleh soal-soal mana yang termasuk dalam kategori mudah, sedang dan sukar. Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Tingkat kesukaran soal dipandang dari kemampuan siswa dalam menjawabnya, bukan dilihat dari sudut guru sebagai pembuat soal. Analisis tingkat kesukaran pada penelitian ini menggunakan satu rumus dan *software* Anates. Hasil analisis Tingkat Kesukaran yaitu:

Tabel 5. Tingkat Kesukaran

No	$P = \frac{\text{jumlah jawaban benar}}{\text{Banyak Siswa}}$		Anates	
	Indeks	Kriteria	Indeks	Kriteria
1	0,797	Mudah	79,69	Mudah
2	0,570	Sedang	57,03	Sedang
3	0,055	Sukar	5,47	Sangat Sukar
4	0,117	Sukar	11,72	Sangat Sukar
5	0,031	Sukar	3,13	Sangat Sukar
6	0,039	Sukar	3,91	Sangat Sukar
7	0,352	Sedang	35,16	Sedang
8	0,914	Mudah	91,41	Sangat Mudah
9	0,758	Mudah	75,78	Mudah
10	0,539	Sedang	53,91	Sedang
11	0,211	Sukar	21,09	Sukar
12	0,234	Sukar	23,44	Sukar
13	0,086	Sukar	8,59	Sangat Sukar
14	0,211	Sukar	21,09	Sukar
15	0,055	Sukar	5,47	Sangat Sukar
16	0,586	Sedang	58,59	Sedang
17	0,297	Sukar	29,69	Sukar
18	0,250	Sukar	25	Sukar
19	0,508	Sedang	50,78	Sedang

No	Indeks	Kriteria	Indeks	Kriteria
20	0,102	Sukar	10,16	Sangat Sukar
$P = \frac{\text{jumlah jawaban benar}}{\text{Banyak Siswa}}$				
			Anates	
No	Indeks	Kriteria	Indeks	Kriteria
21	0,180	Sukar	17,97	Sukar
22	0,313	Sedang	31,25	Sedang
23	0,078	Sukar	7,81	Sangat Sukar
24	0,383	Sedang	38,28	Sedang
25	0,516	Sedang	51,56	Sedang
26	0,672	Sedang	67,19	Sedang
27	0,766	Mudah	76,56	Mudah
28	0,820	Mudah	82,03	Mudah
29	0,672	Sedang	66,41	Sedang
30	0,867	Mudah	86,72	Sangat Mudah
31	0,125	Sukar	12,5	Sangat Sukar
32	0,070	Sukar	7,03	Sangat Sukar
33	0,586	Sedang	58,59	Sedang
34	0,594	Sedang	59,38	Sedang
35	0,484	Sedang	48,44	Sedang
36	0,242	Sukar	24,22	Sukar
37	0,445	Sedang	44,53	Sedang
38	0,258	Sukar	25,78	Sukar
39	0,281	Sukar	28,13	Sukar
40	0,281	Sukar	28,13	Sukar

Hasil analisis yang dilakukan terhadap 40 butir item tes hasil belajar dengan menggunakan rumus menunjukkan bahwa terdapat 6 butir soal (15%) dalam kategori mudah, 14 butir soal (35%) dalam kategori sedang, dan 20 butir soal (50%) dalam kategori sukar. Butir soal yang termasuk dalam kategori mudah yaitu butir soal nomor 1, 8, 9, 27, 28, dan 30. Butir-butir item yang termasuk kategori item berkualitas baik, dalam arti derajat kesukaran itemnya cukup atau sedang (tidak terlalu sukar dan tidak terlalu mudah) yaitu butir soal nomor 2, 7, 10, 16, 19, 22, 24, 25, 26, 29, 33, 34, 35, dan 37. Adapun butir soal yang termasuk kategori sukar adalah butir soal nomor 3, 4, 5, 6, 11, 12, 13, 14, 15, 17, 18, 20, 21, 23, 31, 32, 36, 38, 39, dan 40. Rumus ini mempunyai pendekatan dengan cara mencari persentase siswa yang menjawab salah. Artinya semakin banyak siswa yang menjawab salah, maka derajat tingkat kesukaran butir soal akan semakin sukar.

Berdasarkan hasil analisis dengan menggunakan *software* anates), diperoleh informasi bahwa terdapat 4 butir soal (10%) dalam kategori mudah, 2 butir soal (5%) dalam kategori sangat mudah, 10 butir soal (25%) dalam kategori sangat sukar, 14 butir soal (35%) dalam kategori sedang, dan 10 butir soal (25%) dalam kategori sukar.

Penggunaan *software* anates dalam menentukan tingkat kesukaran menghasilkan nilai yang tidak jauh berbeda dengan rumus. Perbedaan hanya terletak pada penafsiran yang

cenderung lebih banyak dalam *software* anates. Berbeda dengan IBM SPSS, *software* anates tidak begitu dikembangkan, dapat dilihat dari versi anates terakhir yang dikeluarkan oleh pihak *development* (pengembang) yaitu *software* anates versi V4 pada 13 Februari 2004.

Berdasarkan pembahasan di atas dapat disimpulkan bahwa butir-butir soal matematika yang berbentuk *multiple choice* memiliki kualitas soal yang kurang baik karena belum memiliki keseimbangan. Asumsi yang digunakan untuk memperoleh kualitas soal yang baik, di samping memenuhi validitas dan reliabilitas adalah adanya keseimbangan dari tingkat kesulitan soal tersebut. Keseimbangan yang dimaksud adalah jumlah soal-soal yang tergolong mudah, sedang dan sukar seimbang. Tingkat kesukaran soal perlu dilihat dari kemampuan siswa dalam menjawab soal yang diberikan, bukan dilihat dari sudut pandang guru sebagai pembuat soal. Perbandingan antara soal yang mudah-sedang-sukar dapat dibuat 3-4-3 atau 3-5-2, yang diartinya adalah 30% soal kategori mudah, 40% soal kategori sedang, dan 30% soal kategori sukar; atau 30% soal kategori mudah, 50% soal kategori sedang dan 20% soal kategori sukar.

Butir-butir soal yang masuk dalam kategori sedang sebaiknya segera dicatat dalam buku bank soal. Selanjutnya, butir-butir soal tersebut dapat dikeluarkan lagi dalam tes hasil belajar pada waktu yang akan datang. Untuk butir soal yang sukar ada tiga kemungkinan tindak lanjut, yaitu: (1) butir soal tersebut dibuang dan tidak dikeluarkan lagi dalam tes hasil belajar yang akan datang, (2) diteliti ulang, dilacak dan ditelusuri sehingga dapat diketahui faktor apa yang menyebabkan butir item yang bersangkutan sulit dijawab oleh *testee*. Setelah dilakukan perbaikan, butir-butir soal tersebut dikeluarkan lagi dalam tes hasil belajar yang akan datang, (3) butir soal yang sukar sebaiknya digunakan dalam tes seleksi yang sifatnya ketat karena kondisi tersebut mendukung penggunaan soal yang sukar dengan asumsi bahwa *testee* yang berkemampuan cukup memadai akan lolos dalam seleksi yang diadakan. Butir soal yang mudah juga memiliki tiga kemungkinan tindak lanjut yaitu: (1) Butir soal tersebut dibuang dan tidak dikeluarkan lagi dalam tes hasil belajar yang akan datang. (2) diteliti ulang dan ditelusuri secara cermat guna mengetahui faktor yang menyebabkan butir soal tersebut dapat dijawab betul oleh seluruh *testee*; setelah dilakukan perbaikan butir soal yang bersangkutan dikeluarkan lagi pada tes berikutnya untuk mengetahui apakah derajat kesukaran butir soal menjadi lebih baik daripada sebelumnya atau tidak, (3) butir-butir soal yang mudah mempunyai manfaat yaitu butir-butir soal dapat dimanfaatkan pada tes seleksi yang bersifat longgar, dalam arti bahwa sebagian besar dari *testee* akan dinyatakan lulus dalam tes seleksi tersebut. Dalam kondisi ini, pemberian butir soal yang mudah akan memberikan kesempatan bagi banyak *testee* untuk lolos dalam tes seleksi atau ujian yang diadakan.

Analisis daya beda artinya mengkaji soal-soal tes dari segi kesanggupan tes tersebut dalam membedakan siswa yang termasuk kedalam kategori rendah dan kategori tinggi. Daya pembeda butir adalah kemampuan suatu butir tes untuk dapat membedakan antara *testee* yang berkemampuan tinggi dan berkemampuan rendah. Analisis daya beda pada penelitian ini menggunakan rumus dan *software* Anates. Hasil analisis daya beda yaitu:

Tabel 6. Daya Beda

No.	$D = PA - PB$	ANATES
-----	---------------	--------

	Indeks	Kriteria	Indeks	Kriteria
1	0,200	Kurang Baik	14,29	Kurang Baik
2	0,200	Kurang Baik	14,29	Kurang Baik
3	0,086	Kurang Baik	2,86	Jelek Sekali
4	0,057	Kurang Baik	8,57	Jelek Sekali
5	-0,029	Jelek Sekali	-5,71	Jelek Sekali
6	0,086	Kurang Baik	8,57	Jelek Sekali
7	0,257	Cukup	28,57	Cukup
8	0,286	Cukup	28,57	Cukup
9	0,514	Baik	42,86	Baik
10	0,343	Cukup	31,43	Baik
11	0,171	Kurang Baik	17,14	Kurang Baik
	D = PA – PB		ANATES	
No.	Indeks	Kriteria	Indeks	Kriteria
12	0,114	Kurang Baik	14,29	Kurang Baik
13	-0,114	Jelek Sekali	11,43	Jelek Sekali
14	0,029	Kurang Baik	2,86	Jelek Sekali
15	-0,057	Jelek Sekali	-5,71	Jelek Sekali
16	0,486	Baik	42,86	Baik
17	0,229	Cukup	22,86	Cukup
18	0,486	Baik	42,86	Baik
19	0,571	Baik	62,86	Baik Sekali
20	-0,114	Jelek Sekali	-8,57	Jelek Sekali
21	0,029	Kurang Baik	5,71	Jelek Sekali
22	0,114	Kurang Baik	14,29	Kurang Baik
23	0,000	Kurang Baik	0	Jelek Sekali
24	0,371	Cukup	40	Baik
25	0,229	Cukup	20	Kurang Baik
26	0,486	Baik	45,71	Baik
27	0,600	Baik	60	Baik Sekali
28	0,457	Baik	45,71	Baik
29	0,114	Kurang Baik	11,43	Kurang Baik
30	0,257	Cukup	25,71	Cukup
31	0,086	Kurang Baik	8,57	Jelek Sekali
32	-0,086	Jelek Sekali	-2,86	Jelek Sekali
33	0,514	Baik	54,29	Baik Sekali
34	0,743	Baik Sekali	62,86	Baik Sekali
35	0,314	Cukup	34,29	Baik
36	0,229	Cukup	20	Kurang Baik

37	0,371	Cukup	37,14	Baik
38	0,486	Baik	45,71	Baik
39	0,171	Kurang Baik	25,71	Cukup
40	-0,171	Jelek Sekali	-17,14	Jelek Sekali

Berdasarkan hasil analisis dengan menggunakan rumus (lampiran 15), diperoleh informasi bahwa soal dalam kategori baik sekali yaitu 1 butir soal atau 2,50%, soal dalam kategori baik yaitu 9 butir soal atau 22,50%, soal dalam kategori sedang yaitu 10 butir soal atau 25%, soal dalam kategori kurang baik yaitu 14 butir soal atau 35%, dan sisanya 6 butir soal atau 15% tergolong kategori jelek sekali. Butir soal yang termasuk dalam kategori baik sekali adalah butir soal nomor 34. Butir soal yang termasuk dalam kategori baik adalah butir soal nomor 9, 16, 18, 19, 26, 27, 28, 33, dan 38. Butir soal yang dikategorikan sedang adalah butir soal nomor 7, 8, 10, 17, 24, 25, 30, dan 36. Selanjutnya, butir soal yang termasuk dalam kategori kurang baik adalah soal nomor 1, 2, 3, 4, 6, 7, 8, 11, 12, 14, 21, 22, 23, 29, 31, dan 39, sedangkan butir soal yang dikategorikan jelek sekali adalah butir soal nomor 5, 13, 15, 20, 32, dan 40.

Berdasarkan analisis yang peneliti lakukan dengan menggunakan *software* anates, diperoleh informasi bahwa soal tersebut terdapat 28% butir dalam kategori baik atau 11 butir soal, 25% kategori sedang atau 10 butir soal, 35% butir soal dalam kategori kurang baik atau sebanyak 14 butir soal dan 13% butir soal dalam kategori Jelek sekali atau 5 butir soal (lampiran 16). Penggunaan *software* anates dalam menentukan indeks daya beda suatu butir soal menurut peneliti adalah dapat untuk digunakan karena *software* anates memiliki hasil yang tidak jauh berbeda dengan rumus namun alangkah baiknya menggunakan rumus yang ada karena *software* anates sudah tidak dikembangkan lebih lanjut. Hal ini dibuktikan dengan versi terakhir yang dikeluarkan oleh pihak pengembang yaitu versi V4 yang dikeluarkan pada 13 Februari 2004. *Database software* tidak mengalami pembaharuan dan tidak adanya interpretasi yang digunakan pada *software* anates dalam memberikan penafsiran daya beda suatu butir soal sehingga peneliti harus memberikan penafsiran secara mandiri terhadap butir soal yang dianalisis.

Adapun tindak lanjut atas hasil analisis mengenai daya pembeda item tes hasil belajar tersebut adalah soal yang memiliki daya pembeda yang baik hendaknya dimasukkan dan dicatat dalam buku bank soal. Butir soal tersebut dapat dikeluarkan lagi dalam tes selanjutnya karena kualitasnya sudah cukup memadai. Butir soal yang daya pembedanya masih rendah memiliki dua kemungkinan tindak lanjut yaitu: (1) ditelusuri dan diperbaiki sehingga nanti dapat diajukan lagi dalam tes hasil belajar yang akan datang, serta perlu dianalisis kembali apakah daya pembeda meningkat atau tidak, (2) dibuang dan tidak dikeluarkan kembali pada tes yang akan datang. Khusus butir soal yang angka indeks diskriminanya bertanda negatif sebaiknya tidak perlu dikeluarkan pada tes hasil belajar selanjutnya atau direvisi kembali.

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan, dapat diperoleh kesimpulan yaitu terdapat 22 butir soal (55%) dinyatakan valid dan 18 butir soal (45%) dinyatakan tidak valid dari 40 butir soal tes bentuk obyektif. Reliabilitas soal tergolong baik berdasarkan perhitungan menggunakan beberapa rumus dan *software* yang menunjukkan nilai koefisien $r_{11} > r_{tabel}$. Tingkat kesukaran butir soal kurang baik karena keseimbangan soal yang mudah,

sedang dan sukar belum proposional, keseimbangan yang proposional yaitu 3-5-2 atau 3-4-3. Terdapat 1 butir soal (2,50%) baik sekali, 9 butir soal (22,50%) tergolong baik, 10 butir soal (25%) tergolong sedang, 14 butir soal (35%) tergolong kurang baik dan 6 butir soal (15%) tergolong jelek sekali berdasarkan analisis daya beda.

DAFTAR PUSTAKA

- Effendi, Sofian., dan Singarimbun, Masri. (2011). *Metode Penelitian Survei*. Jakarta: LP3ES.
- Novalia, dan Syazali, Muhammad. (2014). *Olah Data Penelitian Pendidikan*. Bandar Lampung: AURA.
- Sudijono, Anas. (2013). *Pengantar Evaluasi Pendidikan*. Bandung: Alfa Beta.
- Sugiyono. (2012). *Statistika Untuk Penelitian*. Bandung: Alfa Beta.
- Sudjana, Nana. (2009). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Sukardi. (2011). *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara.
- Suprananto, Kusaeri. 2012. *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu
- Syah, Muhibbin. (2006). *Psikologi Pendidikan dan Pendekatan Baru*. Bandung: PT Remaja Rosdakarya.