

Perbandingan Bentuk Tes Pilihan Ganda dan Teknik Penskoran Terhadap Reliabilitas Tes Mata Pelajaran Kimia

Eva Ariyanti^{1*}, dan Yoga Budi Bhakti²

¹ STAI Al Aqidah Al Hasyimiyah

² Universitas Indraprasta PGRI

* E-mail: 3va.ariyanti@gmail.com

Abstrak

Tujuan dari penelitian ini adalah untuk mengetahui perbedaan koefisien reliabilitas tes kimia antara pilihan ganda biasa dan pilihan ganda asosiasi. Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental. Penelitian Hypothesis diuji menggunakan analisis varians (ANOVA) satu cara. Penelitian ini dilakukan di empat sekolah. Sampel dari penelitian ini adalah 120 siswa yang dipilih secara acak replikasi. Penelitian menyimpulkan bahwa: (1) ada perbedaan koefisien reliabilitas tes kimia antara pilihan ganda bentuk tes dan metode penilaian, (2) koefisien reliabilitas tes kimia Asosiasi beberapa pilihan bentuk dengan berat metode penilaian lebih tinggi daripada koefisien reliabilitas bentuk beberapa pilihan biasa dengan metode penilaian yang benar, (3) koefisien reliabilitas tes kimia Asosiasi beberapa pilihan bentuk dengan metode penilaian yang benar lebih rendah dari Koefisien reliabilitas bentuk pilihan ganda biasa dengan metode penilaian bobot, (4) koefisien reliabilitas uji kimia Asosiasi bentuk pilihan ganda dengan metode penilaian yang benar lebih tinggi dari kehandalan koefisien bentuk beberapa pilihan biasa dengan metode penilaian yang benar, dan (5) koefisien reliabilitas uji kimia pilihan bentuk ganda biasa dengan metode penilaian bobot lebih tinggi dari koefisien reliabilitas bentuk beberapa pilihan biasa dengan metode penilaian yang benar. Berdasarkan hasil penelitian, menggunakan pilihan multi pilihan tes diselesaikan dengan metode scoring, dapat meningkatkan akurasi hasil dan dapat diandalkan. Akhirnya, ia memiliki implikasi positif pengembangan sistem evaluasi peningkatan, baik untuk evaluasi formatif atau summatif evaluasi.

Kata kunci: bentuk tes pilihan ganda, teknik penskoran, uji reliabilitas kimia.

PENDAHULUAN

Perkembangan dalam bidang pendidikan, khususnya di sekolah sudah mulai dilakukan. Seperti, penggunaan TI (teknologi informasi) dengan munculnya *e-learning*, *e-school*. Proses pembelajaran yang dilaksanakan melalui TI, hasilnya lebih dipastikan lebih unggul karena formulasi pola pembelajaran sudah dibuat lebih fleksibel sesuai dengan kebutuhan peserta didik. Dengan demikian, peserta didik yang mengikuti kegiatan belajar mengajar (KBM) tidak mengalami kejenuhan dengan adanya inovasi yang guru aplikasikan dalam setiap kegiatan belajar.

Seseorang dinyatakan melakukan kegiatan belajar bila terjadi perubahan tingkah laku, misalnya dari tidak tahu menjadi tahu, dan tidak mengerti menjadi mengerti. Dalam hal ini Syukur (2008: 6) menjelaskan belajar merupakan proses perubahan tingkah laku berkat interaksi dengan lingkungan, dan perubahan bersifat pengetahuan (kognitif), keterampilan (psikomotorik), serta menyangkut nilai dan sikap (afektif). Oleh karena itu, dalam penilaian hendaknya diperiksa sejauh mana perubahan tingkah laku peserta didik telah terjadi selama atau setelah pembelajaran berlangsung.

Perubahan tingkah laku, perubahan pengetahuan, nilai dan sikap, hasilnya dapat menjadi tujuan pembelajaran. Dengan mengetahui tingkat ketercapaian tujuan pembelajaran, guru dapat mengambil tindakan perbaikan pembelajaran atau melakukan bimbingan bagi peserta didiknya baik secara klasikal maupun individu. Sehingga siswa juga merasa diperhatikan, terutama siswa yang mempunyai kelemahan dalam bidang mata pelajaran tertentu agar hasil belajar yang diperoleh makin baik.

Penilaian hasil belajar menggunakan berbagai metode dan teknik penilaian yang beragam

sesuai dengan tujuan pembelajaran, pada umumnya teknik penilaian yang digunakan di sekolah yaitu tes tertulis yang digunakan sebagai tes formatif atau tes sumatif. Menurut Muslich (2010: 33) bahwa tujuan dan pengalaman belajar tertentu mungkin cukup efektif dinilai melalui tes tertulis (paper-pencil test), sedangkan tujuan dan pengalaman belajar yang lain efektif dinilai dengan tes praktik.

Muslich (2010: 86) menambahkan bahwa penilaian melalui tes tertulis dilakukan untuk mengungkap penguasaan peserta didik dalam aspek/ranah kognitif mulai dari jejang pengetahuan, pemahaman, penerapan, analisis, sintesis, sampai evaluasi. Di samping itu tes yang diberikan kepada peserta didik kadang-kadang juga mengacu pada kompetensi yang dinilai, sehingga tes yang digunakan mampu mengukur kemampuan siswa. Melalui tes guru dapat memperoleh informasi tentang berhasil tidaknya peserta didik dalam menguasai tujuan-tujuan (standar kompetensi, kompetensi dasar dan indikator) yang telah ditetapkan dalam kurikulum. Melalui tes guru dapat dengan mudah mendeteksi peserta didik yang sudah menguasai atau yang belum menguasai materi. Selain itu tes juga dapat digunakan guru untuk mendeteksi berhasil tidaknya pembelajaran yang telah dilakukan.

Aspek penting lain yang menentukan keberhasilan evaluasi hasil pendidikan adalah teknik pengukuran. Wiersma dan Jurs (1990: 7) mengemukakan bahwa pengukuran merupakan kegiatan untuk mendapatkan informasi secara kuantitatif atau dengan kata lain merupakan prosedur untuk menentukan skor siswa atau menentukan nilai suatu objek atau harga suatu objek, dan untuk dapat menentukan suatu nilai atau harga suatu objek tersebut diperlukan adanya suatu ukuran atau kriteria. Mulyana (2004: 13) menambahkan kegiatan pengukuran sifat suatu objek merupakan suatu kegiatan menentukan kuantitas sifat suatu objek melalui aturan-aturan tertentu yang diperoleh benar-benar mewakili sifat dari suatu objek yang dimaksud, dan dalam pengukuran juga dilakukan proses menentukan angka untuk individu atau menentukan karakteristik individu menurut aturan tertentu yang hasilnya berupa data kuantitatif. Masidjo (1996: 12) mengemukakan hal yang sama bahwa kuantitas yang diperoleh dari suatu pengukuran sifat suatu objek disebut skor. Naga (2009: 25) menambahkan bahwa pengukuran digunakan sebagai pemberian bilangan kepada atribut dari subjek menurut aturan, data atribut dapat berbentuk bilangan atau berbentuk numerik, dan data atribut berbentuk numerik dikenal sebagai skor. Proses pemberian angka seperti dilakukan dalam tes merupakan suatu bentuk pengukuran.

Pengukuran dengan menggunakan alat ukur berupa tes tertulis harus menggunakan bentuk tes bervariasi yang dapat mengasah kemampuan berpikir siswa menjadi lebih kritis, misalnya guru dapat memodifikasi bentuk tes pilihan ganda menjadi pilihan ganda asosiasi, pilihan ganda sebab-akibat, bentuk tes menjodohkan, bentuk tes *true-false* dan sebagainya. Sesuai yang dikemukakan Zainul dan Nasution (dalam Moeis, 2006: 31-38) bahwa informasi yang diberikan oleh tes pilihan ganda lebih kaya, artinya butir soal ini dapat memberikan informasi tentang siswa lebih banyak kepada guru atau penyusun tes lainnya, terutama bila butir soal ini memiliki homogenitas yang tinggi.

Bentuk tes pilihan ganda yang biasa digunakan oleh guru yaitu bentuk pilihan ganda biasa atau konvensional, di mana siswa dapat memilih jawaban paling tepat dari beberapa alternatif jawaban yang ada. Alasan guru yang jarang memodifikasi bentuk tes adalah keterbatasan waktu untuk membuat soal dengan bentuk yang bervariasi, dan guru merasa kesulitan dalam membuat bentuk tes pilihan ganda model lain seperti pilihan ganda asosiasi, atau pilihan ganda sebab-akibat. Padahal dengan model pilihan ganda asosiasi atau kompleks dapat membuat siswa mampu berpikir kritis, di mana dalam menjawab butir soal siswa tidak asal menebak jawaban yang benar, dan siswa semakin familiar melihat bentuk-bentuk pilihan ganda yang dimodifikasi, sehingga ketika siswa mengikuti tes masuk perguruan tinggi atau kejuaraan olimpiade siswa sudah terbiasa dengan bentuk tes pilihan ganda yang bervariasi. Kenyataannya guru lebih senang menggunakan bentuk tes uraian atau bentuk pilihan ganda biasa, padahal guru harus menggunakan soal yang valid dan reliabel untuk mengukur kemampuan kognitif siswa. Salah satu bentuk yang dapat dikembangkan adalah bentuk tes pilihan ganda asosiasi.

Hasil penelitian Moeis (2006: 31-38) menyatakan bahwa bentuk soal pilihan ganda kompleks

mempunyai tingkat kesulitan yang tinggi, sehingga menuntut siswa untuk menguasai dan memahami materi pelajaran secara menyeluruh. Seperti yang dikemukakan oleh Sax (dalam Moeis, 2006: 31-38) bahwa soal pilihan ganda asosiasi, analisis kasus, dan pilihan ganda menggunakan tabel atau peta memerlukan adanya kemampuan sampai pada tingkat analisis, sintesis dan evaluasi dari peserta tes. Dengan demikian, bentuk tes pilihan ganda asosiasi mampu digunakan untuk mengukur kemampuan siswa sampai pada tahap evaluasi dan dapat melatih siswa untuk menguasai materi secara kompleks dan menyeluruh.

Keberhasilan pengukuran dan tes juga ditentukan oleh metode penskoran yang digunakan. Menurut Djaali dan Muljono (2004: 120) penskoran merupakan suatu proses pengubahan jawaban instrumen menjadi angka-angka yang merupakan nilai kuantitatif dari suatu jawaban terhadap butir dalam instrumen. Penskoran yang digunakan untuk memberikan skor atau nilai dari jawaban siswa, harus menggunakan prosedur penskoran yang benar, sehingga skor tersebut dapat memberikan informasi yang akurat mengenai kemampuan kognitif siswa.

Penskoran bentuk tes berbeda-beda, sesuai tipe soal yang digunakan. Teknik penskoran yang biasa digunakan pada bentuk tes pilihan ganda yaitu dengan menghitung jumlah jawaban yang benar, sedangkan pada bentuk uraian guru biasa memberikan pembobotan sesuai kompleksitas butir tersebut. Pada hakikatnya sebelum menentukan skor terlebih dahulu harus menganalisis soal dari tingkat kedalaman materi sampai tingkat kesukaran soal, analisis soal ini dapat digunakan untuk mengetahui butir mana yang valid atau tidak valid. Dengan demikian, dengan tes yang valid akan menghasilkan skor yang akurat dan dapat dipercaya.

Sesuai pendapat Frary (dalam Ndalichako dan Rogers, 1997: 580-589) yang menyatakan bahwa ada beberapa alternatif dalam metode penskoran antara lain, *option weighting*, *answer-until-correct*, *multiple correct options*, *confidence testing*, dan *subtest selection*. Kenyataannya di sekolah, penilaian suatu tes yang dilakukan guru sebagian besar masih menggunakan penskoran tanpa menggunakan pembobotan, terutama bentuk tes pilihan ganda. Bahkan lebih memprihatinkan lagi, proses penskoran biasanya tidak dibedakan satu sama lain, hal ini sering kali menimbulkan hasil yang kurang objektif. Padahal kenyataannya, kemampuan kognitif siswa berbeda-beda. Oleh karena itu guru juga harus memperhatikan teknik penskoran yang benar dan tepat. Dengan demikian, skor yang diperoleh objektif dan siswa memperoleh hak atas nilai yang sesuai dengan kemampuan kognitif siswa.

Pendekatan yang digunakan di sekolah sampai saat ini adalah dengan skor benar atau skor yang diperoleh dengan tidak menghitung jawaban salah. Sesuai dengan pendapat Surapranata (2007: 67) bahwa bentuk tes objektif menggunakan penilaian objektif, yaitu apabila jawaban benar diberi skor 1, salah diberi skor 0. Tes objektif sering pula disebut tes dikotomi, yaitu penilaian 0-1 (*dichotomously scored* butir). Kelemahan skor yang diperoleh dengan menghitung jumlah jawaban benar dapat diatasi, salah satunya dengan pendekatan pemberian bobot pada butir-butir sesuai dengan kontribusinya, atau disebut dengan skor bobot. Pendekatan yang digunakan dalam penelitian ini berupa teknik penskoran dengan skala kategori berdasarkan tingkat kesukaran butir pada bentuk pilihan ganda biasa dan pilihan ganda asosiasi, sehingga tes dapat digunakan untuk mengukur kemampuan/kompetensi yang dimiliki siswa. Jika penilaian hanya melihat jawaban benar dengan tidak memberikan bobot pada setiap butir, hal ini bisa merugikan siswa dan tes yang digunakan menjadi tidak berfungsi dengan baik karena tidak mampu untuk mendiagnosis kemampuan siswa.

Pembobotan dapat dilakukan berdasarkan tingkat kesukaran dari soal yang diujikan, sehingga siswa merasa bahwa dirinya memperoleh penghargaan dari nilai yang diperoleh, dan siswa memperoleh kepuasan atas nilai yang diperoleh. Bobot terhadap soal dapat dilakukan dengan pengkategorian soal yang memiliki tingkat kesukaran sukar, sedang, dan mudah. Dengan demikian, siswa akan mengerjakan soal dengan tidak menerka atau akan mengerjakan dengan sebaik-baiknya karena setiap butir soal memiliki bobot yang berbeda sesuai dengan tingkat kesukaran dari setiap butir, di mana prosentase dari setiap butir-butir dengan kategori sukar, sedang dan mudah sama.

Diperkuat oleh Surapranata (2005: 92) bahwa semakin sukar soal-soal dalam perangkat tes

akan semakin besar reliabilitas tes tersebut, sebaliknya, semakin rendah tingkat kesukaran suatu soal semakin kecil pula reliabilitasnya. Suatu tes yang diberikan kepada reponden harus mempunyai tingkat kesukaran yang beragam dan tingkat kesukaran akan mempengaruhi validitas soal. Karena skor yang valid diperoleh dari tes yang valid. Jika tes yang digunakan valid, maka tes tersebut juga harus reliabel, yaitu apabila tes digunakan untuk melakukan pengukuran dari waktu ke waktu akan menghasilkan skor yang sama atau relatif sama atau ajeg. Masalah reliabilitas adalah masalah yang berkaitan dengan skor yang dihasilkan dari suatu pengukuran. Reliabilitas memberikan konsistensi yang membuat terpenuhinya syarat utama, yaitu validnya suatu hasil skor instrumen. Reliabilitas tinggi menunjukkan kesalahan varian yang minim (Musliha, 2019)

Secara umum masalah yang dikaji dalam penelitian ini adalah: “Apakah terdapat perbedaan koefisien reliabilitas antara bentuk-bentuk tes pilihan ganda dengan penggunaan teknik penskoran? Secara operasional masalah tersebut diuraikan sebagai berikut: (1) Apakah terdapat perbedaan koefisien reliabilitas antara pilihan ganda asosiasi dengan skor bobot dan koefisien reliabilitas pilihan ganda asosiasi skor benar? (2) Apakah terdapat perbedaan koefisien reliabilitas antara pilihan ganda asosiasi dengan skor bobot dan koefisien reliabilitas pilihan ganda biasa dengan skor bobot? (3) Apakah terdapat perbedaan koefisien reliabilitas antara pilihan ganda asosiasi dengan skor bobot dan koefisien reliabilitas pilihan ganda biasa dengan skor benar? (4) Apakah terdapat perbedaan koefisien reliabilitas antara pilihan ganda asosiasi dengan skor benar dan koefisien reliabilitas pilihan ganda biasa dengan skor bobot? (5) Apakah terdapat perbedaan koefisien reliabilitas antara pilihan ganda asosiasi dengan skor benar dan koefisien reliabilitas pilihan ganda biasa dengan skor benar? (6) Apakah terdapat perbedaan koefisien reliabilitas antara pilihan ganda biasa dengan skor bobot dan koefisien reliabilitas pilihan ganda biasa dengan skor benar?

Hasil penelitian ini diharapkan berguna untuk: (1) bentuk tes yang dihasilkan dari penelitian ini dapat dijadikan sebagai tes formatif atau sumatif, (2) memberikan masukan untuk Dinas Pendidikan Nasional bagian BNSP (Badan Nasional Standar Pendidikan) dalam variasi bentuk tes pada ujian nasional, (3) dijadikan sebagai landasan bagi para pengambil kebijakan dalam bidang pendidikan, terutama kebijakan yang berhubungan dengan evaluasi pembelajaran.

METODE/EKSPERIMEN

Pendekatan yang digunakan dalam penelitian ini adalah metode eksperimen dengan desain *one-way* atau Anava satu jalan. Variabel terikat adalah koefisien reliabilitas, variabel eksperimen adalah bentuk tes pilihan ganda yang terdiri atas bentuk pilihan ganda biasa dan bentuk pilihan ganda asosiasi, dan variabel atribut adalah teknik penskoran yang terdiri atas skor bobot dan skor benar. Penelitian dilakukan di 4 sekolah yaitu SMA Negeri 54, SMA Negeri 44, SMA Negeri 103 dan SMA Negeri 89 Jakarta Timur. Penarikan sampel diambil dengan cara teknik *simple random sampling* dengan menggunakan *software Minitab* untuk repikasi sampel.

Eksperimen dilakukan dengan cara memberikan tes formatif bentuk pilihan ganda biasa maupun bentuk pilihan ganda asosiasi kepada siswa kelas X SMA Negeri 54, SMA Negeri 44, SMA Negeri 103 dan SMA Negeri 89 Jakarta Timur. Pemberian tes dilakukan 1x secara serentak. Normalitas data diuji dengan menggunakan uji Liliefors, sedangkan homogenitas variabs diuji dengan menggunakan uji Bartlett. Hipotesis penelitian diuji dengan menggunakan teknik analisis varians (ANAVA) satu jalan, dilanjutkan dengan uji Tukey untuk membandingkan kelompok-kelompok data.

Tabel 1. Desain Penelitian *One-Way*

Variabel bebas	X ₁	X ₂	X ₃	X ₄
Variabel terikat				
Reliabilitas (Y)	(r ₁ , r ₂ ,... r ₃₀)	(r ₁ , r ₂ ,... r ₃₀)	(r ₁ , r ₂ , ... r ₃₀)	(r ₁ , r ₂ , ... r ₃₀)

Keterangan:

X_1 = Pilihan ganda asosiasi dengan skor bobot

X_2 = Pilihan ganda asosiasi dengan skor benar

X_3 = Pilihan ganda biasa dengan skor bobot

X_4 = Pilihan ganda biasa dengan skor benar

Y = Reliabilitas

r_1, r_2, \dots, r_{30} = koefisien reliabilitas dengan replikasi sebanyak 30 kali

HASIL DAN PEMBAHASAN

Hasil

Data hasil statistik deskriptif yang dihitung dengan *microsoft excel* terlihat pada tabel 2.

Tabel 2. Hasil Statistik Deskriptif

No	Kelompok Data	Ukuran Deskriptif	Nilai
1	X_1	Modus	0.7100
		Median	0.7305
		Mean	0.7127
		Simpangan baku	0.0799
		Varians	0.0064
2	X_2	Modus	0.5980
		Median	0.7310
		Mean	0.7216
		Simpangan baku	0.0676
		Varians	0.0045
3	X_3	Modus	0.7510
		Median	0.7325
		Mean	0.7213
		Simpangan baku	0.0668
		Varians	0.0045
4	X_4	Modus	0.6840
		Median	0.7895
		Mean	0.7714
		Simpangan baku	0.0646
		Varians	0.0042

Hasil penelitian hipotesis melalui analisis varians satu jalan dapat dirangkum pada tabel 3 berikut:

Tabel 3. Hasil Perhitungan Anava Satu Jalan

Sumber Varians	JK	Db	RJK	F_{hit}	$F_{tabel} (\alpha=0.05)$	Keterangan
Antar	0,0644	3	0,0215	4,3878	2,6828	Signifikan
Dalam	0,5677	116	0,0049			
Total	0,6321	119				

Berdasarkan hasil perhitungan ANAVA satu jalan pada tabel 3 di atas, dapat dianalisis bahwa terdapat perbedaan koefisien reliabilitas antara bentuk tes pilihan ganda dan teknik penskoran. Pemberian bentuk tes pilihan ganda yang berbeda dengan teknik penskoran yang berbeda memberikan konsekuensi pada perbedaan koefisien reliabilitas. Hasil analisis menunjukkan adanya perbedaan yang signifikan, sehingga dilanjutkan uji Tukey. Hasil uji Tukey dapat dirangkum dalam Tabel 4.

Tabel 4. Hasil Uji Tukey

Kelompok	Q_{hit}	Q_{tabel}	Keputusan	Keterangan
x_1 dan x_2	0,699	3,850	Terima H_0	Tidak signifikan
x_1 dan x_3	0,671		Terima H_0	Tidak signifikan
x_1 dan x_4	4,596		Tolak H_0	Signifikan
x_2 dan x_3	0,029		Tolak H_0	Signifikan

Kelompok	Q_{hit}	Q_{tabel}	Keputusan	Keterangan
x_2 dan x_4	3,897		Tolak H_0	Signifikan
x_3 dan x_4	3,925		Tolak H_0	Signifikan

Pengujian hipotesis pertama menunjukkan bahwa, terdapat perbedaan rata-rata koefisien reliabilitas antara bentuk tes pilihan ganda dan teknik penskoran. Hasil ini sesuai dengan hasil analisis dengan menggunakan Anava satu jalan. Penggunaan bentuk tes pilihan ganda yang terdiri dari pilihan ganda asosiasi dan pilihan ganda biasa, yang diberikan masing-masing teknik penskoran dengan skor bobot dan skor benar menyebabkan koefisien reliabilitas yang diperoleh berbeda.

Bentuk tes yang berbeda menyebabkan peserta tes mengerjakan dengan teknik analisis yang berbeda, sehingga dapat mempengaruhi koefisien reliabilitas (Sarea & Ruslan, 2019). Bentuk tes pilihan ganda menunjukkan bahwa penilaian dilakukan secara objektif, serta waktu tes yang digunakan lebih efisien jika dibandingkan dengan bentuk essay sehingga mempengaruhi koefisien reliabilitas, selain itu siswa dituntut untuk menganalisis soal dan berpikir kritis. Hal ini sesuai dengan pendapat Ruggiero (dalam Moeis, 2006) bahwa untuk memiliki pemikiran yang kritis, seseorang harus dilatih dengan butir soal pilihan ganda yang tidak hanya dapat menemukan kebenaran tetapi harus diikuti oleh pendekatan yang menjurus pada penemuan kebenaran, dengan menggunakan modifikasi butir soal pilihan ganda seperti analisis suatu kasus, gambar, tabel, peta atau karikatur, atau butir pilihan ganda dengan jawaban yang dipilih siswa harus disertai dengan penjelasan atau alasan.

Secara umum teknik penskoran digunakan untuk memberikan angka pada jawaban yang direspon siswa, pemberian skor harus memperhatikan kaidah atau aturan dalam hal ini guru tidak sembarangan memberikan angka pada hasil jawaban siswa. Dengan demikian menyebabkan terdapat perbedaan rata-rata koefisien reliabilitas antara bentuk tes pilihan ganda dan teknik penskoran. Hal ini diperkuat oleh Nitko (dalam Busnawir, 2011) selain faktor skoring, beberapa faktor lain yang perlu diperhatikan sehubungan dengan kestabilan koefisien reliabilitas, antara lain adalah: (1) ukuran sampel, (2) variasi dalam group responden, dan (3) perbedaan metode pendugaan reliabilitas. Aiken, Bennet dan Ward (dalam Ndalichako dan Rogers, 1997) menambahkan bahwa tes pilihan ganda mempunyai beberapa keunggulan yaitu, mudah untuk dinilai, skor objektif, mudah dalam analisis butir, dan reliabilitas yang tinggi.

Pengujian hipotesis kedua menunjukkan ketidaksesuaian hipotesis dan hasil analisis. Hasil analisis bahwa koefisien reliabilitas pilihan ganda asosiasi dengan skor bobot lebih rendah daripada koefisien reliabilitas pilihan ganda asosiasi dengan skor benar. Ketidaksesuaian ini dilihat berdasarkan hasil observasi ketika tes berlangsung yaitu sekolah yang digunakan berbeda, sehingga kemampuan analisis siswa juga berbeda. Di lapangan anak tidak terbiasa menggunakan skor bobot, sehingga siswa cenderung mengerjakan butir yang dibobot dengan butir tanpa bobot hampir sama, dan kelemahannya pada butir yang diberi bobot 3 (sukar) siswa tidak mengerjakan butir tersebut. Bentuk tes pilihan ganda asosiasi skor bobot sudah dibuat sesuai hasil uji coba berdasarkan tingkat kesukaran butir soal, sesuai dengan yang dikemukakan Sudijono (2008: 160), bahwa bobot skor tiap soal bisa ditentukan sebelum tes digunakan, yaitu berdasarkan tingkat kompleksitas atau kesulitannya, yang kompleks atau sulit diberi bobot yang lebih tinggi dibanding dengan yang lebih mudah. Dengan demikian valid dan reliabel tes dapat dipertanggungjawabkan, sehingga instrumen yang digunakan reliabel.

Ketidaksesuaian hasil analisis dengan hipotesis dikarenakan kelas yang kurang kondusif ketika tes berlangsung, sehingga ada beberapa siswa mungkin menanyakan jawaban pada teman yang lain, padahal dalam satu meja sudah diberikan bentuk tes pilihan ganda yang berbeda peneliti sulit mengontrol satu kelas tersebut. Sehingga terjadi ketidakwajaran pengukuran saat tes berlangsung. Diperkuat Hulin *et al.* (dalam Ahiri, 2018) penyebab lain dari ketidakwajaran pengukuran adalah kecurangan dalam tes, misalnya siswa-siswa yang menyontek, baik yang melirik pada lembar

jawaban peserta di sampingnya ataupun yang membawa ringkasan materi. Lembar jawaban bagi siswa yang curang dalam tes mungkin akan tampak ganjil, kelompok jawaban benar akan bercampur dengan kelompok jawaban yang hampir semuanya tidak benar, akibatnya siswa-siswa seperti ini akan memperoleh skor yang tidak tepat, yakni tidak sesuai dengan kemampuan mereka yang sebenarnya.

Pengujian hipotesis ketiga menunjukkan ketidaksesuaian hipotesis dan hasil analisis. Hasil analisis bahwa koefisien reliabilitas pilihan ganda asosiasi dengan skor bobot lebih rendah daripada koefisien reliabilitas pilihan ganda biasa dengan skor bobot. Ketidaksesuaian ini dilihat berdasarkan hasil observasi siswa merasa kesulitan mengerjakan bentuk tes pilihan ganda asosiasi, karena belum pernah menemukan bentuk tes pilihan ganda asosiasi sebelumnya. Sehingga ketika menjawab soal siswa cenderung merasa cemas dan khawatir dalam menjawab soal, selain itu siswa menjadi tidak percaya diri dalam menjawab setiap butir soal. Faktanya di sekolah yang digunakan sebagai penelitian terjadi miskonsepsi pada materi hidrokarbon yang diberikan guru, sehingga ada butir soal yang dijawab tidak sesuai dengan kunci jawaban. Selain itu, guru cenderung memberikan tugas LKS atau latihan soal, namun guru tidak sering melakukan pembahasan atau melakukan penilaian diagnostik terhadap latihan yang sudah diberikan. Hal ini dapat dikatakan terjadi ketidakwajaran dalam pengukuran.

Sesuai dengan pendapat Levine dan Rubin (dalam Ahiri, 2018), ketidakwajaran pengukuran (*Inappropriateness Measurement*), yang digunakan untuk menunjukkan seberapa tepat tes dapat mengukur karakteristik peserta tes. Ahiri (2018) menambahkan ketidakwajaran pengukuran hanya terbatas pada keanehan pola jawaban siswa dalam tes, ada jawaban yang benar terhadap butir-butir tes yang sukar dan ada jawaban salah terhadap butir-butir tes yang mudah. Dalam hal ini, pola jawaban yang dihasilkan oleh peserta tes kadang-kadang tidak normal. Ada sejumlah jawaban benar terhadap butir-butir sulit pada seperdua tes yang pertama dan ada sejumlah jawaban salah terhadap butir-butir tes yang mudah pada seperdua tes berikutnya, atau peserta tes yang kreatif mungkin memberikan penafsiran yang berbeda terhadap beberapa butir tes yang mudah.

Nitko, Wiersma dan Jurs (dalam Ahiri, 2018) menambahkan bahwa tekanan mental peserta tes, seperti perasaan cemas, khawatir, takut akan gagal, kekurangmampuan dalam menulis, dan perasaan tidak disukai, menyebabkan peserta tes tidak berhasil menjawab benar butir tes. Ahiri (2018) menambahkan kecemasan tes bersumber dari aspek psikologis siswa. Dalam menghadapi tes, setiap siswa akan merasakan tingkat kecemasan tertentu baik berupa kecemasan tinggi, sedang, ataupun rendah. Akibat kecemasan tersebut, peserta tes mungkin akan memperoleh skor tes yang tidak sesuai dengan kemampuannya, kendatipun semua butir tes valid. Peserta tes dapat saja memperoleh skor yang lebih tinggi dari skor yang seharusnya diperoleh berdasarkan kemampuannya atau sebaliknya dapat pula memperoleh skor yang lebih rendah dari kemampuannya.

Menurut teori pilihan ganda asosiasi menuntut peserta tes mengerjakan dengan analisis yang kritis, dan akurat sedangkan pada bentuk pilihan ganda biasa peserta tes akan dengan mudah untuk menebak jawaban. Walaupun kedua bentuk tes diberikan bobot berdasarkan kompleksitas butir, peserta tes akan mengerjakan dari butir mudah sampai sukar tetapi jika pilihan ganda asosiasi yang kompleks diberikan pembobotan hasilnya lebih akurat, namun pada pilihan ganda biasa walaupun diberi bobot karena masih adanya faktor *lucky* dalam memilih jawaban menyebabkan rendahnya koefisien reliabilitas.

Bentuk tes pilihan ganda asosiasi merupakan bentuk tes *tipe High Order Thinking Skills* yang melatih anak mempunyai kemampuan berpikir lebih baik (Widoyoko, 2019) dibandingkan pilihan ganda biasa. Hal ini dikarenakan dengan bentuk tes pilihan ganda asosiasi siswa akan menemukan beberapa pilihan dan perlu dianalisis lebih lanjut untuk menjawab pertanyaan dengan benar, sedangkan pilihan ganda biasa siswa akan mempunyai kesempatan untuk menebak jawaban karena hanya satu jawaban yang benar. Sesuai dengan pendapat Sax (dalam Moeis, 2006) bahwa bentuk soal pilihan ganda kompleks mempunyai tingkat kesulitan yang tinggi, sehingga menuntut siswa untuk menguasai dan memahami materi pelajaran secara menyeluruh, selain itu soal pilihan ganda

asosiasi, analisis kasus dan pilihan ganda menggunakan tabel atau peta memerlukan adanya kemampuan sampai pada tingkat analisis, synthesis dan evaluasi dari peserta tes.

Pengujian hipotesis keempat menunjukkan bahwa koefisien reliabilitas pilihan ganda asosiasi dengan skor bobot lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa dengan skor benar. Pilihan ganda asosiasi yang termasuk jenis pilihan ganda kompleks mempengaruhi peserta tes dalam mengerjakan butir soal, kemudian didukung dengan teknik penskoran yang diberikan pembobotan pada setiap butir menurut tingkat kesukaran soal. Hal ini mempengaruhi kehati-hatian peserta dalam menjawab soal. Sedangkan pilihan ganda biasa yang sering digunakan membuat peserta tes mengerjakan lebih mudah dan bisa menerka karena hanya ada satu jawaban paling benar kemudian pilihan ganda biasa tidak diberikan bobot, maka faktor *guessing* bisa mempengaruhi rendahnya koefisien reliabilitas (Bhakti, 2015).

Hal ini sesuai dengan Azwar (2007: 114) bahwa pada tes pilihan ganda terdapat peluang menjawab benar sekalipun jawaban itu dipilih hanya dengan cara menebak (*lucky guess*), padahal seharusnya mereka yang tidak tau jawaban yang benar tidaklah berhak untuk mendapatkan angka. Demikian pula yang dikemukakan Surapranata (2007: 133) pengurangan faktor menebak akan meningkatkan reliabilitas dan validitas, sepanjang alternatif jawaban itu bagus (*plausible*) dan soalnya dibuat dengan baik. Dengan demikian koefisien reliabilitas pilihan ganda asosiasi dengan skor bobot lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa skor bobot.

Pengujian hipotesis kelima menunjukkan bahwa koefisien reliabilitas pilihan ganda asosiasi dengan skor benar lebih rendah daripada koefisien reliabilitas pilihan ganda biasa dengan skor bobot. Kesesuaian hipotesis dan hasil analisis dikemukakan bahwa pilihan ganda asosiasi yang memiliki lebih dari satu pilihan jawaban menuntut siswa untuk berpikir secara kritis namun dengan butir yang tanpa bobot peserta yang mengerjakan butir soal dengan skor benar mengerjakan secara berurutan, ketika menemukan butir yang sukar siswa akan terus mencari sampai mendapat jawaban sehingga menyebabkan waktu *testing* berkurang atau habis. Skor dari jawaban siswa yang dikerjakan pada butir yang valid dan diselesaikan dengan cermat dan baik akan mempengaruhi reliabilitas. Sedangkan pilihan ganda biasa yang diberikan pembobotan membuat peserta mengerjakan dengan mengurut butir berdasarkan bobot yang diberikan.

Pilihan ganda biasa dengan menggunakan skor bobot digunakan untuk mengatasi penilaian yang masih konvensional yaitu pemberian angka berdasarkan skor benar (*correct score*). Hal ini dikarenakan dengan teknik penskoran yang sesuai siswa akan merasa dihargai dari hasil yang dikerjakan. Teknik penskoran skor bobot siswa cenderung mengerjakan butir yang memiliki bobot rendah (mudah) kemudian naik ke level butir soal yang lebih sukar.

Kestabilan koefisien reliabilitas lebih tinggi pada teknik penskoran dengan skor bobot karena skor hasil tes dengan penskoran yang diberi bobot lebih reliabel dibandingkan dengan skor siswa yang diperoleh dengan skor benar. Dengan demikian, bentuk tes yang diberikan pembobotan bersifat ajeg dan jika digunakan untuk pengukuran berulang-ulang akan menghasilkan skor yang konstan. Hal ini diperkuat oleh Ahiri (2018) bahwa reliabilitas didefinisikan seberapa besar konsistensi skor tes yang dicapai peserta tes pada pengujian ulang, definisi ini akan memuaskan jika skor tes dapat menggambarkan kemampuan peserta tes, jika tidak maka skor tes tidak sistematis, tidak dapat diulangi atau tidak terikat. Reliabilitas juga diartikan sebagai indikator ketidakhadiran kesalahan acak. Jika kesalahan acak dapat diperkecil maka skor tes akan lebih konsisten dari suatu pengujian ke pengujian berikutnya.

Pengujian hipotesis keenam menunjukkan bahwa koefisien reliabilitas pilihan ganda asosiasi dengan skor benar lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa dengan skor benar. Secara umum bentuk tes pilihan ganda dapat meningkatkan kemampuan siswa untuk berpikir kritis. Bentuk tes pilihan ganda termasuk bentuk tes yang banyak digunakan untuk mengukur kemampuan siswa, sehingga dalam penelitian ini dapat terlihat di lapangan yaitu siswa yang mengerjakan bentuk tes pilihan ganda asosiasi lebih *interest* daripada pilihan ganda biasa. Walaupun ketika tes itu diberikan siswa merasa kesulitan dengan teknik menjawabnya, namun diakhir tes mereka merasakan

bahwa dengan bentuk tes pilihan ganda asosiasi siswa seperti mengerjakan soal dalam bentuk uraian karena mereka harus menguraikan setiap pilihan yang diberikan. Dengan demikian pilihan ganda asosiasi dibandingkan dengan pilihan ganda biasa mempengaruhi siswa dalam menjawab pertanyaan.

Kemampuan guru dalam mengkonstruksi butir soal itu yang menjadi kendala terutama bentuk tes pilihan ganda asosiasi. Kendala yang guru hadapi agar dalam setiap tes bisa memberikan seperangkat butir dengan model bervariasi adalah waktu, mengkonstruksi bentuk tes pilihan ganda asosiasi memerlukan waktu yang lama dan sulit dalam membuat pilihan apalagi pilihan jawaban lebih dari satu. Tes yang diberikan kepada siswa harus memiliki bentuk tes yang bervariasi agar siswa memiliki kemampuan untuk menjawab bukan sekedar menerka jawaban, siswa akan memiliki kemampuan untuk berpikir dengan logika dan kritis. Oleh karena itu, dalam penelitian ini memberikan bentuk tes pilihan ganda asosiasi yang jarang digunakan di sekolah. Sedangkan bentuk pilihan ganda biasa memudahkan siswa untuk menerka jawaban, sehingga siswa akan cenderung menebak bukan menganalisis jawaban.

Bentuk tes pilihan ganda asosiasi dan biasa keduanya masih memungkinkan anak untuk menebak jawaban, namun pada bentuk tes pilihan ganda asosiasi sulit untuk menebak jawaban karena memerlukan analisis dan berpikir secara kritis untuk menjawab soal tersebut. Namun, pada bentuk tes pilihan ganda biasa karena hanya satu jawaban yang benar akan mudah bagi siswa untuk menebak jawaban. Hal ini menyebabkan koefisien reliabilitas pilihan ganda asosiasi dengan skor lebih tinggi daripada pilihan ganda biasa dengan skor benar. Hal ini sesuai pendapat Sukati (dalam Moeis, 2006) bahwa sebenarnya tes bentuk soal pilihan ganda kompleks lebih dapat dijawab benar oleh siswa yang berprestasi atau berkemampuan tinggi daripada bentuk soal pilihan ganda biasa, sebab mereka dapat memberikan jawaban dengan sedikit kemungkinan tebakan.

Tes dengan waktu yang dibatasi kemudian setiap butir diberikan bobot berdasarkan tingkat kesukaran dibandingkan dengan butir yang tidak diberikan kategori tingkat kesukaran, maka siswa akan mengerjakan tes dengan santai dan percaya diri atau tidak terjadi kecemasan saat tes. Kecemasan terjadi karena siswa memperhatikan waktu yang diberikan dengan butir soal yang sedang diselesaikan, sehingga dengan adanya pembobotan butir soal memudahkan siswa untuk mengerjakan soal dengan menggunakan waktu yang efisien. Kegiatan pembelajaran juga akan semakin efektif karena siswa mampu menyelesaikan setiap butir dengan menganalisis sesuai dengan kemampuannya. Hal ini diperkuat oleh Hopkins dan Antes (dalam Ahiri, 2018) pembelajaran berhubungan dengan persiapan siswa secara akademik dan secara emosional untuk menyelesaikan tes disertai dengan aktivitas untuk membantu siswa membangun rasa percaya diri, menumbuhkan kesadaran akan kemampuannya, dan menimbulkan optimisme untuk mencapai tujuannya.

Pengujian hipotesis ketujuh menunjukkan bahwa koefisien reliabilitas pilihan ganda biasa dengan skor bobot lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa dengan skor benar. Hal ini sesuai antara hipotesis dan hasil analisis, karena jika bentuk tes diberikan pembobotan peserta akan menggunakan waktu secara efisien karena mereka mengerjakan butir soal berdasarkan bobot yang diberikan pada setiap butir, sedangkan yang tanpa bobot mereka cenderung mengerjakan butir dengan waktu yang kurang efisien.

Pilihan ganda biasa dengan skor bobot diberikan berdasarkan tingkat kesukaran butir, sehingga butir yang diberikan sudah dilakukan validasi empiris sebelumnya, dengan demikian butir yang diberikan kepada peserta sudah memiliki bobot berbeda-beda. Dengan pemberian bobot pada setiap butir menyebabkan peserta dapat menjawab butir dengan baik dan menggunakan waktu dengan efisien, selain itu mengurangi tingkat kecemasan karena siswa akan mengerjakan dari butir mudah sampai butir yang sukar. Hal ini mengakibatkan siswa yang mengerjakan butir tidak sembarang menjawab butir soal, dan mengurangi faktor menebak. Koefisien reliabilitas lebih stabil jika dalam butir soal peserta mengerjakan dengan analisis atau tidak menerka jawaban. Senada dengan pendapat Crocker dan Algina (1986: 106) bahwa tinggi rendahnya reliabilitas dipengaruhi oleh beberapa faktor, seperti faktor menebak, keadaan saat tes, kesalahan administrasi, konten

sampel, kesalahan dalam penskoran, dan fluktuasi peserta saat ujian.

PENUTUP

Berdasarkan hasil pengujian hipotesis dan pembahasan penelitian, maka beberapa kesimpulan sebagai berikut, (1) Terdapat perbedaan koefisien reliabilitas antara bentuk tes pilihan ganda dengan teknik penskoran. (2) Koefisien reliabilitas pilihan ganda asosiasi dengan skor benar lebih rendah daripada koefisien reliabilitas pilihan ganda biasa dengan skor bobot. (3) Koefisien reliabilitas pilihan ganda asosiasi dengan skor benar lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa dengan skor benar. (4) Koefisien reliabilitas pilihan ganda asosiasi dengan skor benar lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa dengan skor benar. (5) Koefisien reliabilitas pilihan ganda biasa dengan skor bobot lebih tinggi daripada koefisien reliabilitas pilihan ganda biasa dengan skor benar.

DAFTAR PUSTAKA

- Ahiri, J. (2018). Keefektifan Bimbingan Tes Dalam Mengatasi Ketimpangan Skor Tes Akuntansi Ditinjau Dari Kecemasan Siswa. *Jurnal Teknodik*, 21(3), 140-163.
- Azwar, S. (2007). *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.
- Bhakti, Y. B. (2015). Pengaruh Jumlah Alternatif Jawaban dan Teknik Penskoran Terhadap Reliabilitas Tes. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, 5(1).
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Florida: Holt, Rinehart, and Winston, Inc.
- Djaali dan Pudji Muljono. (2004). *Pengukuran dalam Pendidikan*. Jakarta: PPs UNJ.
- Masidjo, Ign. (1996). *Penilaian Pencapaian Hasil Belajar*. Jakarta: Kanisius.
- Moeis, Syarif. (2006). "Pengaruh Bentuk Tes Objektif terhadap Kemampuan Berpikir Kritis dalam Pendidikan Sejarah Ditinjau dari Sikap Siswa." *Percikan Iman*. Vol.70, pp. 31-38.
- Mulyana, et al. (2004). *Evaluasi Pendidikan*. Jakarta: Lembaga Akta Mengajar UNJ.
- Muslich, Masnur. (2010). *Authentic Assessment: Penilaian Berbasis Kelas dan Kompetensi*. Bandung: Refika Aditama.
- Musliha, M. (2019). *Analisis butir soal tes objektif bentuk multiple choice buatan guru penilaian akhir semester I mata pelajaran pendidikan agama Islam dan budi pekerti (PAI & BP) kelas X tahun pelajaran 2018/2019 di SMA N 2 Pemalang dan SMA PGRI 1 Taman Pemalang* (Doctoral dissertation, UIN Walisongo).
- Naga, Dali. S. (2009). *64 Rumus Terapan Probabilitas dan Sekor pada Hipotesis Statistika*. Jakarta: Grasindo.
- Ndalichako, Joyce L. dan W. Todd Rogers. (1997). "Comparison of Finite State Score Theory, Classical Test Theory, and Item Response Theory in Scoring Multiple-Choice Items." *Educational and Psychological Measurement*. Vol. 57 (4), pp. 580-589.
- Sarea, M. S., & Ruslan, R. (2019). Karakteristik Butir Soal: Classical Test Theory Vs Item Response Theory?. *Didaktika: Jurnal Kependidikan*, 13(1), 1-16.

- Surapranata, S. (2005). *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*. Bandung: Rosdakarya.
- Surapranata, S. (2007). *Panduan Penulisan Tes Tertulis Implementasi Kurikulum 2004*. Bandung: Remaja Rosdakarya.
- Syukur, F. (2008). *Teknologi Pendidikan*. Semarang: RaSAIL Media Group.
- Widoyoko, S. E. P. (2019, March). Penilaian Pembelajaran Berbasis Hots untuk Meningkatkan Kemampuan Berpikir Kritis Siswa. In *Seminar Nasional Pendidikan Dasar* (Vol. 1, No. 01).
- Wiersma, William dan Stephen G. Jurs. (1990). *Educational Measurement and Testing*. New York: The University Teledo Allyn and Bacon.