

PERBANDINGAN EFEKTIFITAS PENGEMBANGAN INSTRUMEN TES SUMATIF FISIKA MENGGUNAKAN METODE *SELF* DAN *FEEDBACK REVISION*

Kenny Anindia Ratopo, Sutadi Waskito, Dewanto Harjunowibowo

Universitas Sebelas Maret

Jl. Ir. Sutami 36 A Surakarta, Telp/Fax (0271) 648939

E-mail : kenny_anindia@yahoo.com

Abstrak

Tujuan dari penelitian ini adalah membandingkan efektifitas dua metode revisi untuk mengembangkan instrumen tes sumatif Fisika yang sesuai dengan karakteristik tes Fisika yang baik dan baku.

Jenis penelitian ini adalah penelitian pengembangan (*Research and Development*). Pengumpulan data dilakukan melalui teknik tes dan non tes. Sumber data berupa pola respon siswa pada tes. Analisis data pada kelompok kecil dilakukan secara kualitatif yang meliputi materi, konstruksi dan bahasa serta kuantitatif dengan menggunakan program MicroCat ITEMAN versi 3.00 untuk mengetahui taraf kesukaran, daya beda dan efektifitas pengecoh. Pengembangan tes dilakukan menggunakan tahapan: analisis kebutuhan, penyusunan spesifikasi tes, penulisan soal tes, penelaahan tes secara kualitatif, revisi I, uji kelompok kecil, analisis butir soal secara kuantitatif, revisi II dan uji kelompok besar. Pada tahapan revisi II digunakan dua metode revisi, yaitu *self revision* dan *feedback revision*.

Berdasarkan hasil analisis data yang dilakukan dapat disimpulkan bahwa dari kedua metode revisi tersebut efektifitas metode *feedback revision* lebih tinggi daripada *self revision*. Hasil reliabilitas dari 20 butir soal revisi isi kedua metode sama-sama tergolong tinggi, yaitu 0,668 untuk metode *feedback revision* dan 0,633 untuk metode *self revision*. Untuk taraf kesukaran soal dari 20 butir soal yaitu dengan metode dengan metode *feedback revision* terdapat 2 butir soal tergolong mudah, 12 butir soal tergolong sedang dan 6 soal tergolong sukar. Sedangkan pada metode *self revision* terdapat 13 butir soal tergolong sedang dan 7 butir soal tergolong sukar. Disamping itu, taraf kesukaran untuk 6 butir soal yang direvisi dengan kedua metode tersebut diperoleh hasil metode *feedback revision* lebih unggul dibandingkan dengan metode *self revision*. Pada metode *feedback revision* diperoleh 5 soal kategori sedang dan 1 soal kategori sukar. Sedangkan pada metode *self revision* diperoleh hasil 4 soal kategori sedang dan 2 soal kategori sukar. Sedangkan daya beda untuk 6 butir soal yang direvisi dengan kedua metode tersebut diperoleh hasil lebih unggul dengan menggunakan metode *feedback revision* dibandingkan dengan menggunakan metode *self revision*. Hasil daya beda dari 20 soal revisi isi untuk kedua metode tersebut yaitu dengan metode *self revision* terdapat 1 soal kategori sangat baik, 10 soal kategori baik dan 9 soal kategori cukup. Dengan metode *feedback revision* terdapat 10 soal kategori baik, 9 soal kategori cukup dan 1 soal kategori jelek.

Kata kunci : pengembangan, tes sumatif, *self revision*, *feedback revision*.

I. PENDAHULUAN

Berdasarkan hasil supervisi dan evaluasi keterlaksanaan KTSP 2009 dan masukan pada pelaksanaan bimtek oleh Direktorat Pembinaan SMA diketahui bahwa masih banyak ditemukan guru yang belum memahami dan mampu mengembangkan soal sesuai dengan mekanisme dan prosedur penilaian. Kondisi tersebut mengakibatkan hasil penilaian peserta didik belum sepenuhnya menggambarkan tingkat pencapaian kompetensi siswa yang sesungguhnya (Juknis Analisis Butir Soal SMA, 2010:118).

Fakta-fakta di atas menunjukkan bahwa masih banyak guru yang mengalami kesulitan dalam mengembangkan soal, menganalisis serta menginterpretasikannya. Meskipun, jika dilihat secara umum guru dengan pengalamannya mengajar serta pengetahuannya terhadap karakteristik peserta didiknya memiliki kemampuan untuk memahami dan mengembangkan soal. Namun, karena minimnya sosialisasi tentang prosedur penyusunan instrumen tes, seringkali guru mengembangkan tes tanpa menggunakan prosedur yang benar.

Maka sangat perlu untuk mengembangkan instrumen tes sumatif yang dapat dipertanggungjawabkan. Proses penyusunan tes sumatif ini disesuaikan dengan prosedur penyusunan tes yang baku, sehingga informasi yang didapatkan melalui tes sumatif dapat mencerminkan data yang sebenarnya.

Tujuan dari penelitian ini adalah mengetahui keefektifitasan mengembangkan instrumen tes sumatif Fisika yang sesuai dengan karakteristik tes Fisika yang baik dan baku menggunakan metode revisi yang berbeda.

II. PEMBAHASAN

2.1 Metode Penelitian

Penelitian pengembangan ini bertujuan untuk mendapatkan produk alat penilaian pembelajaran berupa instrumen tes sumatif Fisika yang memenuhi kriteria validitas dan reliabilitas yang baik. Dalam penelitian ini digunakan model prosedural yaitu bersifat deskriptif dan menggariskan langkah-langkah sistematis yang harus diikuti untuk menghasilkan produk yang baik (Tim Puslitjaknov, 2008:8).

Penelitian ini mengembangkan alat ukur untuk menghasilkan instrumen tes yang memiliki karakteristik internal dan tingkat validitas serta reliabilitas yang baik. Instrumen ini disusun dalam bentuk tes pilihan ganda (*multiple choice*). Penelitian ini menggunakan jenis penelitian pengembangan (*Research and Development*). Jenis penelitian *Research and Development* mengungkapkan pola dan perurutannya melalui tahapan perubahan sebagai fungsi waktu. Alasan penggunaan jenis ini didasarkan pada pemikiran bahwa *Research and Development* ditunjukkan untuk menentukan pola pembahasan dalam rangka meramalkan produk di masa yang akan datang. Dalam kaitan ini, perolehan model lewat uji coba merupakan bagian penting dalam penelitian pengembangan yang dilakukan. Tujuannya agar model yang dibuat dapat diuji coba dan diperbaiki sehingga dapat digunakan di sekolah.

Prosedur penelitian memaparkan prosedur yang ditempuh oleh peneliti dalam membuat produk mengadopsi dari prosedur penyusunan tes hasil atau prestasi belajar menurut Djemari Mardapi (2008: 88), tahapan dari prosedur tersebut adalah: (1) menyusun spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan ujicoba tes, (5) menganalisis butir soal, hingga tahap (6) memperbaiki tes.

Pada tahap penyusunan spesifikasi tes yaitu mencakup kegiatan : menentukan tujuan tes, menyusun kisi-kisi, menentukan bentuk tes, menentukan panjang tes. Tahap penulisan soal merupakan langkah penjabaran indikator menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan perincian pada kisi-kisi yang telah dibuat. Dalam penulisan instrumen menggunakan pedoman yang nantinya digunakan dalam penelaahan butir soal yang meliputi aspek materi, konstruksi dan bahasa.

Pada tahap telaah butir soal, yaitu menelaah butir soal secara deskriptif, ini dilakukan berdasarkan kaidah penulisan butir soal, yaitu telaah dari segi materi, konstruksi, bahasa, kebenaran kunci jawaban serta pedoman penskorannya yang dilakukan oleh beberapa penelaah. Pada penelitian ini validasi dilakukan oleh ahli yaitu dosen pembimbing dan guru bidang studi. Dengan telaah soal ini diharapkan dapat memperbaiki kualitas soal yang dibuat. Setelah soal tersebut divalidasi oleh ahli maka soal-soal yang kurang baik

direvisi tahap pertama. Revisi tahap pertama ini meliputi revisi dari segi materi, konstruksi, bahasa, kebenaran kunci jawaban serta pedoman penskorannya. Kemudian soal tersebut diujicobakan ke kelompok kecil. Uji coba dilakukan untuk memperbaiki kualitas soal.

Hasil dari uji coba kelompok kecil dianalisis secara kuantitatif. Analisis butir soal dilakukan untuk mengetahui informasi penting tentang kualitas isi soal antara lain reliabilitas dan analisis kualitas butir tes (daya pembeda soal, taraf kesukaran, keefektifan pengecoh). Apabila hasil dari nilai taraf kesukaran, daya pembeda, efisiensi distraktor, dan reliabilitas telah memenuhi standar baik, maka instrumen tes yang telah dibuat dapat dikatakan memiliki karakteristik yang baik.

Sesudah analisis kelompok kecil tersebut soal yang harus direvisi isi dilakukan revisi menggunakan dua metode, yaitu *self revision* dan *feedback revision*. Metode *self revision* biasa dilakukan dengan cara merevisi soal yang perlu direvisi berdasarkan rekomendasi hasil analisis ITEMAN. Sedangkan pada metode *feedback revision*, soal yang dinyatakan harus direvisi sesuai hasil analisis ITEMAN, kemudian direvisi berdasarkan hasil penggalan informasi menggunakan angket kepada obyek uji coba. Angket tersebut berisi pertanyaan bagi siswa, bagian soal mana sajakah berdasarkan rekomendasi hasil analisis ITEMAN yang dijawab salah oleh siswa dan alasannya.

Langkah dari metode *feedback revision* yaitu hasil analisa dari kelompok kecil dengan nomor soal yang dinyatakan direvisi bagian isi ditulis pada angket yang akan diberikan peserta tes yang telah mengikuti tes uji kelompok kecil. Peserta tes memberikan alasan kenapa menjawab salah pada nomor soal yang terdapat diangket tersebut. Dari angket hasil jawaban peserta tes tersebut penulis dapat mengetahui bagian mana soal yang perlu direvisi ataupun tidak. Dengan metode *feedback revision* diharapkan soal yang direvisi sesuai dengan pemahaman peserta tes pada kelompok kecil dan dapat mewakili peserta tes selanjutnya.

Setelah soal selesai direvisi, dilakukan uji coba ke kelompok besar dengan jumlah peserta tes yang sama untuk masing-masing metode revisi. Soal yang sudah diujicoba kemudian dianalisis kembali agar dapat diperoleh informasi penting tentang kualitas

soal antara lain reliabilitas dan analisis kualitas butir tes (daya pembeda soal, taraf kesukaran). Berdasarkan hasil analisis tes tersebut, dapat diketahui mana saja instrumen tes yang diterima, perlu direvisi, ataupun ditolak.

2.2 Hasil dan Pembahasan

Sebuah tes dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, yaitu memiliki : 1) validitas, 2) reliabilitas, 3) obyektivitas, 4) praktikabilitas, 5) ekonomis (Suharsimi Arikunto, 2011:57). Untuk mendapatkan tes pilihan ganda dengan kualitas yang baik perlu dilakukan analisis butir soal. Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Analisis butir soal dilaksanakan dalam dua langkah, yakni analisis secara kuantitatif dan kualitatif seperti yang diungkapkan oleh Anastasi dan Urbina yang dikutip Depdiknas (2008: 1) “Dalam melaksanakan analisis butir soal, para penulis soal dapat menganalisis secara kualitatif, dalam kaitan dengan isi dan bentuknya, dan kuantitatif dalam kaitan dengan ciri-ciri statistiknya.”

Telaah secara diskriptif ini dilakukan berdasarkan kaidah penulisan butir soal, yaitu telaah dari segi materi, konstruksi, dan bahasa yang dilakukan oleh beberapa penelaah. Pada penelitian ini telaah soal dilakukan oleh ahli materi. Tabel 1 adalah rangkuman hasil telaah diskriptif soal.

Tabel 1. Rangkuman Telaah Diskriptif Soal yang Tidak Memenuhi Kriteria

Aspek yang ditelaah	Nomor Soal	Jumlah
Materi	-	-
Konstruksi	3, 18, 21, 22	4
Bahasa	-	-

Soal nomor 3 tidak memenuhi kriteria pada aspek konstruksi karena pokok soal tidak dirumuskan dengan singkat jelas dan tegas. Untuk soal nomor 18 tidak memenuhi aspek konstruksi karena panjang rumusan pilihan jawaban tidak relatif sama. Sedangkan nomor soal 21 dan 22 tidak memenuhi aspek konstruksi karena pokok soal tidak dirumuskan dengan singkat, jelas dan tegas.

Setelah melakukan telaah secara kualitatif, hal yang dilakukan selanjutnya adalah merevisi instrumen tes tersebut. Revisi instrumen tes ini dilakukan berdasarkan hasil

koreksi dari ahli yang mencakup aspek materi, konstruksi, dan bahasa. Revisi akan terus dilakukan sampai para ahli menilai instrumen tes tersebut layak untuk diujicobakan.

Uji coba kelompok kecil dilakukan setelah instrumen tes yang telah dibuat mendapat persetujuan dari para ahli untuk diujicobakan. Uji coba pertama ini dilakukan di kelas XI IPA 4 SMA Negeri 7 Surakarta sebanyak 26 orang siswa. Uji coba ini dapat digunakan sebagai sarana memperoleh data empiris tentang tingkat kebaikan soal yang disusun. Aspek yang akan dianalisis sebagai berikut :

a. Reliabilitas

Besarnya reliabilitas diinterpretasikan sesuai dengan interpretasi besarnya koefisien korelasi sebagai berikut :

$0,800 < r_{11}$	1,00	(sangat tinggi)
$0,600 < r_{11}$	0,800	(tinggi)
$0,400 < r_{11}$	0,600	(cukup)
$0,200 < r_{11}$	0,400	(rendah)
$0,000 < r_{11}$	0,200	(sangat rendah)

(Suharsimi Arikunto, 2011:71)

b. Taraf kesukaran

Taraf kesukaran soal dalam analisis data menggunakan ITEMAN ini ditunjukkan dengan besarnya P (*Prop. Correct*). Sesuai dengan simpulan Thorndike dan Hagen (1961) bahwa tingkat kesukaran yang baik adalah 0,3 sampai 0,7 atau kategori sedang, taraf kesukaran dibawah 0,3 dianggap sukar dan taraf kesukaran diatas 0,7 butir soal dianggap mudah (Sudijono, 2005: 327).

c. Daya beda

Analisis daya beda soal bertujuan untuk mengetahui dapat tidaknya suatu soal membedakan kelompok atas dan kelompok bawah. Perhitungan daya pembeda soal pada program ITEMAN menggunakan rumus point biserial. Pengambilan keputusan menggunakan acuan pengelompokan indeks daya pembeda berikut:

Tabel 2. Kriteria Daya Pembeda Soal

Nilai	Kategori	Interpretasi
0,00 – 0,19	<i>Poor</i> (jelek)	Daya pembeda lemah sekali (jelek), dianggap tidak memiliki daya pembeda yang baik
0,20 – 0,39	<i>Satisfactory</i> (cukup)	Memiliki daya pembeda yang

		cukup (sedang)
0,40 – 0,69	<i>Good</i> (baik)	Memiliki daya pembeda yang baik
0,70 – 1,00	<i>Excellent</i> (baik sekali)	Memiliki daya pembeda yang baik sekali
Bertanda negatif	Jelek sekali (ditolak)	Daya pembedanya jelek

Sumber: Anas Sudijono, 2005 : 389.

Keputusan item soal dapat diterima jika daya beda bernilai 0,3. Apabila daya beda bernilai 0,1- 0,3 maka butir soal tersebut perlu direvisi dan daya pembeda bernilai 0,1 maka butir soal ditolak.

d. Keefektifan pengecoh

Alternatif jawaban yang merupakan pengecoh (distraktor) yang baik harus memiliki koefisien korelasi yang negatif dan tinggi (Azwar, 2002:151). Karena hal ini mengindikasikan bahwa pengecoh (distraktor) dipilih oleh siswa dari kelompok bawah.

Analisis menggunakan program ITEMAN daya beda pilihan jawaban ditunjukkan oleh *prop endorsing* atau proporsi pemilih jawaban dan nilai *alternative biserial* dimana distraktor dikatakan baik jika *prop endorsing* bernilai lebih dari 0,05 atau minimal dipilih oleh 5% testee serta *alternative biserial* bernilai negatif tinggi. Nilai negatif pada *alternative biserial* menunjukkan bahwa pemilih alternatif jawaban tersebut sebagian besar dari kelompok bawah.

Rangkuman hasil analisis soal menggunakan program MicroCat ITEMAN versi 3.00 disajikan dalam Tabel 3.

Tabel 3. Rangkuman Hasil Uji Kelompok Kecil

Kriteria	Soal	Ket
Jumlah soal	40	
Jumlah peserta tes	26	
Koefisien reliabilitas	0.730	Tinggi (0,6-0,8)
Rata-rata tingkat kesukaran	0.485	Sedang (0,3-0,7)
Rata-rata daya beda semua soal	0.304	Cukup (0,2-0,39)

Berdasarkan Tabel 3 tentang hasil uji coba yang dilakukan dalam kelompok kecil ini maka dapat ditentukan bahwa soal yang di uji coba ini sudah baik. Hasil ini terlihat pada nilai

koefisien reliabilitas, rata-rata tingkat kesukaran dan rata-rata daya beda semua soal. Namun berdasarkan analisis tiap butir soal terdapat 20 butir soal yang harus direvisi isi (Tabel 4).

Tabel 4. Rangkuman Keputusan Uji Kelompok Kecil

Kategori	Nomor Soal	Jml soal
Diterima	12, 17, 20, 40	4
Direvisi	Distraktor 1, 2, 6, 8, 10, 15, 16, 21, 22, 26, 27, 28, 30, 31, 34, 35	16
	Isi 3, 4, 5, 7, 9, 11, 13, 14, 18, 19, 23, 24, 25, 29, 32, 33, 36, 37, 38, 39	20
Ditolak	-	-
Jumlah		40

Hasil uji pada Tabel 4 digunakan sebagai bahan untuk dilakukan revisi isi soal. Soal yang diperbaiki ini adalah soal yang akan digunakan untuk tes pada uji coba yang kedua yaitu uji coba kelompok besar. Revisi soal dilakukan dengan menggunakan dua metode, yaitu *feedback revision* dan *self revision*.

Uji kedua dilaksanakan dengan peserta uji berasal dari kelas XI IPA 1 dan XI IPA 2 SMA Negeri 7 Surakarta untuk metode *feedback revision*, sedangkan dari kelas XI IPA 3 dan XI IPA 5 SMA Negeri 7 Surakarta untuk metode *self revision*.

Pada uji coba kedua ini, jumlah soal yang digunakan sebanyak 40 butir soal dengan waktu 90 menit. Pengerjaan tiap butir soal diperkirakan 2-3 menit. Seperti pada analisis uji coba kelompok kecil, analisis uji coba kelompok besar ini juga menggunakan program MicroCat ITEMAN versi 3.00. Rangkuman ITEMAN uji kelompok besar disajikan pada Tabel 5.

Tabel 5. Rangkuman ITEMAN Uji Kelompok Besar

Kriteria	<i>feedback</i>	ket	<i>self</i>	ket
Jumlah soal	40		40	
Jumlah peserta tes	55		55	
Koefisien reliabilitas	0.769	Tinggi (0,6-0,8)	0.818	Sangat tinggi (0,8-1)
Rata-rata tingkat kesukaran	0.401	Sedang (0,3-0,7)	0.434	Sedang (0,3-0,7)

Rata-rata daya beda semua soal	0.315	Cukup (0,2-0,39)	0.348	Cukup (0,2-0,39)
--------------------------------	-------	---------------------	-------	---------------------

Analisis pada uji coba kelompok besar meliputi aspek reliabilitas, taraf kesukaran, dan daya beda. Dari 40 soal yang diujikan berdasarkan Tabel 5 didapat hasil reliabilitas pada metode *self revision* lebih unggul dibandingkan dengan metode *feedback revision*. Untuk rata-rata tingkat kesukaran lebih unggul metode *self revision* dibandingkan dengan metode *feedback revision*, tetapi kedua metode sama-sama tergolong sedang. Dan untuk rata-rata daya beda lebih unggul metode *self revision* dibandingkan dengan metode *feedback revision*, tetapi kedua metode sama-sama berkategori cukup.

Untuk mengetahui efektifitas dari kedua metode revisi, maka analisis yang dilakukan hanya pada butir soal yang mengalami revisi isi pada keputusan uji coba kelompok kecil yaitu sebanyak 20 butir soal. Rangkuman hasil ITEMAN untuk 20 butir soal dapat dilihat pada Tabel 6.

Tabel 6. Rangkuman ITEMAN untuk 20 butir soal

Kriteria	<i>feedback</i>	ket	<i>self</i>	ket
Jumlah soal	20		20	
Jumlah peserta tes	55		55	
Koefisien reliabilitas	0.656	Tinggi (0,6-0,8)	0.647	Tinggi (0,6-0,8)
Rata-rata tingkat kesukaran	0.411	Sedang (0,3-0,7)	0.411	Sedang (0,3-0,7)
Rata-rata daya beda semua soal	0.354	Cukup (0,2-0,39)	0.360	Cukup (0,2-0,39)

Berdasarkan Tabel 6 didapat hasil reliabilitas pada metode *feedback revision* lebih unggul dibandingkan dengan metode *self revision*, tetapi sama-sama kategori tinggi. Untuk rata-rata tingkat kesukaran kedua metode sama, yaitu tergolong sedang. Dan untuk rata-rata daya beda lebih unggul metode *self revision* dibandingkan dengan metode *feedback revision*, tetapi kedua metode sama-sama berkategori cukup.

Dari 20 soal yang dinyatakan direvisi bagian isi, pada metode *feedback revision* soal yang direvisi hanya 6 soal karena untuk soal yang lain peserta tes banyak yang menjawab lupa rumusnya. Dan penulis menganggap bahwa soal tersebut tidak perlu direvisi dan dapat diujikan langsung ke kelompok besar. 6 butir soal tersebut yaitu nomor soal 9, 14, 18,23, 25 dan 36.

Dari hasil yang telah didapat dari uji kelompok besar perbandingan antara metode *self revision* dan *feedback revision* untuk keenam butir soal tersebut dapat dilihat pada Tabel 7.

Tabel 7. Perbandingan jumlah jawaban siswa pada metode *self revision* dan *feedback revision*

Jenjang Soal	No. soal	Feedback Revision		Self Revision	
		Jawaban benar	Jawaban salah	Jawaban benar	Jawaban salah
C3	9	32	23	22	33
C2	14	5	50	23	32
C1	18	21	34	24	31
C2	23	30	25	13	43
C3	25	30	25	8	47
C1	36	31	24	34	21

Berdasarkan 6 soal yang di revisi secara *feedback* dan *self revision* didapat hasil yaitu 3 soal lebih baik dengan menggunakan metode *feedback revision* dengan nomor soal 9 jenjang C3, nomor soal 23 jenjang C2 dan nomor soal 25 jenjang C3. Sedangkan 3 soal lainnya hasil yang didapat lebih baik menggunakan metode *self revision* dengan nomor soal 14 jenjang C2, nomor soal 18 jenjang C1 dan nomor soal 36 jenjang C1. Dari hasil tersebut dapat dilihat bahwa pada metode *feedback revision* memiliki jenjang soal yang lebih tinggi dibandingkan dengan metode *self revision* meskipun keduanya memiliki jumlah soal yang sama.

Berikut adalah hasil analisis untuk masing-masing aspek:

a. Taraf kesukaran item

Hasil analisis taraf kesukaran untuk 20 soal diperlihatkan tabel berikut:

Tabel 8. Hasil Analisis Taraf Kesukaran

Klasifikasi	Feedback revision		Self revision	
	Nomor soal	Jml	Nomor soal	Jml
Mudah (0,7 < P < 1,00)	38, 39	2	-	-
Sedang (0,3 < P < 0,7)	4, 9, 13, 18, 23, 24, 25, 29, 32, 33, 36, 37	12	4, 9, 11, 13, 14, 18, 24, 32, 33, 36, 37, 38, 39	13
Sukar (P < 0,3)	3, 5, 7, 11, 14, 19	6	3, 5, 7, 19, 23, 25, 29	7
Jumlah		20		20

Sedangkan perbandingan taraf kesukaran untuk keenam soal antara kedua metode tersebut dapat dilihat pada Tabel 9.

Tabel 9. Perbandingan taraf kesukaran pada metode *self revision* dan *feedback revision*

Nomor soal	Self Revision		Feedback Revision	
	P	Ket	P	Ket
9	0,400	sedang	0,582	sedang
14	0,418	sedang	0,091	sukar
18	0,436	sedang	0,382	sedang
23	0,218	sukar	0,545	sedang
25	0,145	sukar	0,545	sedang
36	0,618	sedang	0,564	sedang

Berdasarkan Tabel 9 diperoleh hasil untuk metode *feedback revision* memiliki taraf kesukaran soal yang sedang sedabnyak 5 butir soal. Sedangkan untuk metode *self revision* memiliki taraf kesukaran soal yang sedang sebanyak 4 butir soal.

Kusaeri (2012:174) berpendapat “Tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Misalnya, untuk keperluan ujian semester digunakan butir soal yang mkemiliki tingkat kesukaran sedang, untuk keperluan seleksi digunakan butir soal yang memiliki tingkat kesukaran tinggi atau sukar, dan untuk keperluan diagnostik biasanya diogunakan butir soal yang memiliki tingkat kesukaran rendah atau mudah.”

Sehingga dapat disimpulkan bahwa untuk analisis taraf kesukaran pada 20 butir soal revisi isi setelah diujicobakan kelompok besar dari kedua metode revisi, maka lebih unggul dengan menggunakan metode *self revision* dibanding metode *feedback revision*. Sedangkan untuk 6 butir soal dari kedua

metode revisi, maka lebih unggul dengan menggunakan metode *feedback revision* dibanding metode *self revision*.

b. Daya pembeda item

Rangkuman hasil analisis daya pembeda soal disajikan pada Tabel 10 dan keputusan hasil analisis daya pembeda disajikan pada Tabel 11.

Tabel 10. Hasil Daya Pembeda Soal

Klasifikasi	Feedback		Self	
	Nomor	Jml	Nomor	Jml
Excellent (0,70–1,00)	-	-	13	13
Good (0,40–0,69)	5, 9, 11, 13, 23, 32, 33, 36, 37, 38	10	3, 4, 5, 14, 23, 24, 32, 33, 38, 39	10
Satisfactory (0,20–0,39)	3, 7, 14, 18, 19, 24, 25, 29, 39	9	7, 9, 11, 18, 19, 25, 29, 36, 37	9
Poor (0,00–0,19)	4	1	-	0
Jelek sekali (< 0,00)	-	-	-	0
Jumlah		20		20

Tabel 11. Keputusan Hasil Daya Pembeda Soal

Klasifikasi	Feedback		Self	
	Nomor	Jml	Nomor	Jml
Diterima	3, 5, 7, 9, 11, 13, 14, 18, 19, 23, 24, 29, 32, 33, 36, 37, 38, 39	18	3, 4, 5, 7, 9, 11, 13, 14, 18, 19, 23, 24, 25, 29, 32, 33, 36, 37, 38, 39	20
Direvisi	4, 25	2	-	0
Ditolak	-	-	-	0
Jumlah		20		20

Sedangkan untuk perbandingan daya beda soal dari keenam soal tersebut antara metode *self revision* dan *feedback revision* dapat dilihat pada Tabel 12.

Tabel 12. Perbandingan daya beda soal pada metode *self revision* dan *feedback revision*

Nomor soal	Self Revision		Feedback Revision	
	B	Ket	B	Ket
9	0,372	Cukup	0,415	Baik
14	0,651	Baik	0,355	Cukup
18	0,389	Cukup	0,386	Cukup
23	0,628	Baik	0,582	Baik

25	0,328	Cukup	0,278	Cukup
36	0,354	Cukup	0,487	Baik

Berdasarkan Tabel 12 diperoleh hasil untuk metode *feedback revision* memiliki daya beda soal yaitu 3 soal kategori baik dan 3 soal kategori cukup. Sedangkan metode *self revision* memiliki daya beda soal yaitu 2 soal kategori baik dan 4 soal kategori cukup. Dari hasil tersebut dapat dilihat bahwa daya beda soal untuk 20 butir soal revisi isi metode *self revision* lebih unggul dibanding metode *feedback revision*. Sedangkan untuk perbandingan dengan 6 butir soal revisi pada metode *feedback revision* memiliki daya beda soal yang lebih baik dibandingkan dengan metode *self revision*.

Dari hasil dan analisa yang telah dilakukan secara keseluruhan dapat disimpulkan bahwa metode *feedback revision* memiliki hasil yang lebih baik dibandingkan dengan metode *self revision*. Meskipun secara hasil dari 20 butir soal yang diperoleh metode *self revision* sedikit lebih unggul dibanding dengan metode *feedback revision*. Akan tetapi jika dilihat pada soal-soal yang sama-sama dilakukan revisi secara *feedback* dan *self revision*, maka hasil yang diperoleh yaitu metode *feedback revision* lebih unggul dibandingkan dengan metode *self revision*.

Keuntungan pada metode *feedback revision* ini yaitu dapat mengetahui peta kekurangpahaman siswa, ketepatan dalam melakukan revisi soal, dan sebagai evaluasi model pembelajaran yang telah digunakan pada kegiatan belajar mengajar. Sedangkan untuk kelemahan pada metode *feedback revision* yaitu dibutuhkan waktu dan biaya tambahan untuk membuat angket pertanyaan.

III. Kesimpulan

Dari hasil dan analisa yang telah dilakukan dapat disimpulkan bahwa metode *feedback revision* memiliki hasil yang lebih baik dibandingkan dengan metode *self revision*. Dan jika dilihat pada soal-soal yang sama-sama dilakukan revisi secara *feedback* dan *self revision*, maka hasil yang diperoleh metode *feedback revision* lebih unggul dibandingkan dengan metode *self revision*. Dengan demikian bisa disimpulkan bahwa metode *feedback*

revision lebih efektif dibandingkan metode *self revision*

IV. Daftar Pustaka

- Arikunto, Suharsimi. 2011. *Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.
- Azwar, S. (2002). *Tes Prestasi Fungsi Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.
- Depdiknas. 2008. *Panduan Analisis Butir Soal*. Diperoleh 27 Januari 2014. Dari: http://teguhsasmitosdp1.files.wordpress.com/2010/05/analisis_soal1.pdf
- Direktorat Pembinaan SMA. 2010. *Juknis Analisis Butir Soal*. Diperoleh 27 Januari 2014. Dari: http://brawy.files.wordpress.com/2011/04/34-juknis-analisis-butir-soal-isi-revisi_2911.pdf
- Kusaeri dan Suprananto. 2012. *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu.
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen tes dan Non Tes*. Yogyakarta: Mitra Cendikia Press.
- Sudijono, Anas. 2005. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Tim Puslitjaknov. 2008. *Metode Penelitian Pengembangan*. Diperoleh 19 Agustus 2014. Dari <http://www.infokursus.net>
- Wayne Patience. 1990. *MicroCAT^d Testing System Version 3.0*. Diperoleh 28 Agustus 2014. Dari <http://www.jstor.org/discover/10.2307/1434769?uid=3738224&uid=2&uid=4&id=21104565372157>

Notulensi Tanya Jawab

Pertanyaan :

1. Bagaimana jika soal yang dibuat merupakan soal isian singkat bukan pilihan ganda? bagaimana kriteria soal bagus untuk soal isian singkat?
2. Apakah ada perbandingan untuk jenjang kesukaran soal antara C1, C2, C3 dan C4?
3. Jika sampelnya beda apakah hasilnya beda?
4. Penelitian ini bertujuan untuk meneliti soal atau siswa?

5. Kenapa memilih software iteman?
6. Apakah ada peneliti yang menggunakan program iteman?

Jawaban :

1. Tes ini merupakan tes sumatif sehingga materinya banyak, jika kita menggunakan soal uraian ditakutkan siswa tidak bisa mengerjakan soal tersebut. Kalau membuat soal isian singkat daya serap siswa susah dan terlalu mudah
2. Untuk distribusi soal C1 dan C2 terdistribusi sama. Untuk tes formatif kebanyakan menggunakan soal C3 dan C4
3. Saya belum melakukan penelitian untuk sekolah lain. Walaupun dilakukan penelitian disekolah lain mungkin hasilnya akan berbeda
4. Kita gunakan metode feedback supaya tahu apakah siswa itu mengerti atau memang soalnya yang bermasalah sehingga disini kita mencari tahu apakah siswa atau soal mempengaruhi hasil belajar siswa
5. Karena software ini lebih mudah digunakan, lebih cepat dan efisien waktu.
6. Penelitian lain sudah pernah menggunakan software iteman ini